DATA ASSIMILATION BY REPEATED INSERTION INTO A FORECAST MODEL - PRINCIPLES, PRACTICE, PROBLEMS AND PLANS

A. C. Lorenc
Meteorological Office
Bracknell, U.K.

Abstract

The repeated insertion assimilation method is a reasonable approach to reconciling the principles of data assimiation with operational needs at the Met Office, which requires among other things, high resolution analyses for short period forecasts of rainfall.

During research and development of the method, various systems using it have been used for practical purposes related to FGGE:- for OSSE before FGGE, for making level IIIa analyses, for analysis system intercomparisons and for OSEs. A global and a limited area fine mesh version have also been developed for operational use. Examples from these systems show that the method can achieve its objective of fitting the observations while maintaining an approximate balance consistent with the sophisticated model used, including diabatically forced flow and high resolution frontal structures.

However, problems do remain in the practical implementation of the method, in its more fundamental design, and in important areas common to all methods. Plans to tackle some of these are based on a careful design of the method of relaxing the assimilating model towards the observations. With this and the iterative nature of repeated insertion, the total scheme can be made to approximate more theoretically optimal methods.

1. INTRODUCTION

The UK Met Office has for some years been doing research into four-dimensional data assimilation. The main objective of this is to develop systems for operational use, but there is naturally also an interest in the basic principles of data assimilation. These, and the characteristics needed for operational use which have influenced our choice of this system, are discussed in section 2.

During development of the repeated insertion data assimilation method, various versions have been used for research into related problems, many to do with the First GARP Global Experiment (FGGE). Lorenc (1976) performed OSSE to evaluate possible special observing systems for FGGE. A system was developed to make near real time analyses during FGGE (Lyne et al 1982). These analyses were used for general circulation studies; they have been compared with the ECMWF FGGE IIIb analyses by Lorenc and Swinbank (1984). The FGGE system in a slightly modified form was used for Observing System Experiments (OSE) (Barwell and Lorenc 1985), and numerical weather prediction analysis intercomparisons (Hollingsworth et al 1985). Following the success of the FGGE system, the method was recoded and modified for global operational use (Lyne et al 1983). This version has also been used for OSE's (Bromley 1984). A high resolution limited area version is currently being tested. These various systems are described, and some examples presented, in section 3.

This experience has identified a number of important problem areas. To tackle some of these the assimilation routines are currently being rewritten. These problems and plans are discussed in section 4.

2. PRINCIPLES

The basis of any analysis must be the observations, and the first principle is thus clear:-

1. The analysis must fit the observations to within their estimated observational errors

Note the important qualification admitting the existence of observational errors.

Unfortunately in practice we do not usually have enough observations to define the atmospheric state uniquely via simple interpolation, and our experience of atmospheric structures must be used. This can be expressed in words:-

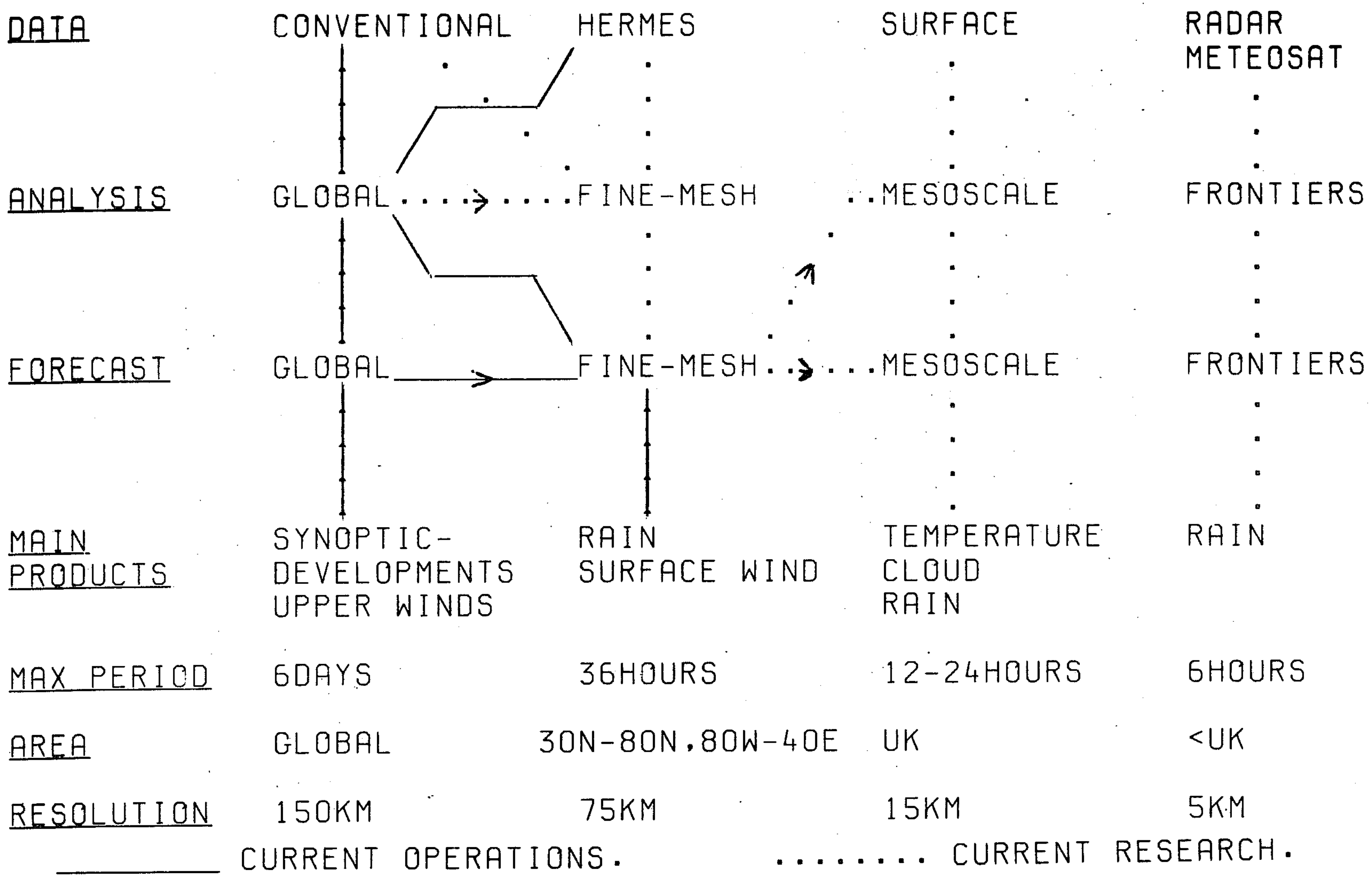
2. The analysed fields must be internally consistent, matching the structure scale and balance of the atmosphere

This is often incorporated implicitly into analysis schemes without being clearly stated or quantified. Some aspects can be expressed as relationships (such as geostrophy) which the atmosphere approximately obeys, and which can be used as constraints on the analysis. Even these two principles are not enough to define well the atmospheric state, particularly in data gaps. Operational forecasting systems have long been organized to use information from earlier times to help fill these. Thus we have our third basic principle:-

3. The analysis must be near the forecast based on earlier observations, unless current observations indicate otherwise

The Met Office has a wide range of forecasting activities, from extended range monthly forecasts to nowcasting. Some of these are shown in figure 1. The practical compromises made and emphasis given when implementing these three principles will depend on the type of forecast to be made. The scheme described in this paper was developed for a global data assimilation system, but it was always envisaged that it would also be used for the fine-mesh shorter period forecasts, and perhaps in the future for mesoscale forecasts. These requirements affect our implementation of all three principles. For the first the crucial aspect is the assumed observational errors, which are made up of instrumental components, and components due to the unrepresentativeness of the observations for the scale we wish to analyse. For example it is appropriate for ECMWF to assign an error of 5 m/s for an upper level aircraft wind, because it is not very representative of the synoptic scales they wish to analyse and forecast. However the actual instrumental error is much less than this, and if you compare one aircraft observation with that of the next along the same flight path the differences are also less. Because of the operational use made of our analyses themselves, it is desirable that they should fit the observations as closely as reasonably possible.

The small scale atmospheric phenomena important for fine-mesh forecasts are not easy to characterize mathematically in order to apply the second principle. Many of the techniques used for initializing larger scale models are less valid, although a fine-mesh analysis must of course still represent



1. **Simplified diagrams of data, analyses, models, and products at the Met. Office.**

well the larger scale motions. Many important rain-making processes are highly non-linear. Despite these difficulties, it is through the application of the second principle that improvements in automatic analysis techniques are to be made; human analysts using conceptual models can do so successfully. Our approach is in two parts:-

2a. Assume that the forecast model is the best available model of the atmosphere, and that atmospheric motions are usually slowly varying. Together these assumptions give the rule:- Slowly varying states of the forecast model are possible atmospheric states.

2b. Iterate to achieve principles 1 and 2a. At each iteration linearize about the current model state and apply simplified relationships such as geostrophy to the (hopefully) small differences between it and the observations.

Finally, for the third principle, a high resolution in space requires similar precision in time. The scheme described in section 4 should be able to assimilate observations at their actual valid time.

3. PRACTICE

3.1 Basic method

The basic repeated insertion method is to nudge a forward running model over a period of time towards the observations. This relaxation has two advantages compared to a single insertion:-

(1) It avoids exciting rapidly varying model modes, with periods short compared to the relaxation period. This satisfies principle 2a.

(2) It allows for interactions between the observed information, via the effect that each has had on the model fields at earlier steps.

The method of nudging has for computational economy to be simple, but it must attempt to maintain a meteorologically likely structure in the model fields. (Principle 2b).

Development of the data assimilation scheme has continued over several years, as described in section 1. In this section I describe and give examples from two of these systems, passing quickly over problems; these are discussed further in section 4. Only selected points are made from each example; references to a fuller discussion of each case are given.

3.2 FGGE system

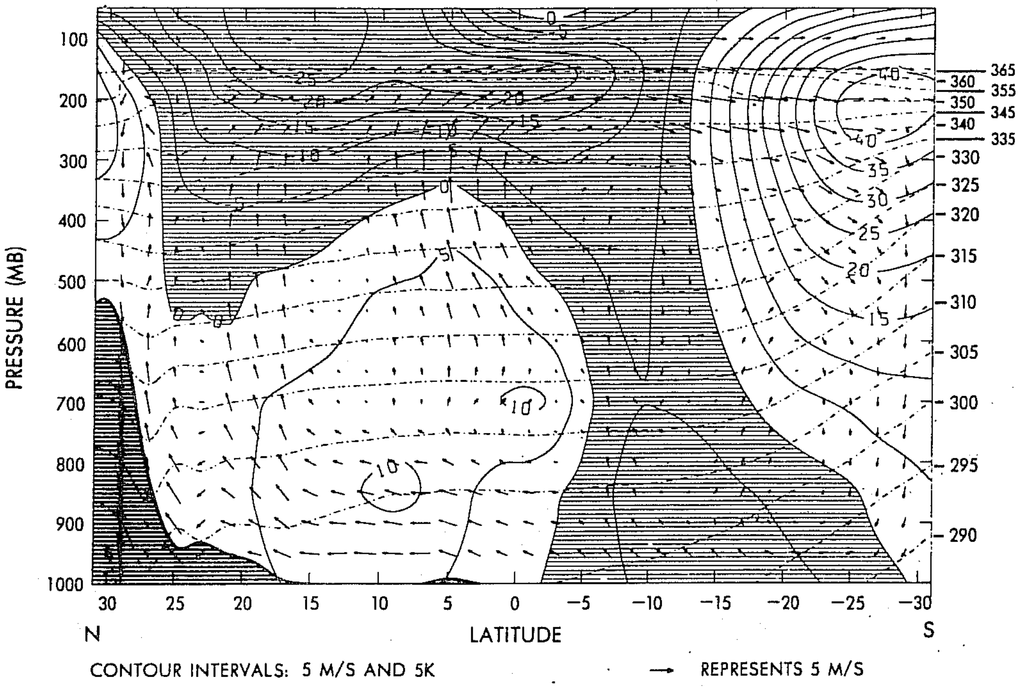
This used a global assimilation model with 11 sigma levels and a horizontal resolution of about 200 km (Saker 1975). Data assimilation was performed in six-hourly cycles, data from ±3 hours being used without any correction to allow for actual time of observation. Each time-step the model was relaxed towards the observed values, using weights pre-calculated using two-dimensional univariate OI. (I use OI as the accepted name for the statistical interpolation equations (Gandin 1963), preferring this to Optimal Interpolation since the use made of the weights is in no way mathematically optimal).

$$\psi_k(t) = \psi_k^*(t) + \lambda(t) \sum_i W_{ki}(\psi_i^o - \psi_i^*(t)) \qquad (1)$$

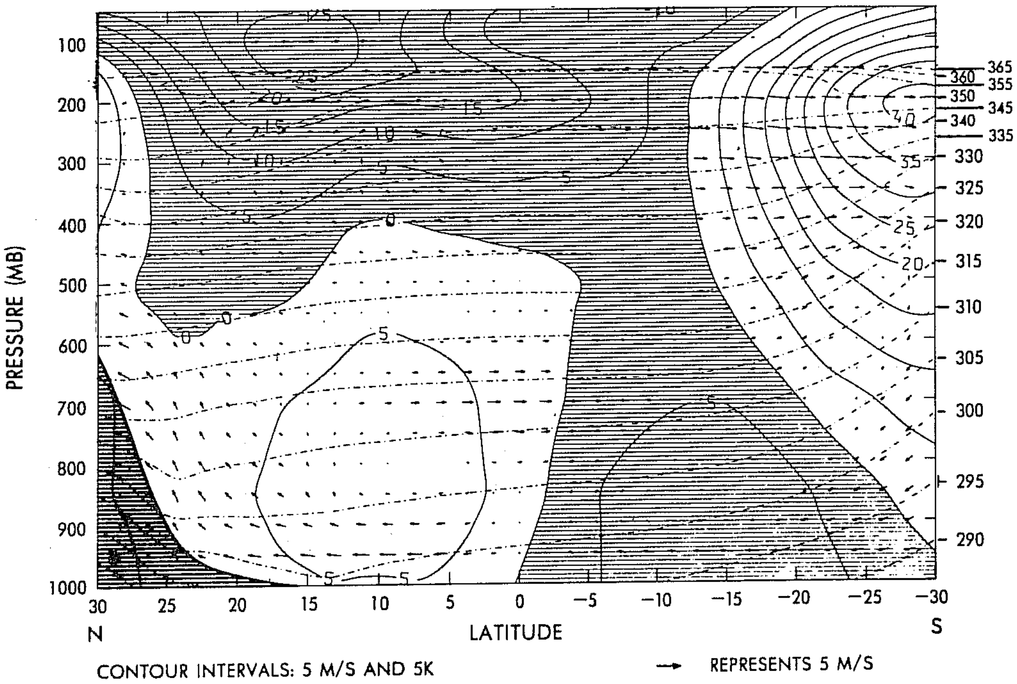
where $\psi_k^*(t)$ and $\psi_k(t)$ are grid point values at time t before and after insertion, ψ_i^o and $\psi_i^*(t)$ are nearby observed and equivalent interpolated values, and W_{ki} are the OI weights. The relaxation coefficient $\lambda(t)$ varied over the assimilation period between 0 and 0.5. Further details are given by Lyne (1981).

An example from the FGGE IIIa analyses made using this system is shown in figure 2, from Lorenc and Swinbank (1984). The point which I wish to illustrate here is that the analysed vertical velocity (computed directly from divergence) is consistent with the diabatic heating over this region during the monsoon period, unlike that from the ECMWF adiabatic non-linear normal mode initialization (figure 3).

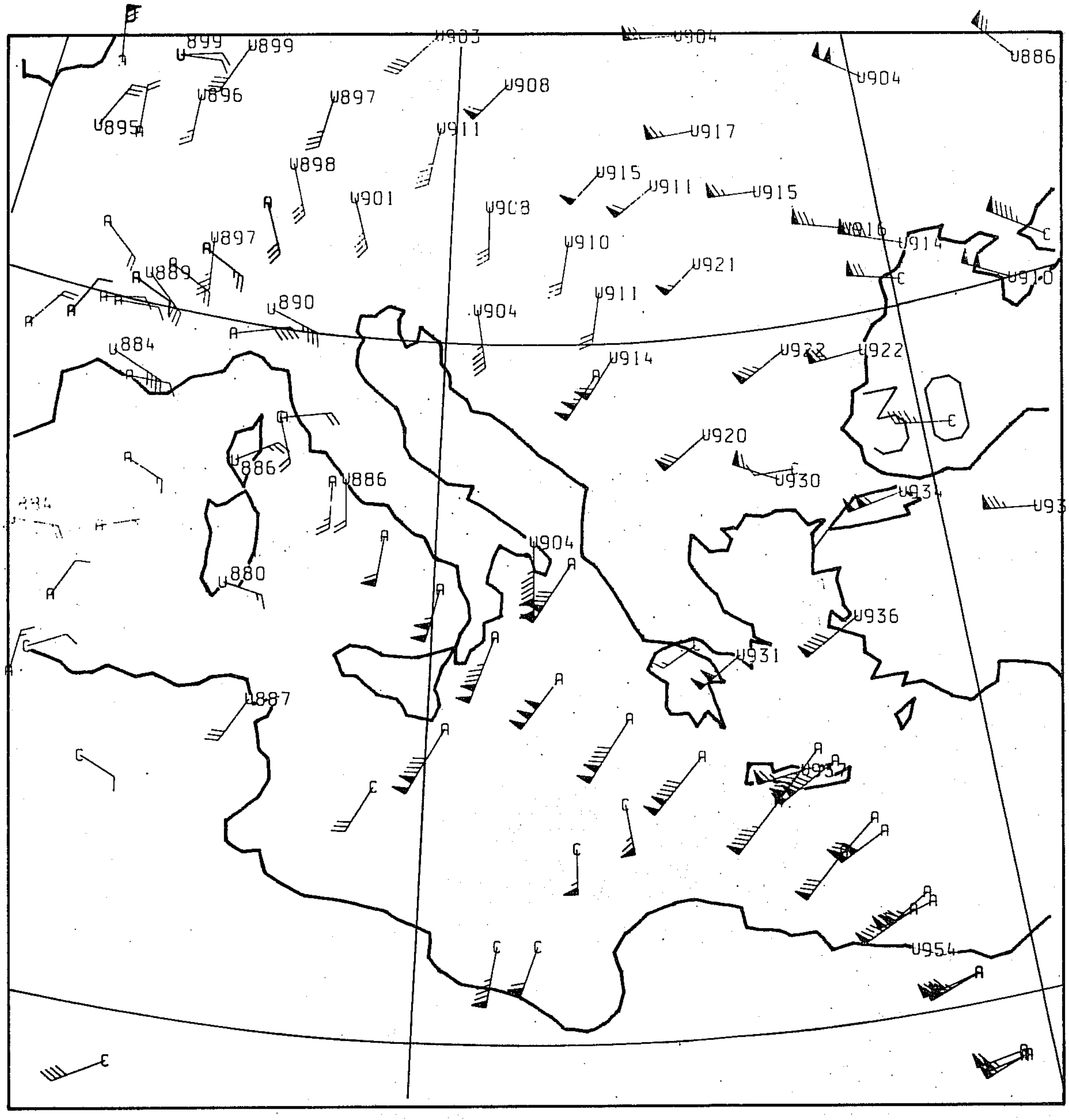
The FGGE analysis system was also used for OSE (Barwell and Lorenc 1985) and analysis system intercomparisons (Hollingsworth et al 1985) using FGGE level IIb observations. Various parameters were modified in this revised system, including the OI assumed error statistics and the relaxation coefficient λ.



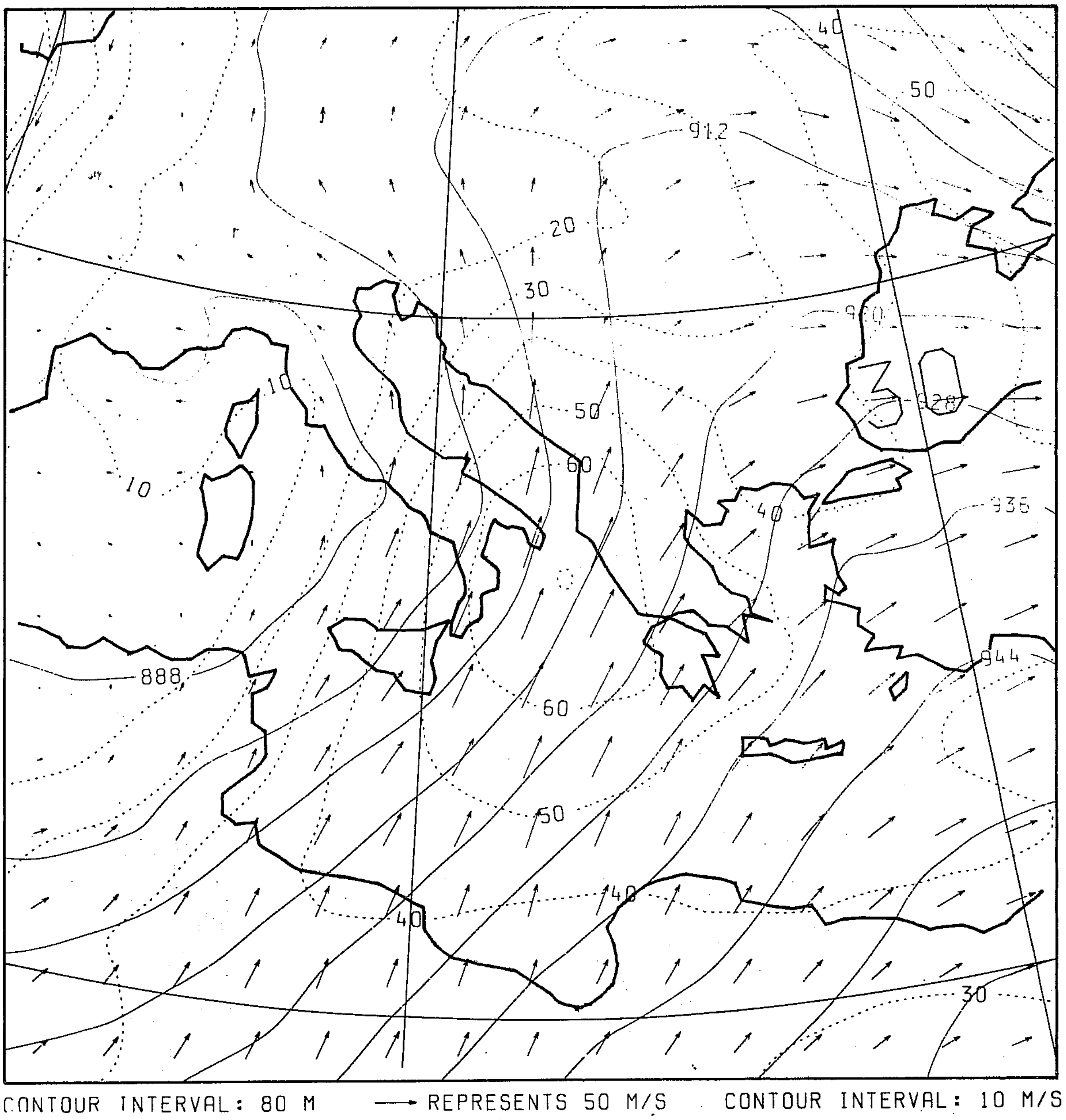
2. North-south vertical cross section from the Himalayas across the Bay of Bengal and the Indian Ocean, meaned for July 1979 and longitudes 80 E to 100 E, using the Met Office FGGE IIIa analyses. Showing u-component (solid contours, negative shaded), v and w (arrows), potential temperature (pecked), and model topography (heavy shading).



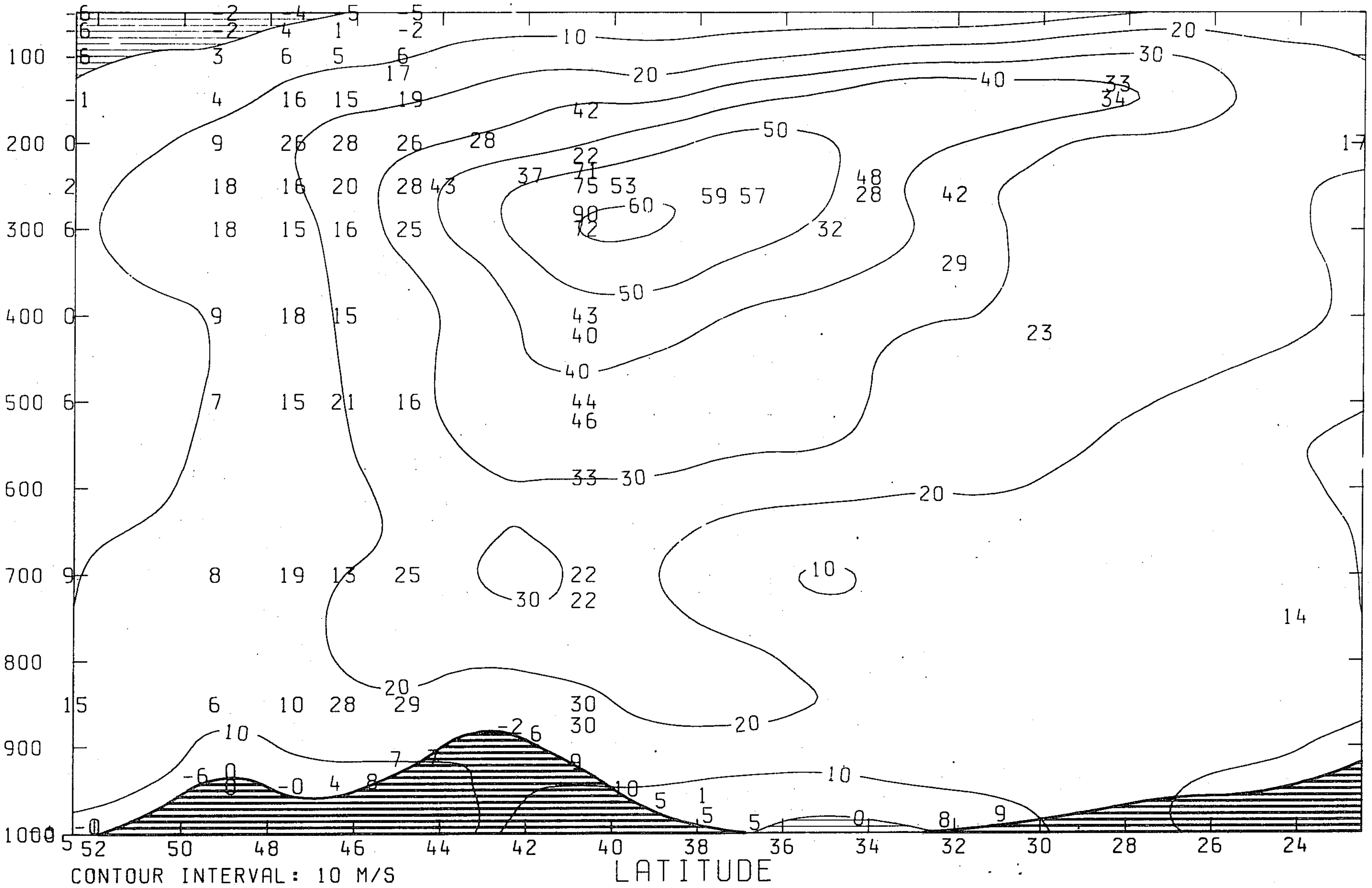
3. As 2 for ECMWF IIIb analyses.



4. Observational data over the central Mediterranean in the six-hour period centred on 00 GMT, Feb 17 1979 between 250 and 350 mb.



5. The Met Office FGGE IIIb analysis at 300 mb corresponding to figure 4. Solid lines are geopotential contours, dotted lines are isotachs, and arrows indicate wind direction and strength.



6. Cross-section of meridional wind component, averaged between longitudes 17E and 22E, and relevant observational values, corresponding to figures 4 and 5.

An example from the latter study is shown in figures 4,5,6. Figure 4 shows the upper air observations for a case where there was an active trough over the Mediterranean. There is some disagreement between observations of different type. The analysis (figure 5) has fitted most observations quite closely, and achieved an approximate dynamic balance more complex than geostrophy:- the conversion from potential to kinetic energy (which can be judged from the angle between the wind vectors and the geopotential contours) balances the acceleration (which can be judged from the distance between isotachs along the direction of flow). Figure 6 is a cross-section through the analysis; this also shows the rather good fit to somewhat contradictory data. This and similar sections show that the flow over the mountains is consistent with what might reasonably be expected.

3.3 Operational system

An operational version using the 15-level model on the Cyber 205 computer was completely recoded, using the same basic design as the FGGE system, but taking the opportunity to add various refinements to the process of relaxing the model towards the observations. These were:-

(1) Three-dimensional OI equations were used. However in practice this is restricted by data selection criteria which mean that soundings are used two-dimensionally.

(2) Multivariate OI equations were used. However in practice this was found not to be cost-effective, and is only used for wind components near the poles.

(3) From the correction to the surface pressure, corrections to the potential temperature fields in the troposphere were calculated using the hydrostatic equation in such a way that the net change to the geopotential in the lower stratosphere was zero. This was found to aid the assimilation of surface pressure observations. The correction was made in addition to that from temperature observations.

(4) From the observed corrections to surface pressure and temperature, in the extra tropics, a balancing correction to the wind fields was calculated using the geostrophic relationship. A fraction of this was added in addition to that from wind observations. However it was found that, because of the

roughness of the temperature correction fields (itself caused by practical limitations on the search radius for observations), only a small fraction (about $^1/_4$) of this could be added without generating noisy analyses.

(5) The model equations were modified by a term

$$\frac{\partial v}{\partial t} = \ldots\ldots\ldots\ldots -k\underline{\nabla}\,(\underline{\nabla}.\underline{v})$$

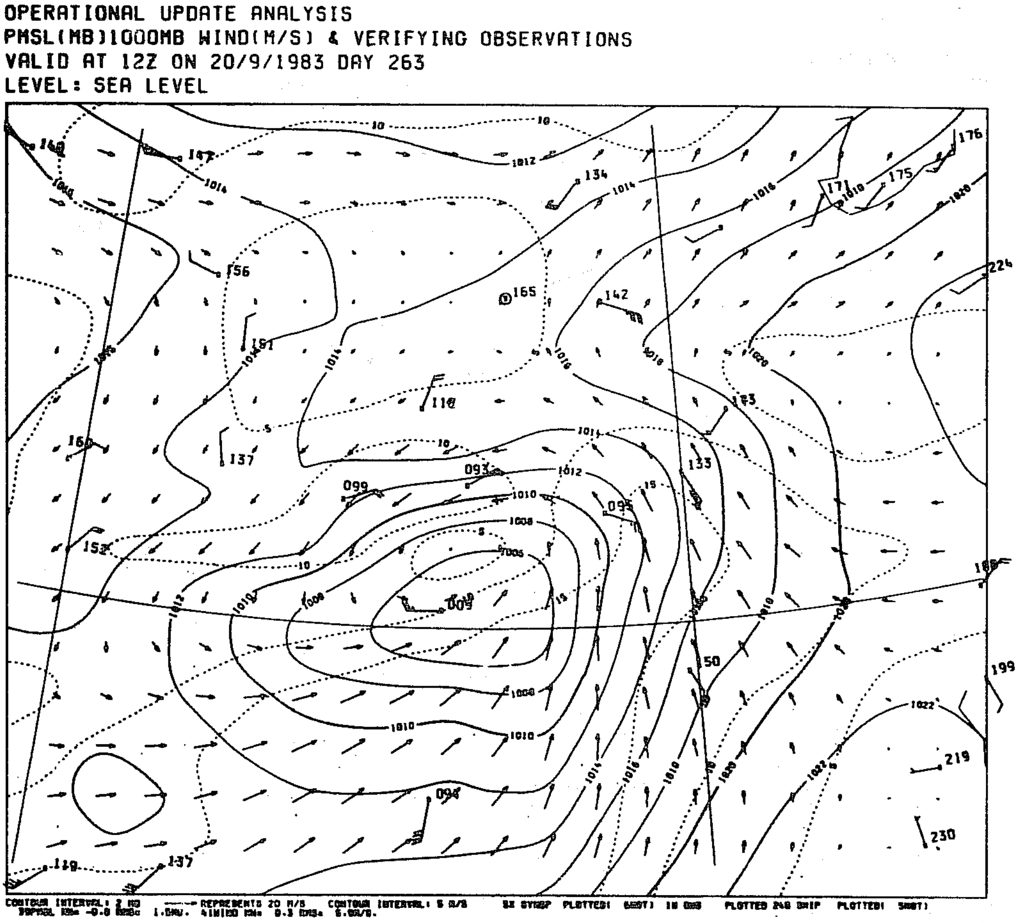
to damp out excessive divergent modes excited by the assimilation.

Additionally the time-stepping, relaxation coefficieints, and OI assumed error statistics were all modified. Further details may be found in Lyne et al (1983).

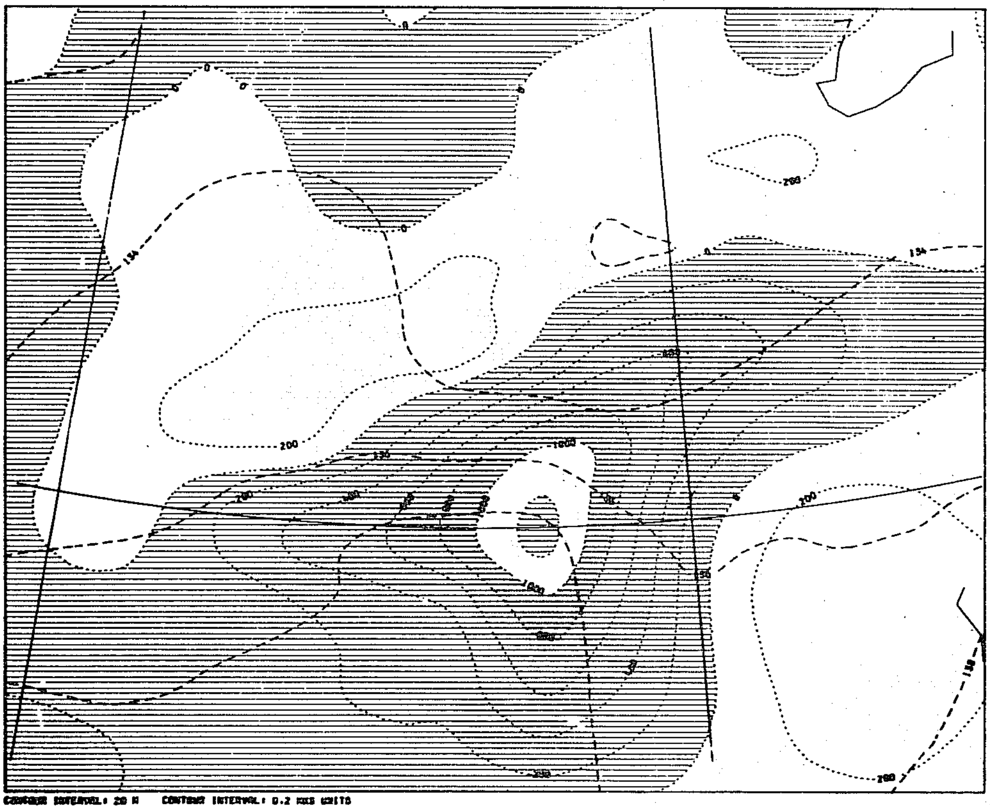
This system has been used for OSE's using FGGE data (Bromley 1984). An example analysis from operations is shown in figures 7 and 8. This case was originally chosen for study because the operational forecast was poor (Young 1984). I use it here simply to illustrate the more marked frontal structure in our analyses, compared for instance to those from ECMWF (figures 9 and 10). This characteristic has been observed in many other cases.

A fine-mesh version of the assimilation system is currently being tested. At present we do not envisage having a completely separate fine mesh assimilation cycle, but rather to start each run from an interpolated field from the global cycle field valid 12 hours earlier. Thus 12 hours of assimilation of high resolution observations (in 3-hourly batches) into the fine-mesh model preceeds each forecast. Figures 11 and 12 show a test run of this system for the case just discussed. The frontal structure is even more pronounced. This analysis lead to a significantly better forecast.

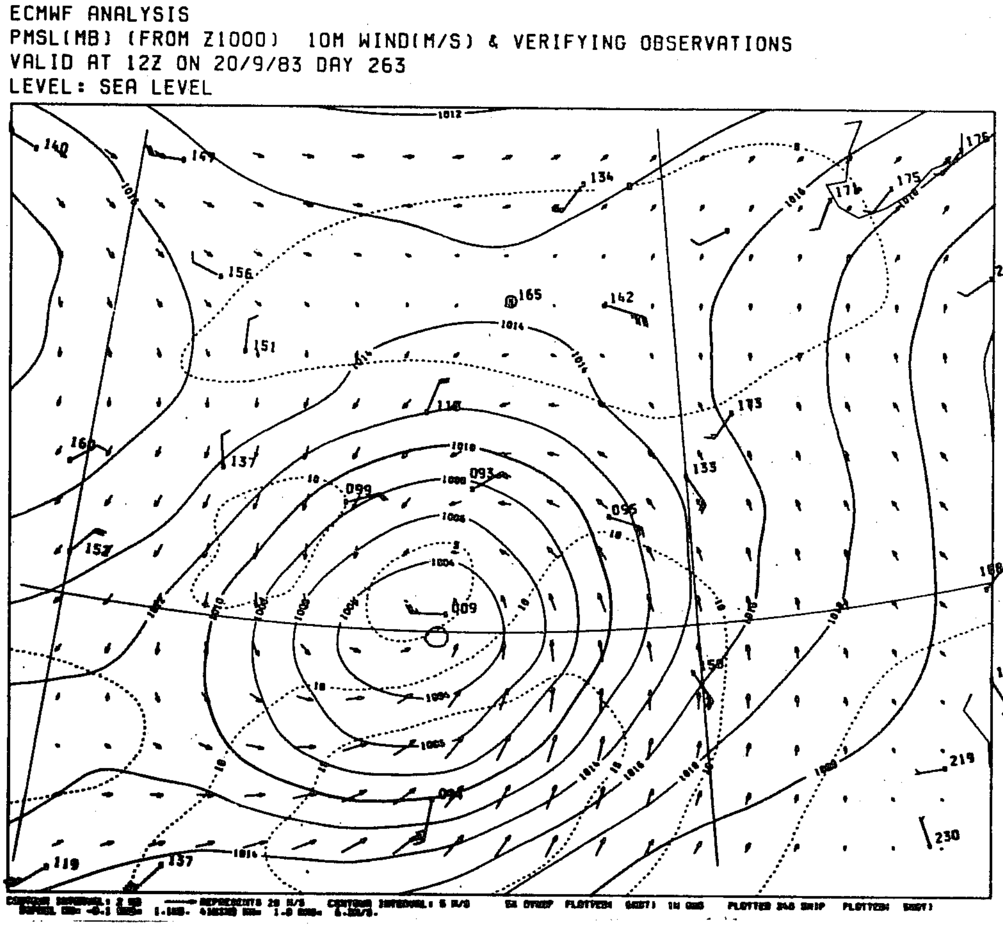
Case studies such as these are the best way of getting an insight into the effects of different aspects of an analysis scheme. However they are time consuming, and might not be representative. So for my last example I use a different approach, and just present gross statistics for a recent case chosen at random. Figure 13 shows the results of verification against observations for Met Office and ECMWF background fields and analyses.



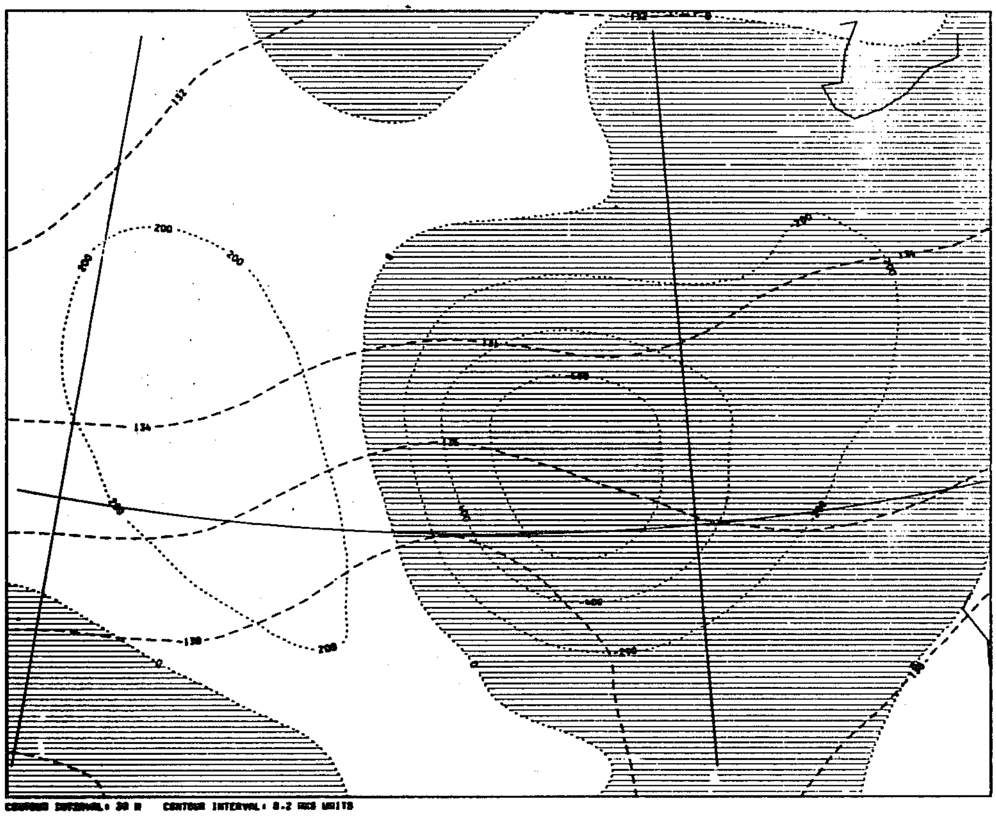
7. Surface observations, and corresponding Met Office operational analysis, for a depression approaching the SW UK at 12 GMT 20 Sept 1983.



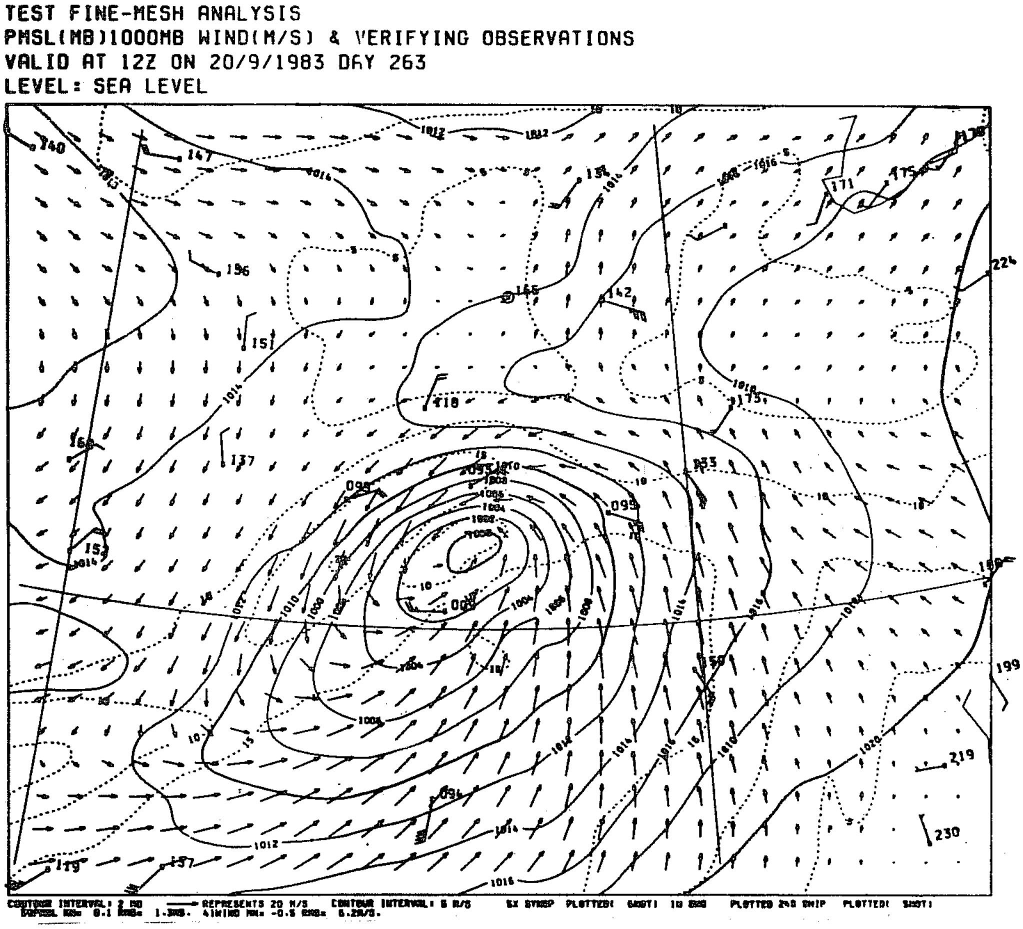
8. 1000–850 mb thickness, and 700 mb vertical motion (upward motion shaded), for the analysis shown in figure 7.



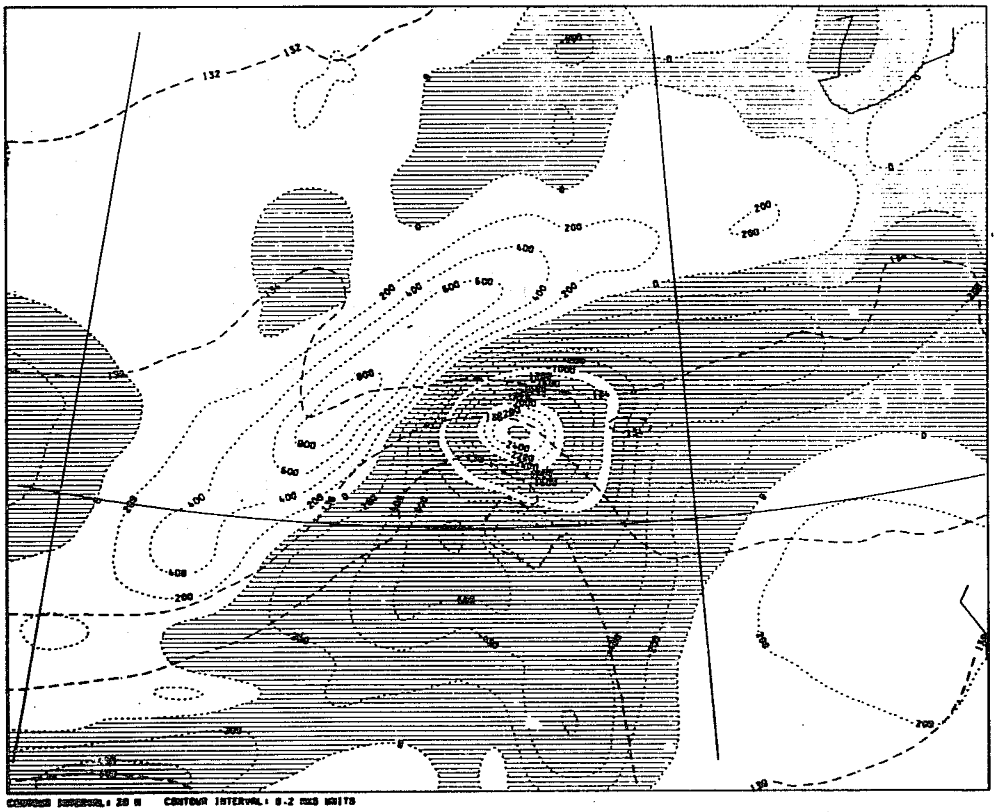
9. As figure 7 for operational ECMWF analysis.



10. As figure 8 for operational ECMWF analysis.



11. As figure 7 for test fine-mesh analysis.



12. As figure 8 for test fine-mesh analysis.

RADIOSONDES			RMS				MEAN				
LEVEL			6HR FC		ANAL.		6HR FC		ANAL.		
MB		N	MO	EC	MO	EC	MO	EC	MO	EC	
100	Z	509	60	50	62	38	32	17	44	7	M
250	Z	571	46	38	45	26	28	16	34	3	M
500	Z	590	30	21	26	14	18	7	17	-1	M
1000	Z	564	22	22	15	19	8	-4	2	-5	M
250	T	573	2.2	2.0	1.3	1.7	-.3	.2	.0	.1	K
500	T	591	1.7	1.7	1.3	1.4	.3	.2	.6	-.1	K
700	T	583	1.7	1.7	1.2	1.6	.4	-.0	.5	-.2	K
850	T	556	2.2	2.4	1.5	2.2	.3	-.3	.5	-.5	K
500	RH	579	25	29	18	25	-8	-16	-6	-14	%
700	RH	577	21	23	15	21	1	-10	0	-8	%
850	RH	553	20	22	13	19	2	-10	0	-6	%
SATEM (REF LEVEL 1000MB)											
-100	DZ	617	39	37	30	34	16	8	14	4	M
-500	DZ	617	33	23	24	19	18	9	16	5	M
250 WIND			VECTOR RMS				MEAN SPEED				
TEMP		532	8.4	7.5	4.8	5.3	.0	.6	.4	.4	M/S
T.SHIP		7	12.1	9.6	5.4	7.8	-.4	1.9	.3	2.6	M/S
PILOT		91	10.5	8.9	6.1	5.0	-.1	-.2	-.2	-.2	M/S
AIREP		464	11.2	9.9	8.7	9.1	.7	1.1	.0	.3	M/S
SATOB		51	9.9	8.8	7.2	6.0	-1.0	-1.6	-.1	-.9	M/S

GLOBAL VERIFICATION AGAINST OBSERVATIONS 12Z 7JUNE'84
6HOUR FORECAST BACKGROUND, & ANALYSIS FROM MET.O & ECMWF

13. Table showing mean and rms differences between observations valid at 12 GMT 7 June 1984 and the corresponding Met Office and ECMWF 6 hour forecast background fields and analyses.

4 PROBLEMS & PLANS

	PROBLEMS	PLANS
1.	QUALITY CONTROL OF OBSERVATIONS	SEPARATE FROM ASSIMILATION
2.	BIASED OBS & MODELS	CORRECT
3.	DATA SELECTION	
	HIGH DENSITY OBS	SIMPLIFY METHOD
	HIGH RESOLUTION IN SPACE & TIME	&
	COMPLEXITY OF VECTORIZED CODE	USE MORE DATA
	CUMBERSOME PRECALCULATED WEIGHTS	
4.	BEST USE OF SATELLITE SOUNDINGS	FILTER INCREMENTS
	'NOISY' ANALYSES	NON-DIVERGENT, BALANCED
	NON-METEOROLGICAL MODES	INCREMENTS
	UNREALISTIC INITIAL RAINFALL	SELECTIVE DIVERGENCE DAMPING
5.	HUMIDITY ANALYSIS	

14. Table summarizing problems and plans.

Statistics are only for one date, and only a crude preliminary quality control of observations was used, but some interesting features are still apparent:-

(1) For many observation types mean differences are significant. For height, thickness, temperature and relative humidity the comparison with ECMWF is dominated by these.

(2) The Met Office analysis fits the upper wind observations better, however this is not true of the background fields. It seems that the wind information is lost in the early stages of the forecast from the Met Office analysis. This effect was studied in detail by Barwell and Lorenc (1985). It has also been noted by Bromley (1984).

4. PROBLEMS AND PLANS

4.1 Summary

The major percieved problems in the analyses, and plans to deal with them, are summarized in figure 14. Problems 1 and 2 are largely unconnected to the repeated insertion analysis method; 5 is included for completeness and will not be discussed again in this paper. Both the practical problems of 3 and the more fundamental problems of 4 are being tackled by a redesign of the method of relaxing towards the observations. This new design is described briefly below. I then go on to give examples of and discuss the problems in more detail.

4.2 Planned new scheme

Little advantage is gained by using OI to relax towards the observations. Since the background contains observed information from earlier insertions, the OI assumed error statistics used to calculate the weights are inappropriate, and because of the factor λ the weights are anyway not used optimally. Use of the OI method leads to many of the practical difficulties listed in figure 14.2. The new design is therefore based on a simple correction method; the correction at each grid point is a weighted average of the deviations of the observations from the model, and the weights for each observation are dependent only on the position and type of that observation.

Because of the repeated insertion the analysis scheme is therefore rather similar to the successive correction method, (Bergthorsson and Doos, 1955; Cressman 1959). The simpler weights calculation allows us to do away with the complicated data selection which takes over half of the analysis computer time in the current scheme. Instead each observation influences all nearby gridpoints. This is achieved efficiently in 3 dimensions by first calculating a vertical column of increments at each observation position, then spreading these 2-dimensionally at each level.

4.3 Examples and discussion of problems

Figure 14 listed the problems seen in the analyses; I shall go through the list giving examples from the cases already presented in section 3.

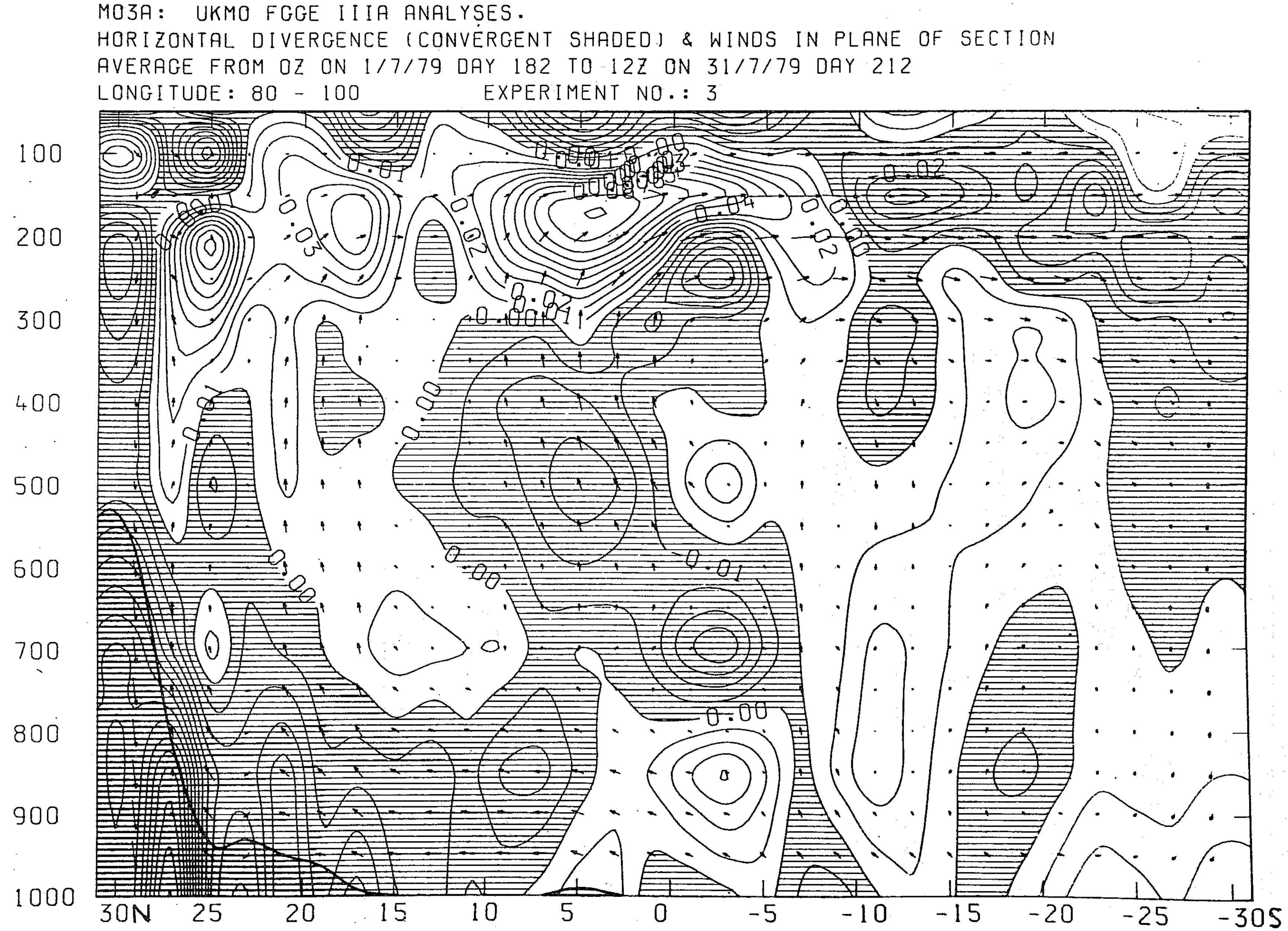
Quality control of the observations is an all pervading problem. Many of the cloud wind speeds shown in figure 4 were probably too low; the Met Office analysis which drew closely to them gave a much worse forecast than those from other schemes which did not. The main difference between the analyses of figures 7 and 9 seems to be associated with the acceptance or rejection of pressure and wind data from the ship near the centre of the low and that to the south of it. The repeated insertion analysis method cannot be readily modified to perform quality control, as can OI, so our longer term plans are to develop a quality control scheme using OI and independent of the analysis scheme.

Figure 13 showed that objective verification of mass and humidity data is dominated by mean differences. Some of these are instrumental; the Met Office analysis system applies correction factors to radiosondes because of this. These corrections are not applied in this verification table. Their calculation is difficult as the models themselves hae significant biases which affect both background and analysis fields. For instance a comparison of the Met O and ECMWF analyses for this case show zonal mean differences of about 2 dam in the 500 mb height, 4 m/s in the upper tropospheric tropical easterlies, 2 m/s in the N hemisphere jet and 4 m/s in the S hemisphere jet, 4 m/s in the upper branch of the Hadley cell, 2K in the tropical tropopause temperature, 1 K in the mid latitude tropospheric temperature, and 15% in the tropical humidity. These differences are almost the same in the background fields. Use of observations and backgrounds with different biases can lead to erroneous gradients in the analysis which make calculations of 'balance' more difficult.

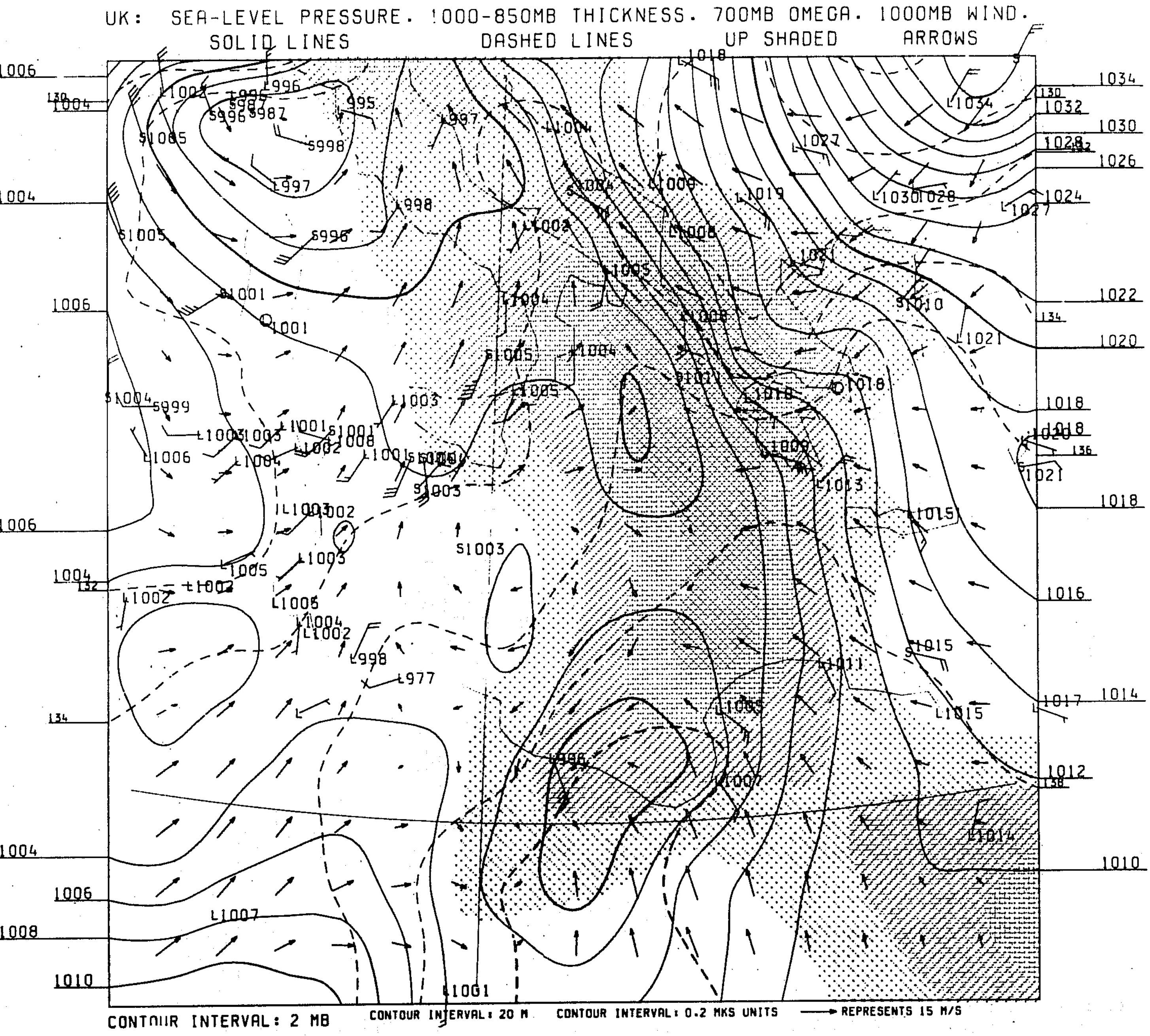
Near the UK there can sometimes be a large number of observations which might influence the analysis at a grid point, and now observing systems such as the Hermes temperature soundings (resolution 50 km) will increase this. Unfortunately much of this high density data is biased or for only one level, and there is a need to ensure more distant radiosonde observations are still used to correctly analyse synoptic scales. There are thus large practical difficulties in including all the data into an OI analysis scheme; either complex search algorithms or a very large matrix must be used. As discussed in 4.2, our approach to this is to go for a very simple method, which can be applied to all data, and to iterate using repeated insertion to get the interactions between observed information that OI calculates explicitly.

One problem that OI can tackle is the systematic error structure in satellite temperature soundings. By assuming correlation structures in the vertical and horizontal for observational errors, OI can calculate weights that should still extract useful information from the observations. For Hermes data these systematic errors can be significant, however since the data density is similar to the model grid we can use the OI formalism and assumed error structures to pre-calculate filters to the observational increment field which have an equivalent effect to a full OI analysis.

This filtering of increments should be useful in tackling some of the other problems of figure 14.4. For instance Barwell and Lorenc (1985) showed that it aids assimilation of wind data if most of the information goes into non-divergent modes. One problem of the repeated insertion method is that slowly varying modes of the model are permitted even if we do not believe that they are likely atmospheric modes. This was seen in the small equivalent depth inertial-gravity modes studied by Barwell and Lorenc (1985). It can also be seen in Figure 15, which shows the July mean cross section of divergences equivalent to figure 2. There appear to be standing waves in the divergence, perhaps associated with the mountains. Another example is given



15. North-south vertical cross-section of divergence and wind corresponding to figure 2.



16. Surface observations and analysis corresponding to figure 5.

in figure 16, which shows the surface analysis corresponding to figure 5. The strong vertical motions induced by the strong jet entrance and the nearby mountains has excited a two-gridlength wave in the model's surface pressure which is stationary under the model's finite difference scheme. In this case this is clearly un-meteorological, although it does not actually disagree with any observations.

Although, by filtering and balancing of increments, we can hope to reduce the excitation of these modes during the data insertion process, we cannot completely prevent it since we wish to be able to analyse real small scale phenomena with similar characteristics. More study is needed of which modes are in fact 'unmeteorological', and how the atmosphere actually damps them. Until recently forecast models ran from initialized initial conditions and lacked the small scale forcing to generate such modes, hence parametrization in models of the mechanisms by which the atmosphere achieves balance was not necessary, and is lacking. One such parametrization is the divergence damping of equation 2. We plan to modify this to make it more mode-selective.

References

Barwell, B R and A C Lorenc 1985 A study of the impact of aircraft wind observations on a large scale analysis and numerical weather prediction system. Quart. J.R. Met. Soc. 111 (to appear).

Bergthorsson P and Doos B.R. 1955 'Numerical weather map analysis' Tellus 7 329-340.

Bromley, R. 1984 Observing system experiments using the Meteorological Office's 15-level model.
ECMWF Seminar on Data Assimilation Systems and Observing System Experiments.

Cressman, G.P. 1959 'An operational objective analysis scheme'. Mon. Wea. Rev 87 367-374.

Gandin, L.S. 1963 'Objective Analysis of Meteorological Fields'. Gidrometeorologicheskoe Izdatel 'stwo, Leningrad. Translated from Russian, Israeli Program for Scientific Translations, Jerusalem, 1965, 242pp.

Hollingsworth, A., Lorenc, A.C., Tracton, M.S. Arpe, K., Cats, G., Uppala, S., Kallberg, P.
'The response of Numerical Weather Prediction, Systems to FGGE IIb Data, Part I: Analyses. Quart. J.R. Met. Soc 111 (to appear).

Lorenc, A.C. 1976 'Results of observing system simulation experiments for the First GARP Global Experiment'.
The GARP program on Numerical Experimentation Report No. 10. WMO/ICSU pp. 37-68.

Lorenc, A.C. and R Swinbank 1984 On the accuracy of general circulation statistics calculated from FGGE data - A comparison of results from two sets of analyses.
Quart J Roy Met Soc 110 (to appear).

Lyne, W H 1981 'The Meteorological Office's data assimilation scheme for FGGE'.
International Conference on FGGE data analysis and Results (Bergen).
ICSU WMO GARP, Geneva 222-227.

Lyne, W.H., Little, C.T., Dumelow, R.K. and Bell, R.S. 1983 The operational data assimilation scheme. Met O 11 Tech Note 168.

Lyne, W.H., Swinbank, R. and Birch, N.T. 1982 A data assimilation experiment and the global circulation during the FGGE special observing periods. Quart J R Met Soc 108 575-594.

Saker, N.J. 1975 An 11-layer general circulation model. Met O 20 Tech Note II/30.

Young, M 1984 A case study of course mesh fields from data assimilation experiments carried out for 12Z 20 September 1983. Met O 11 Tech Note 192.