THE RESPONSE OF NUMERICAL WEATHER PREDICTION SYSTEMS TO FGGE IIB DATA

PART I : ANALYSES

A. Hollingsworth, A.C. Lorenc, M.S. Tracton, K. Arpe,
G. Cats, S. Uppala, P. Kallberg
ECMWF, UKMO, NMC

ABSTRACT

We discuss an intercomparison of analyses of the main FGGE IIB data set with three advanced analysis systems. The purposes of the work are to estimate the magnitude of the differences between the analyses, to identify the reasons for the differences, and finally to estimate the significance of the differences for forecasting purposes.

We discuss objective evaluations of analysis quality, such as fit to observations, statistics of analysis differences, and analysis energetics. However, we give substantial weight to a series of case studies that have been selected to illustrate the importance of different aspects of the analysis procedures, such as quality control, data selection, resolution, and the role of the assimilating model. In Part II, Arpe et al (1983), we consider the impact of the analysis differences on forecast quality.

We confine ourselves to a consideration of the extra-tropical analyses. The subject of the tropical analyses merits separate treatment.

In general the analysis systems draw reasonably well to the data, although each system has its own characteristics in this regard. The root mean square differences between the analyses are of the expected order of magnitude, although there are clear differences in the closeness of agreement of different pairs of analyses. Systematic differences arising from particular components of the assimilation suites can be identified.

A consideration of a set of case studies forms a large part of this paper. The discussion of these cases has tried to highlight those areas where differences of approach to the analysis problem have led to significant differences in the analyses. In choosing some of the cases we have used the forecast models as selective amplifiers of analysis differences. Some of the case studies suggest strongly that analysis differences in the vicinity of active baroclinic zones are of particular importance. In order to validate these suggestions in one of the cases we present the results of an experiment where we transplant one analysis into another, and show that we can then attribute large differences in the medium range forecasts to localised differences in the analyses.

The results from the case studies, together with the forecast verification results of Part II suggest that in some flow regimes, uncertainties in the analysis are a major contributor to the loss of forecast skill in the medium range.

1. INTRODUCTION

This paper is concerned with the results of an intercomparison of analyses of five days of data from February, 1979 performed with advanced numerical weather prediction systems at the European Centre for Medium Range Weather Forecasts (EC), at the National Meteorological Centre Washington (US), and at the United Kingdom Meteorological Office (UK). All three systems were presented with the same observational data from the main FGGE IIB data-set.

The objectives of the study are:

1. To quantify, as far as possible, the extent of the differences between the analyses;

2. To relate these differences to the procedures for quality control, interpolation, and dynamical control used in the analysis systems.

In Part II we present statistics on, and discuss the significance of, the analysis differences for the short and medium range predictive skill of forecasts based on the analyses.

Several different techniques were used to compare the analysis systems. These included:

a) fit of the analyses to the data

b) energetics and mean features of the analyses

c) internal consistency of the analyses

d) skill of the forecasts based on the analyses.

With regard to criterion c) we have found that this is best discussed in terms of the individual case studies. There is no single overall measure which summarises the various requirements of balance and consistency which three dimensional analyses should satisfy. We are thinking here of conceptual models of atmospheric behaviour such as geostrophic or gradient balance, the structure of baroclinic waves or the vertical circulation of fronts. An analysis that differed radically from our expectations in these aspects would be regarded with scepticism. We shall touch on these points in the case studies and summarise the results in the last section.

Assessments of the skill of forecasts are based on case studies in this paper. A more extensive discussion of the forecast experiments is presented in Part II.

Intercomparison studies of analyses by different systems have been undertaken several times in the recent past (cf. Jarvenoja (1982), Rosen and Salstein (1980), Lau and Oort (1981, 1982), Arpe (1980), Parker (1980), Trenberth and Paolino (1980)). These studies have concentrated mainly on the means and variances of the differences over periods of a month or longer. The main interest of the authors was to evaluate the differences for the purposes of climatological studies.

Our focus is rather different in that our main interest is in the source of analysis differences, and in the impact of the analysis differences on forecast quality. The work of Otto-Bliesner et al (1977) was concerned solely with the differences between analyses. We are aware of only one earlier set of studies where the performance of analysis/forecast systems have been compared on the same set of initial data. This was the set of studies based on the Data Systems Test data (cf. Desmarais et al (1978), Atlas (1979), Ghil et al (1979), Tracton et al (1980), Tracton et al (1981)). The main interest of these studies was in estimating the impact of satellite temperature retrievals on forecast quality.

In Sect. 2 we describe the analysis systems and discuss in qualitative terms the likely implications of the differences between them. In Sect. 3 we discuss some of the more important results on the verification of the analyses against observations. Generally speaking, all the systems fit the observations reasonably well; but each system has idiosyncracies in the way in which it uses the data. It should be borne in mind that a good fit to the data is merely a minimum requirement of an analysis system. The data must be interpolated and extrapolated in a physically consistent manner so that as much information as possible is conveyed to the forecast model. In Sect. 4 we present some differences between the mean analysis fields. These have interesting implications on the effect of the assimilating model on the analyses. In Sect. 5 we present a set of case studies which were chosen on two grounds, either to illustrate the general discussion and the results in earlier sections, or because there were large differences in the subsequent forecasts. One case study was chosen to illustrate a major finding of this study, that analysis differences can lead to considerable differences in the medium-range forecasts through a process of down-stream propagation and amplification, such as that discussed by Simmons and Hoskins (1979). In order to justify this interpretation of the case study we have transplanted one analysis into the other in the area suspected of being the source region of the medium-range forecast differences. A forecast from this composite analysis amply justifies our interpretation. We conclude in Sect. 6 with a summary of our results.

The work presented here has concentrated mainly on the analysis of the mass and wind fields of the extra-tropics. We have devoted no space to consideration of the humidity analyses and rainfall forecasts nor to the tropical analyses and forecasts. This is because of pressure of space and time, and because known biases and lower levels of skill make intercomparison less fruitful. It is clear from a preliminary survey that there is much to discuss in both these important areas.

Finally, it should be understood that the results discussed below were not produced by the current operational analysis systems. The EC analyses were produced in 1980 with the FGGE IIIb production system. The UK analyses were produced with a research system used by Lyne et al (1982). This system was a precursor of the current operational system at the UKMO. The US analyses were produced by a pre-operational version of the current operational system. All three operational systems have benefitted from important development work since the time these analyses were produced.

2. NWP SYSTEMS AND DATA USED

Some details of the systems used are given in Table 2.1. Further details can be found in the references given, and the systems are described in Sect. 2.1, 2.2 and 2.3 below. In 2.4 differences significant to this study are discussed.

All the systems were designed using the concept of four-dimensional data assimilation; data are used to correct a forecast "first-guess" provided by a sophisticated numerical forecast model. This first-guess is usually rather accurate and all the analysis systems take account of this, using the "optimal interpolation" (OI) technique, and gave it considerable weight. Thus the properties of the forecast models used will affect the analyses.

	EC		US		UK
	Analysis	Forecast	Analysis	Forecast	Analysis and Forecast
Levels:	p (mb)	σ	p (mb)	σ	σ
	10	.025		.025	.022
	20	.077			
	30	.132			
	50	.193	50		
	70	.260	70	.075	
	100	.334	100	.125	.089
	150	.415	150	.175	.157
	200	.500	200	.225	.230
	250	.589	250	.275	.317
	300	.678	300	.338	.436
	400	.765	400	.438	.577
	500	.845	500	.575	.718
	700	.914	700	.725	.843
	850	.967	850	.862	.937
	1000	.996	1000	.963	.987
Resolution :	1.875°	1.875°	3.5°	Rhomboidal 30	2°
Grid :	Lat. Long.	Lat. Long.	Quasi-homogeneous	Spectral	Quasi-homogeneous
Variables :	ϕ,u,v,q	T,u,v,p_s,q	ϕ,u,v,RH	T,u,v,p_s,q	T,u,v,p_s,q

TABLE 2.1a SUMMARY OF NWP SYSTEMS USED

	EC	US	UK
Analysis method	3-dimensional multivariate OI	3-dimensional multivariate OI	2-dimensional univariate OI
Data used for each point	$\leq$ 191	$\leq$ 20	$\leq$ 8
Update interval	6 hours	6 hours	6 hours
References	Lorenc (1981)	Bergman (1979) MacPherson et al. (1979)	Lyne et al. (1982)
Initialisation method			
	Non-linear normal mode. 5 vertical modes. Adiabatic.	Non-linear normal mode. 4 vertical modes. Adiabatic.	Repeated insertion during 6-hour forecast with increased diffusion and time-filtering
References	Machenhauer (1977) Temperton and Williamson (1982) Williamson and Temperton (1982)	Machenhauer (1977)	

TABLE 2.1b: PRINCIPAL FEATURES OF THE ANALYSIS SYSTEM

FORECAST	EC	US	UK
Horizontal scheme	2nd order staggered	Spectral	2nd order flux form
Time scheme	Semi-implicit	Semi-implicit	Leapfrog
Diffusion	Linear 4th order	Linear 4th order	Non-linear
Orography	Medium smooth	Smoothed 30 wave	Almost full resolution
Diurnal cycle	None	None	Included
Surface exchanges	Included	Included over sea. Only drag over land.	Included
Radiation	Interactive clouds	None	Climatological clouds
Latent heating	Included	Included	Included
Convection	Included (Kuo)	Included(Kuo)	Included
References	Hollingsworth et al. (1981)	Sela (1980)	Saker (1980)

TABLE 2.1c PRINCIPAL FEATURES OF THE FORECAST MODELS

INTERPOLATION	EC	US	UK
Forecast first guess	Cubic spline in log p	linear in log p; spectral	None
Analysis forecast	Cubic spline in log p	Linear in log p for analysis increments; spectral	None
Analysis output	None	$p \rightarrow \sigma \rightarrow p$	Cubic spline in log p

TABLE 2.1d PRINCIPAL FEATURES OF THE INTERPOLATION SCHEME

Analyses were exchanged between the NWP systems as output fields on standard pressure levels, and were studied using the processing software available at the three Centres. This involved some interpolation, which may have affected some detailed results; however our tests have not revealed any significant differences due to the interpolations.

Test forecasts were made at each centre using models similar but not identical to those used for the data assimilations. Each centre applied its own interpolation and initialisation techniques before the forecasts.

Five days from 15th to 19th February, 1979 were chosen for study. This was during the first special observing period of the First GARP Global Experiment (FGGE). Because of the special observing systems deployed for FGGE and the delayed mode (level IIb) data collection, observational coverage was generally better than in current operational practice. Selected maps of the EC analyses and the observation coverage for this period have been published by Bjorheim et al (1981). The main FGGE observational (level IIb) dataset was used by each system (Bengtsson et al 1982a) without manual intervention or manual quality control, other than some data selection decisions discussed in Sect. 2.4 below.

2.1 The EC system

The data assimilation system used to produce the IIIb analyses is practically identical to the ECMWF's operational system as it existed in 1980. It is an intermittent data assimilation system consisting of a multivariate optimum interpolation analysis, a non-linear normal mode initialisation (Machenhauer 1977, Andersen 1977) and a high resolution model which produces a forecast which is used as a background field for the subsequent analysis. Data are assimilated with a frequency of 6 hours. The analysis consists of two parts, one for simultaneous analysis of surface pressure, geopotential height and horizontal wind, and another part for analysis of humidity. Table 2.1 summarises the systems.

The EC system analyses for a large volume of atmosphere using many data simultaneously in order:

a) to ensure the enforcement of the linear constraints of non-divergence on the wind increments, hydrostatic balance of geopotential increments and near-geostrophy between height and wind increments in the extra-tropics.

b) to exploit fully the multivariate relationships in the data (Lorenc 1981).

Systems of equations as large as order 192 may be inverted in the analysis of a region covering 660 km square and a third of the atmosphere deep.

During the analysis, the observations are subject to a four step quality control containing: (i) checks of observation positions, internal consistency and climatological reasonability, (ii) check against the first guess forecast, (iii) checks of consistency between neighbouring observations, (iv) check against a preliminary analysis at the position of the observation, excluding the observation itself, but using all other data. Those observations that deviate more than a certain threshold value, depending on the observation error and the first guess error, are rejected. Detailed descriptions of the analysis and initialisation schemes can be found in Lorenc (1981), Temperton and Williamson (1981) and Williamson and Temperton (1981).

The EC model used for the forecasts in this intercomparison study is the ECMWF operational forecast model as it existed in 1982. It differs from that used for the assimilations in some details, most significantly in that its topography is less smooth (cf. Arpe (1982) , for further details).

2.2 The US system

The US analyses were produced by a multivariate optimum interpolation procedure, with the first guess provided by the NMC global spectral prediction model (Sela, 1980). The update interval in the data assimilation cycle was 6h, with a nonlinear normal mode initialisation preceding each six-hour forecast.

Height and wind residuals (observation minus guess) are analyzed simultaneously on 12 standard pressure surfaces from 1000-50 mb. Relative humidity residuals are analyzed univariately on 6 standard pressure surfaces from 1000-300 mb. Up to twenty height and wind observations and 8 relative humidity observations are used in the analysis at each point of a quasi equal area Gaussian grid with a horizontal resolution approximately 3.5 degrees. Output fields however are on a 2.5 degree mesh after spectral representation in sigma coordinates with a rhomboidal-30 truncation. Data are selected on the basis of distance to the gridpoint (in height/height correlation space) and are subject to a gross error check (with respect to the first guess) and a "buddy" check (with respect to each other) before being passed to the analysis.

After the analysis step is complete, the analyzed height, wind and relative humidity residuals on isobaric surfaces are converted to residuals of temperature, wind and specific humidity in the sigma coordinate of the global spectral prediction model. The updating of the first guess is then performed in this sigma coordinate. Preceding the six-hour forecast which provides the firsts guess for the subsequent analysis, two iterations of the Machenhauer (1977) normal mode initialisation are performed, with four vertical modes.

The forecast model used in the data assimilation cycle is a 12-layer, primitive equation, hydrostatic model employing a global spherical harmonic representation of the variables in the sigma coordinate system. The spectral model has a rhomboidal-30 truncation. The physical effects include the influence of orography, position-dependent surface friction, and subscale horizontal dissipation parameterized by diffusion. The moisture cycle is based on a mixing ratio formulation with large-scale precipitation and Kuo (1965)-type convection, and evaporation from the oceans. Sensible heating from underlying water is also included. There is no inclusion of the thermal effects of radiation or the diurnal cycle.

The prediction model used to produce medium range forecasts was the same as that employed in the data assimilation system except that the spectral truncation was reduced to rhomboidal 24 beyond two days and the vertical resolution diminished to six layers at 3.5 days. This version of the model is equivalent to that used operationally by NMC at the time of these experiments and was dictated by computational constraints.*

2.3 The UK system

The UK system is based on a repeated interpolation and insertion of data at each timestep of a forward running forecast model. This method was first used for idealized data network studies by Lorenc (1976). It was then developed for practical use during FGGE where it was used to produce global analyses for general circulation studies during the two Special Observing

---

*Tests comparing this version of the model with the full 30 wave, 12 layer version through 10 days shows a marked improvement in the latter relative to the former (NMC Office Note, Tracton, 1982).

Periods (Lyne et al, 1982). The system used for the present study was a slightly modified version of the FGGE system. The method has since been considerably developed and recoded for operational use at the UK Meteorological Office.

An 11-layer general circulation model was used to assimilate the observations in 6 hour batches. These were first interpolated vertically to the model's levels and transformed if necessary to the model's variables. The surface pressure and the vertical shears from the analysis 6 hours earlier are used to define the pressures at sigma levels, and to move single level observations to the nearest model level. Data were then quality controlled by rejecting those which disagreed grossly with the previous analysis and then by comparing each with a value interpolated from nearby data using horizontal univariate OI. The normal prognostic equations of the model, represented schematically by

$$\psi_t = M(\psi_{t-\Delta t})$$

were modified to give

$$\psi_t^* = M^1(\psi_{t-\Delta t})$$

and

$$\psi_t = \psi_t^* + \lambda \sum_i (\psi_{i,obs} - \psi_t^*) w_i$$

M represents the model equations modified by adding extra diffusion and damping of high frequency modes. The weights w_i were calculated by univariate horizontal OI and kept constant over the 6 hour period. The factor λ was increased linearly from 0 at the beginning of the period to 0.5 at the end.

The major modification from the system used during FGGE is the length of the insertion period; here it is 6 hours ending at the nominal analysis time, then it was 3 hours centred on the nominal analysis time.

2.4 Significant differences between analysis systems

a. Approach to slow manifold

A useful concept when discussing the principles of objective analysis is that of the slow manifold (e.g. Daley and Puri, 1980). This asserts that out of all possible states the atmosphere is always in or near a small sub-set characterized by slow variations with time. The task of an objective analysis procedure is to choose a state which fits the current observations to within the likely observational error, which is close to a forecast based on earlier observations, and which is on or near the slow manifold. All three conditions are necessary because of the incompleteness of data coverage at any one analysis time. The EC and US systems differ from the UK system in the approach used to achieve this third condition. The EC and US systems assume that a good approximation to the slow manifold is the manifold of Rossby modes and a better approximation is that achieved by non-linear normal mode initialisation. Because of their initialisations the forecast first-guess is near the slow manifold, and the analyses make changes to this which are close to geostrophic balance and hence close to Rossby modes. This is achieved in their multivariate OI by using geostrophically consistent structure functions and by using the same data for analysis of mass and wind fields, for as large a volume as possible (Lorenc, 1981, Phillips, 1982). The UK system on the other hand, approaches the slow manifold by repeatedly inserting the observational information into the numerical model during a six hour assimilation period. Modes of the model with periods less than or about six hours are thus less excited by the data than are the slowly varying modes. Moreover, the numerical model is modified to damp high frequency modes.

Both approaches only approximate the idealized slow manifold, which itself is only an approximate description of real atmospheric behaviour. The UK system discriminates solely on the basis of frequency; internal gravity wave modes with complex vertical structures have periods much longer than six hours, as have horizontal 2 grid length waves in the model, so the UK system can generate such modes to fit isolated or inconsistent data. In many such cases the analyses achieved are further from balance than we believe the atmosphere to be. On the other hand, the UK approach may represent features which are not geostrophic, such as flow around mountains or fronts; whether or not such motions remain on the slow manifold may still be an open question.

b. Quality control, selection and weighting of data

A major part of the effort of building an analysis system for operational use is expended on the design of methods for choosing which data to leave out. Two types of data need to be identified; those which are grossly incorrect or misleading (quality control), and those which carry little extra information over other data which are being used and which can therefore be disregarded to save time (data selection). Because all data have errors of observation or representativeness, and because of intrinsic or explicit assumptions about the smoothness of fields, the analysis has to be a compromise between the various selected observed values and the first-guess; the compromise is specified by the data weights.

All the schemes had a number of externally imposed quality control decisions. For instance the UK and US systems did not use land surface wind data, the US system did not use satellite temperature soundings over land, and the EC system did not always use the reported cloud wind levels, recalculating them, where possible in the upper troposphere, from reported temperatures. These all caused occasional analysis differences. Quality control of individual observations on a case by case basis is also essential, and all the systems had automatic methods for doing this. (No human intervention was allowed for this study). Such a quality control is only possible if there is information redundancy, and, since observed data are in many cases too sparse for this, it is necessary also to use information from a forecast together with knowledge of the likely structure of atmospheric motions. The EC system should be best at this, since it uses its full analysis method to check each datum against an analysis made not using it.

The US scheme has a separate comparison with the forecast, with strict limits which, in some cases studied here, reject some data which the others accept. (Wrong decisions in such cases can be crucial since it is the data which disagree with the forecast and are correct which carry most new information). This check is followed by comparisons between close observations designed to identify and reject those that disagree with several neighbours. These comparisons are univariate and two-dimensional, so little knowledge of atmospheric structures is used, other than that they are smooth and continuous.

The UK scheme also has a comparison with a forecast field (for these studies a six hour "persistence" forecast) followed by a univariate two dimensional comparison with neighbours, in this case using OI, with the forecast as first-guess. The rejection limits used are rather lax, in order that "noise" in the rather unbalanced forecast should not cause rejection of good data. The UK system occasionally accepts and draws to data rejected by the others. This is exacerbated by the relative lack of checks in the UK system on things like message formats and observation positions, compared with the other systems, which were developed for operational use.

The data selection methods also had a number of externally imposed decisions which differed between the systems. For instance, within one radiosonde report the height and temperature data convey largely redundant information, so after using this redundancy for quality control it is unnecessary to had automatic methods for doing this. (No human intervention was allowed for this study). Such a quality control is only possible if there is information redundancy, and, since observed data are in many cases too sparse for this, it is necessary also to use information from a forecast together with knowledge of the likely structure of atmospheric motions. The EC system should be best at this, since it uses its full analysis method to check each datum against an analysis made not using it.

The US scheme has a separate comparison with the forecast, with strict limits which reject, in some cases studied, some data which the others accept. (Wrong decisions in such cases can be crucial since it is the data which disagree with the forecast and are correct which carry most new information). This check is followed by comparisons between close observations designed to identify and reject those that disagree with several neighbours. These comparisons are univariate and two-dimensional, so little knowledge of atmospheric structures is used, other than that they are smooth and continuous.

The UK scheme also has a comparison with a forecast field (for these studies a six hour "persistence" forecast) followed by a univariate two dimensional comparison with neighbours, in this case using OI, with the forecast as first-guess. The rejection limits used are rather lax, in order that "noise" in the rather unbalanced forecast should not cause rejection of good data. The UK system occasionally accepts and draws to data rejected by the others. This is exacerbated by the relative lack of checks in the UK system on things like message formats and observation positions, compared with the other systems, which were developed for operational use.

The data selection methods also had a number of externally imposed decisions which differed between the systems. For instance, within one radiosonde report the height and temperature data convey largely redundant information, so after using this redundancy for quality control it is unnecessary to select both. The UK system uses temperatures, the EC and US heights. From satellite temperature soundings the UK system uses temperatures, the EC inter-level thicknesses and the US system heights, calculated using a preliminary surface analysis as reference level. These differences affect the effective weights given to the data and the fit of the analyses to the observations.

Quality control and data selection procedures, which give certain data zero weight, are of greater practical importance than the differences between weights assigned to observations by the systems' respective OI schemes. The estimated observational error variances which are used in the OI method to determine the relative weights of observations were similar in the three systems.

c. Resolution

The effective resolution can be limited in two ways: by the grid used to represent the fields, and by constraints on the smoothness of the fields. The UK system is limited solely by the former; the analysis is performed directly on the model's grid, and there are few constraints on smoothness. Thus all features which can be represented by the model can be analysed. The EC system has a slightly higher grid resolution than the UK system, however it selects and uses many more data for the analysis of each grid point value than the others, thus the resolution of features changed during the analysis is limited by the scale of the prediction error correlation structure functions used (Lorenc 1981). Moreover the interpolations between sigma and pressure co-ordinates have a smoothing effect. The US system has a coarser horizontal grid than the others, and its resolution is further limited by spectral smoothing of analysis increments and fields.

d. The assimilating model

A study by Leith (1981) indicates that there is a synergy between the accuracy of the analysis and the accuracy of the assimilating model. Using energy budget arguments he demonstrates that a model which gives a more accurate first-guess, leads to a more accurate analysis. There are practical as well as theoretical reasons why this should be the case. With a more accurate first-guess, quality control can be made more stringent and so more effective. Secondly, the linear constraints used by most analysis systems will be more accurate when applied to smaller amplitude differences between observational data and the first guess. If the first-guess is very inaccurate then the analysis should consider non-linear wind laws, such as the gradient wind relation, in areas where it has to make large increments. No satisfactory procedure for doing this has been worked out. We shall see examples of both these practical considerations in the case studies.

3. VERIFICATION OF THE ANALYSES AGAINST OBSERVATIONS

The most basic evaluation of an analysis is the degree to which it fits the observational data. All three systems recognise that the observational data are themselves erroneous and should not therefore be fitted exactly. However it is an essential requirement for a good analysis system that the data should be fitted to within their assumed observational error.

In what follows we shall see that the systems do not fit the data to the same degree, although the differences are not large. All the significant differences between the analysis systems discussed in 2.4 above can affect these results. An analysis is less likely to fit closely those data it does not select. The imposition of constraints of smoothness, in either the vertical or horizontal or of multivariate relationships such as geostrophy or non-divergence make it less easy for analyses to fit the data closely.

Our verifications of the analyses against observational data will be affected by a bias introduced through the selection of verifying data. We verified the analyses against those data which were accepted by the EC system, from which a record of rejected data was available in convenient form; the verification was done on standard pressure levels. We do not believe that biases of this nature affect our conclusions. Statistics for two case studies using manually controlled data are presented later (cf. Table 5.1); the exclusion of manually controlled data in these cases did not significantly change the results.

Fig. 3.1 shows the results of verifying the analyses of geopotential against all radiosondes for the analyses from Feb. 15 12Z to Feb. 19 12Z. On the right we show the bias, i.e. the mean difference between the analyses and the observations, while on the left we show the standard deviation (SD) of the analysis-observation differences. The square root of the sum of the squares of these two terms gives the root mean square (rms) difference between analyses and observations.

There is a marked bias in the fit to the radiosonde data for the UK system. This would suggest that the assimilating model is cold relative to the radiosondes in the troposphere. The EC and US systems show little bias in the troposphere.An interesting aspect of these results is that the day to day variability of the SD of the fit is rather small for all three systems, while the corresponding variability of the bias of the fit is relatively larger. In order to facilitate comparisons we have produced Fig. 3.2, which makes the pair-wise comparisons of the three sets of curves on Fig. 3.1. The EC and US analyses appear to fit the data equally well, with the EC analyses showing a tighter fit in the troposphere and the US analysis giving a tighter fit in the lower stratosphere. Both these analyses fit the data more closely than the UK analyses.

As pointed out earlier, the UK system uses temperature as the mass variable in the free atmosphere. We therefore compared, in Fig. 3.3, the fit of the

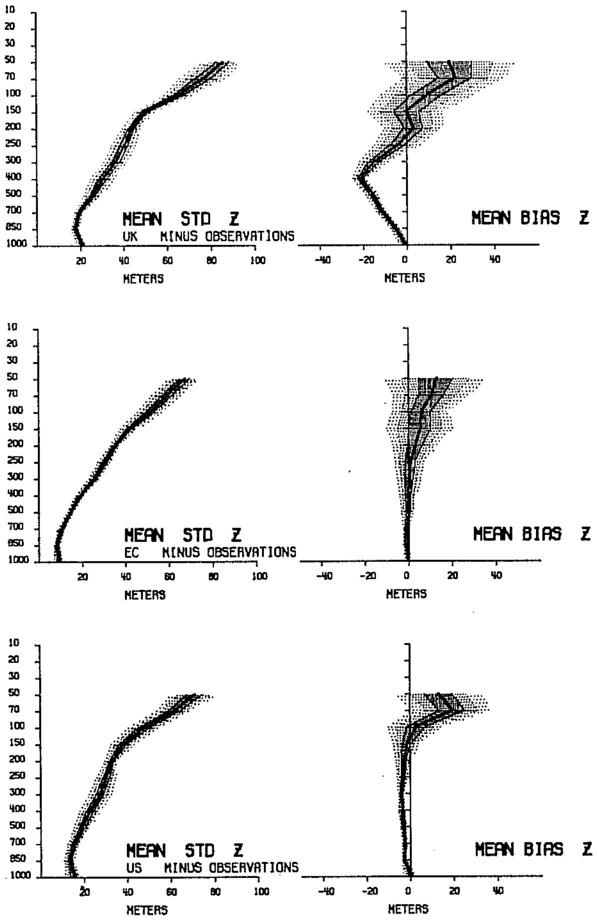


Fig. 3.1

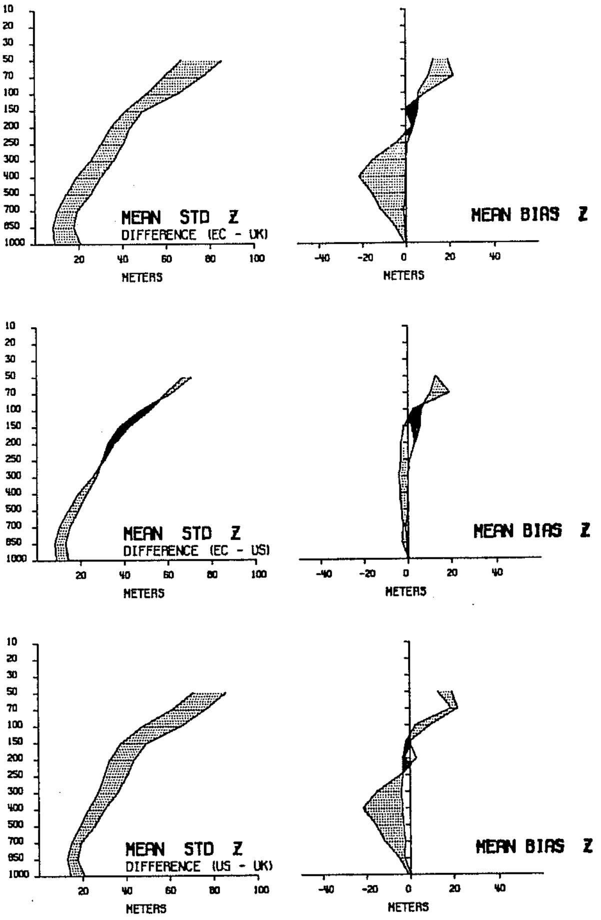


Fig. 3.2

Fig. 3.1 Solid lines : Standard Deviation (left) and Bias (right) of the fit of the UK (top), EC(middle), and US (bottom) analyses of height to radiosonde reports of height, as accepted by the EC system. Only standard level data is used in the calculations. The results are averages for all radiosondes and all analysis times. The shading indicates the day to day variability of the results for individual analysis times, the shading corresponding to one, two, or three standard deviations of the results for a single analysis time.

Fig. 3.2 Comparison of the results in Figure 3.1 for the average fit to height observations. The pairs of analyses compared are EC/UK (top), EC/US (middle), and US/UK (bottom). The shading convention is that lighter shading indicates that the first-named analysis of the pair is closer to the origin, while dark shading indicates the opposite. The units are meters.

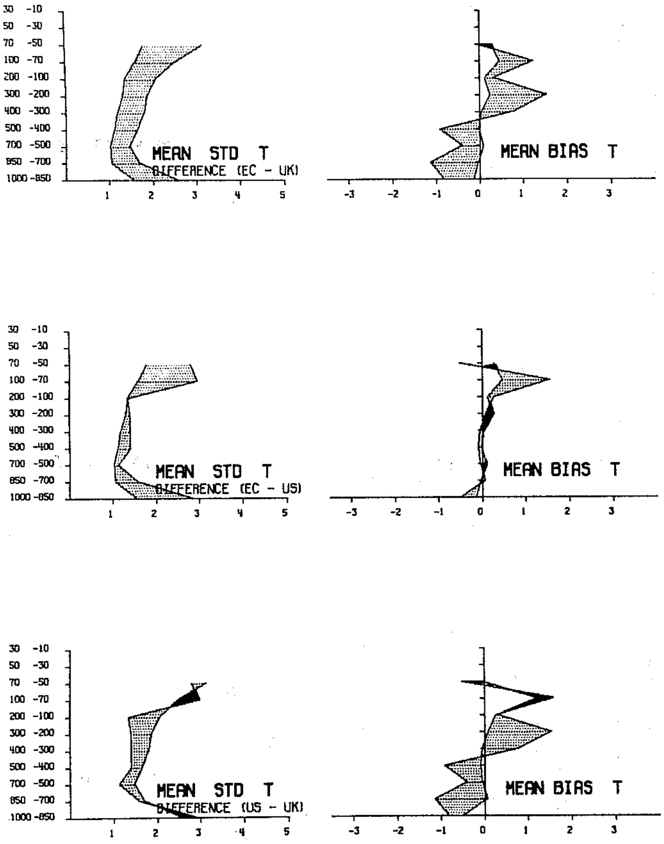


Fig. 3.3

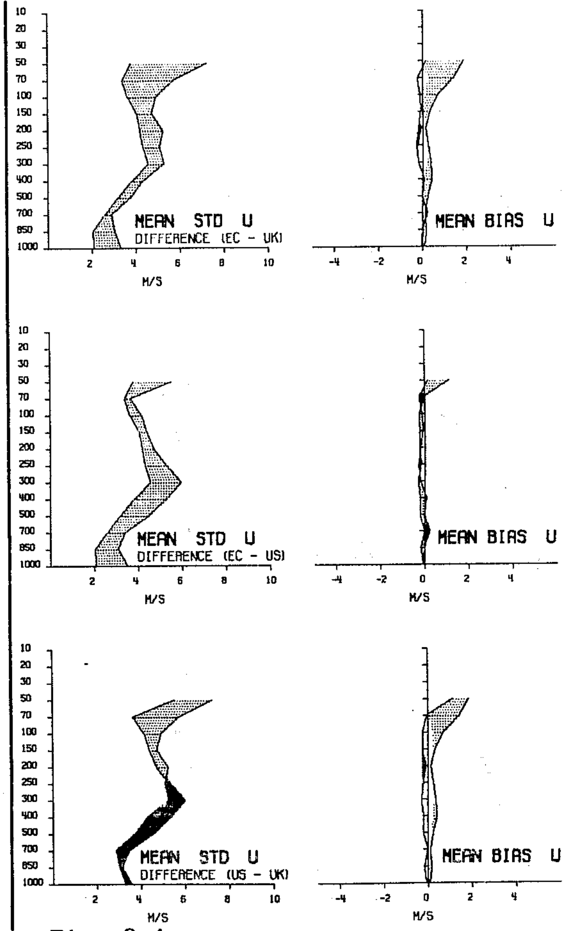


Fig. 3.4

Fig. 3.3 As Figure 3.2 for the fit to virtual temperature based on standard level thicknesses from radiosonde reports. The unit is one Kelvin.

Fig. 3.4 As Figure 3.2 for the comparison of the fits of the analyses to the radiosonde reports of the zonal wind component. The units are m/s; light shading indicates that the first of the pair is closer to the origin.

analyses to the radiosonde-thickness virtual temperatures. The UK system actually used the reported radiosonde temperature, which is not the same as the virtual temperature derived from thickness. The EC analyses show the closest fit to the data although the differences between them and the US results are modest except near and above the tropopause. It would appear that bias contributes significantly to the UK results.These latter analyses appear to be, on average, 1K too cold in the lower troposphere and 1K too warm in the upper troposphere.

The fit of the analyses to the radiosonde winds are compared in Fig. 3.4. We show only the u-component, as the results for the v-component are essentially the same. The order of tightness of fit is EC, UK, US in the troposphere and EC, US, UK in the lower stratosphere. Near jet levels the standard deviations (SD) of the fits vary between 4 m/s and 5.5 m/s. Bias effects in these calculations are small for all three systems.

Having considered the conventional data we now turn to some of the observing systems especially commissioned for FGGE. We begin with the fit to the Tiros-N satellite temperature retrievals, SATEMs. We verify the fit to this data in the extra- tropical southern hemisphere. The reason for this choice of area is that the US system did not use SATEM data over land, the UK system used it over land only if the report gave a value for the 1000-850 mb layer, while the EC system used it everywhere. For these reasons it was felt that the southern hemisphere extra-tropics offered the most suitable area for verification. A further point to note is that in this area there is very little mass data to compete with the SATEM data.

Fig. 3.5 compares the fit of the analyses to the SATEM clear path data in this area. The US and UK analyses show approximately the same SD of the fit. The UK SD is slightly lower than the US SD below 400 mb, while the reverse is true in the upper troposphere. It is noticeable that the bias term is as large as

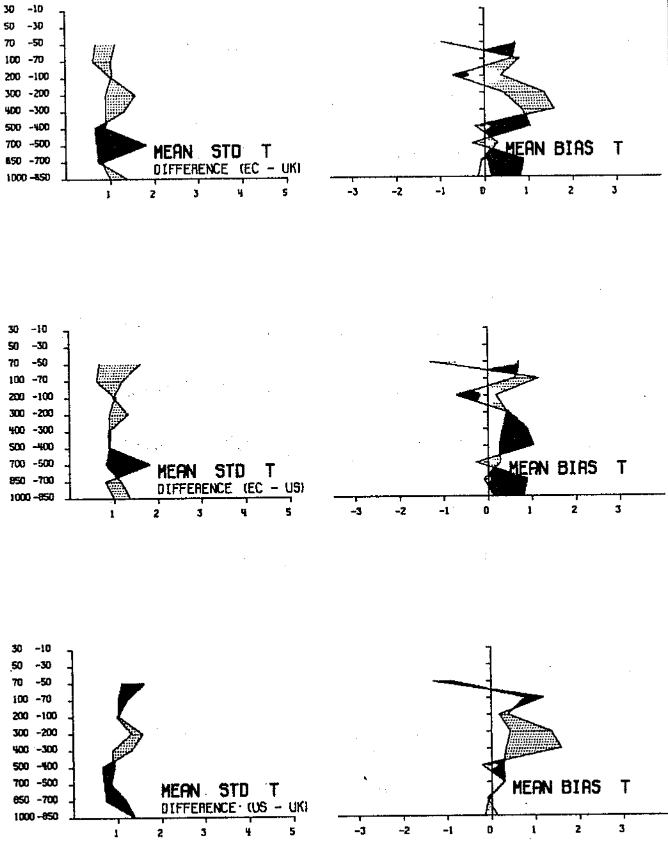


Fig. 3.5

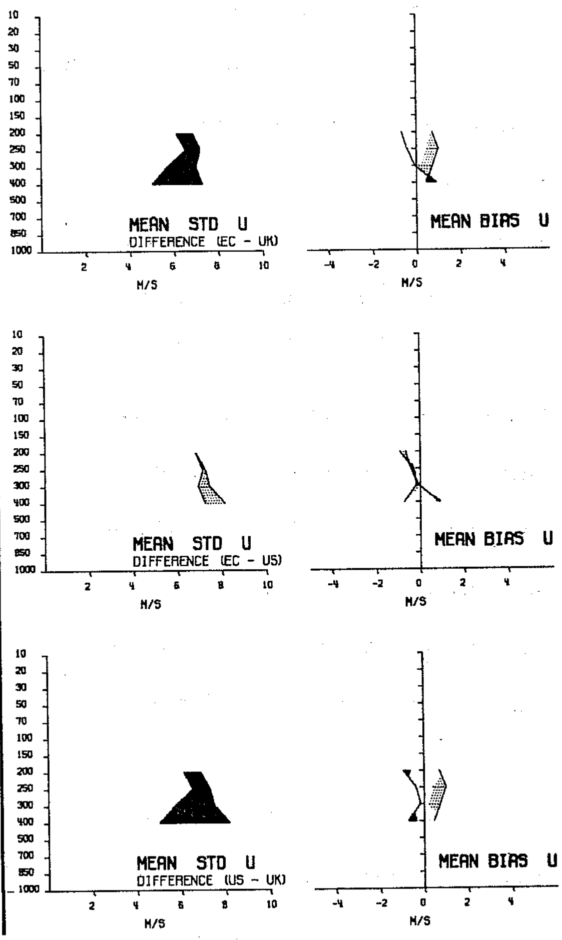


Fig. 3.6

Fig. 3.5 As Figure 3.2 for the fit of the analyses to the Clear-path satellite retrievals over the extra-tropical southern oceans. The variable is virtual temperature derived from thicknesses between standard levels. The unit is one Kelvin. Light shading indicates that the first-named analysis of the pair is closer to the origin.

Fig. 3.6 As Figure 3.2 for the fit of the analyses to all ordinary aircraft wind reports (AIREPS). The unit is m/s. Light shading indicates that the first-named analysis of the pair is closer to the origin.

the SD term for the UK analyses. The SDs of the fit for the EC system are somewhat tighter than for the other systems above 400mb. However, the EC system clearly has a problem with the 700-500 mb thickness. This is attributable to the unsatisfactory way in which analysis volumes were being overlapped in the vertical. This problem affected the FGGE IIIb analyses produced at ECMWF from the beginning of the FGGE year to the beginning of the second special observing period (SOP II May 5-June 5, 1979). The US analyses show the smallest biases, relative to the SATEM data, while the EC analyses are intermediate between the other two in this regard.

We turn now to the aircraft data. Fig. 3.6 shows the fit of the analyses to all AIREPs, for the u component. The UK system fits the data closer than either of the other two systems. The bulk of the data occurs at 250 and 200 mb, where the difference between the fit of the UK analyses and the others varies between 0.5 and 0.75 m/s. The difference between the EC and US systems is small at these levels.

Comparing Fig. 3.6 for the AIREPs with Fig. 3.7, which shows the corresponding results for the AIDS/ASDAR data (data taken by the inertial navigation systems on specially equipped wide-body aircraft), we see that all systems fit the AIDS/ASDAR data more closely than they fit ordinary AIREP data. The UK system again fits this single-level data more closely than do the other two systems. Differences between the EC and US systems are small at 250 and 200 mb where the bulk of the data occurs. The manner in which single level data is handled is discussed by Lorenc (1982).

Finally we consider, in Fig. 3.8, the fit of the analyses to cloud-wind data, SATOBs. We use the data processed by the Dept.of Meteorology at the Univ. of Wisconsin. This data was chosen because it had significant quantities of both upper and lower level data during the period in question. We discuss only the zonal component since it shows some typical problems for the cloud-wind data.

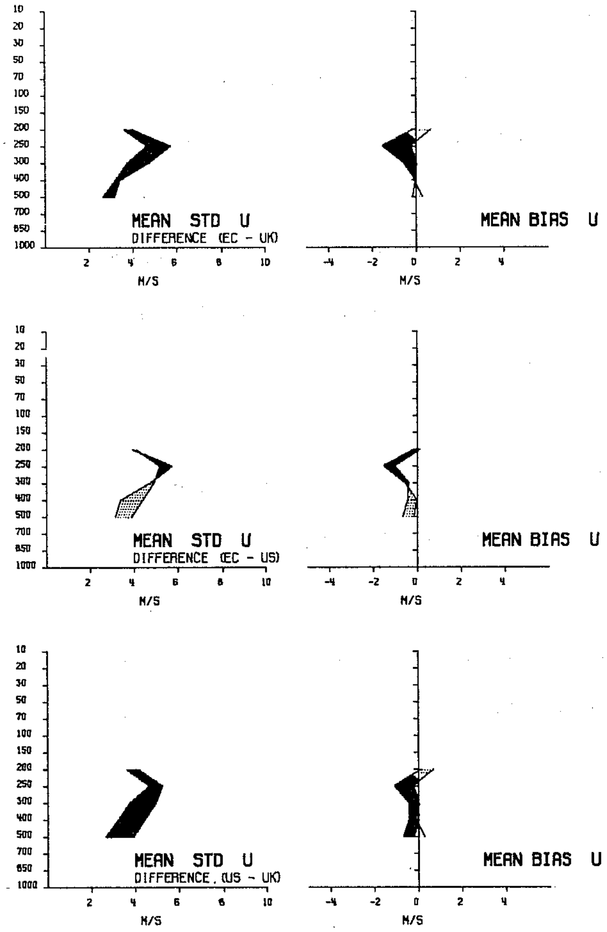


Fig. 3.7

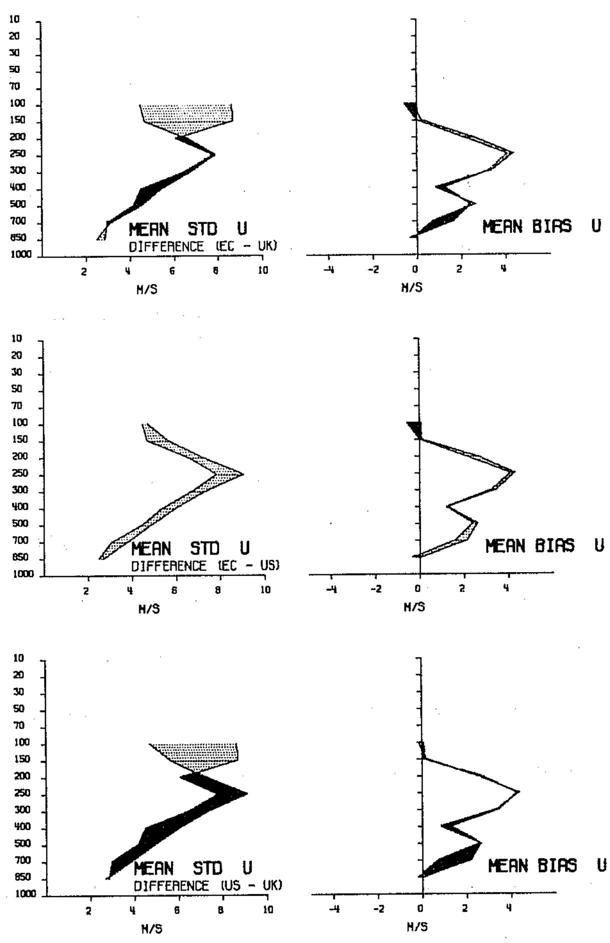


Fig. 3.8

Fig. 3.7 As Figure 3.2 for the fit of the analyses to the specially equipped wide-bodied jets which supplied high-density observations, either through a real-time satellite link, or in delayed mode on casettes (ASDAR/AIDS). The unit is m/s. Light shading indicates that the first-named analysis of the pair is closer to the origin. The bulk of the reports were in the 200-300 mb layer.

Fig. 3.8 As Figure 3.2 for the cloud drift winds (SATOBS) as processed by the University of Wisconsin. Most of the reports occurred in the 150-250 mb layer and at 850 mb. The unit is m/s. Light shading indicates that the first-named analysis of the pair is closer to the origin; dark shading indicates the opposite.

The main problem is that there is a large bias component in the fit for all three analysis systems in the upper tropospheric layers. This effect has been discussed by Gustafsson and Pailleux (1981) among others and is discussed in a case study in Sect. 5.1 . The effect arises from a tendency for the upper level cloud-winds to under-estimate the winds relative to other data. Cloud-wind data is ascribed larger observational error than either radiosonde or aircraft data in all three systems. As a consequence cloud-wind data is fitted much less tightly than other wind data. Given the fact that none of the systems fit this data tightly, the differences in the degree of fit are modest. However, once again the UK system fits this single-level data more tightly than do the other systems.

To summarise, all three systems fitted the data reasonably well, each system having its own peculiarities. The EC system fitted the radiosonde data tightest but it clearly had problems in using the SATEM data correctly in the 700-500 mb layer. The UK system fitted the single-level wind data most closely but had a biased fit to the mass field as measured by the radiosondes and SATEMs. The US system fitted all the data reasonably well in the troposphere but seemed to have some bias problems relative to the radiosonde temperatures near the tropopause.

4. STATISTICAL COMPARISON OF THE ANALYSES

In this section we study the statistics of the differences between the three sets of analyses, and try to identify some of the reasons for the differences.

In exploring the reasons underlying the analysis differences we have found it useful to examine in detail the various steps of the assimilation procedure, viz. the increments due to the analysis, the increment due to the initialisation, and the increment due to the combination of these steps. An earlier study along these lines (Hollingsworth and Arpe, 1982) was quite informative. Detailed investigation along these lines was only possible for the EC and US systems; the UK system does not have separate forecast, analysis and initialisation steps.

We begin our discussion with a general discussion of the magnitude of the analysis differences. We continue with more detailed studies of the extra-tropics of both hemispheres. We conclude with a discussion of some of the zonally averaged statistics.

4.1 Overview

Table 4.1 provides some general statistics on the area averaged RMS differences between the three sets of analyses (first three columns), for the northern and southern hemisphere extra-tropics and for the tropical regions. The remaining six columns provide information on the magnitudes of the changes made by the analysis, the initialisation, and their combined, or net, effect for the EC and US systems. In the northern hemisphere it is clear that the EC and US analyses are closer to each other than either is to the UK analysis. Areas of high terrain make the main contribution to differences in 1000 mb height between the UK and the other analyses. These numbers therefore reflect mainly the differences in extrapolation methods between the UK and the other systems.

At 500 mb the RMS difference of 19.4m compares well with the value of 20m calculated by Jarvenoja (1982) for the EC IIIb analyses and the US IIIa analyses for the winter season. The reasons for the differences between the UK analyses and the others are discussed below. The general level of the differences between the EC and US analyses agree reasonably well with the error estimates produced by the analysis systems as a by-product of the analysis calculation (Gandin, 1963).

NORTHERN HEMISPHERE 20°N - 90°N

MEAN RMS - DIFFERENCES OF THE DAILY FIELDS

	ECA-USA	ECA-UKA	USA-UKA
1000 mb Z (dm)	1.54	4.38	4.74
850 mb V (m/s)	3.62	4.19	4.77
500 mb Z (dm)	1.94	2.52	2.75
200 mb V (m/s)	5.43	6.63	7.40
TROPICS 25°S to 25°N			
850 mb V (m/s)	3.56	3.69	4.26
200 mb V (m/s)	6.18	7.46	7.79
SOUTHERN HEMISPHERE 20°S to 90°S			
1000 mb Z (dm)	3.04	7.54	6.74
850 mb V (m/s)	5.01	5.57	6.64
500 mb Z (dm)	2.96	2.91	3.53
200 mb V (m/s)	7.88	7.68	9.16

TABLE 4.1 (a) RMS values for the differences between the analyses. The areas are the northern hemisphere between 20 and 90N (top), the tropics, 25S to 25N (centre) and the southern hemisphere between 20 and 90S (bottom). The variables are winds at 850 mb and 250 mb in all three regions, and heights at 1000 mb and 500 mb for the extratropical regions.

NORTHERN HEMISPHERE 20°N - 90°N

MEAN RMS - DIFFERENCES OF THE DAILY FIELDS

	ECG-ECA	ECA-ECI	ECG-ECI
1000 mb Z (dm)	1.47	1.15	1.43
850 mb V (m/s)	2.89	1.54	2.91
500 mb Z (dm)	1.61	.95	1.50
200 mb V (m/s)	5.11	2.06	4.79
		TROPICS 25°S to 25°N	
850 mb V (m/s)	2.40	1.46	2.39
200 mb V (m/s)	5.06	2.71	4.58
		SOUTHERN HEMISPHERE 20°S to 90°S	
1000 mb Z (dm)	1.46	1.36	1.79
850 mb V (m/s)	2.69	1.20	2.68
500 mb Z (dm)	1.92	1.09	1.95
200 mb V (m/s)	4.27	1.77	4.08

TABLE 4.1 (b) RMS magnitudes for the analysis, initialisation and net increments of the EC system. The areas are the northern hemisphere between 20 and 90N (top), the tropics, 25S to 25N (centre) and the southern hemisphere between 20 and at 850 mb and 250 mb in all three regions, and heights at 1000 mb and 500 mb for the extratropical regions.

NORTHERN HEMISPHERE 20°N - 90°N

MEAN RMS - DIFFERENCES OF THE DAILY FIELDS

	USG-USA	USA-USI	USG-USI
1000 mb Z (dm)	1.68	1.19	1.39
850 mb V (m/s)	2.75	1.20	2.58
500 mb Z (dm)	1.05	1.32	1.52
200 mb V (m/s)	4.92	2.14	4.77
		TROPICS 25°S to 25°N	
850 mb V (m/s)	2.49	1.22	2.23
200 mb V (m/s)	4.65	2.37	4.27
		SOUTHERN HEMISPHERE 20°S to 90°S	
1000 mb Z (dm)	1.87	1.66	1.54
850 mb V (m/s)	2.68	1.14	2.55
500 mb Z (dm)	2.34	1.81	1.59
200 mb V (m/s)	5.05	2.11	4.86

TABLE 4.1 (c) As Table 4.1(b) for the US system. The areas are the northern hemisphere between 20 and 90N (top), the tropics, 25S to 25N (centre) and the southern hemisphere between 20 and 90S (bottom). The variables are winds at 850 mb and 250 mb in all three regions, and heights at 1000 mb and 500 mb for the extratropical regions.

The ordering of the differences between the analyses, with the EC and US showing closer agreement than either of them with the UK system, is found also in the tropics and in the lower troposphere of the southern hemisphere. At upper levels in the southern hemisphere the closest agreement is between the EC and UK analyses.

The magnitude of the wind differences between the analyses in the tropical upper troposphere is larger than in the (winter) upper troposphere of the northern hemisphere. In view of the generally lower variability of the tropical atmosphere this must be considered a disappointing result. Many of the special observing systems of FGGE were designed to improve the definition of the tropical analyses. A preliminary survey of the results of the three analysis systems in the tropics suggests that the subject needs an extensive set of case studies. Some further comments on the tropical analyses are made in the summary at the end of this section.

Turning now to the day to day variability of the analysis differences, Fig. 4.1 shows the RMS differences between the EC and US analyses for the geopotential at 500 and 1000 mb and the 500 mb wind, at twelve hourly intervals; it also shows the contributions of the long and medium waves to these terms. The geopotential fluctuations show a clear diurnal variation, being larger at 00Z than at 12Z. A good deal of this variability at 1000 mb is due to extrapolation of mass information below topography.

Partitioning of the differences into wave number groups indicates that there is larger uncertainty in the analysis of the medium waves than in the analysis of the long waves. This would be even more pronounced if we had normalised the RMS differences by the climatological variance of the respective wave-number group. The contribution to the total difference by the short waves (zonal wave numbers 10 to 20) is comparable with that of the lower wave numbers, despite their lower amplitudes.

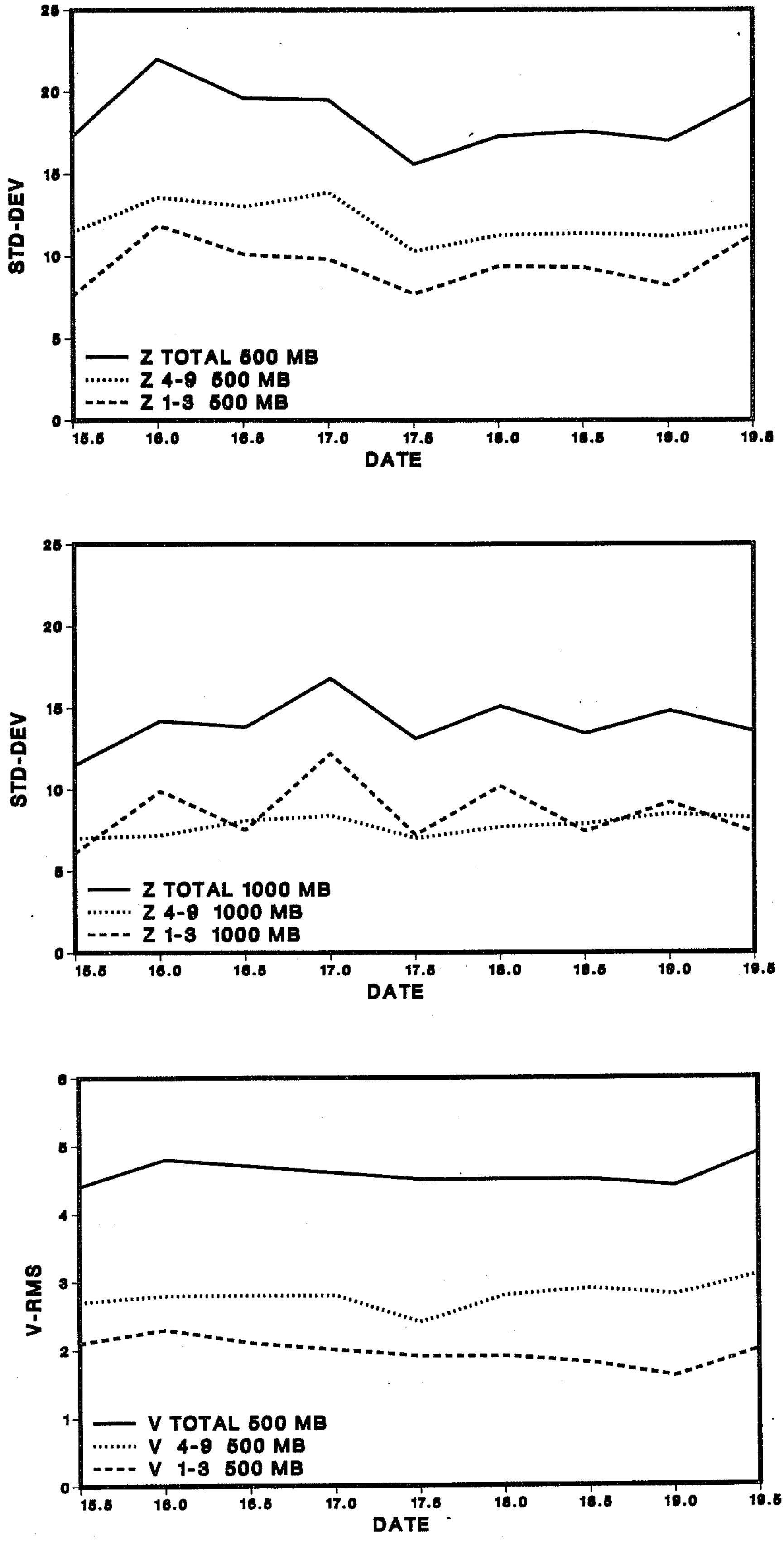


Fig. 4.1 Time series of the RMS differences between the EC and US analyses for 500 mb height (top), 1000 mb height (centre), and 500 mb wind averaged over the northern hemisphere between latitudes 20 N and 60 N. The time series run from Feb 15 12Z to Feb 19 12Z. Within each panel we show the total RMS of the field together with the contributions from the zonal wave-number groups 1-3 and 4-9.

A similar comparison between the operational analyses from NMC Washington and DWD Offenbach for February, 1976 (Arpe, 1980) gave much larger differences (30m cf. 19.4m) for the 500 mb height. Jarvenoja (1982) showed that for the winter season the differences between the (winter) ECMWF IIIb analyses and the IIIa analyses from several other operational centres varied between 17.9m and 29.7m. Secular changes in the RMS level of analysis differences averaged over seasons can arise from changes in data availability as well as from changes in analysis technique. Variations in the level of difference between pairs of analyses using the same observational data are wholly attributable to the differences in analysis method.

We conclude this general overview with a discussion of the magnitudes of the analysis, initialisation and net increments in the EC and US systems. As background for this discussion it is necessary to consider Table 4.2 which provides some statistics on the accuracy of the respective six-hour forecasts. It shows verifications of these forecasts against observations during the period of interest. There is some bias in these results, as the verification is against those data accepted by the EC system, but this is not thought to be serious. The results in Table 4.2 are a small sub-set of those available, but they are typical in suggesting that when verified against most observation types, the EC first guess is more accurate than the corresponding US field. This implies that, all things else being equal, the analysis increments should be slightly smaller in the EC analyses. This expectation is borne out for the height field at 1000 mb and 500 mb by the results of Table 4.1. However the wind increments of the EC system are generally somewhat larger than those of the US system, except in the southern hemisphere upper troposphere.

When we examine the net increments (the combined effect of the analysis and initialisation) in Table 4.1 we see that the net impact of the data on the EC

Instrument	Variable	EC	US
Radiosonde	Z_{500}	24.2	25.0
Radiosonde	U_{300} V_{300}	6.1 5.6	7.0 6.6
ASDAR	U_{250} V_{250}	6.5 6.0	7.2 6.7
AIREP	U_{250} V_{250}	8.0 8.0	8.4 8.2
SATEM CLEAR	$Z_{200}-Z_{300}$ $Z_{700}-Z_{850}$	1.5 1.7	1.6 1.8
Rawinsonde	$Z_{200}-Z_{300}$	2.0	2.0
Cloud wind	U_{150}	6.9	6.75
	V_{150}	6.3	6.3
	U_{850}	3.5	3.4
	V_{850}	3.4	3.4

TABLE 4.2 Selected statistics of the verification against observations of the first guess fields from the EC and US systems. The various levels for the wind verifications were chosen where the data was most abundant for that particular system. For the radiosondes we chose the 300 mb level because the winds, and the forecast errors, were largest there. The results are averages over all accepted data for all the 00Z and 12Z analyses.

system is somewhat larger than on the US system. This is an unexpected result since the EC first guess is more accurate. It implies that the EC system is more responsive to data than the US system. It is unlikely that this result can be attributed to any single feature of the analysis systems. It is not possible to measure the response of the UK system to data in this way since its algorithms are so different.

4.2 Northern Hemisphere differences

In this section we consider the geographical distribution of the differences between the analyses. Figs. 4.2 and 4.3 show the RMS differences between the three pairs of analyses for the 500 mb height field and for the 200 mb wind field respectively. As expected from Table 4.1 the differences are smallest for the EC-US pair of analyses. Not surprisingly, these two analyses show generally larger differences over the oceans than over land. There are, however, some large differences over north Africa and over the Himalayas. Over the pole there are also important differences which probably arise from a programming error in the US system.

The RMS differences between the EC and US 500mb height fields have a significant bias component as may be seen from Fig. 4.4b. The major troughs in the mean field, especially those over the Atlantic and the Mediterranean tend to be deeper in the US analyses as compared to the EC analyses. Test results, which are discussed more fully in the Appendix, indicate that differences in the vertical interpolation schemes contribute to the mean differences. In the EC system as used for the FGGE analyses, the analysed fields were interpolated back to sigma coordinates in order to resume the assimilation. Since then Talagrand (personal communication, 1981), following a suggestion by Rutherford, revised the vertical interpolation from pressure coordinates to sigma coordinates by interpolating only the analysis changes, or increments, to sigma coordinates and adding them to the sigma coordinate first guess. This preserved the model's boundary layer structure and

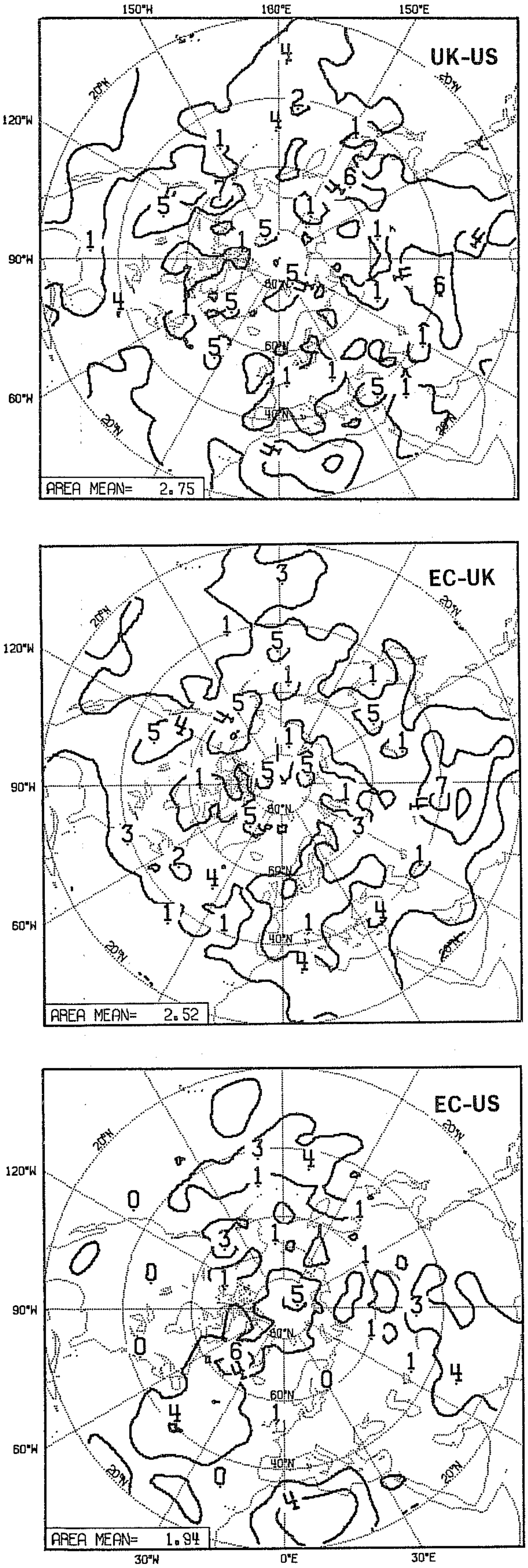


Fig 4.2

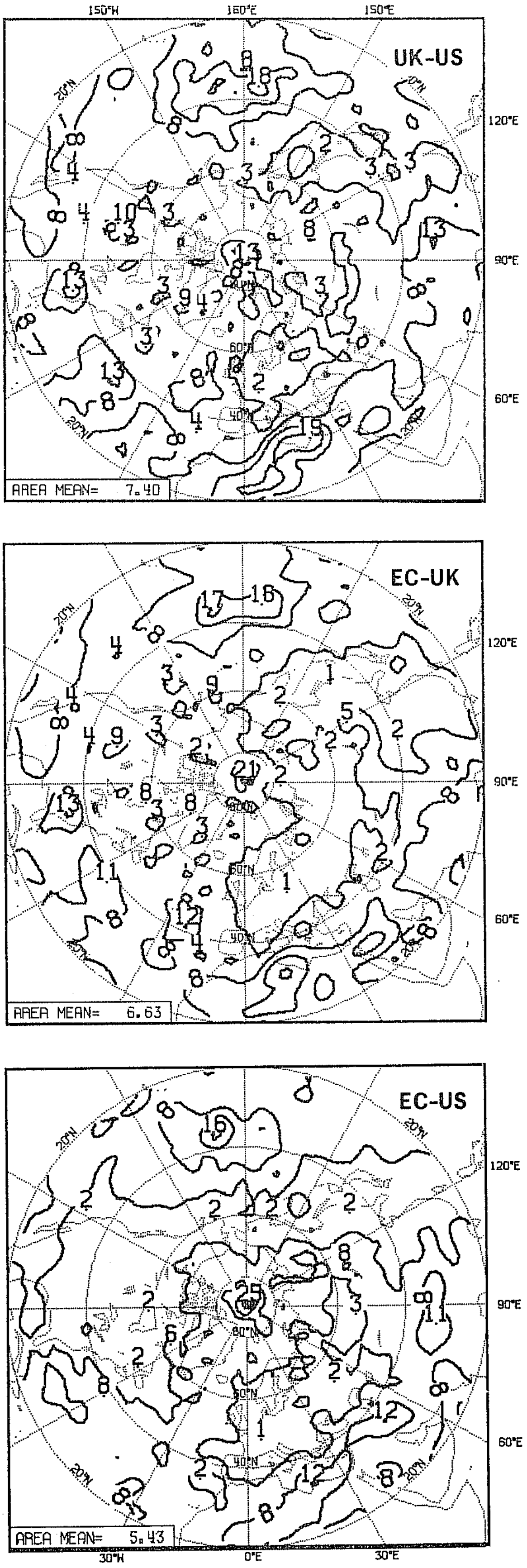


Fig. 4.3

Fig. 4.2 Plots of the RMS height differences between pairs of analyses for the Northern Hemisphere poleward of latitude 20 N: UK/US (top), EC/UK (centre) and EC/US (bottom). The area mean value is indicated on each plot. The contour interval is 2 dam.

Fig. 4.3 As Fig. 4.2 for the 200 mb wind. The contour interval is 4 m/s.

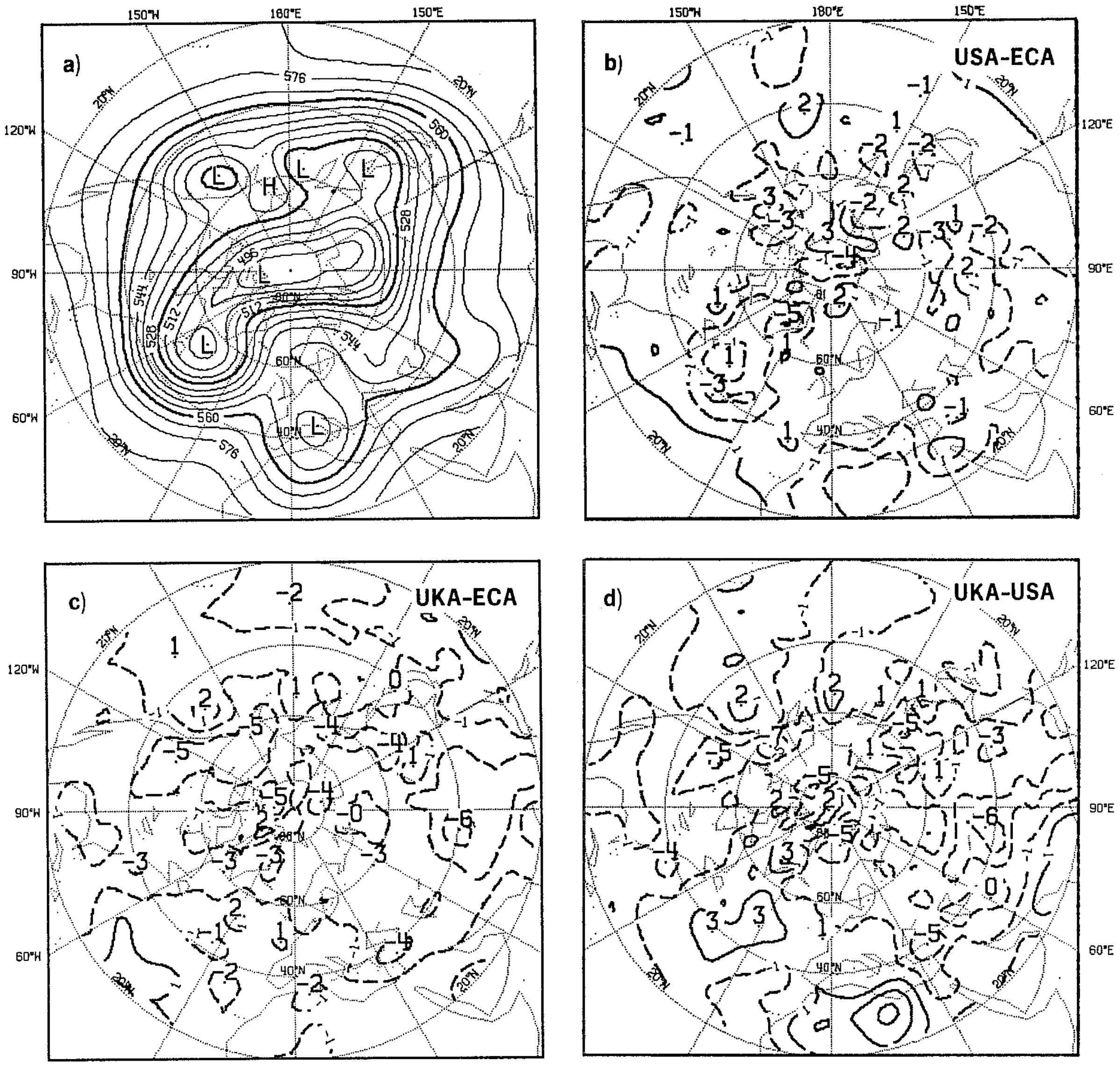


Fig. 4.4 a) The mean 500 mb field for the period from the US analysis. The contour interval is 8 dam. The remaining panels b) to d) show the differences between the mean analyses for the period. The contour interval is 2 dam with contours at ± 1 dam. The difference fields are respectively :

b) US minus EC

c) UK minus EC

d) UK minus US

Positive contours are solid, negative contours are dashed.

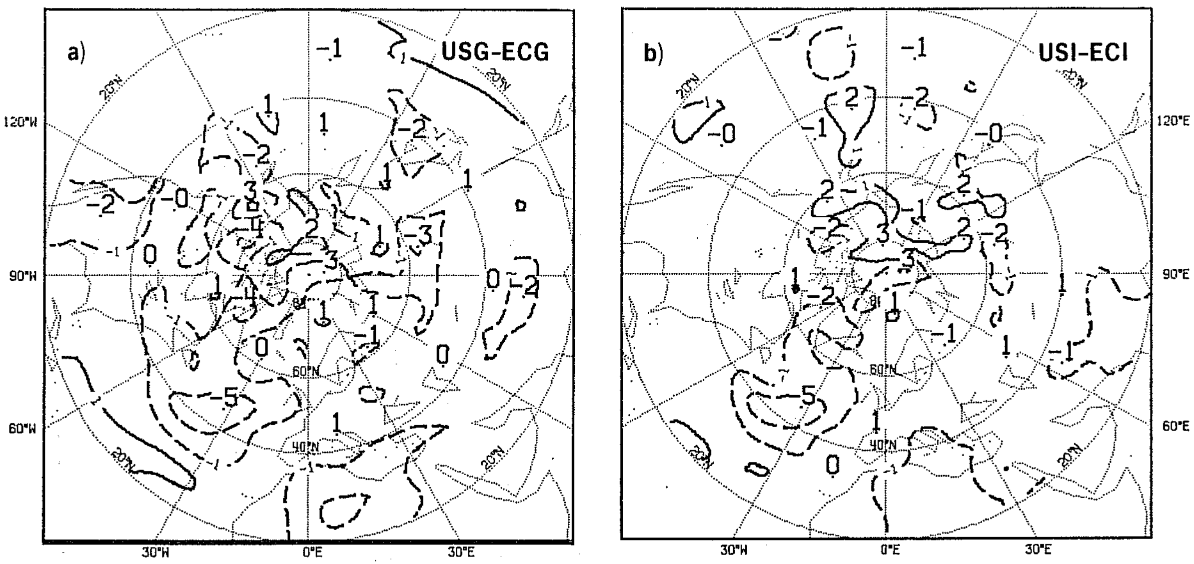


Fig. 4.5 The mean differences between the first guess fields, US minus EC (right), and the mean differences in the initialised fields, US minus EC (left). The contour interval is 2 dam with contours at ± 1 dam. Positive contours are solid, negative contours are dashed.

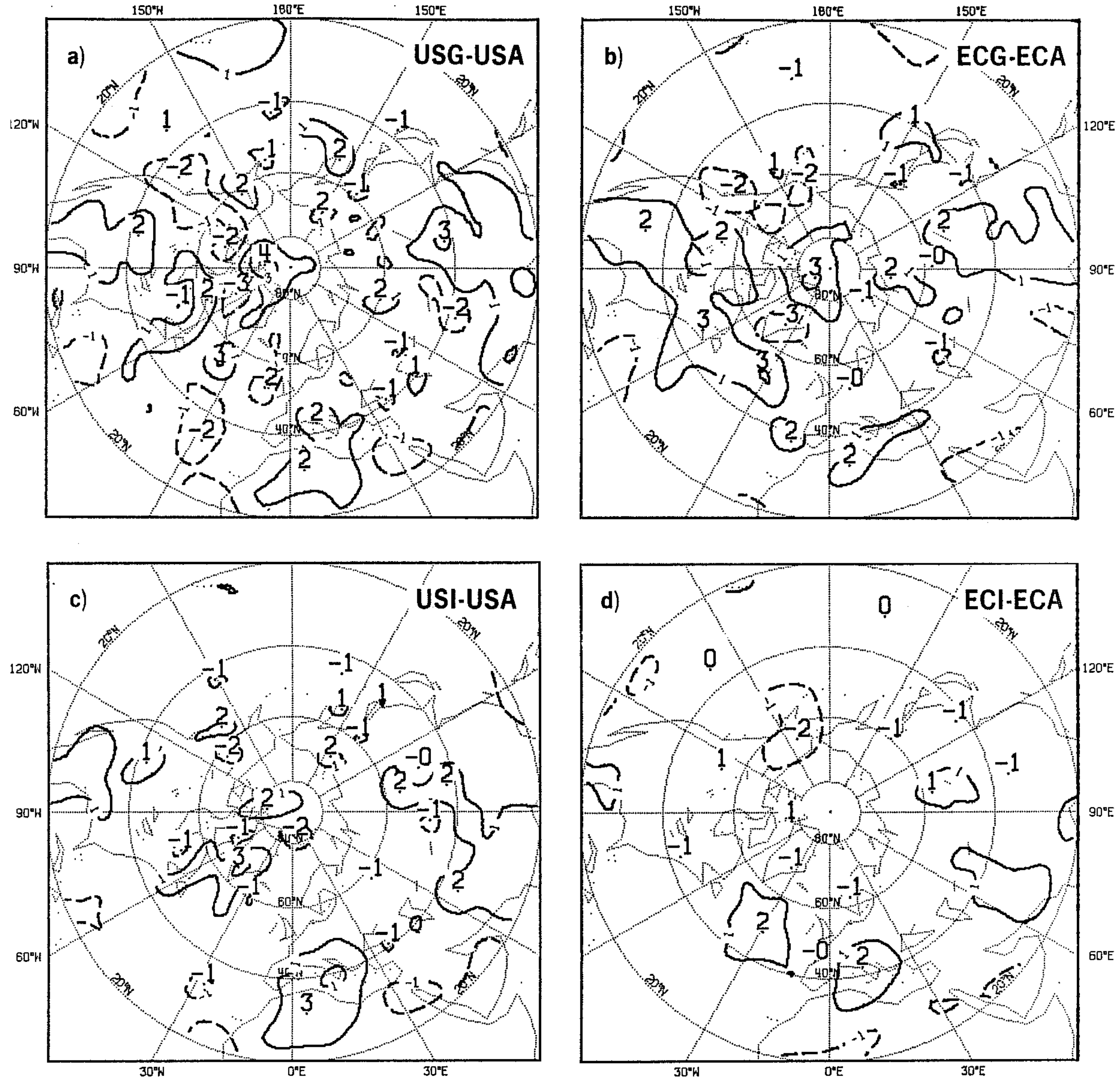


Fig. 4.6 Mean analysis and initialisation increments for the US and EC assimilations. The analysis increments are shown with negative sign.

a) Mean US analysis increment (with negative sign)

b) Mean EC analysis increment (with negative sign)

c) Mean US initialisation increment

d) Mean EC initialisation increment

The contour interval is 2 dam. Positive contours are solid, negative contours are dashed.

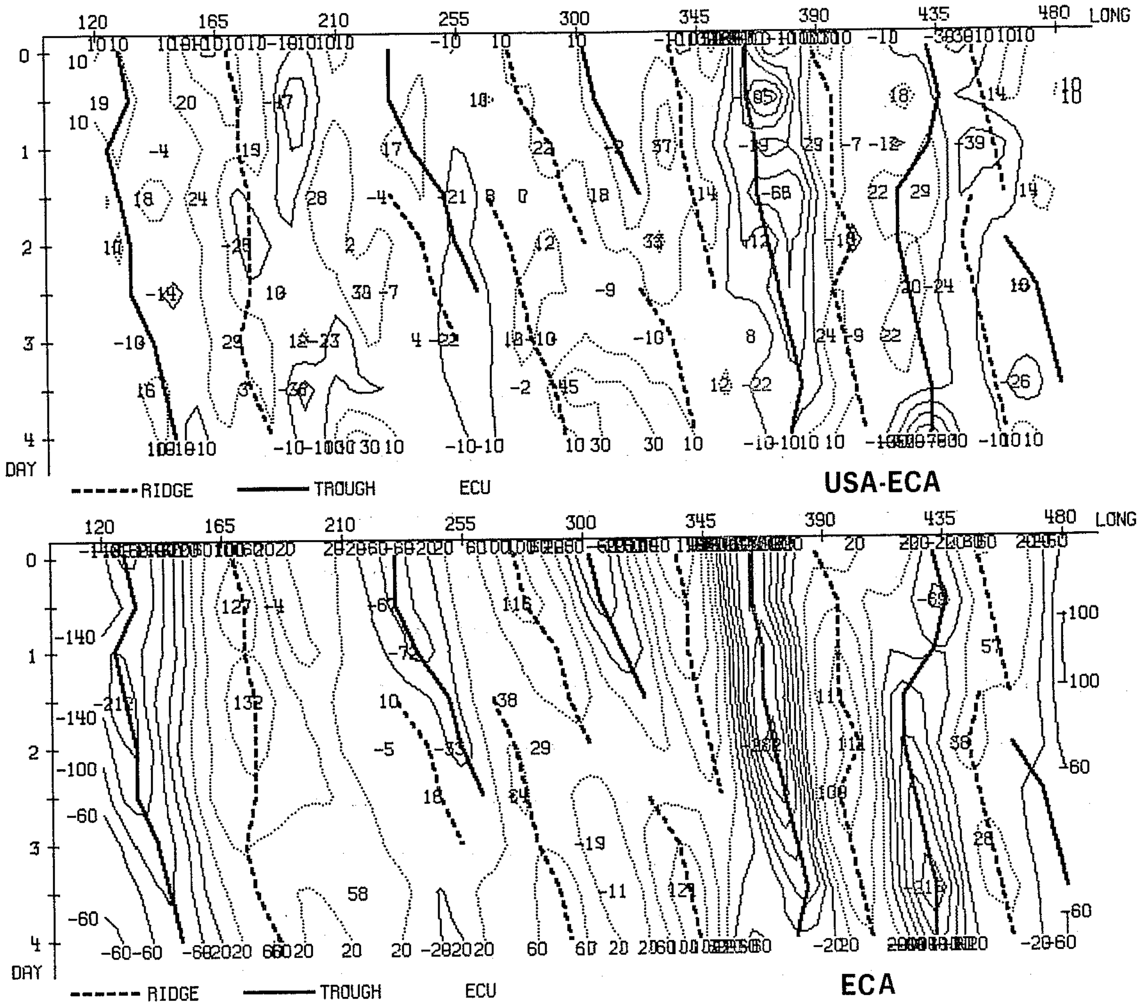


Fig. 4.7 (Bottom) Longitude time plot of the EC analysis of 500 mb height averaged over a 5 degree band of latitude centred on 30N, with the zonal mean removed. Negative contours are solid. Positive contours are dotted. The heavy solid and dashed lines indicate troughs and ridges respectively. The time axis is marked in 12 hour interval from Feb 15 12Z to Feb 19 12Z.

(Top) Corresponding plot for the US minus EC mean analysis difference along 30N. The trough and ridge lines from the bottom panel have been copied onto this panel.

considerably reduced the spin-up time for the physical processes. The US analysis results discussed here used this procedure. The results discussed in the Appendix concern a rerun of the analyses with a more recent version of the EC operational system which incorporates Talagrand's procedure for interpolation of increments. The results indicate a significant reduction in the magnitude of the mean differences between the EC and US analyses. However the difference in vertical interpolation method does not account for all the differences.

If we compare the mean differences between the first-guess fields, the analysed fields, and the initialised fields for the EC and US systems (Figs. 4.4b, 4.5) we see that the mean differences in the first guess and initialised fields over the Atlantic are significantly larger (and very similar to each other) than the mean differences in the analysed fields. This implies that the data insertion at the analysis step reduces the first guess differences but that this effect is largely cancelled by one or other of the initialisation procedures. Fig. 4.6 shows the mean analysis and initialisation increments at 500 mb in the two systems. Over the Atlantic the mean initialisation increment in the EC system tends to cancel the mean analysis height increment, which is not the case for the US system. However, one can see that, overall, more of the mean 500 mb height increment tends to be cancelled by the mean initialisation increment in the US system than in the EC system. For example, over north Africa the mean US initialisation increment tends to over-cancel the mean analysis increment while a marked cancellation effect is also evident in the US cycle over the Himalayas. The relative magnitudes of the analysis and initialisation increments is an important consideration for the overall efficiency of an assimilation system (Hollingsworth and Arpe, 1982).

The main remaining areas of difference between the EC and US analyses are over the Sahara and the central Pacific. Fig. 4.7 shows trough-ridge diagrams of the EC analysis and of the difference between it and the US analysis, averaged over a five degree strip centred along 30N. The largest differences at this latitude occur over north Africa with lesser differences over the the mid-Pacific.

In the Sahara there is a clear diurnal period to the differences in the analyses with larger differences at 00Z. This is consistent with the diurnal variation of radiosonde availability in the area , with reduced coverage at 00Z. The fact that EC used SATEMs over land while the US did not may also contribute. During the period in question the SATEM data in the area tended to be clear path retrievals at 12Z and microwave retrievals at 00Z. The differences between the analyses in this area are further discussed in a case study in Sect. 5.1.

There is a clearly defined mean difference between the analyses of the ridge near the date-line. The US analysis is higher,and the phase of its ridge is further west than the ridge in the EC analysis. There are also clearly marked intermittent increases in the differences in mid-Pacific. This may seem surprising since they occur in the vicinity of a radiosonde station at Midway Island (28N, 177W). They arise because of differences in the treatment of aircraft data and SATEM data in the vicinity of the radiosonde report. A case study in Sect. 5.2 discusses the analysis in the mid-Pacific at a somewhat more northern latitude.

Space does not permit a thorough discussion of the the mean analysis and initialisation increments in the height and wind fields at other levels. However it is worthwhile to mention one result (cf. Arpe 1983) concerning the mean increments in the 850 mb wind field in the vicinity of the Himalayas. In both analyses there is the strong suggestion that the flow in the first guess is too strong on the northern slopes of the Tien-Shan and Altai mountains. Both analysis systems reduce the strength of the low-level westerlies in this

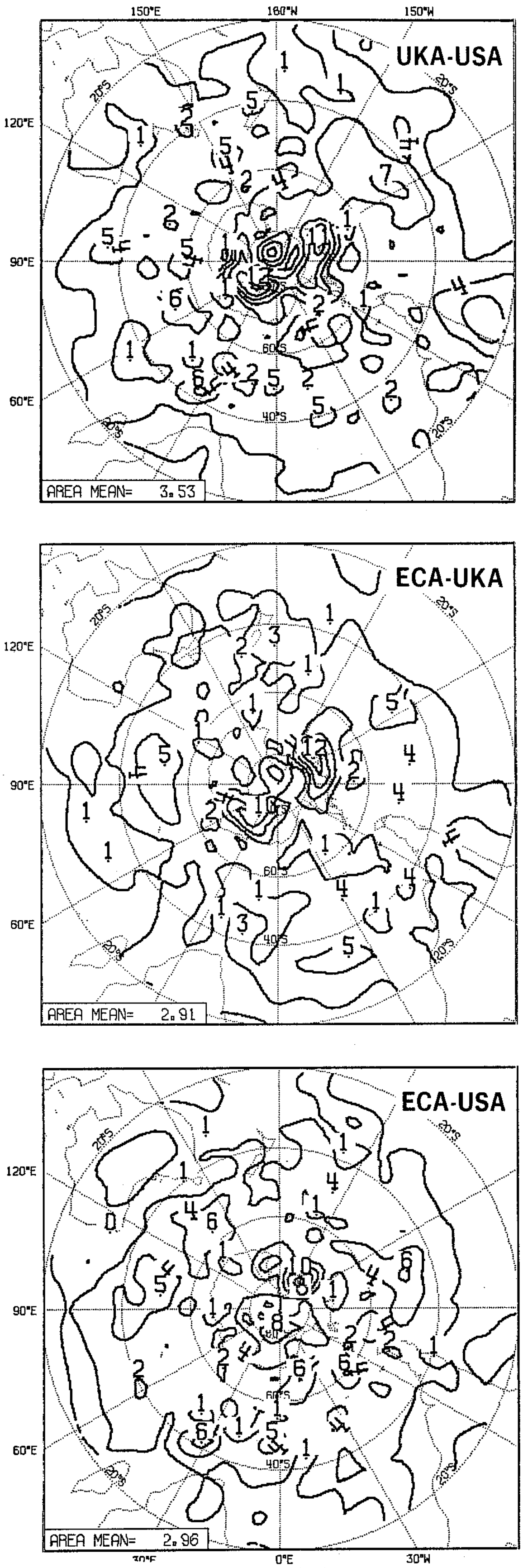


Fig. 4.8.

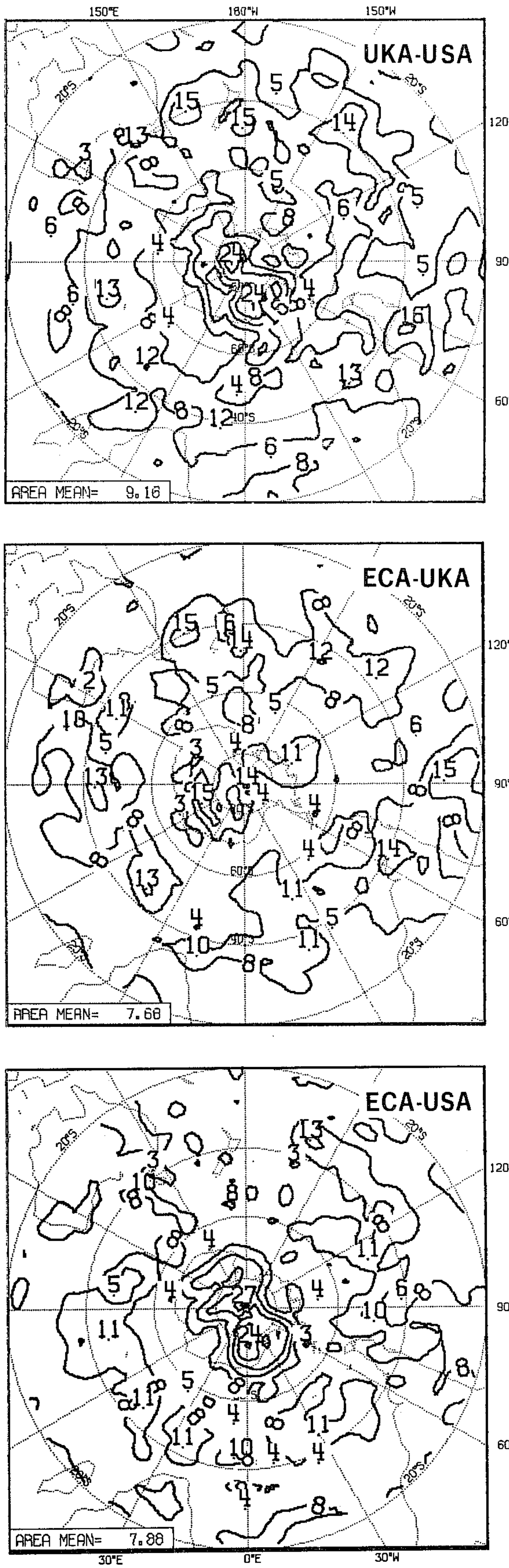


Fig. 4.9

Fig. 4.8 Plots of the RMS 500 mb height differences between pairs of analyses for the southern hemisphere poleward of latitude 20 S; UK/US (top), EC/UK (centre) and EC/US (bottom). The area mean value is indicated on each plot. The contour interval is 2 dam.

Fig. 4.9 As Fig. 4.8 for the 200 mb wind. The contour interval 4/ms.

area, but both initialisation procedures tend to restore the first-guess values. This is an indication that the representation of topography in the models is inadequate, as recently suggested by Wallace, Simmons and Tibaldi (1983). In both first-guesses there tends to be a barotropic and negative error in the geopotential on the upwind side of the main Himalayan massif, with a positive barotropic error on the down-wind side. Arpe (1983) has shown that this typical model error can be linked through the mountain torque to systematic errors in the zonal mean winds.

If we turn now to comparisons between the UK system and the other two systems (Fig. 4.2) we see that at 500 mb the RMS differences are as large over land as over sea, which is rather surprising. If we examine the mean differences between the UK analyses and, say, the EC analyses we see that the RMS differences over land arise mainly from a bias effect. Geopotential height is not an analysis variable in the UK system, and the bias over land is caused in this system by a number of effects in the post-processing of the analyses and from the tendency of the UK model to cool in the lower troposphere, especially over land. The bias effect is significantly smaller over the oceans. In ocean areas the level of the RMS differences between the UK analyses and either of the others is of the same general level as the oceanic differences between the EC and US analyses. The large differences near the pole are due to an error in the UK system.

4.3 Southern hemisphere

As shown in Table 4.1 for the southern hemisphere, the RMS differences between the the UK and US analyses (35m at 500 mb) are significantly larger than the differences between either of these and the EC analyses (29.5m). The reasons for this are not clear, but may be due to difficulties in the analysis near the South Pole, as is suggested by Fig. 4.8. Similarly large differences are also visible in the 200 mb wind field (Fig. 4.9). Part of these differences may be attributable to decisions on whether or not to use SATEM data over

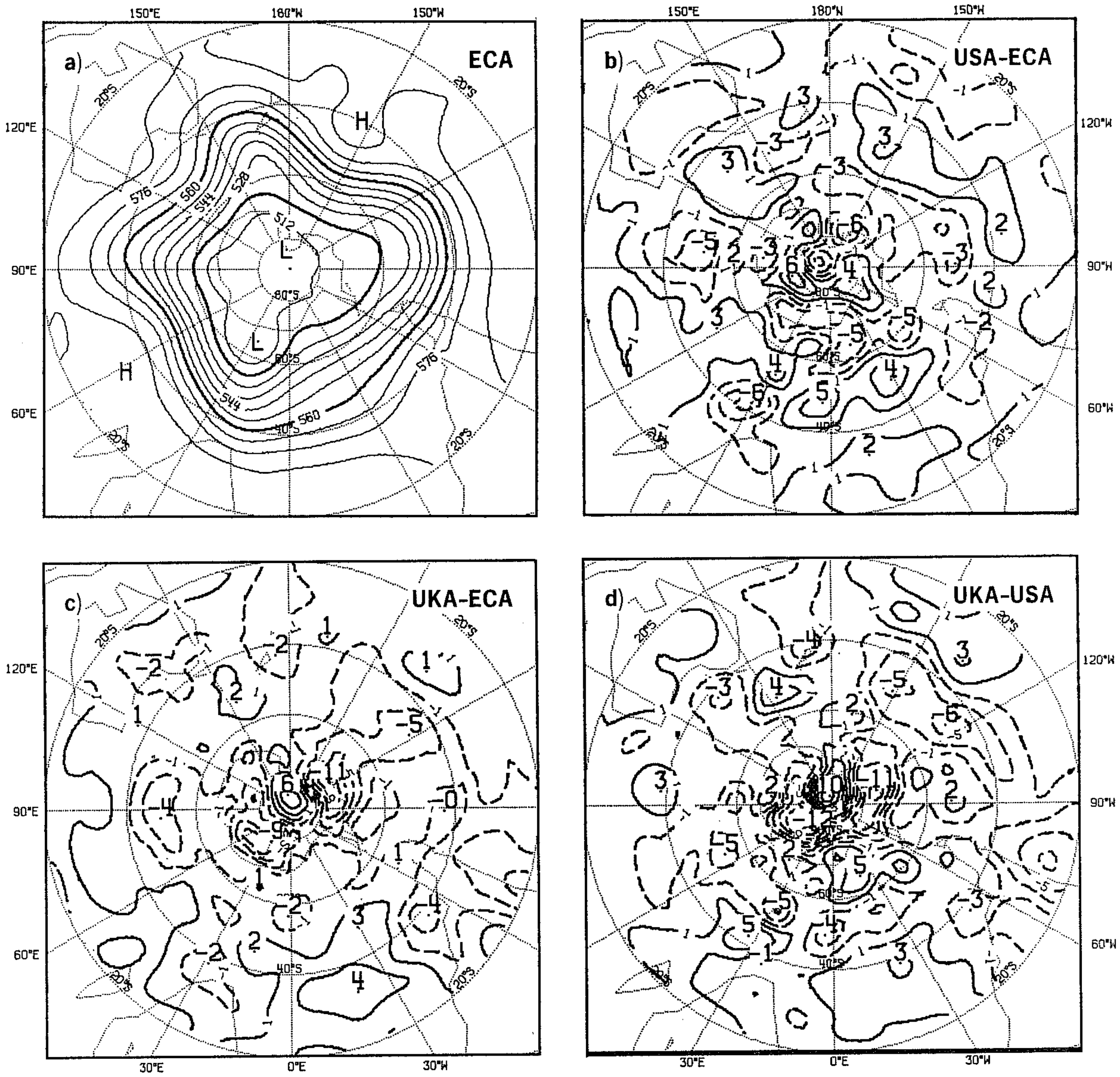


Fig 4.10 a) The mean 500 mb field for the period from the EC analysis. The contour interval is 8 dam. The remaining panels b) to d) show the differences between the mean analyses for the period. The contour interval is 2 dam with contours at +, -1 dam. The difference fields are respectively :

b) US minus EC

c) UK minus EC

d) UK minus US

Positive contours are solid, negative contours are dashed.

Antarctica, but may also be due to programming problems near the poles in the UK and US systems.

The contribution of the mean differences to the RMS differences at 500 mb is significant in this hemisphere also (Fig. 4.10). In most of the oceanic troughs we find the same kind of difference between the EC and US analyses that we have seen already in the northern hemisphere. We may presume that the difference in vertical interpolation is playing a role.

Away from the polar region it is noticeable that the mean differences between the UK and EC analyses are smaller than the corresponding differences of either of them with the US analysis. There is some suggestion that the mean differences between the US analysis and either of the others tends to be large near isolated island stations such as Isle Bouvet (specially manned for FGGE at 50S, 5E), Marion Island (45S, 38E), or Macquarie Island (45S, 150E). On the other hand there is rather little evidence to suggest noticeably large differences in the vicinity of Kerguelen Island (45S, 75E). The problem of the analysis of sparse data in the southern hemisphere is the subject a case study in Sect. 5.3. There we show that different ways of combining SATEM and radiosonde data can lead to differences in the analyses.

One of the areas where the three analyses differ most from each other in the presence of radiosonde data is along the west coast of Latin America. During this time the radiosonde at Valparaiso was reporting generally southerly winds in the middle and upper troposphere. None of the systems drew for these reports very well; they all tended to return a southwesterly. For the two systems for which a first guess was available it appears that the forecast models could not sustain the ridging upstream, and troughing downstream of the Andes that would be necessary to accommodate these reports. The US scheme drew closest to these coastal radiosonde reports while the UK scheme drew least, as suggested by Fig.4.10; the initialisation in the US system tended to cancel most of this analysis increment. Operational experience has also indicated that in certain flow configurations near the Andes the analyses will have some difficulty in accommodating the reported winds from west coast stations. The matter clearly needs further investigation.

4.4 Zonal mean statistics

Our last special topic in this section is a brief examination of some zonally averaged statistics of the analyses. Fig.4.11 shows the zonal mean zonal flow averaged over the period of study for the EC analyses together with the differences between this field and the corresponding fields for the other analyses. This presentation was chosen in order to make the differences more readily visible.

The largest differences are found in the lower stratosphere of both hemispheres, just above the sub-tropical jet streams; the differences are as large as 7m/s. The SD and bias figures for the fit of the analyses to the radiosonde reports of zonal winds at 50 mb between latitudes 30N and 30S are as follows:

	SD	bias
EC:	4.8	0.03 m/s
US:	7.2	2.81 m/s
UK:	10.4	2.92 m/s

Both the US and UK analyses show significant positive biases in the zonal components. For the US the bias of the first guess is larger again. These results on the bias of the analyses relative to the tropical radiosondes in the lower stratosphere suggest that the US and UK analyses overestimate the westerlies when averaged over the entire tropical belt.

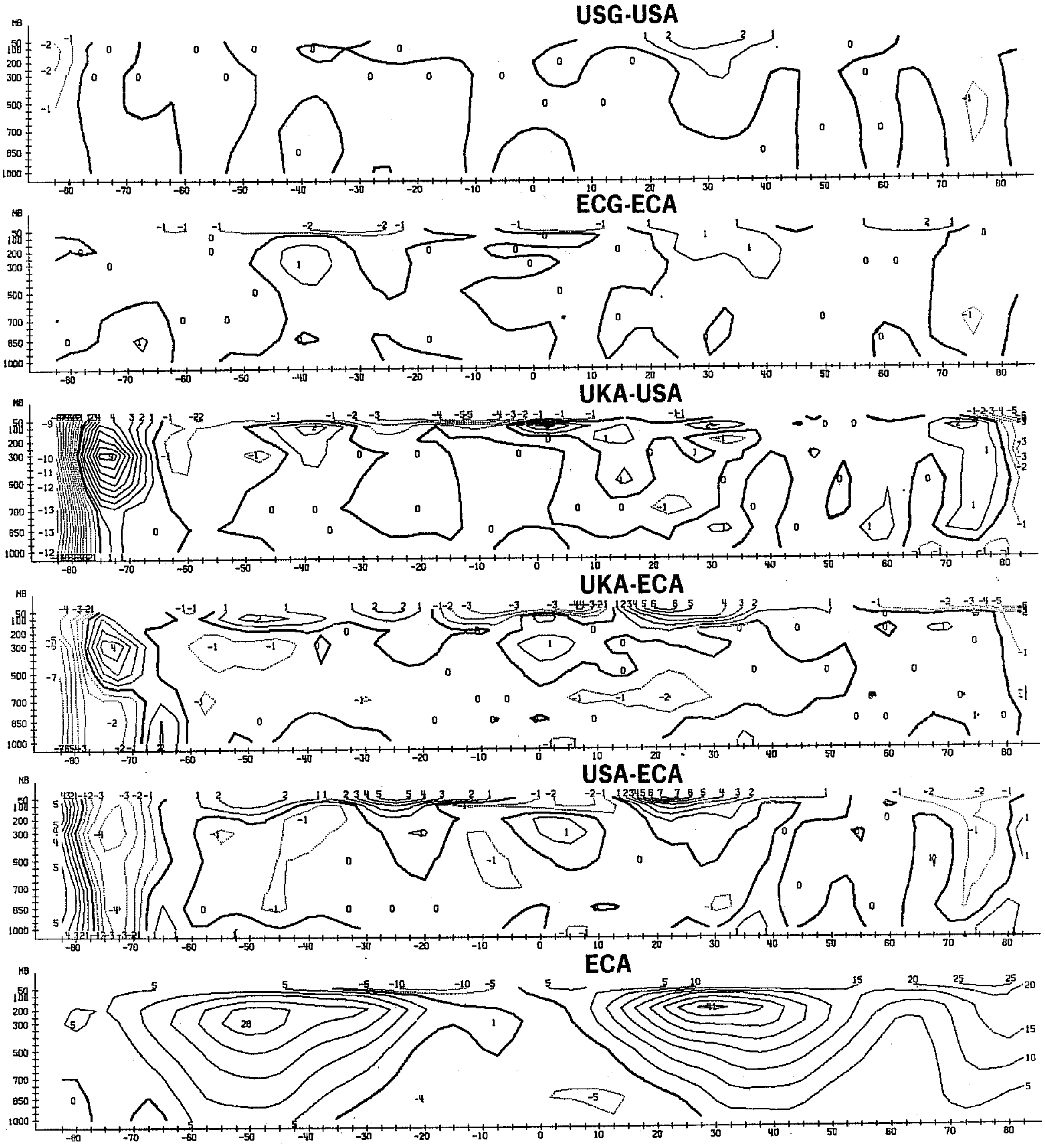


Fig. 4.11 The time and zonally averaged zonal wind as a function of latitude and height for the EC analysis (bottom); the contour interval is 5m/s. Panels a) and b) show the values of this quantity in the US and EC analysis increments (with negative sign). Panels c) to e) show the UK minus US (panel c), UK minus EC (panel d) and US-EC analysis differences; for panels a) through e) the contour interval is 1 m/s.

Differences in the zonal mean temperature are rather large near the tropical tropopause and just above it, in the lower stratosphere. Both the US and UK analyses show thickness temperatures which are up to 4K warm relative to the EC analyses in the 70-100mb and up to 4K cold relative to the EC analyses in the 50-70mb layer. Table 4.3 shows the mean differences between the reported thicknesses (expressed as temperatures) and the analysed thicknesses for both radiosondes and SATEMs for the region between 30N and 30S, averaged over all analysis times.

Table 4.3a Sondes

Level	EC	US	UK
50-70	-0.27	-1.49	48
70-100	0.73	3.83	2.80
100-200	0.21	0.4	2.16

Table 4.3b SATEMS

Level	EC	US	UK
50 70	0.50	-1.91	0.70
70-100	0.58	3.97	3.26
100-200	-1.46	0.82	83

These results suggest the differences in the zonal mean temperature arise from differing treatments of the data in the analyses. In the mean the EC system draws closest to the radiosonde temperatures and closest to the SATEMs, at least in the lower stratosphere. There is a marked meridional gradient in the differences in the zonally averaged thicknesses which suggests that the temperature differences in the tropics are geostrophically related to the differences in the zonal flow noted above. This is consistent

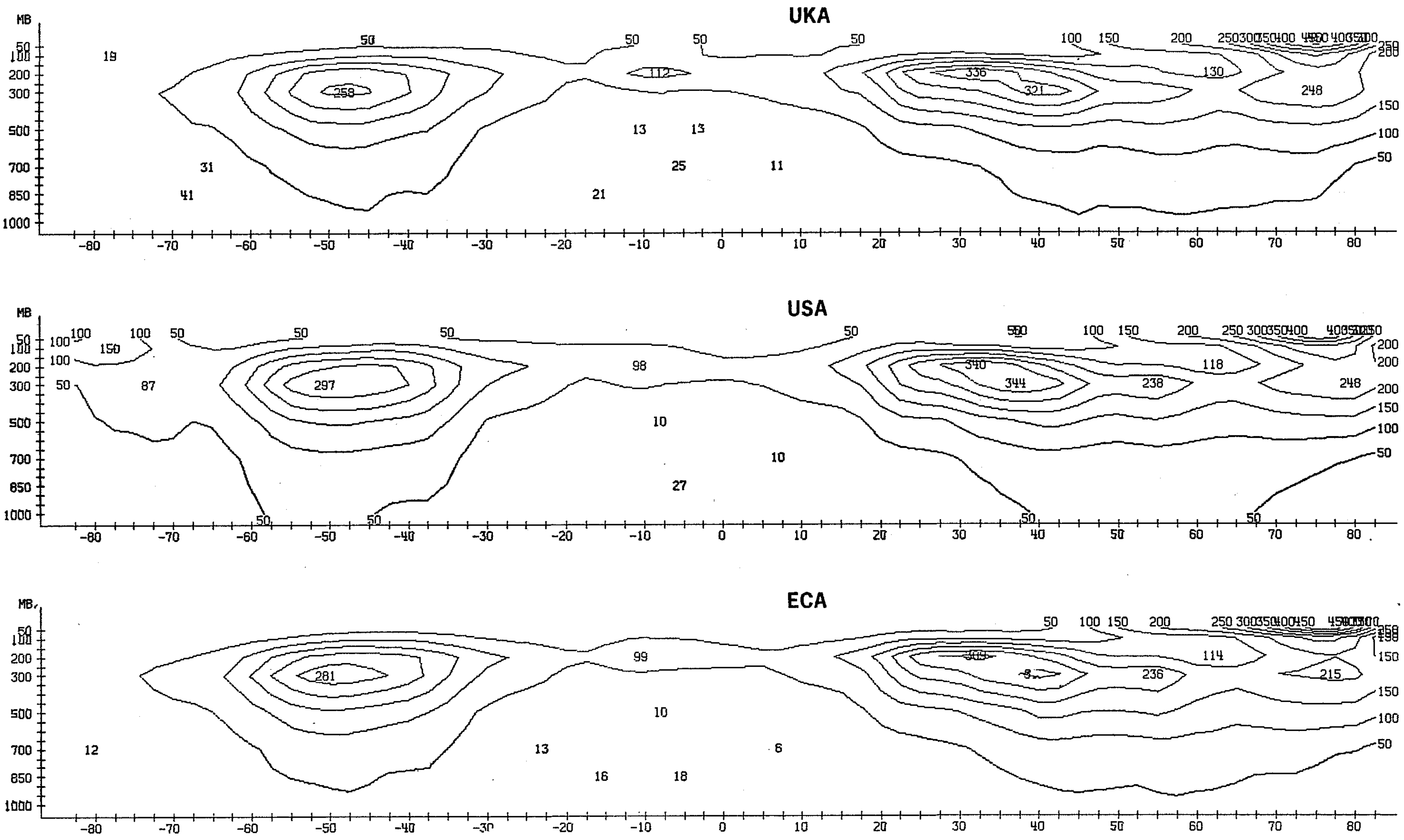


Fig. 4.12 Eddy Kinetic energy in zonal wave-numbers 1-20 for the UK, US, and EC analyses, averaged over the period.

with the fact that the zonal flow differences above the sub-tropical jet persist for over five-days in the forecasts from the respective datasets (see Part II).

For the US system the bias of the first guess relative to the observations in tha areas in question is larger than that shown in Table 4.3.

Given that all three systems have a six hour frequency for insertion of data, we may ask why these biases persist. A possible suggestion is the well known tendency of many numerical models to produce a too strong upward extension of the sub-tropical jets into the lower stratosphere. If this tendency is stronger in the US and UK models than in the EC model then it could lead to the present differences. Some support for this supposition may be found by comparing the mean increments in the zonal mean zonal flow, which are also shown in Fig.4.11. This shows that in the area in question the mean increment is a reduction of the zonal flow by over 2m/s in the US system and a smaller reduction in the EC system of between 1 and 2m/s. If the suggestion that model bias is the source of the differences in the zonal-mean thermal and zonal wind fields then it provides a further example of a bias in the analyses arising from a component of the assimilation system.

We turn our attention now to the variance of the wind analyses by considering the eddy kinetic energy for the three analyses, in Fig.4.12. In both the northern and southern hemispheres the largest values occur in the US analyses; for example the northern hemisphere US analyses show a maximum of 3440 kJ m^2 bar^{-1} compared to 3210 in the UK and 3150 in the EC analyses. The differences arise mainly in zonal wave-numbers 4 to 9. The effect of the difference in vertical interpolation on these results is unclear, but the present results are consistent with the differences we have noted already in the mean height fields. The only further suggestion to explain the differences is the fact that the EC system draws slightly closer to the cloud-drift wind data than the US system. Given the known bias towards underestimating wind-strengths in this data, it may be possible that this data can explain some of the differences in the eddy-kinetic energy.

The US analyses also show larger kinetic energies than either the EC or UK analyses at 1000 mb. This result is considered a little questionable because of the fact that where the 1000 m surface intersects the model topography we take the lowest sigma level model wind as the 1000 mb wind. Given the differences in the sigma-levels this effect will cause some bias in our calculation of the 1000 mb kinetic energy. In addition it has been found that the EC system uses the lowest level radiosonde reports in an inconsistent way which tends to reduce the intensity of the analysed wind.

The eddy momentum flux has been shown to be sensitive to the analysis method in earlier studies such as Arpe (1980), Jarvenoja (1982). For our short period of record we find reasonably good agreement between the three sets of results in the northern hemisphere, with the US system showing the highest values and the UK system the lowest. In the southern hemisphere there is good agreement between the EC and US systems while the values for the UK system are significantly lower.

4.5 Summary

The differences between the EC and US analyses are of the expected magnitude in the northern hemisphere; the magnitude of the differences agree with estimates of the accuracy of the analyses made by the analysis systems themselves. Differences between the UK analyses and the other two appear to be rather large in the height field, apparently due to a cold bias in the the UK model.

The southern hemisphere differences between the analyses are larger than the northern hemisphere differences.

The differences in the upper level tropical winds are as large as in the mid-latitudes. This is an indication that further attention needs to be devoted to the tropical analysis problem.

Various components of the assimilation system can contribute to the important mean differences between the analyses. The full-field vertical interpolation between pressure and sigma-levels seems to have had a smoothing effect on the EC analyses, while model bias dominates in the subtropical lower stratosphere of the US and UK analyses.

Because of the similarity of their general structure it was possible to make a detailed investigation of the behaviour of the main components of the EC and US systems. In general, the EC analysis system had more accurate first-guess fields, so that the analysis increments should in general be smaller in the EC system than in the US system. In fact the EC analysis increments tend to be smaller in the mass field and of comparable magnitude in the wind field. More of this information seems to be retained after initialisation by the EC system.

5. CASE STUDIES

In comparing the analyses in this study one invariably finds, somewhere on the charts, rather large local differences in the values for each and every analysed variable. In studying these differences our first approach was to identify the origin of the larger differences, and to try to relate them to differences in the analysis systems. This was the motivation for our first case study (Sect. 5.1) which discusses the differences in the analyses for an important depression over the Mediterranean which produced heavy and destructive rains over the Balkans. However, tests with the forecast models soon showed that in this case the rather large analysis differences , though important for the short range (one-day) forecasts had only a modest effect on the medium range forecasts.

The second case study (Sect.5.2) was chosen because analysis differences in the mid-Pacific led to very large differences in the forecasts, with all three forecast models responding in the same way to the analysis differences. In this case a crucial quality control decision led to an incorrect baroclinic development in forecasts with all the models from one analysis.

The third case (Sect.5.3) is a southern hemisphere case. The data coverage in this case is significantly sparser than in the first two cases. This case serves to highlight some of the difficulties involved in the analysis of sparse observations. The differences in the analyses again led to consistent differences in the forecasts from all three models.

The fourth case (Sect.5.4) is concerned with the later evolution of the forecasts for an intense storm in the western Atlantic known as the President's day storm, which has been the subject of a number of other studies. Forecasts made with the same model from different analyses, though agreeing quite well out to day 4, show a rapid divergence thereafter. We introduce an experimental technique which we call a transplant to show that this divergence of the medium range forecasts is due to a rapid downstream propagation and growth of analysis differences which were initially a long distance upstream of the relevant weather system. This experiment demonstrates in a dramatic fashion the impact of the analysis on forecast performance; and this impact is firmly based on theoretical as well as on experimental grounds.

5.1 Mediterranean 00GMT 17th February: A complex situation, illustrating quality control, internal constraints and resolution.

The complex Mediterranean cut-off low to be discussed here was the most active system in the northern hemisphere on Feb 17. There was good data

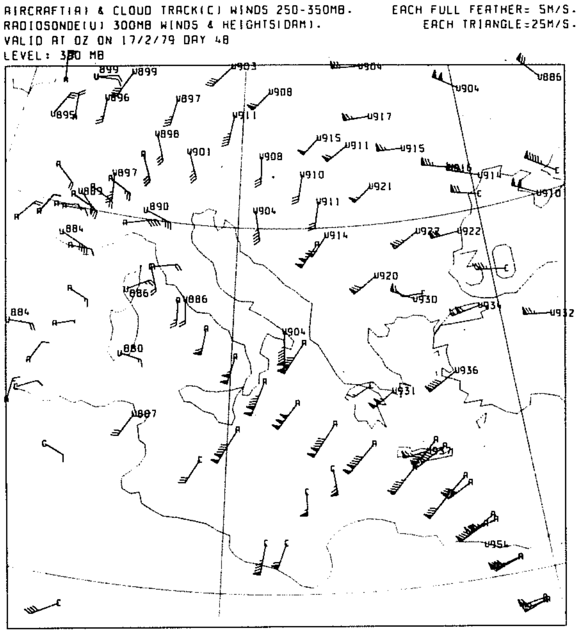
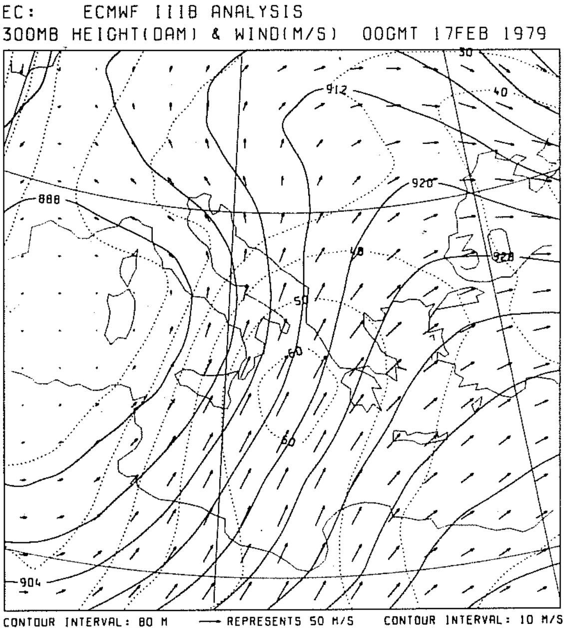
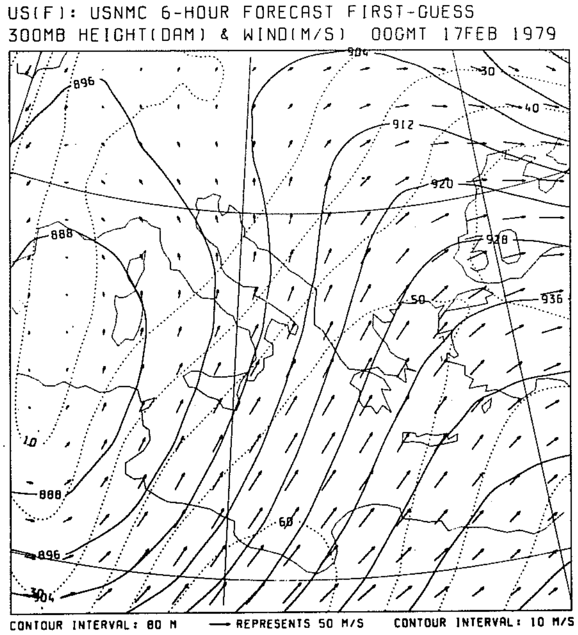

AIRCRAFT(A) & CLOUD TRACK(C) WINDS 250-350MB. EACH FULL FEATHER= 5M/S.
RADIOSONDE(U) 300MB WINDS & HEIGHTS(DAM). EACH TRIANGLE=25M/S.
VALID AT 0Z ON 17/2/79 DAY 48
LEVEL: 300 MB

Fig. 5.1.1


EC: ECMWF IIIB ANALYSIS
300MB HEIGHT(DAM) & WIND(M/S) 00GMT 17FEB 1979

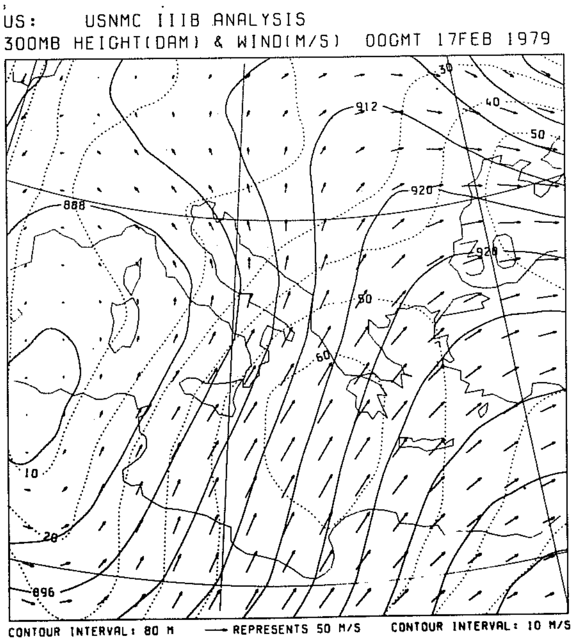
CONTOUR INTERVAL: 80 M —— REPRESENTS 50 M/S CONTOUR INTERVAL: 10 M/S

Fig. 5.1.2


US(F): USNMC 6-HOUR FORECAST FIRST-GUESS
300MB HEIGHT(DAM) & WIND(M/S) 00GMT 17FEB 1979

CONTOUR INTERVAL: 80 M —— REPRESENTS 50 M/S CONTOUR INTERVAL: 10 M/S

Fig. 5.1.3


US: USNMC IIIB ANALYSIS
300MB HEIGHT(DAM) & WIND(M/S) 00GMT 17FEB 1979

CONTOUR INTERVAL: 80 M —— REPRESENTS 50 M/S CONTOUR INTERVAL: 10 M/S

Fig. 5.1.4

Fig. 5.1.1 Observational data over the Central Mediterranean in the six-hour period centred on 00Z, Feb 17 1979, between 250 and 350 mb. Radiosonde winds and heights are identified by 'U' at the observation position, aircraft winds by 'A'; and cloud winds by 'C'. The heights are given in dam; the winds are plotted in the usual meteorological convention: a full barb is 5 m/s, a half barb is 2.5 m/s, and a solid feather is 25 m/s.

Fig. 5.1.2 The EC analysis at 300 mb over the Central Mediterranean at 00Z on Feb 17. Solid lines are geopotential contours (interval 8 dam), dotted lines are isotachs (interval 10 m/s), while arrows indicate wind direction and strength. The conversion from potential to kinetic energy may be subjectively judged from the angle between the wind vectors and the geopotential contours, and compared with the acceleration shown by the isotach gradient along the flow.

Fig. 5.1.3 As Figure 5.1.2 for the US first -guess.

Fig. 5.1.4 As Figure 5.1.2 for the US analysis.

coverage near the surface and at jet level. Nevertheless there were significant differences in the analyses at the surface and particularly at upper levels. The 300 mb analyses for this case illustrate the effects of differing approaches to mass-wind balance, data selection, quality control, and resolution, as discussed in Sect.2.4.

Although the analysis differences were rather large and affected the short range forecasts, they did not have a marked effect on the medium range forecasts, perhaps because the situation over the Mediterranean was rather stationary.

At the surface there was a cold front extending from the Adriatic southwards into North Africa, with strong warm advection and a sharp upper ridge ahead of it, a cut-off low aloft, and surface lows in the cold air over the Western Mediterranean and North Africa. This active, rather complex situation gave heavy rains, which caused extensive damage in the Balkans.

Most of the region has a good coverage of observations; those near 300 mb are shown in Fig.5.1.1. Within each observation type there is quite good consistency between nearby observations, although the sharp ridge and troughs cause some large horizontal shears. However, between different observation types there is less good agreement. The cloud motion winds give generally lower speeds than other types; they are occasionally extremely low such as those over Greece. The cloud wind directions over the Mediterranean and North Africa in the lower centre of Fig.5.1.1 are southerly, in good agreement with Brindisi radiosonde in the centre of the figure, but in disagreement with the south-westerly aircraft reports which occur between the Brindisi sonde and the cloud-wind reports.

The EC heights and winds, (Fig.5.1.2) are close to geostrophic balance, with sub-geostrophic wind speeds in the troughs, super-geostrophic wind speeds in

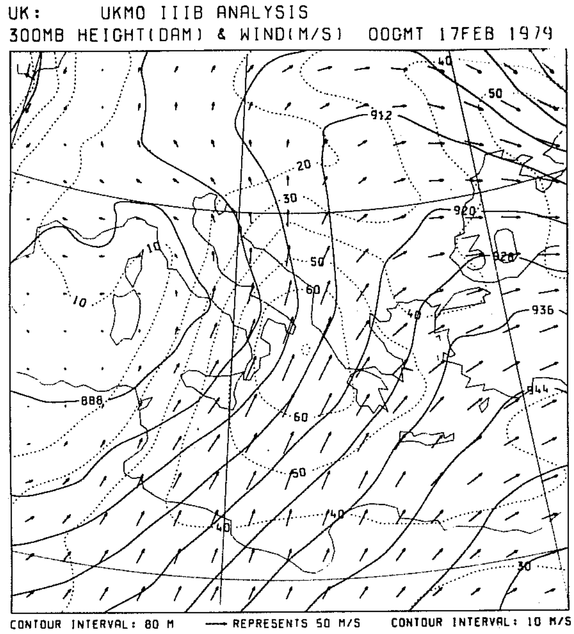


Fig. 5.1.5 As Fig. 5.1.2 for the UK analysis.

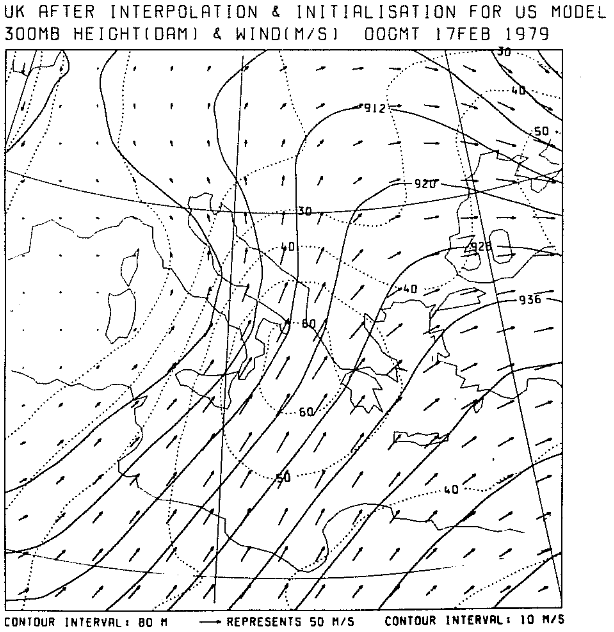


Fig. 5.1.6 As Figure 5.1.2 for the UK analysis after interpolation and initialisation for the US model.

the ridge, and slight cross-contour flow in the regions of acceleration and deceleration, all of which appear dynamically consistent with the slow manifold concept that the patterns shown change rather slowly.
Considering its geostrophic constraint , and moderate smoothing, the EC system has drawn quite closely to all the data, except for the extremely weak cloud winds. The RMS fits of the analyses to the observations plotted in this figure are given in Table 5.1.

The US six-hour forecast first-guess (Fig.5.1.3) did not predict a wind maximum over the Mediterranean, and hence the US system rejected the strong winds at Brindisi, as well as many of the cloud motion winds. The US first-guess has a dynamically consistent up-gradient flow in the decelerating jet over the Mediterranean. The US analysis (Fig.5.1.4) between Italy and Libya does show a wind maximum, which is rather smoother than EC's, and with a sharper trough to the west. Since the analysis technique is designed to make changes to the first-guess which are close to geostrophic balance, it follows that the up-gradient flow of the first-guess is retained in the US analysis. Consequently the jet maximum in the US analysis is decelerating strongly, which appears to be dynamically inconsistent. The initialisation corrects this by relaxing the geopotential trough closer to its value in the first-guess field.

The UK system has no explicit geostrophic relationship and its analysis (Fig.5.1.5) is even further from geostrophic balance. It has attempted to fit closely all the wind data, including the strong wind at Brindisi and the weaker winds to the south and north. The large accelerations and decelerations this implies are dynamically consistent with the analysed large cross contour flow in these regions. Interpolating the UK analysis for the US model, which involves some smoothing, and applying the US initialisation gives Fig.5.1.6. The trough over Brindisi which in the UK analysis fits the Brindisi observation of 904 dam, has been slackened by 8 dam, and the

Analysis field	Observation type & symbol		Number	EC	US	UK	Units
300mb geopotential	Radiosonde	U	46	22.	24.	29.	mb
300mb wind	Radiosonde	U	46	10.	13.	9.	m/s
300mb wind	Aircraft	A	47	10.	10.	9.	m/s
300mb wind	Cloud motion	C	14	24.	29.	22.	m/s
Sea level pressure	Land	L	61	3.5	4.0	43	mb
Sea level pressure	Ship	S	21	4.6	4.8	44	mb

TABLE 5.1 Verification against the plotted observations of the EC, US, and UK analyses shown in Figs. 5.1.2, 5.1.4, and 5.1.5 at 300 mb, and Figs. 5.1.8 to 5.1.10 at 1000mb.

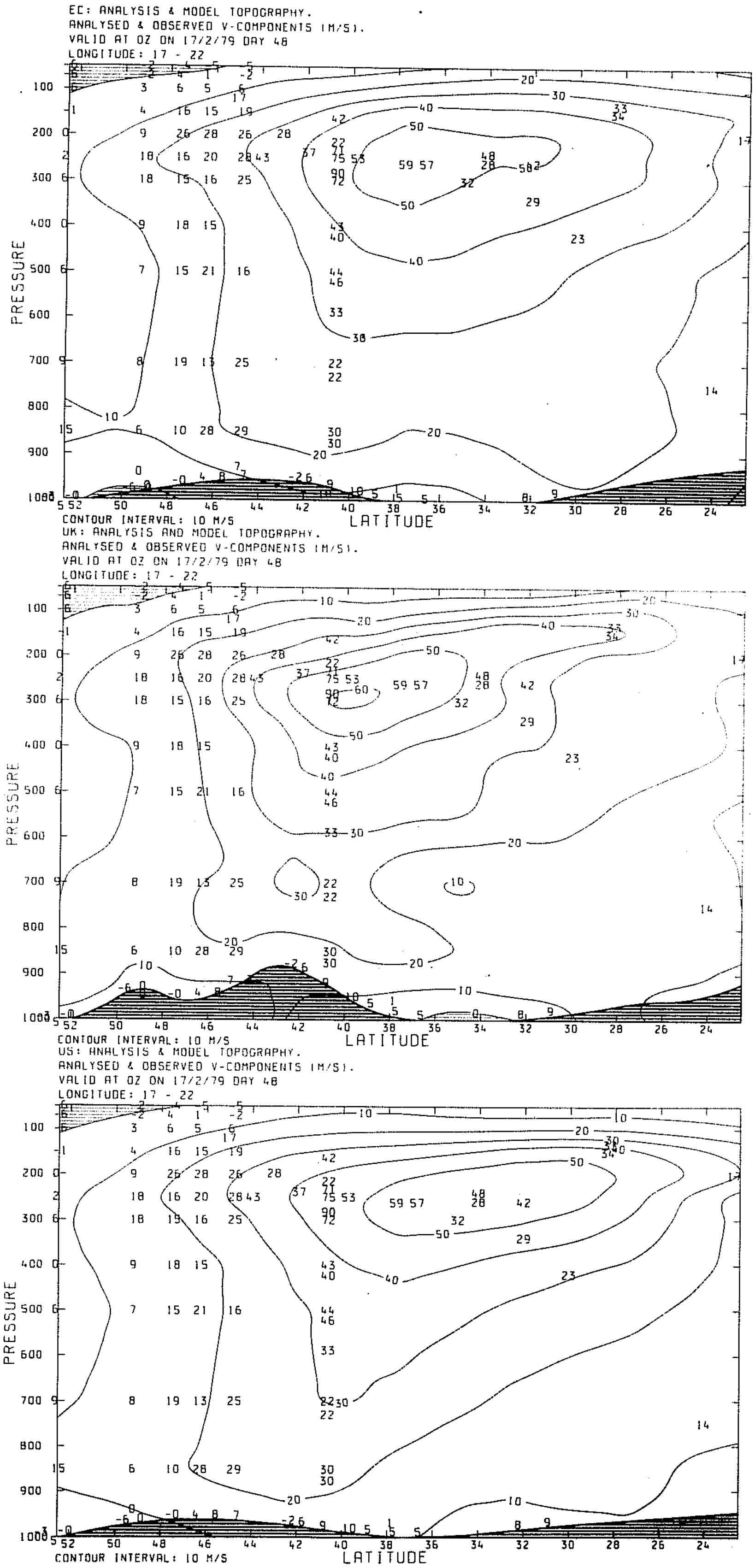


Fig. 5.1.7 Cross-sections of meridional wind component, averaged between longitudes 17E and 22E for the EC (top), UK (centre) and US (bottom) analyses at 00Z on Feb. 17 1979. The contour interval is 10 m/s. Also plotted are the relevant observational data. Radiosonde data is identifiable by the fact that it is plotted in columns.

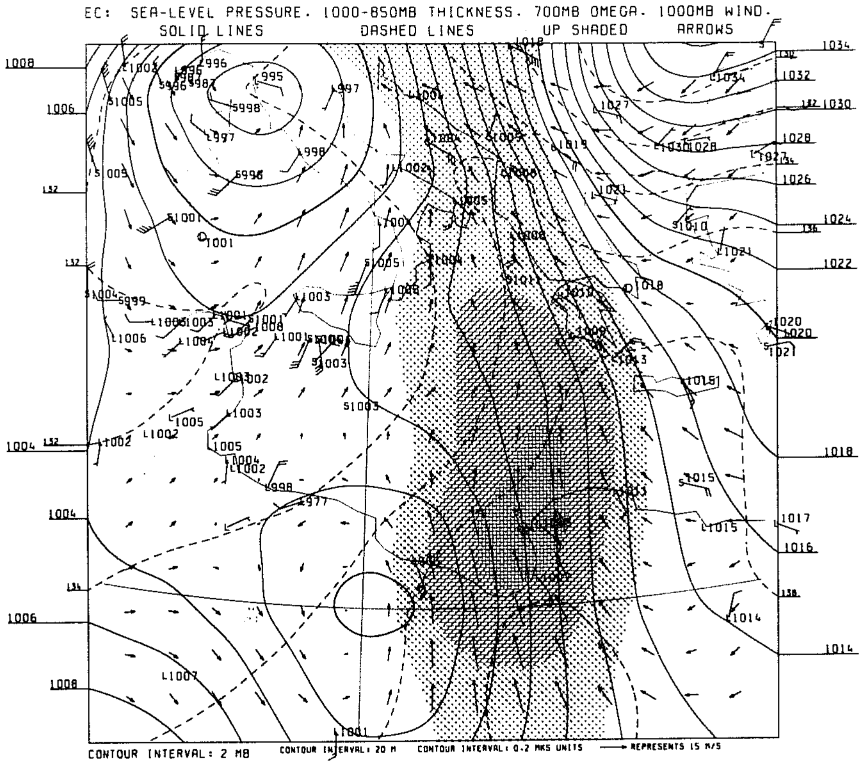


Fig. 5.1.8 The low-level EC analysis over the central Mediterranean for 00Z February 17 1979. Solid lines show sea-level pressure (contour interval 2 mb), dashed lines show the 1000-850 mb thickness (contour interval 20 m or 4.2 K), and the gradations of hatching show the 700 mb vertical velocity with a contour interval of .2 Pa/s (approximately 2 cm/sec). Observations of mean sea-level pressure and surface wind from land (L) and ships (S) are shown in the usual way.

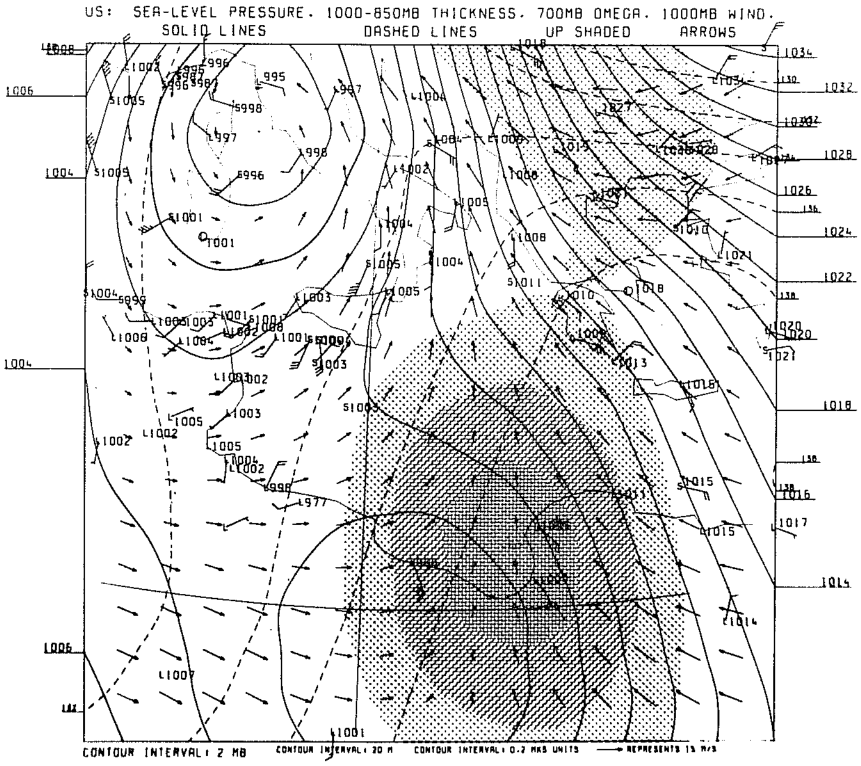


Fig. 5.1.9 As Figure 5.1.8 for the US analysis.

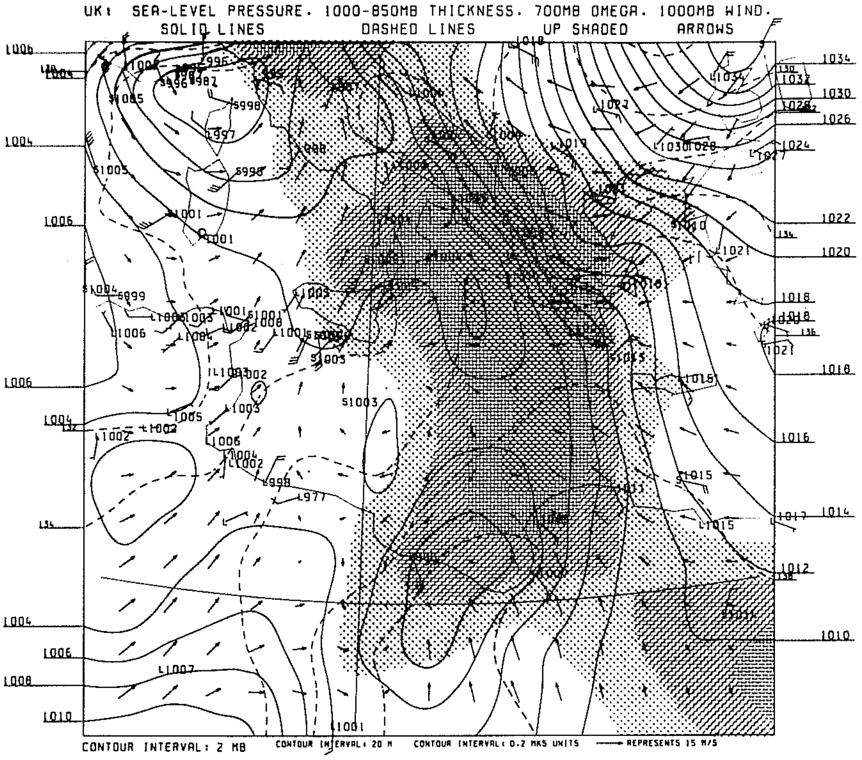


Fig. 5.1.10 As Figure 5.1.8 for the UK analysis.

accelerations, decelerations and cross-contour flow have been reduced. However, apart from these, and the eye catching effect of the smoothing, the initialisation has not altered any of the main features.

Some of the apparent inconsistency between aircraft winds near Brindisi and other data might be due to a real vertical shear, the aircraft being near 250 mb. A strong E-W thermal gradient in the troposphere across the cold front, and a steeply sloping tropopause, are geostrophically consistent with a sharp maximum in southerly wind component just under the tropopause at Brindisi. The UK system analyses such a pattern in the vertical while the other analyses are much smoother in the vertical. However the wind shears reported at Brindisi are impossibly large (52 m/s in 2 mb) and there are no other observations showing this detail, so it is difficult to judge which analysis is more nearly correct.

It is of more relevance to large scale numerical weather prediction to consider analyses which are averaged sufficiently to remove scales which the forecast models cannot accurately handle. Fig.5.1.7 shows North-South cross sections of the v-component averaged from 17 E to 22 E and the southerly wind observations in this band. Nearly all the observed levels are shown for Brindisi at 41 N, some slightly shifted to avoid overplotting; other radiosondes only have standard levels plotted. The US system has rejected the cloud winds between 30 N and 36 N and produced an analysis which is very smooth horizontally. The EC system has given them some weight, and produced an analysis which is rather smooth horizontally and vertically, while the UK analyses fit them closely, with rougher fields.

The upper flow apparently affects the surface analyses, shown in Figs.5.1.8 to 5.1.10. The UK system (Fig.5.1.10) has analysed a maximum upward motion at 700 mb of .95 Pa/sec, associated with the very sharp jet entrance and enhanced by the UK model's steep "envelope" topography in this region. The surface winds are strongly convergent into a small surface low to balance this. Such a feature is meteorologically feasible, but no surface observations support it in this case, and it is below the scale which the model can accurately predict, being part of a two grid-length wave in the surface pressure field. (This feature is not resolvable by the US model, so after interpolation and initialisation for this model the UK analysis has apparently a very different character). On the other hand, some aspects of the surface wind field, such as the deflection round the high ground to the east of the trough, are verified by the land surface observations plotted, though the UK system does not use land surface wind data.

The US analysis (Fig. 5.1.9), associated with the greater southward extent of its jet, has upward motion farther south over the north African coast with maximum .71 Pa/sec. Its ability to resolve the small scale features in this situation is limited, and the analysis is much smoother. In response to the observations of 996 mb and 1001 mb on the African coast it has drawn a broad low with central value 1001 mb over north Africa, but its multivariate analysis has failed to generate a geostrophically consistent wind field. Hence its initialisation fills the low by 4 mb to balance the analyzed winds. This effect of the initialisation in filling the surface low is associated with the corresponding filling of the 300 mb upper trough discussed earlier.

The upper jet in the EC analysis is a dynamically consistent compromise between the conflicting data sources; its associated upward motion is spread along the frontal trough (Fig.5.1.8) with maximum value .66 Pa/sec. It has used the 996 mb observation and its accompanying 20 m/s southerly wind geostrophically to draw the north African low farther west than the other systems.

During the next day the North African low moves north to be centred on Southern Italy, with its associated frontal trough advancing in the south to lie south-east across the Mediterranean to Egypt. The low in the north-west Mediterranean filled and merged with that over Italy. All three analysis systems show this sequence.

The analysis differences just discussed cause errors in predicting these developments which were independent of the forecast system used; forecasts from the US analysis are best, forecasts from the EC analysis are nearly 12 hours slow, forecasts from the UK analysis nearly 24 hours slow, with very little initial movement of the frontal trough and North African low. This suggests perhaps that the cloud winds rejected by the US system were indeed wrong, and that the upper jet was crucial to surface developments.

In summary, there appears to have been a small scale jet maximum to the south of Brindisi; because of conflicts between the radiosonde data, the aircraft data and the cloud wind data, together with some apparent in consistencies in the radiosonde data, it is difficult to be certain of its structure. The EC system produced a balanced, moderately smooth, analysis which fits most of the data quite well. The US system rejects some observations, partly because its forecast did not predict the jet maximum. This and its greater horizontal smoothing means that it fits the observations least well. The UK system fits most wind observations closely, and other types quite well. Its analyses have some realistic looking detail but they are noisy and furthest from geostrophic balance.

Short range forecasts of surface features were judged to be best from the US analysis, possibly because it rejected the cloud winds just north of the African coast, and worst from the UK analysis, possibly because it accepted and drew closely to the same cloud winds. The impact of these analysis differences on later times in the forecasts was not large, unlike the situations to be considered in later sections. It seemed surprising that rather large differences in the analysis of the most vigorous system in the northern hemisphere should have only modest effects on the medium range forecasts. Presumably this was because the system was, to some extent, a closed and rather stationary system.

5.2 Mid-Pacific, Feb 17: The importance of quality control

Our next case was chosen because the forecast models acted as consistent amplifiers of analysis differences. The shorter range forecasts showed a decided sensitivity to analysis differences , particularly those between the US on the one hand and the EC and UK on the other. We generated nine predictions for this case (three analyses, three models). In the three predictions from the US analysis - and in only these three - a spurious surface low was predicted in mid-Pacific by days 2-3. To illustrate, Fig.5.2.1 presents the day 3 mean-sea-level pressure charts produced by the EC model run from the EC, US, and UK analyses, together with the corresponding verification chart. Large differences between the forecasts occur near (160W, 40N) where the US based prediction shows an active depression (low A, 998mb) that appears neither in the EC or UK based forecasts, nor in the verification.

These results suggest that the forecast models are responding to some aspect of the US analysis that sets it apart from the EC and UK analyses. To assess the analysis differences we begin with Fig.5.2.2a which shows the wind data over the mid-Pacific between 200 and 300 mb at 00Z on Feb 17, and the three 250 mb height and wind speed analyses. At this level the data coverage is quite respectable. We focus attention upon the area around the dateline at 40N where a minor surface trough present in all the analyses begins to develop in the forecasts generated only from the US analysis.

Differences among the height analyses at 250 mb appear relatively small and subtle compared to the pronounced differences in the intensity and structure

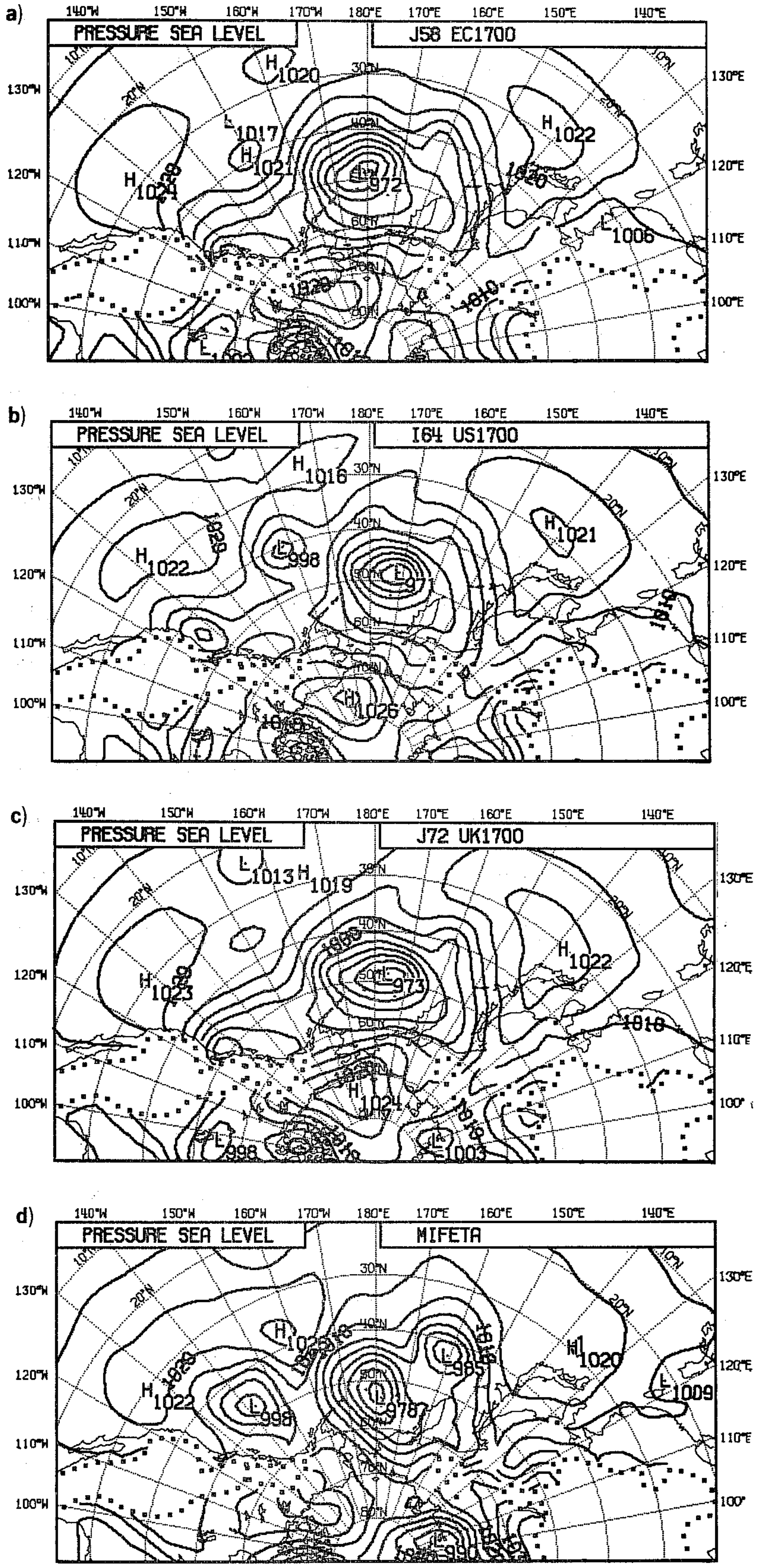
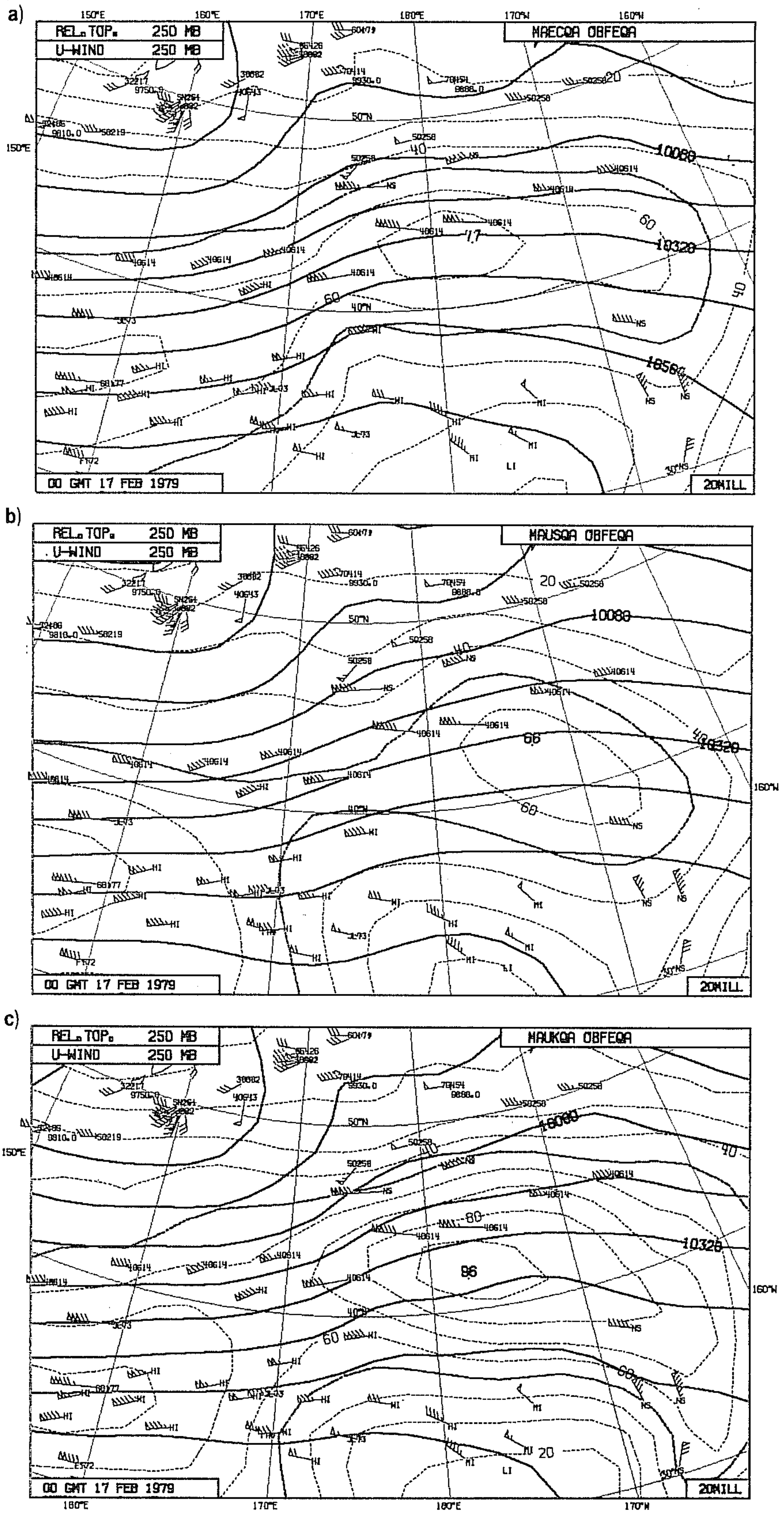
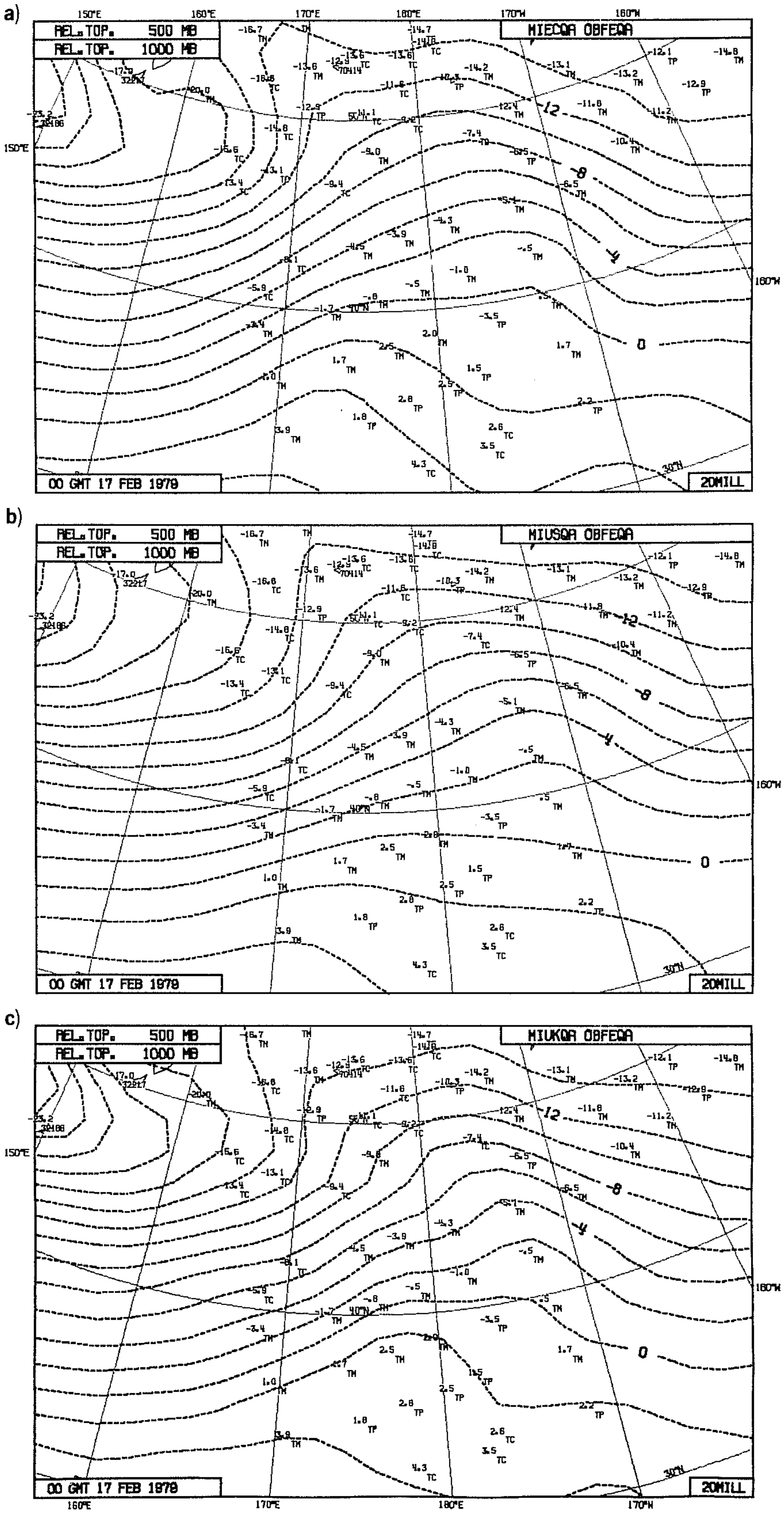


Fig. 5.2.1 a) Mean sea-level pressure three-day forecast valid at 00Z on February 20 1979, made with the EC model from the EC analysis for 00Z February 17 1979 (contour interval 5 mb) b) As panel a) made with the EC model from the US analysis c) As panel a) made with the EC model from the UK analysis d) Verifying analysis.



5.2.2 Analyses of 250 mb height and wind speed at 00Z on February 17 1979 over the mid-Pacific, together with radiosonde (station no.) cloud track wind (HI or NS), and AIREP (call sign) reports. The analyses are a) EC, b) US, and c) UK. The height field is shown with solid contours (interval 12 dam), and the isotach field with dotted contours (interval 10 m/s).



5.2.3 Initialised Analyses of 500-1000 mb thickness at 00Z on February 17 1979 over the mid-Pacific, together with clear-path (TC), partly cloudy (TP) microwave (TM) and radiosonde (station no.) reports. The contour interval for the virtual temperature is 2K. The analyses are a) EC, b) US, and c) UK.

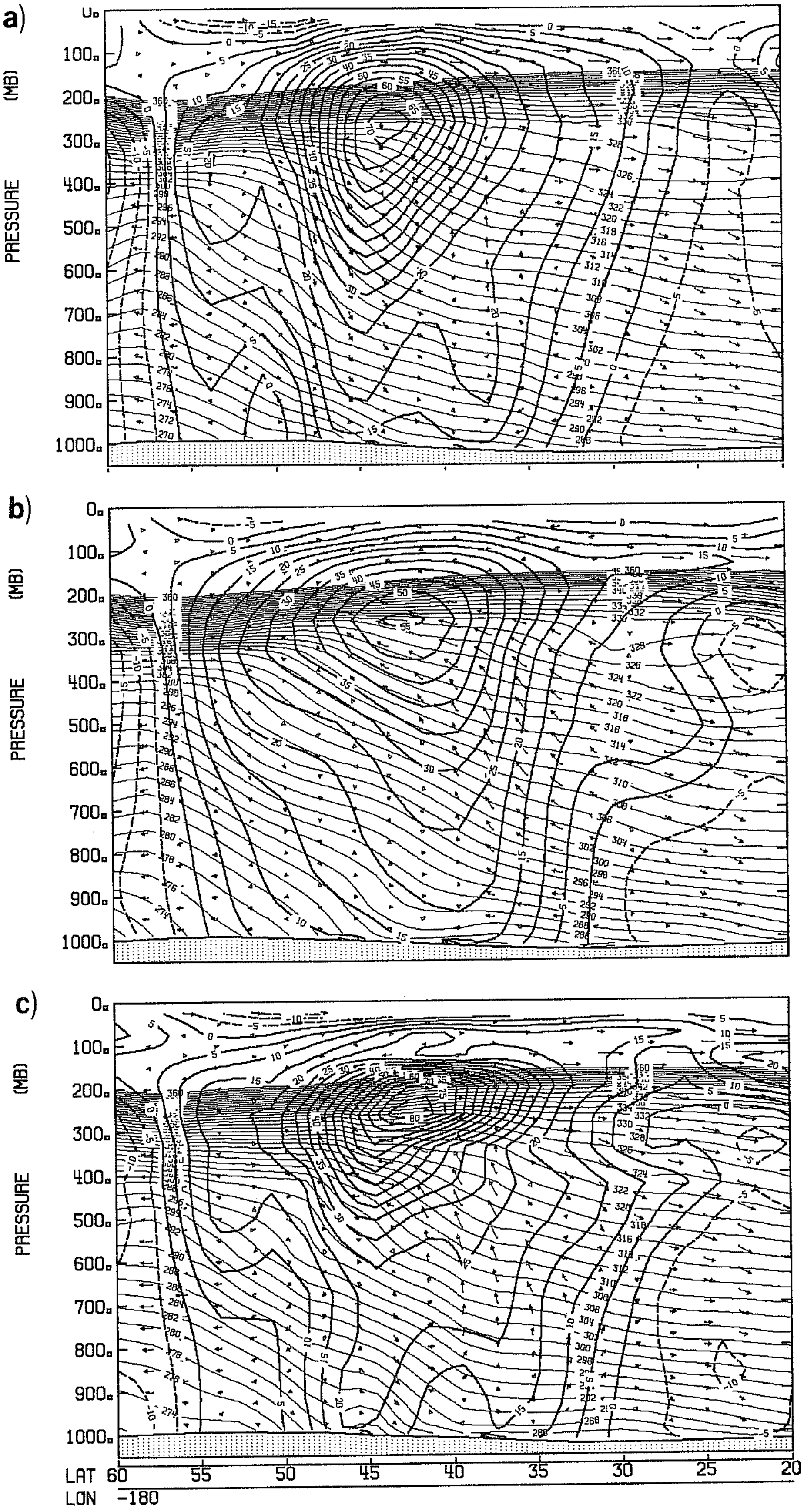


Fig. 5.2.4 Meridional cross-section between 20 N and 60 N of the initialised values of zonal and transverse wind along 180E from the US(a), EC(b), and UK(c) analyses valid for 00Z February 17 1979. The contour interval for zonal wind is 10 m/s.

of the jet. The jet core in the US analysis is notably weaker and displaced downstream relative to the wind maxima in the EC and UK analyses. Inspection of the data reveals that (to a considerable extent) the large differences between the US and other analyses are a consequence of the US disregarding the sequence of wind reports from the aircraft with call sign 40614, together with supporting cloud winds. These aircraft reports were,in fact, rejected in the US gross error check, presumably because of errors in the first guess. The UK draws most closely to these reports, and the EC analysis clearly attempts to accommodate them as well.

In the mid and lower troposphere, the only source of data, other than ships, is a swath of satellite temperature soundings about 20 degrees wide centred approximately on the date line. From Fig.5.2.3 it is apparent that the three data assimilation systems respond quite differently to this data. The temperature structure in the US analysis is noticeably smoother than in the other two. All three show a ridge in the thermal field over the aforementioned surface trough, but only weakly so in the US depiction; the baroclinic zone to the northwest is more intense in the EC and UK analyses when compared with the US analysis. Over the surface trough the satellite data are all microwave retrievals. Visual inspection indicates that in this case the US is most faithful to both partly-cloudy and microwave retrievals, the US draws about as closely as UK to non-microwave retrievals, but less so to microwave data, and EC is least accommodating to the satellite observations, regardless of type.

A vertically more complete picture of the differences in the (initialised) analyses is shown by the north-south cross-sections along the date-line of Fig.5.2.4. Large differences in the three initialised analyses are apparent. The weaker jet core in the US analysis is due to the rejection of the 250 mb aircraft data already referred to. The lower tropospheric differences between the analyses reflect the differences of the response of each system to the satellite temperature soundings, both directly in the thermal field and indirectly in the multivariate effects on the wind field. Note, for example, the southward shift relative to EC of the vertical axis of the US jet and corresponding mid to low-level baroclinity. Although the magnitude of the vertical shear and strength of the baroclinic zone is generally less in the US than EC, that shift (coincidentally or otherwise) places the jet and baroclinicity just above the surface trough that develops in the predictions from the US analysis.

Without an exhaustive study it is difficult to say whether the differences in the mass fields or the differences in the wind fields contributed most to the development of the surface low in the US based forecasts. We present some quantitative results below, but first we consider a synoptic point of view.

The US first-guess, analysed and initialised fields were very similar to each other at this time; the wind data were rejected by the US system because they deviated significantly from the first-guess. These data indicate a strong but narrow jet across the north of Fig.5.2.2. There were waves moving rapidly along this jet with associated surface troughs. The distinct break in the jet in the US analysis leads to a larger scale slower moving difluent-trough/confluent-ridge pattern favourable to the development of one of these features into the spurious surface low. The thermally indirect circulation driven by the jet-exit near 35N combines with the thermally direct circulation associated with the jet-entrance near 45N to give a large area of rising motion near 40N (Fig.5.2.4), where the spurious low developed. The EC and UK analyses also show rising motion in the area but it is more localised and transitory. Thus the synoptician's rule of thumb, relying mainly on the wind field, would indicate significant development potential in the area where it indeed occurred.

To assess the differences between the analyses we have calculated quasi-geostrophic geopotential tendencies and also the generation term of the vorticity equation. The input to the quasi-geostrophic calculations consists only of geopotential (Tracton, 1978) so that the results reflect only the mass field differences; moreover, we only consider the combined effects of thermal and differential vorticity advections, and neglect diabatic effects which may also be important. The results show initial pressure tendencies in the vicinity of the trough, of -0.4, -0.2, and +0.5 mb/hour in US, EC, and UK analyses respectively. Thus the differences in the thermal field translate directly into the greater development potential in the US analysis implied by the forecast results.

Computations of the development term of the vorticity equation also show greater potential in low-level development in the US than in the EC or UK analyses. The calculations were based on cross-sectional values of absolute vorticity and divergence at 950 mb. The relative difference between the US and EC (9.4 versus 3.9 10^{-10}sec^{-2} are comparable to those implied by the quasi-geostrophic computations. Unlike the quasi-geostrophic results, which indicated filling of the surface trough, the UK analysis suggests development almost as strongly (8.9 10^{-10}sec^{-2} as for the US. This is probably symptomatic of the more general lack of consistency between mass and wind fields in the UK analyses. In the present example the UK is significantly rougher and further from thermal wind balance than the other analyses, but was closer to the upper level observations. A comparison of Fig.5.2.2 and 5.2.4 at 250 mb shows that the smoothing and balancing effect of initialisation has brought the UK analysis closer to the EC analysis, which may explain why the forecasts from the UK and EC analyses were similar.

In summary, there are large differences between the three analyses over a surface trough that develops erroneously only in forecasts produced from the US analyses. The most obvious aspect of the US analysis which sets it apart is the weaker upper-level jet that results from the rejection of good aircraft and SATOB data. Diagnostic computations suggest, however, that the differing influence of satellite soundings upon the low-level mass field and the effects of the first-guess on the low level wind field are also of considerable significance for the erroneous surface development.

Apart from the different treatment of Low A (998 mb), all three forecasts in Fig. 5.2.1 show reasonable agreement. Low B, the major system in the mid-Pacific is forecast reasonably well in all three forecasts. Low C, the system which was wrongly predicted by all three forecasts to move onshore and decay along the west coast of North America, in fact slowed down and began to re-develop in the verifying analysis. At day 5 Low A in the US based forecast coincides with observed low C, whose evolution is treated poorly in all three forecasts. It is noteworthy that at day 5, the spurious low in the US based prediction gave rise to better objective verification scores (see Part II).

Because of these active waves and their associated cloud, the satellite temperature soundings were of variable type and quality; in the crucial area near the jet the only soundings available were microwave retrievals. Their gradients gave no indication of the strong jet shown by the wind data.

This case illustrates the vital importance of quality control decisions for the validity of the analyses and for the subsequent forecast. It also indicates that for small scale systems in mid-latitudes the availability of wind data, and the correct analysis of that data is of particular importance. In this case the UK analysis was much further from balance than the EC analysis; the forecasts were similar because the initialisation procedures removed the imbalance in the UK analysis by modifying the mass and wind field. Note, however, that initialisation is not essential; forecasts with the UK model without initialisation were similar to those with the EC and US models after initialisation. Finally, this case illustrates the dangers of

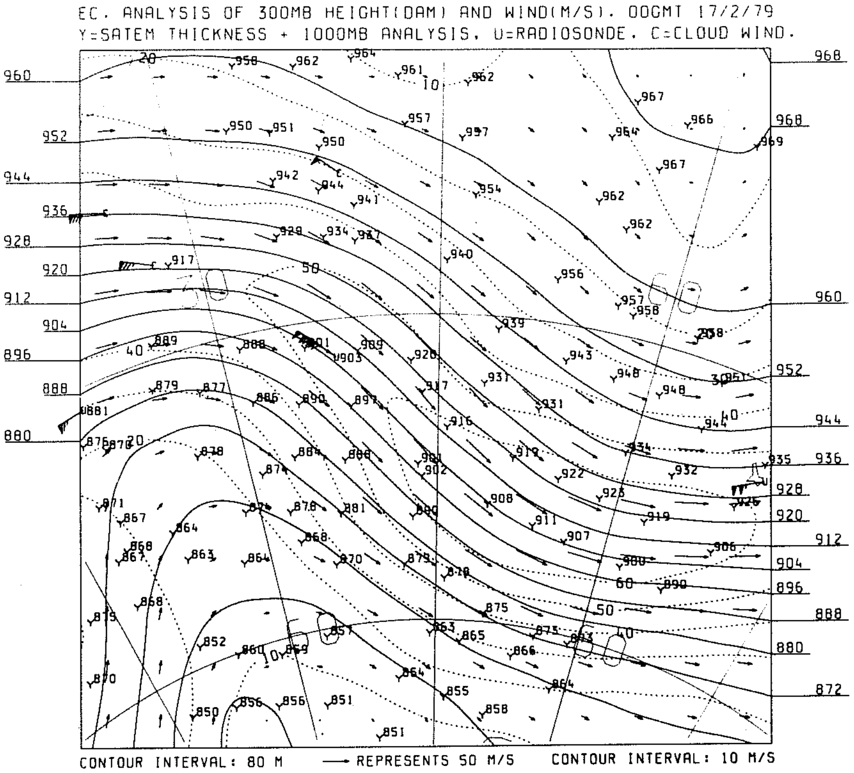


Fig. 5.3.1 EC 300 mb Analysis in the Southern Ocean, southeast of Africa valid at 00Z on Feb 17 1979. Solid lines indicate geopotential (contour interval 8 dam), dotted lines indicate isotachs (contour interval 10 m/s). The available data are also plotted 'Y' indicates SATEM, 'U' indicates radiosonde, 'C' indicates cloud wind.

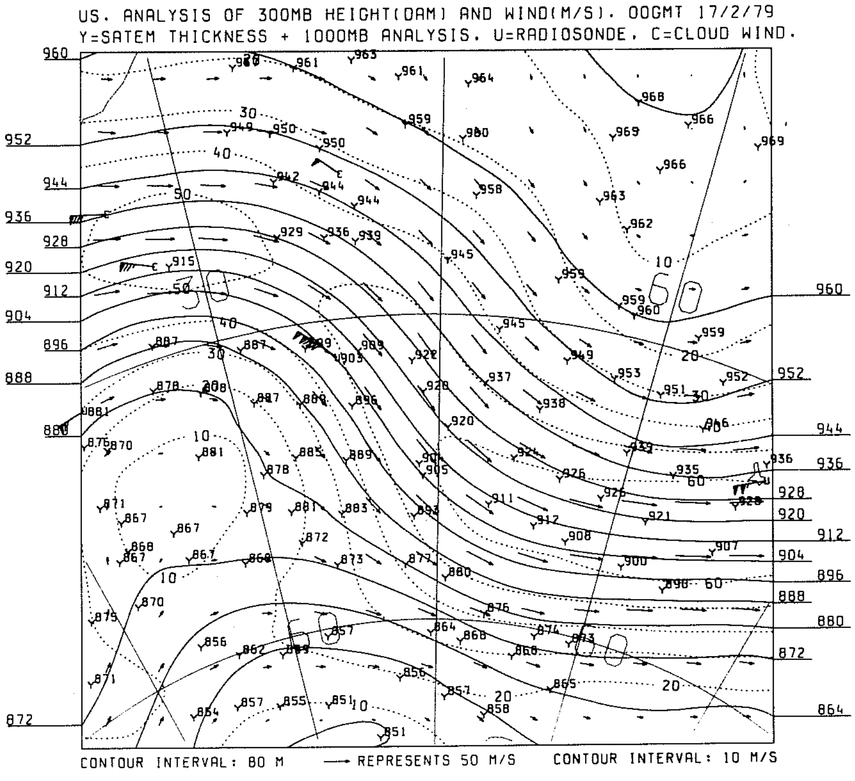


Fig. 5.3.2 As Figure 5.3.1 for the US analysis.

mis-interpretation that can arise when using objective scores of medium range forecasts to evaluate analysis quality: the medium-range scores for the US based forecasts were significantly higher than for the others, and this better score arose mainly because of an error in the analysis.

5.3 A Southern Hemisphere depression. 00GMT 17th. Internal consistency data selection and vertical resolution

Thanks to the advent of satellite observing systems automatic NWP is now possible in the Southern Hemisphere. However, even with the enhanced FGGE observing system, data are still sparse, and it is more important than in the Northern Hemisphere to make best use of every good observation, and to interpolate meteorologically consistent features for variables and regions lacking in data. Our example from the Southern Hemisphere is a frontal wave depression near Marion Island (47S 38E). Its thermal structure was observed by satellite temperature soundings which, apart from the Marion Island radiosonde sounding and a few cloud motion winds, were the only upper air data available. Analyses made with an earlier version of the UK system without the SATEM soundings did not show the depression (Lorenc, 1981).

The 300 mb height and wind analyses from each system are shown in Figs.5.3.1 to 5.3.3 . The SATEM observations plotted have been calculated using the corresponding 1000 mb analyses as reference level, and hence may differ on the three figures. Marion Island is near the centre; it has an observed wind of 70 m/s which has not been drawn to closely by any of the systems. The US analysis fits least closely, with an upper trough which disagrees also with the observed cloud wind directions. The RMS fits of the analysis to these plotted observations are given in Table 5.2. It appears that the EC system, in trying to fit the wind observations geostrophically, gives little weight to the SATEM thicknesses, and hence fits them less closely than the other systems do. The US system gives less credence to the winds and fits these less well, while the UK system fits both but produces an unbalanced analysis.

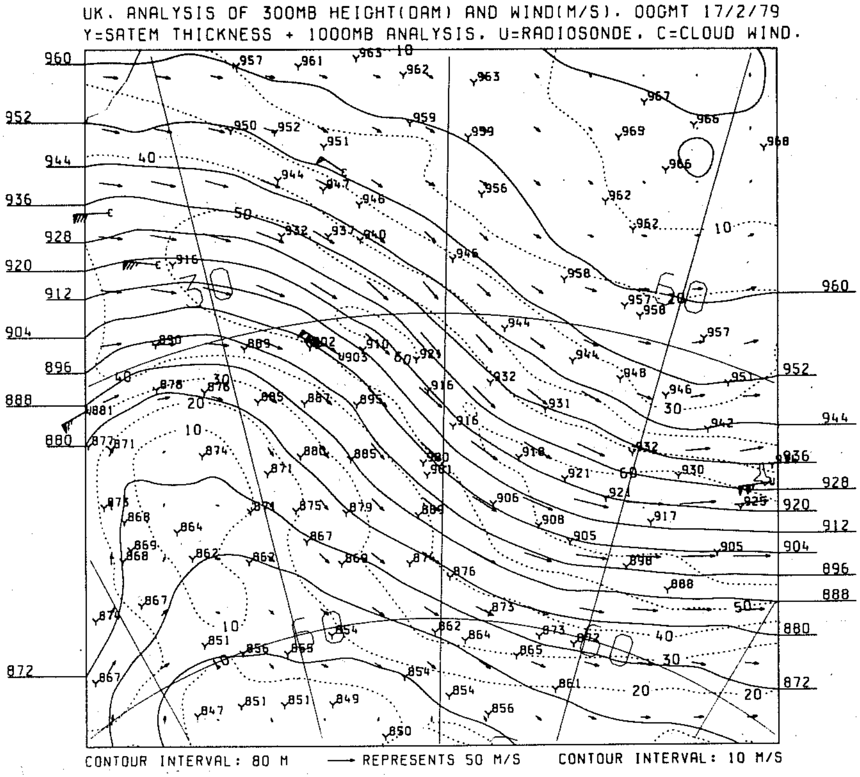


Fig. 5.3.3 As Figure 5.3.2 for the UK analysis.

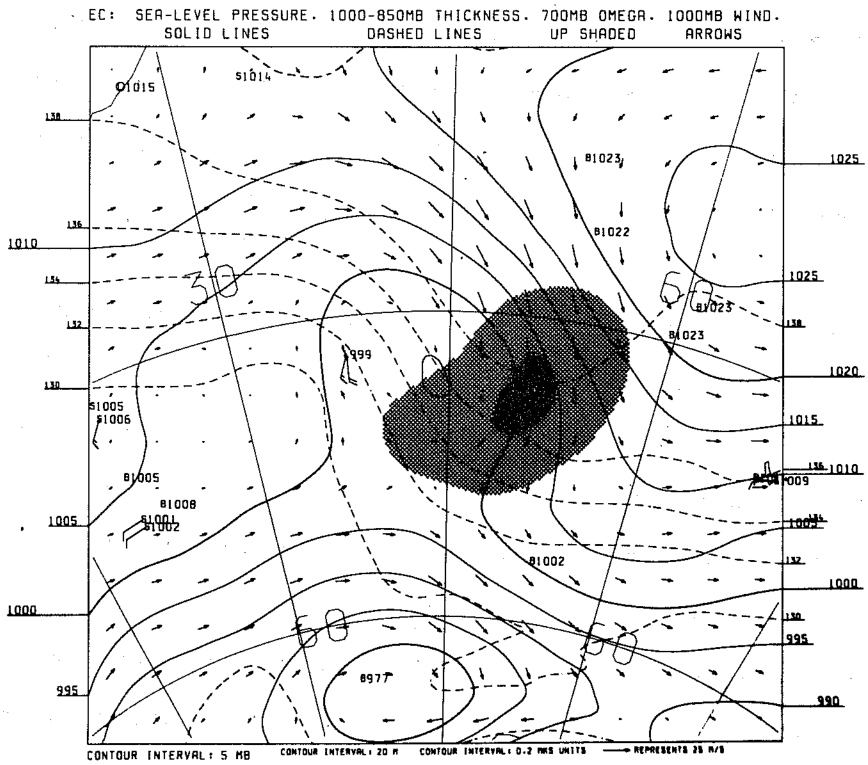


Fig. 5.3.4 The low-level EC analysis over the same area as Figure 5.3.1 for 00Z February 17 1979. Solid lines show sea-level pressure (contour interval 5 mb), dashed lines show the 1000-850 mb thickness (contour interval 20 m or 4.2 K) ; and the gradations of hatching show the 700 mb vertical velocity with a contour interval of .2 Pa/s (approximately 2 cm/sec). Observations of mean sea-level pressure and surface wind from land (L) ships (S) and buoys (B) are shown in the usual way.

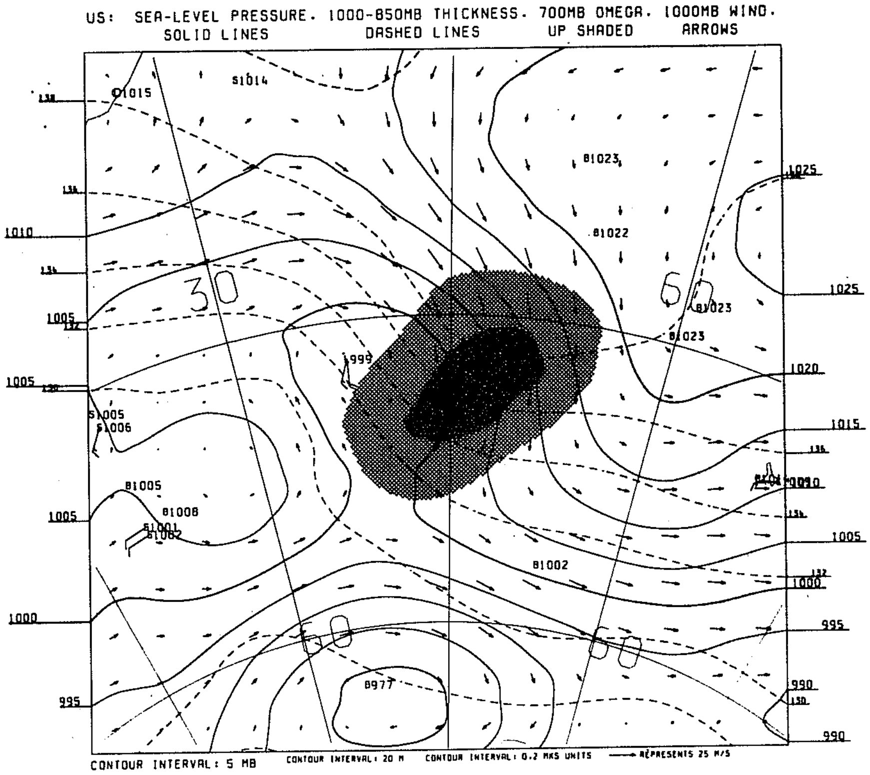


Fig. 5.3.5 As Figure 5.3.4 for the US analysis.

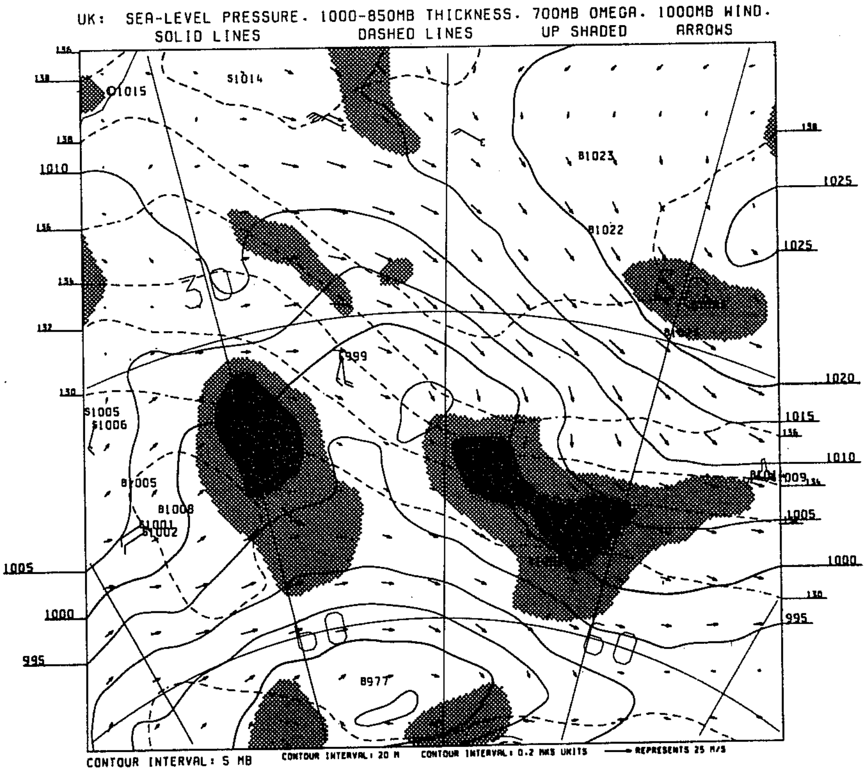


Fig. 5.3.6 As Figure 5.3.4 for the UK analysis.

Analysis field	Observation type & symbol		Number	EC	US	UK	Units
1000mb wind	Land	L	59	4.9	8.6	7.9	m/s
1000mb wind	Ship	S	20	7.8	9.7	7.0	m/s
1000-300mb thickness	SATEM	Y	109	39.	26.	25.	mb
300mb geopotential	Radiosonde	U	2	32.	19.	45.	mb
300mb wind	all	CU	6	8.	14.	9.	m/s
sea level pressure	all	LSB	17	1.8	1.6	2.7	mb
1000mb wind	all	LS	9	3.6	4.9	8.2	m/s
1000-850mb thickness	SATEM (not plotted)		109	5.7	6.7	9.0	mb

TABLE 5.2 Verification against the plotted observation of the EC, US, and UK analyses shown in Figs. 5.3.1 to 5.3.3 at 300 mb, and Figs. 5.3.4 to 5.3.6 at 1000 mb.

The surface analyses are shown in Figs.5.3.4, 5.3.5 and 5.3.6, together with the available surface data. Only the Marion Island observation is near the centre of the low. The EC system has used its wind geostrophically to deduce a low centre to the east. The US system puts a similarly shaped low centre near Marion Island. Unlike the other systems it has accepted and drawn closely to the observation of 1008 mb from a buoy to the south west which was badly calibrated, systematically registering pressures several millibars higher than nearby observations throughout the period. The UK system gives a pressure analysis of a different character to the others. The surface pressure is largely induced during the repeated insertion data assimilation from observations of other variables. This usually gives realistic features, such as sharp troughs along fronts, but fields are rougher and fit the pressure observations less well as can be seen in Table 5.2.

The EC and US wind analyses in Figs.5.3.4 and 5.3.5 are consistent with their respective pressure analyses. Because of the positioning of the low centre and the spurious high in the US analysis, the US winds fit the observations less well. The UK system does not explicitly couple wind and pressure analyses, so less balance is evident in Fig.5.3.6 . Usually the UK system, because of its lack of constraints, is free to fit the observed winds closely. However, in this case the Marion Island wind was rejected because it was a surface report from a land station, while a mistake caused a badly coded missing pressure level indicator to be interpreted as 999 mb in the cloud motion winds shown to the north in Fig.5.3.6 . These winds are actually consistent with cirrus level motions. Thus the UK analysis fits the wind observations of Fig.5.3.4 and 5.3.5 badly. Unlike the 300 mb thicknesses shown in Figs.5.3.1 to 5.3.3, the 1000-850 mb thickness observations (not shown) are not explicitly contradicted by wind observations, and the EC analysis shown in Fig.5.3.4 fits them closely, with its well developed frontal wave in accordance with the observations, although the curvature of the 1320m line near Marion Island is an attempt to fit both its observed 1310m and

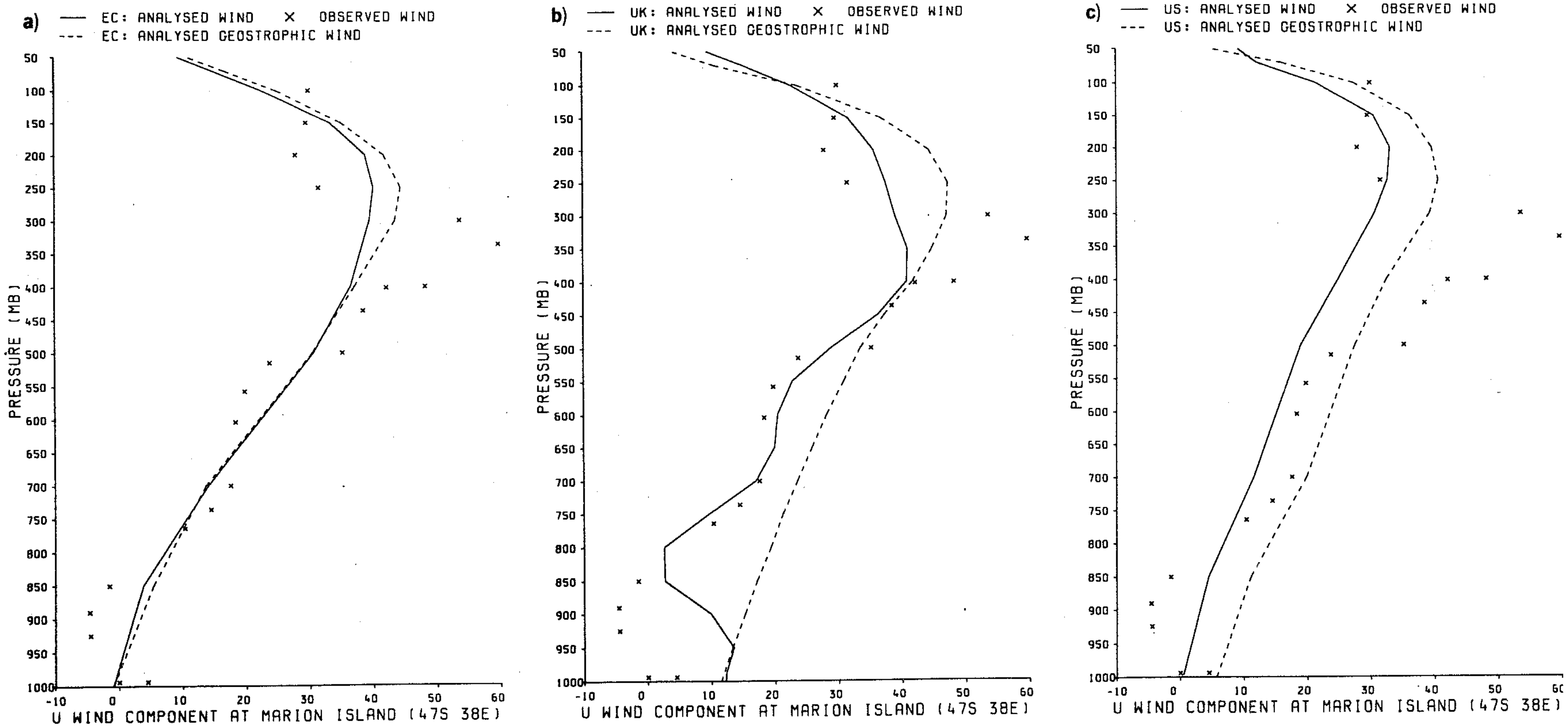


Fig. 5.3.7 Comparison of the reported u-wind (crosses), analysed wind (solid lines) and analysed geostrophic wind (dashed lines) for the analyses at 00Z February, 1979 ar Marion Island (47S, 38E) by (a) EC, (b) UK and (c) US systems.

surrounding SATEM reports about 10m higher. Presumably because the US system has less resolution its frontal thermal gradients (Fig.5.3.5) are less strong than observed and it fits the data slightly less well. The shortcomings noted earlier in the UK wind analyses mean that the warm and cold advection which created the frontal zone are underestimated. This and a general tendency in the UK data assimilation model to cool the lower layers means that its analysis (Fig.5.3.6) fits the thickness observations least well.

Figs.5.3.4, 5.3.5 and 5.3.6 also show shaded regions of upward motion at 700 mb, calculated directly from the initialized winds of the EC and US systems and the data assimilation model of the UK system. The EC and US fields are consistent with simple adiabatic baroclinic instability theory, with upward motion in the region of maximum warm advection. (The EC and US initialisation schemes are adiabatic). In contrast the UK upward motion is more confined to frontal zones, including some at the cold front, and consistent with humidity fields in these areas. However, some of the vertical motion is a response to the inconsistent wind data used rather than actual ascent.

The different response of the three systems to the Marion Island wind sounding is further illusrated in Fig.5.3.7, which shows the eastward wind component at all levels reported in the TEMP, PILOT and SYNOP messages, and the three analyses. The EC system has used the surface and standard level winds with considerable vertical smoothing, reducing the wind shear from that observed but still giving a greater shear than the US system, which apparently gave more weight to the winds implied by the thermal gradients between the SATEM observations and rejected winds in the jet maximum. Both these systems gave analyses close to gradient wind balances; the US analysis shows a trough near Marion Island and hence has sub-geostrophic winds. The UK system gave analyses further from balance. Its temperature analyses fitted the SATEMs, hence the vertical shear of its geostrophic winds is

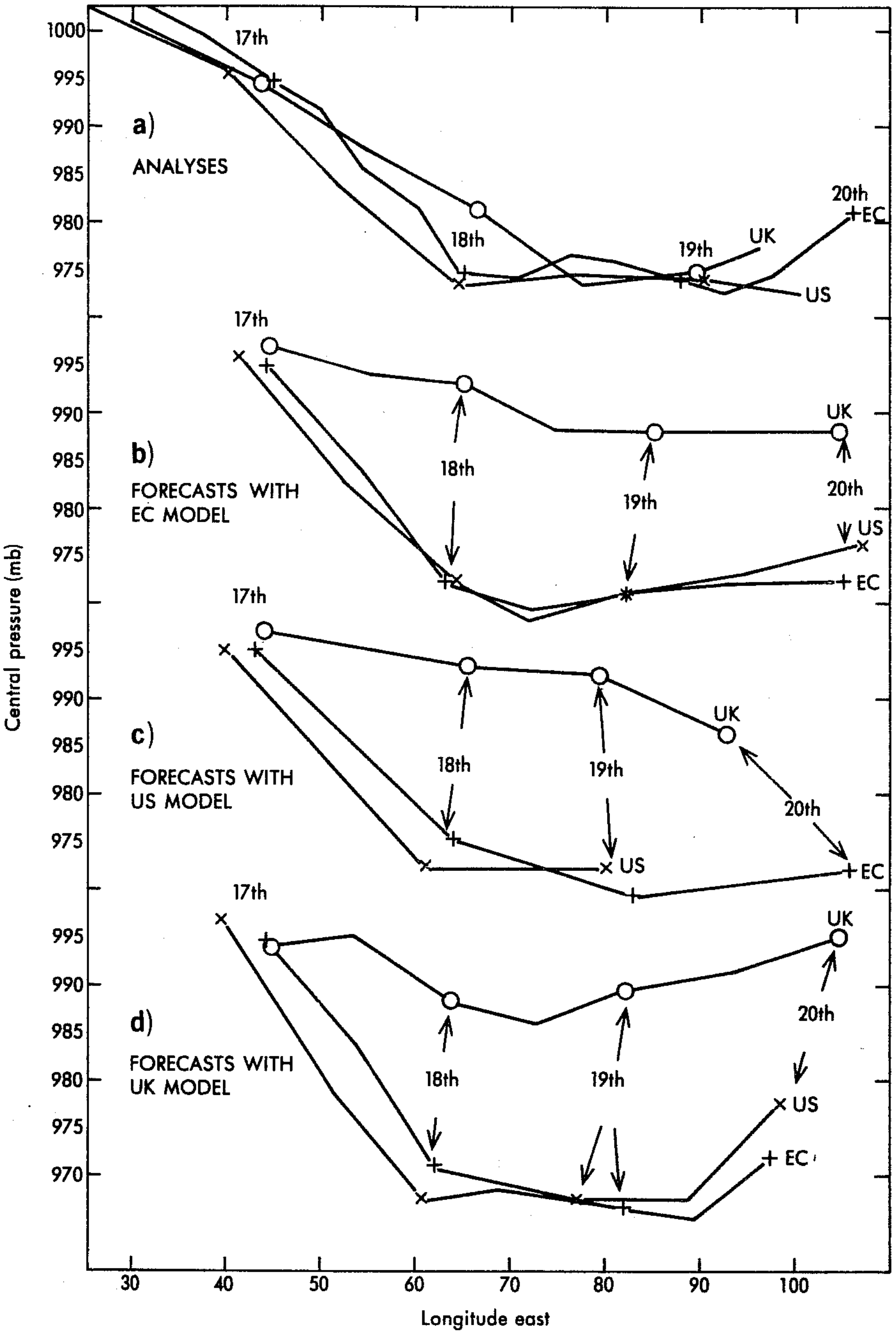


Fig. 5.3.8 Pressure and longitude traces for the analysed and forecast values of the central pressure and position of the low shown in Figures 5.3.4 to 5.3.6 a) Traces from the EC(+), US(x), and UK (o) analyses. b) As a) for the forecasts made with the EC model. c) As a) for the forecasts made with the US model. d) As a) for the forecasts made with the UK model.

similar to that of the US system, while its wind analyses fit those wind data which it did use with less smoothing than the other systems.

During the next day this depression deepened rapidly, and thereafter continued an eastward movement until merging with a slower moving system near 105 E on the 20th. This sequence is shown in Fig.5.3.8a. Data on the precise depth of the low at 00GMT on the 18th is scarce, otherwise the analyses agree well. The response of all three forecast systems to the analysis differences is remarkably similar (Fig.5.3.8b, c, d); forecasts from the UK analysis fail to predict the initial deepening, while forecasts from the EC and US analyses somewhat overdeepen the low. In some of the forecasts from the UK analysis the system is indeed only visible as a trough whose depth and position have been somewhat subjectively determined. Forecasts from the EC and US analyses rapidly sharpen and intensify the rather broad area of vertical motion shown in Fig.5.3.4 and 5.3.5, giving after 12 hours north-east to south-west orientated vertical velocity maxima about three times that analysed for forecasts from the US analysis and about two times for the EC analysis. Forecasts from the UK analysis move the maximum on the warm front in Fig.5.3.6 rapidly eastward away from the low centre. The main features of the UK analysis are unaltered by interpolation and initialisation for the EC and US forecast systems; both the vertical motion field of Fig.5.3.6 and the vertical profiles of wind and geostrophic wind of Fig. 5.3.7c are only slightly smoothed.

To summarize, the EC system gave more weight to the Marion Island radiosonde than to nearby satellite temperature soundings, leading to stronger and straighter upper flow and greater vertical shear. The US system was inaccurate in positioning the surface low, analysed a spurious high, and rejected Marion Island wind observations of the jet. The UK system analysed small scale features both vertically and horizontally across the front, as expected from the discussion of Sect. 2.5. Because of this and poor selection and quality control of low level wind data, it failed to analyse a low level thermal advection field of sufficient north-south extent. The differences between the EC and US analyses had very little effect on the forecasts, while forecasts from the UK analysis failed to predict the rapid baroclinic development. Hence, unlike the previous two examples , it appears that the upper jet stream is not crucial for this forecast, since large differences between EC and US analyses made no difference. The different approach to internal consistency and the slow manifold, which causes the UK system to give the front a narrow N-S extent, seems more important.

5.4 The President's day storm: Impact in the medium range

This case study is concerned with the impact of analysis differences on the forecasts for the later evolution of an intense and much studied storm known as the President's day storm.

On Feb 19, 1979 a short upper trough approached the east coast of North America. At the surface an extremely cold anticyclone began to move off the continent over the Atlantic. As the upper trough neared the coast it triggered the development of a small but very intense system over the Gulf Stream. This storm has come to be known as the President's day storm, and has been studied by Bosart (1981), Bengtsson (1981) and Uccellini et al (1981). The system gave heavy snowfall along the east coast of the U.S.A. at 12Z on Feb 19.

We consider the forecasts from Feb 18 00Z for this system. The 36 hour forecasts in the present study were generally better than the operationally produced forecasts. There were minor differences in the short range forecasts for the system which could be traced to rather large differences in the low level analyses over the Gulf of Mexico. However, these differences proved to be unimportant, as the main dynamical control was exerted by a short wave upper-level trough which was over the Mid-west at the beginning of the forecasts. All authors agree on the crucial importance of this trough for the development.

Our interest lies in the later evolution of this system when there were significant differences between medium range forecasts over the Atlantic, the forecasts being made with the same model from different analyses. We demonstrate that these differences are due to analysis differences over the north east Pacific, well upstream of the system of interest at the initial time.

Apart from relatively minor differences in the first two days, the forecasts with the EC model from the EC and US analyses from Feb 18/00Z for the President's day storm were very similar out to four days, because the upper-air analyses for the controlling short-wave feature were in good agreement. Rather dramatic differences occurred in these same two forecasts for the same surface low between day-4 and day-6, Fig.5.4.1. In one case the low has deepened in situ while in the other it has moved rapidly off to the northeast with much less deepening.

Fig.5.4.2 shows difference maps between forecasts with the EC model from the EC and US analyses at both the surface and 300mb for days 0, 2, 4, 6. The forecast model was the same in both forecasts so the differences are directly attributable to differences in the initial analyses. A study of the difference maps between the forecasts suggests an explanation. In the difference maps at day-4 we see what appears to be a well organised wave-train over the central and eastern part of north America. This wave-train can be traced in the difference maps both forwards and backwards in time, and its amplitude grows in time. If we follow its evolution in time

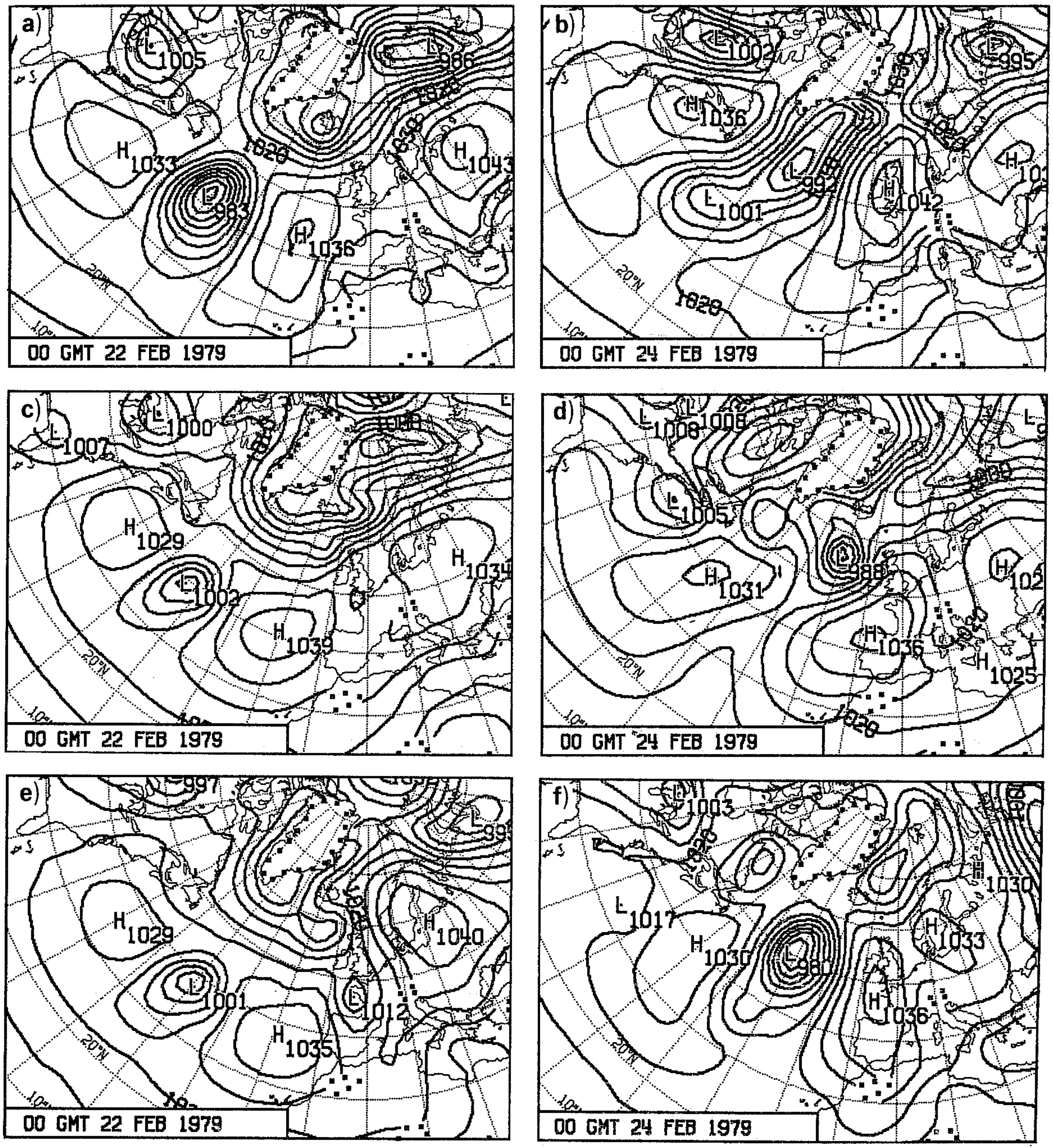


Fig. 5.4.1 Analyses and forecasts for the mean sea level pressure field in the North Atlantic at 00Z on February 22 and 24 1979 ; a) and b) Verifying analyses for the 22nd and 24th (contour interval 5 mb) ; c) and d) Four and six day forecast with the EC model from the EC analysis valid at 00Z on February 18 1979; e) and f) Four and six day forecasts with the EC model from the US analysis valid at 00Z on February 18 1979.

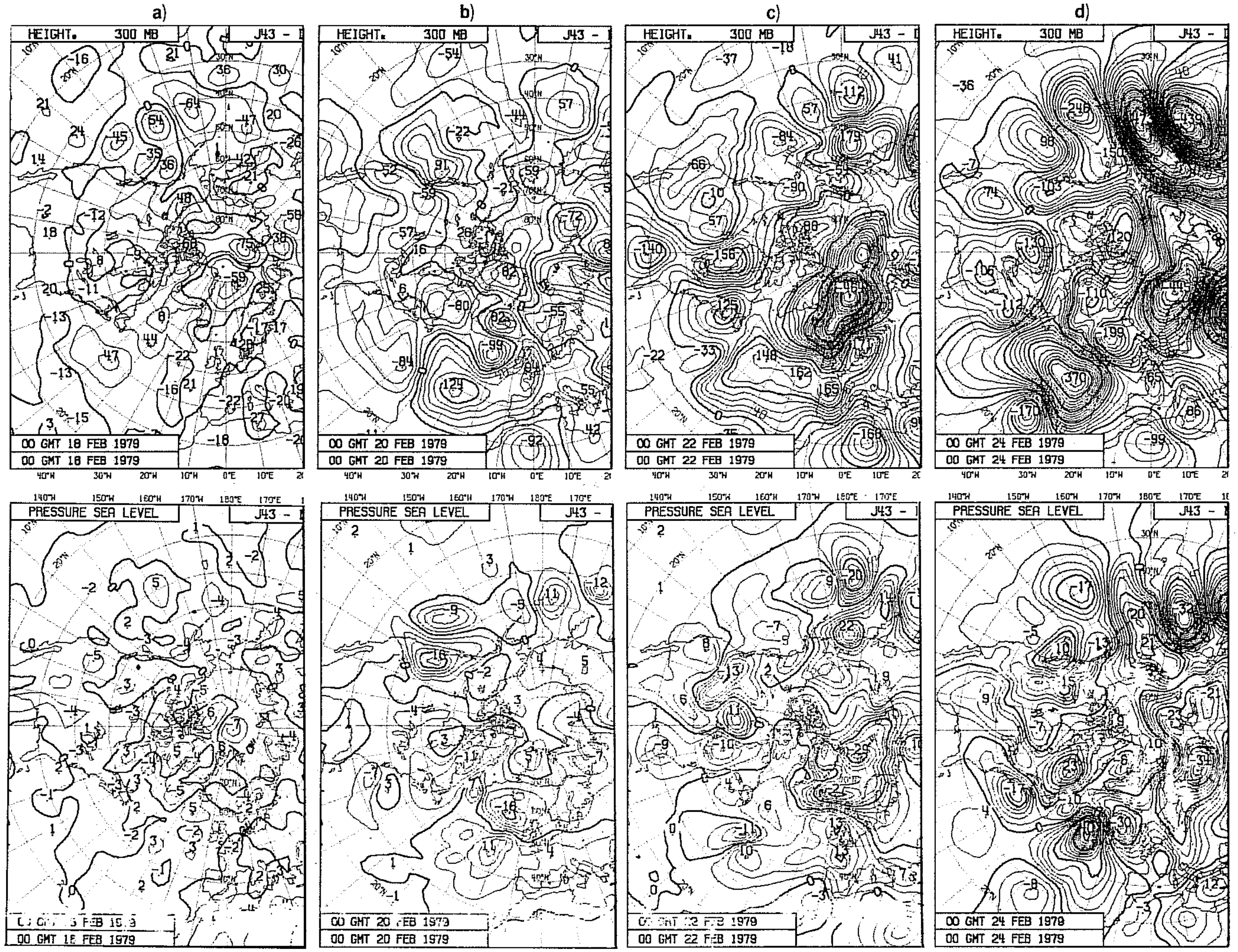


Fig. 5.4.2 Difference maps for 300 mb geopotential (contour interval 20 m) and mean sea-level pressure (contour interval 4 mb) between forecasts with the EC model from the EC and US analyses valid at 00Z February 18, 1979. The difference fields are shown for the initial data (a), day 2 (b), day 4 (c), and day 6 (d). These are referred to in the text as the control differences.

from day-4 through day-6 we see a rapid amplification as the leading edge of the signal in the difference field reaches the central Atlantic.

Amplification of the difference signal at the surface is manifested by the fact that in one forecast the surface low deepens rapidly in situ while in the other forecast the surface low moves rapidly away to the north-east.

This down-stream propagation and amplification of the wave-train in the difference field is strongly reminiscent of the down-stream propagation phenomenon discussed by Simmons and Hoskins (1979). They considered an unstable baroclinic flow on which they superimposed an isolated barotropic perturbation. The perturbation began to amplify locally and also sent a Rossby-wave propagating downstream. This wave generated a new daughter low downstream. This process continued through many generations until disturbances arrived back to the location of the original perturbation.

It may seem surprising to try to apply such a simple experiment to interpret the present results. Experiments to be presented below lend support to the interpretation, which is no more than a re-phrasing of the synoptician's view of what happens in the two forecasts. In the synoptic description one forecast shows a deeper trough than the other. The deeper trough throws up a higher ridge downstream which in turn leads to an extra deepening of the next trough downstream, etc. Essentially we consider differences in the initial data to behave as small perturbations on a more complicated basic flow than that used by Simmons and Hoskins. The basic flow is unstable and the small perturbation grows on the flow while shedding disturbances downstream which amplify in turn.

The wave train is traceable back in time to the small perturbations at time 0 because the initial conditions for both forecasts had the noise, which might otherwise have obscured it, removed by non-linear normal mode initialisation.

In the equivalent forecast pair with the UK model, which had no such initialisation, a similar wave train could be seen, although its behaviour after three days was somewhat different. However, its origin was less easy to trace because unbalanced differences in the initial analyses lead to rapidly changing differences in the short term forecasts.

If one traces the wave-train upstream it appears to originate in the north-east Pacific. In this area the main data sources are the SATEM data (mainly micro-wave), a reasonable coverage of surface ships, some cloud-wind data, some AIREP data and ship PAPA(50N, 145W). The main differences in the analyses at 00Z on Feb 18 and at 12Z on Feb 17 were in the lower troposphere and appeared to arise from a conflict between the thickness field as reported by the SATEMS and by ship PAPA. Fig.5.4.3 shows the 500-700 mb thickness field in the area at 12Z on Feb 17, together with the relevant observations; the coverage by the SATEM data was much sparser at 00Z on Feb 18 when the forecasts started. We discuss the analysis differences at 12Z on 17 Feb (at the time when data was most abundant) because these analyses had a large effect, through the assimilating model, on the Feb 18 00Z analyses (when data was much less abundant).

It is clear that the US analysis has been rather faithful to the SATEM microwave reports; by contrast the EC system has largely ignored them and has drawn reasonably well for the report from ship PAPA and kept the strong frontal zone well to the south. The differences between the analyses at the surface or near 250 mb, where the remaining data types were concentrated, were much less marked. The only remaining arbiter of the accuracy of the analyses in the area was the quality of the short-range (one or two day range) surface forecasts. Examination of these short range forecasts indicated that the forecasts from the EC analysis gave a better short-range forecast. This result would suggest that the EC analysis was more accurate on this occasion,

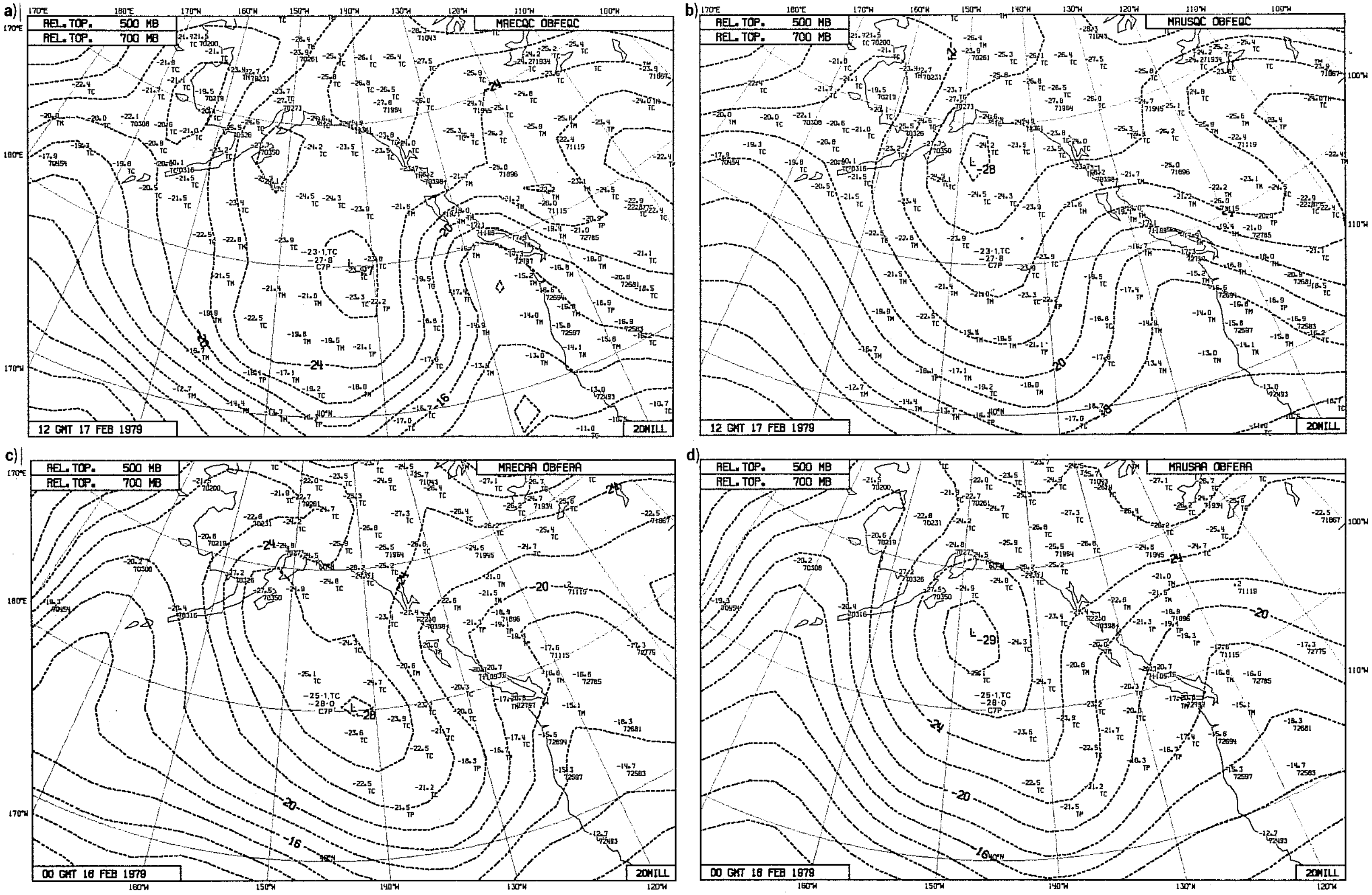


Fig. 5.4.3 Analysis of the 500-700 mb thickness (converted to virtual temperature, contour interval 2K) over the North East Pacific at 12Z February 16 1979 (top panels), and 00Z February 18 1979 (bottom panels) as produced by the EC system (panels on the left) and the US system (panels on right). Also shown are the radiosonde and Satem reports. The different types of satellite retrievals are indicated as 'T' for clear path, 'TP' for partly clear and 'TM' for microwave retrievals.

and would not conflict with the general impression that microwave retrievals are inferior in quality to radio-sonde data or other SATEM data.

The arguments presented above suggest that the differences in the medium-range forecasts in the central and eastern Atlantic were due to the differences in the initial data in the north-east Pacific, even though the trough giving rise to the President's Day storm is already east of the Rockies in the initial data.

In order to justify the validity of this interpretation we performed what we call a transplant experiment. By this we mean that we transplanted the US analysis into the EC analysis in the north east Pacific and ran a forecast from the resulting merged analysis. If the interpretation outlined above is substantially correct then we should expect to reproduce the growth and downstream propagation of the difference field wave-train in the difference field calculated from the forecasts from the EC analysis and the merged, or transplanted, analysis.

The technique used to merge the fields was very simple. In the latitude-longitude rectangle with corners at (120W, 30N) and (180E, 60N) we replaced the un-initialised EC analysis for all variables by the US analysis. The discontinuities were removed by a linear merge, variable by variable and level by level, of the two analyses in a zone 10 degree wide surrounding the rectangle. The resulting merged analysis was then initialised. The initialisation made very little difference to the fields inside the rectangle or outside the border zone; the changes were mainly confined to the border zone where the two analyses had been combined.

Fig.5.4.4a shows the differences between the EC analysis and the merged analysis after the initialisation. Inside the rectangle the differences are

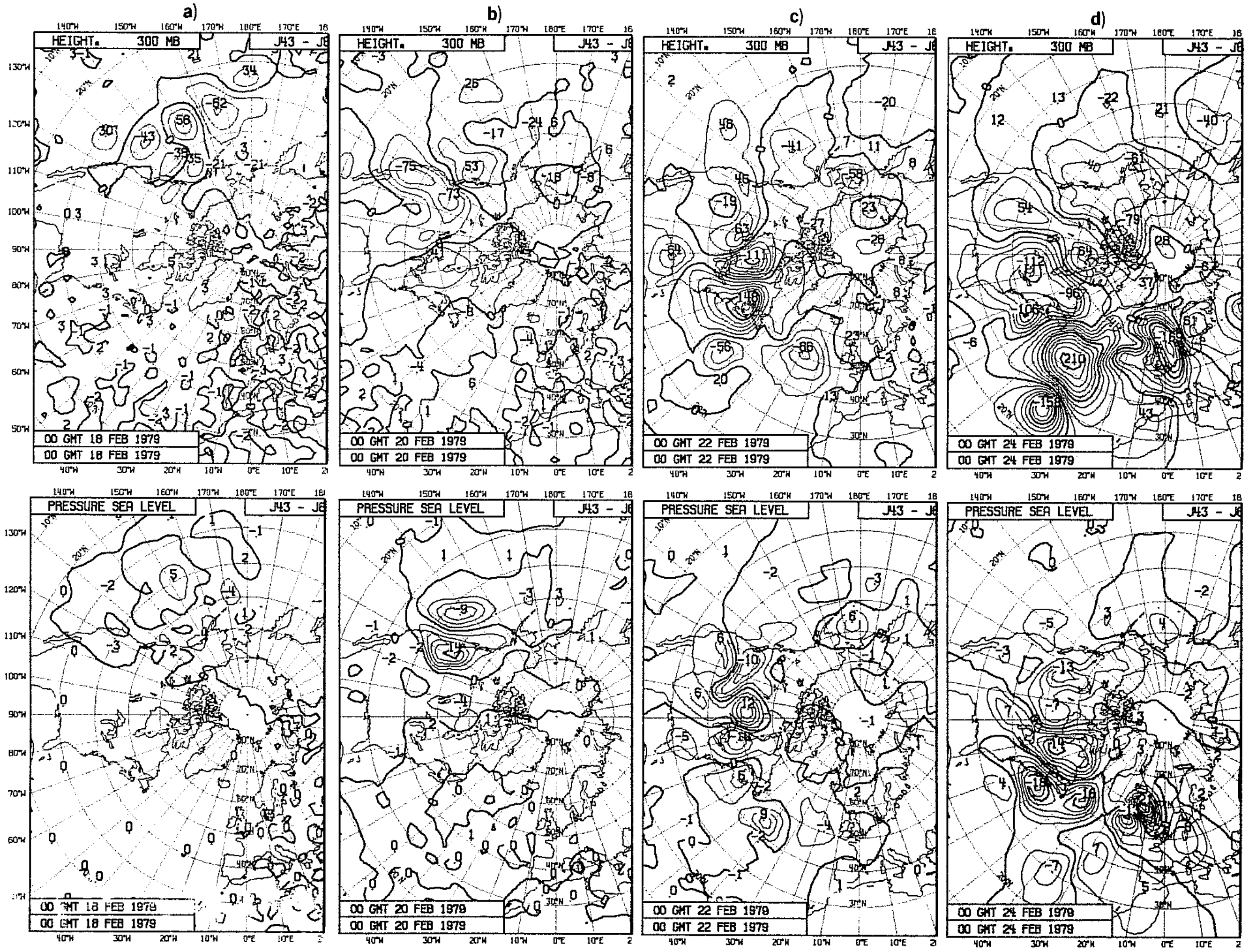


Fig. 5.4.4 Difference maps for 300 mb geopotential (contour interval 20 m) and mean sea-level pressure (contour interval 4 mb) between forecasts with the EC model form a) and the EC analysis valid at 00Z February 18, 1979, and b) the data set formed by transplanting the US analysis over the North East Pacific into the EC analysis. The difference fields are shown for the initial data a), and at day 2 b), day 4 c), and day 6 d). These are referred to in the text as the experimental differences.

well preserved; outside the rectangle the apparently noisy field is due to very small amplitude differences of order 1-2 m.

To simplify the rest of our discussion we introduce some terminology. If we denote the EC based forecast as forecast A, the US based forecast as forecast B and the forecast from the merged data set as forecast C, then we may speak of the A-B differences as the control differences and the A-C differences as the experimental differences.

Fig.5.4.4 shows the experimental differences and may be compared with Fig.5.4.2. The experimental differences are entirely due to the differences between the EC and US analyses in the north-east Pacific, since the same forecast model is used throughout these comparisons. The similarities between the experimental differences and the control differences at day 2 are striking, with all the main features in the experimental differences occurring also in the control differences.

At day 4 the experimental differences at 300 mb have already reached the eastern Atlantic with amplitudes of 86m, but their largest amplitude is over Newfoundland with amplitudes of 146m. The similarity in the structure of the experimental differences and the control differences at this stage of the forecast is quite remarkable. Almost every feature over north America in the experimental field is to be found in the control field, with the same structure and amplitude. The only exceptions in the control differences are the feature over Alaska which originated in the western Pacific, and the feature over the coast of the Gulf of Mexico; this latter corresponds to a weak spurious low which developed in the EC based forecast but not in the US-based forecast. This feature occurred less strongly in the forecast from the merged data-set. The overall agreement is very encouraging, given that the experimental differences appear to be on schedule to cause a large signal in mid Atlantic between day 5 and day 6.

At day 5 (not shown) the leading edge of the experimental differences at 300 mb is over Europe, while the largest amplitudes are to be found in mid-Atlantic. Although the absolute values of the extrema in the experimental and the control differences do not agree over the western and central Atlantic, the patterns, especially the maximum to minimum gradients, agree very well indeed . There is no agreement, however, over the Eastern Atlantic and Europe, where the experimental differences are weak. This may be taken to mean that these latter differences arose from analysis differences outside the transplanted rectangle.

At day 6 the leading edge of the upper level experimental differences are perceptible over Russia (Fig.5.4.4), but the largest amplitudes are still over the mid-Atlantic. Although the quantitative agreement between the experimental difference fields and the control difference fields is not quite as good as it was on day 5 it is nonetheless impressive, given that we are discussing phenomena that are probably well into their non-linear phase of growth.

These results support our view that the medium range differences over the Atlantic were indeed caused by the analysis differences over the north eastern Pacific,even though the system of interest was over the central U.S.A. at the initial time.

We have tested the transplant technique in some other areas using the same initial data sets. The largest differences in Fig.5.4.1 at day 6 occur over the western Pacific. At least half of the amplitude of these differences is due to analysis differences in the polar cap , poleward of latitude 75, as suggested by another transplant experiment. These polar analysis differences had propagated and amplified along the curved jet that travelled southward and then eastward over central Asia.

Our conclusion, that in some circumstances the growth of analysis differences in the forecasts is governed (at least qualitatively) by the pulsed baroclinic theory of Simmons and Hoskins (1979), is therefore supported by two experiments. Cats and Akesson (1983) have applied our transplant technique to the study of the divergence of medium range forecasts made from analyses separated by one day. Their results confirm the relevance of the theory to the amplification of small perturbations in a complex, unstable flow.

The results presented here illustrate some practical limits to predictability arising from analysis uncertainty. Analysis errors in baroclinically unstable regions are particularly important and their influence depends both on their magnitude, and on the strength and direction of the flow in which they occur. The rapidity with which analysis errors can propagate downstream demonstrates the importance of having adequate, accurate, observations over land and sea for the success of medium range forecasts.

6. SUMMARY AND CONCLUSIONS

The main purposes of this study was to assess the nature, cause and significance of the differences between the output of three different advanced analysis systems when they had all been presented with the same input data. The extent to which these differences reflect the real uncertainty about the atmospheric state, and the effect of such uncertainty on practical limits to predictive skill are further discussed in a second paper.

The main conclusions to be drawn from this study are:

1) In general all the analyses fit the data acceptably closely. However large differences do occur between the analyses.

2) Some analysis differences which, a priori, appeared significant had little effect on the forecasts, because the systems in which they occurred were decaying or were isolated from the main baroclinic zones. Other analysis differences significantly affected the subsequent forecast quality, particularily when they occured in crucial regions of baroclinic development. The downstream development theory of Simmons and Hoskins (1979) proved very useful in documenting these effects. In fact, the results suggest that this theory describes one of the essential mechanisms for the loss of forecast skill through the unstable amplification of small analysis errors.

3) The accuracy of the forecast first-guess, quality control and selection of observations, resolution and concepts of balance interact in a complex way, and it is not always possible to assign with certainty a single simple cause to any analysis difference.

4) Many analysis differences were associated with differences in quality control and data selection. In some case studies it appeared that the EC system tended to average inconsistent data, the US system to reject some, and the UK system attempted to fit most data, sometimes in an unbalanced fashion.

5) Differences in the approach to the concept of balance in the UK system caused large differences in the analyses which seemed of little significance to forecasts where both mass and wind field data were available, but which had a large negative effect on a southern hemisphere case where mass data predominated.

6) Biases between the analyses can be identified, and are thought to arise from several parts of the assimilating systems.

The cases presented here show that poor data coverage and inaccurate data can significantly reduce current forecasting skill. It is also clear that there is room for improvement in the current analysis algorithms to exploit the available data more fully. Many improvements have already been made to all three systems, some of which are attributable to the present study.

References

Andersen, J.H. 1977 A routine for normal mode initialisation with non-linear correction for a multi-level spectral model with triangular truncation. ECMWF Internal Report No. 15, 41pp; available from ECMWF.

Arpe, K. 1980 Confidence limits for verification and energetics studies. ECMWF Technical Report No.18; available from ECMWF.

Arpe, K. 1982 Diagnostic Evaluation of analyses and forecasts : Climate of the model. Proceedings of ECMWF Seminar on Interpretation of Numerical Weather Prediction Products; available from ECMWF

Arpe, K., A. Hollingsworth, A.C. Lorenc, M.S. Tracton, G. Cats, P. Kallberg 1983 The response of Numerical Weather Prediction Systems to FGGE IIB Data, Part II: Forecasts. To appear.

Atlas, R. 1979 A comparison of GLAS SAT and NMC High resolution NOSAT forecasts from 19 and 11 February, 1976. NASA Tech. Memo. 80591, Goddard Space Flight Centre.

Bengtsson, L. 1975 4-Dimensional assimilation of meteorological observations. GARP Publications Series No. 15, WMO-ICSU, Geneva.

Bengtsson, L. 1981 The Weather Forecast. Pure & Appl. Geophys., 119, pp.515-537.

Bengtsson, L., M. Kanamitsu, P. Kallberg, S. Uppala 1982a FGGE 4-dimensional data assimilation at ECMWF. Bull. Amer. Met. Soc., 63, p.27.

Bengtsson, L., M. Kanamitsu, P. Kallberg, S. Uppala 1982b FGGE research activities at ECMWF. Bull. Amer. Met. Soc., 63, p.277.

Bergman, K.H., 1979 A multivariate interpolation analysis system for temperature and wind fields. Mon. Wea. Rev., 107, p.1423-1444.

Bjorheim, K., P. Julian, M. Kanamitsu, P. Kallberg, P. Price, S. Tracton, S. Uppala 1981 FGGE IIIB Daily global analyses, Parts I-IV; available from ECMWF.

Bosart, L. F. 1981 The President's Day snowstorm of 18-19 February, 1979: A subsynoptic scale event. Mon. Wea. Rev., 109, pp.1542-1566.

Cats, G., O. Akesson 1983 An investigation into a marked difference between two successive forecasts of September, 1982. To appear in Contrib. Atmos. Phys.

Charney, J.G., R. Jastrow, M. Halem 1969 Use of incomplete historical data to infer the present state of the atmosphere. J.Atmos. Sci., 26, pp.1160-1163.

Desmarais, A., S. Tracton 1978 The NMC report on the Data Systems Test (NASA contract S-70252-A G) U.S. Dept. of Commerce, NOAA, Rev., 107, pp.140-171.

Gustafsson. N., J. Pailleux 1981 On the quality of FGGE data and some remarks on the ECMWF data assimilation system. ECMWF Technical Memo. 37; available from ECMWF.

Hollingsworth, A., Arpe, K. 1982 Biases in the ECMWF assimilation system Tech. Memo. 46; available from ECMWF.

larvenoja, S. 1982 An intercomparison of different numerical 500mb height analyses in the Northern Hemisphere. Report 22, Dept. of Meteorology, Univ. of Helsinki.

Kuo, H. L. 1965 On formation and intensification of tropical cyclones through latent heat release by cumulus convection. J.Atmos. Sci., 22, pp.40-63.

Lau, N-C. and A. H. Oort 1981 A comparative study of observed northern hemisphere circulation statistics based on GFDL and NMC analyses. Part I : The time mean fields. Mon. Wea. Rev., 109, pp.1380-1403.

Lau, N-C. and A. H. Oort 1982 A comparative study of observed northern hemisphere circulation statistics based on GFDL and NMC analyses. Part II: Transient Eddy Statistics and Energy Cycle. Mon. Wea. Rev., 110, pp.889-906.

Leith, C. 1981 Statistical methods for the verification of long and short range forecasts. Proceedings of ECMWF 1981 Seminar on Problems and Prospects in Long and Medium Range Weather Forecasting; available from ECMWF.

Lorenc, A. C. 1976 Results of some experiments assimilating observations from a simulation of the FGGE observing system into a global General Circulation model. Report of JOC study conference on four dimensional data assimilation, Paris 1975; Report 11, GARP program on Numerical Experimentation; WMO/ICSU, pp.358-374.

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

Lorenc, A. C. 1982 Assimilation of single level wind data. Proceedings of ECMWF workshop on Current problems in data assimilation; (to appear) will be available from ECMWF.

Lyne, W. H., R. Swinbank, N. T. Birch 1982 A data assimilation experiment and the global circulation during the FGGE Special Observing Periods. Quart. J. Roy. Met. Soc., 108, pp.575.

Louis, J. F. (Ed.) 1981 The ECMWF forecast model documentation manual by ECMWF Research Dept. ECMWF, Reading, UK.

Machenhauer, B. 1977 On the dynamics of gravity oscillations in a shallow water model, with application to normal mode initialisation. Contrib. Atmos. Phys., 50, pp.253-271.

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

Otto-Bliesner, B., D. P. Baumhefner, T. W. Schlatter, R. Bleck 1977 A comparison of several analysis schemes over a data-rich region. Mon. Wea. Rev., 105, pp.1083-1091.

Parker, D. E. 1980 Climatic change or analysts' artifice - a study of grid-point upper-air data. Meteor. Mag., 109.

Phillips, N. A. 1982 On the completeness of multivariate optimum interpolation for large scale meteorological analysis. Mon. Wea. Rev., 110, pp.1324-1334.

Rosen, D. R., D. A. Salstein 1980 A comparison between circulation statistics computed from conventional data and NMC Hough analyses. Mon. Wea. Rev., 108, pp.1226-1247.

Saker, N. J. 1980 UK Meteorological Office Model in Catalogue of Numerical Atmospheric Models for the FGGE. Ed.: J. Smagorinsky, Joint Scientific Committee Publication, WMO-ICSU, Geneva.

Sela, J. G. 1980 Spectral modelling at the National Meteorological Centre. Mon. Wea. Rev., 108, pp.1279-1292.

Simmons, A. J., B. J. Hoskins 1979 The downstream and upstream development of unstable baroclinic waves. J. Atmos. Sci. 36, pp.1239-1254.

Temperton, C., Williamson, D. L. 1981 Normal mode initialisation for a multi-level grid-point model, Part I : Linear Aspects. Mon. Wea. Rev., 109, pp.729-743.

Tracton, S., A. Desmarais, R. J. van Haaren, R. D. McPherson 1980 The impact of satellite soundings on National Meteorological Centre's analysis and forecast system-The Data Systems Test results. p543-586. pp.543-586.

Tracton, S., A. Desmarais, R. J. van Haaren, R. D. McPherson 1981 On the system-dependency of satellite-sounding impact - Comments on recent impact test results. Mon. Wea. Rev., 109, pp.197-200.

Trenberth, K.E., D.A. Paolino 1980 The northern hemisphere sea-level pressure dataset: Trends, errors and discontinuities. Mon. Wea. Rev., 108, pp.855-872.

Uccellini, L. W., P. Kocin, C. H. Wash, R. A. Petersenand J Paegle 1981 The President's day cyclone 18-19 February 1979: An analysis of upper tropospheric and low level jets prior to cyclogenesis. Submitted to Mon. Wea. Rev.

Williamson, D. L., C. Temperton 1981 Normal mode initialisation for a multi-level grid-point model, Part II: Nonlinear aspects. Mon. Wea. Rev., 109, pp.744-757.

APPENDIX

Impact of interpolation and packing

When planning these experiments, we decided to use pressure level data packed in the FGGE IIIB format as input to the forecast models. This is the standard international archive format and its use ensures comparability of the results. The initial data for the forecasts was prepared by unpacking the data and interpolating the fields to model coordinates. However in operational practice in some centres, the initial data for the forecasts is prepared in a somewhat different manner; the analysis increments are interpolated to model coordinates and added to the first-guess in model coordinates. This method was developed by Talagrand (pers comm. 1981) for the EC operational system and has since been implemented also in the US system; it was used in the US assimilation which produced the data studied bere, but not the EC data. The technique is called interpolation of increments. The previous system in use at EC, which was used to prepare the data studied here, is known as full field interpolation because the analysis increments are added to the pressure first guess, and the full field is interpolated to model coordinates. The interpolation of increments has the benefit of retaining features of the first guess, such as the boundary layer structure, which are destroyed by the full-field interpolation.

We had available to us a set of analyses for the same dates using a revised version of the EC system which incorporated the interpolation of increments technique. We present here some simple comparisons of the two sets of data and of some forecasts from the data. We shall call the data from the revised EC run ECRE.

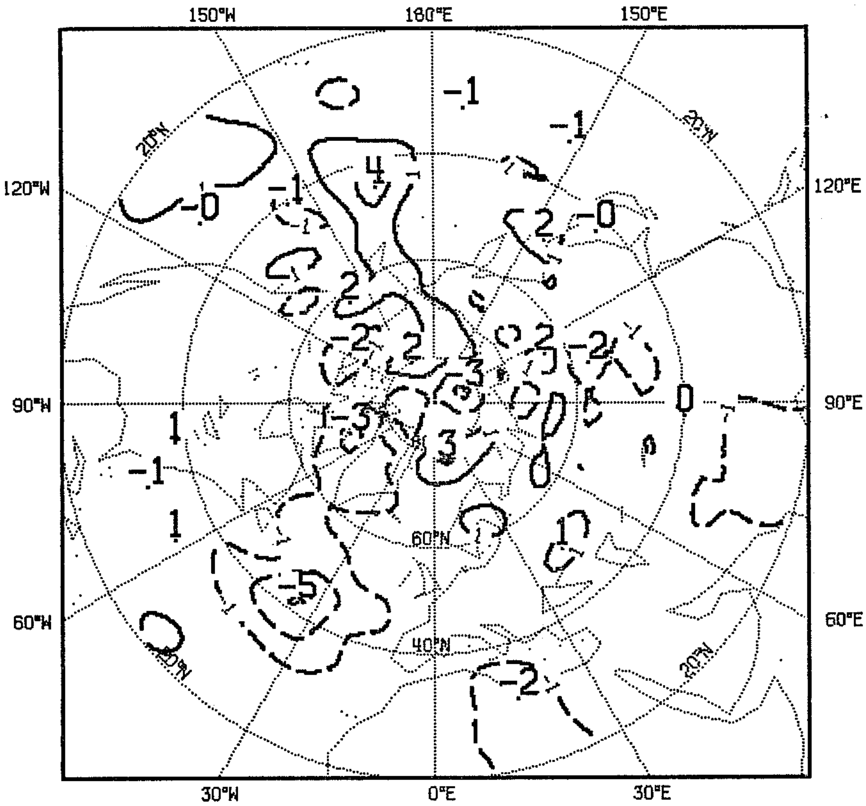


Fig. A.1 The difference between the EC and ECREU 500 mb height fields averaged over the four cases for which forecasts were made (0000 GMT Feb 17 Feb 18 and 1200 GMT Feb 18 and 19).

Figure A.1 shows objective verifications of three forecasts made with the EC model from three initial data sets for Feb18 00Z; EC as decsribed in the main text; ECREP, which was produced by taking the ECRE pressure analysis , packing and unpacking it, and finally interpolating it to model coordinates; and lastly ECREU which was the model coordinate initial data derived from ECRE by using the interpolation of increments technique.

The forecast verifications at days 5 and 6 are sensitive to the differences between the EC and ECRE systems and, are sensitive also, though to a lesser extent, to the packing and interpolation procedures. To explore this latter point further we examine Figure A.2 which shows the initial data ECREU at 500mb together with the differences, ECREU-ECREP, between it and the data which had been through the packing and interpolation algorithms. It is clear that there has been a damping effect on the fields through the interpolation process, as the major troughs are shallower, and the major ridges flatter in the interpolated data. Clearly this effect could be cumulative in an assimilation. To test this possibility we examine Figure A.3, which shows the mean difference between the ECREU and EC analyses, for the four times from which forecasts were run in the main study. It is clear that there is a marked mean difference between the analyses withe similar features to those we noted earlier in our comparisons of the EC and US data. Most striking is the difference in the intensity of the mid-Atlantic trough, but there are also differences in the same sense in the other areas noted in the discussion of the EC-US differences, notably in mid-Pacific and over north Africa. Typically the mean ridges and troughs are stronger in the ECRE analyses as compared to the EC analyses. These results strongly suggest that some of the mean differences we noted earlier in the EC-US comparison stem to a large extent from differences in the vertical interpolation algorithms that were used in the assimilations.