2. ANALYSIS FOR GLOBAL OR HEMISPHERIC MODELS

2.1 Introduction

Many operational and research groups now have the facility of performing NWP using large scale models, each with the common requirement of a satisfactory objective analysis/data assimilation procedure. There is often a commonality in the basic data sources for such analyses, and in the methods of analysis. Consequently, many of the problems, both scientific and technical, which are encountered are common to many groups.

In the review which follows we seek to identify the major problems of large scale extra-tropical analysis which are considered to be of general interest.

In the recommendations that follow the review we address ourselves specifically to the Centre's requirements.

2.2 Review of current problems in large scale analysis

Many of the problems described here have been referred to in the workshop presentations.

2.2.1 Data availability

Overriding all considerations of analysis technique is the problem of data availability. The lack of sufficient good quality data is considered to be the main source of error in large scale analysis.

2.2.2 Statistical requirements for O.I.

O.I. continues to be the most favoured analysis technique. Being a statistical technique, a basic requirement is knowledge of structure functions, first guess error and observational error statistics. Although the technique has been established for several years now, most users of the technique are currently using basic Gaussian horizontal structure functions, gross approximations to vertical structure functions and first approximation estimates of the first guess errors. Observational errors ascribed in different applications of the technique are also seen to have significant differences.

The weaknesses of the Gaussain function are described by Lönnberg and the need for a better specification of vertical correlation profiles is discussed by both Andersen and Lorenc.

It is clear from the work presented at the workshop that an increase in the effective analysis resolution, both in the horizontal and vertical depends on the specification of the structure functions.

Noting that few centres, if any, have yet produced satisfactory revisions of these basic statistical parameters, it may be a little early at this stage to pursue some of the more ambitious algorithms of dynamically varying structure functions. Procedures to modify first guess error variances according to synoptic situations (e.g. areas of rapidly developing lows, jet streams) may be more fruitful.

2.2.3 Quality control

Quality control is a major factor in successful objective analysis. Many of the major discrepancies in different analyses applied to the same data stem from quality control decisions. Whilst analysis techniques are described in the scientific literature, the details of quality control procedures are usually insufficiently publicised. This is particularly true of pre-analysis quality controls. Many of the latter have been developed at individual centres (some examples are extreme checks, lists of representative and reliable surface wind stations, ship monitoring and correcting, alternative station lists).

Most operational centres consider it necessary to perform some manual control of the data.

As regards the data checking within the analysis, a common practice is to provide an independent estimate at an observation location and discard the observation if it departs from this estimate by more than a specified deviation. There may be a case for this deviation to be a function of not only the estimated observational error but also a probability of gross error. The latter might be decreased if the observation has positively satisfied quality control checks such as time continuity.

2.2.4 Fit of the analysis to the data

In particular analysis examples, observations are frequently seen to have a very different fit to the analysis than their ascribed rms error. It may be possible to accomodate, within the data checking procedure of the analysis, a tuning of the first guess error to achieve a fit to the data which is consistent with the ascribed rms error. The mathematical technique of cross validation would be one basis for such a procedure.

2.2.5 Analysis of jet streams

Hollingsworth highlighted the diffficulty of analysing jet streams as evidenced in the disparities arising in three different analysis procedures presented with the same data, and the sensitivity of downstream predictions to the initial state in the vicinity of such features.

A direct (univariate) insertion technique is shown by Lorenc to have conflicting requirements in the vicinity of a jet - the need for a short insertion period to avoid downstream propagation of the data and the need for a long insertion period to achieve a balanced state.

Single level data are not fully exploited in current analysis schemes.

A dichotomy exists in the use of wind data in analysing departures from a model forecast field. Should the wind being analysed be regarded as a mean over a vertical layer (in the model sense) or a value appropriate to particular level, in the sense that it is typically balanced with a geopotential increment ? The problem is relevant when deciding what observational wind data should be presented to the analysis.

2.2.6 Analysis of the PBL

An adequate analysis of the PBL is considered necessary, notwithstanding the lack of sensitivity found in the predictions described by Hollingsworth. Some improvement arising from improved PBL analyses in tropical predictions is reported by Andersen and the increasing attention being paid to surface analyses in large scale models probably requires an adequate description of the PBL. It also has use in diagnostic studies.

2.2.7 Overall vertical resolution of large scale analyses

The isentropic analyses of Reimer show considerable gain in the ability to resolve atmospheric stratifications. This may stem from the enhanced number of levels in the vertical relative to the base model (24 compared with 9). The availability of high vertical resolution data from many radiosondes, together with the need to exploit fully non-standard level data such as SATOBS and AIREPS, suggests that relatively high vertical resolution analyses could be produced to some advantage.

2.2.8 Avoidance of unnecessary vertical interpolations

Several groups in the past have worked with p-level analyses and σ-coordinate prediction models, performing the necessary vertical transformations before and after the analysis, usually using a cruder mathematical from of interpolation than is used in the p-level analysis. Clearly this is undesirable and several groups are developing means of avoiding unnecessary interpolations. It is relatively simple to eliminate the $p \rightarrow \sigma$ interpolation following the analysis, by evaluating the analysis increments directly on the model surface. Avoidance of the $\sigma \rightarrow p$ interpolation of the first guess is more difficult but is thought to be both feasible and desirable.

2.2.9 Orographic considerations

All large scale analysis schemes, be they p-coord, σ-coord or of other vertical form, are thought to have problems in regions of marked orography. Presumably the structure functions appropriate to these regions differ from those in the free atmosphere.

Another common problem in regions of orography is the use of extrapolated information, in particular pmsl. Apart from the inaccuracy introduced by extrapolation of the original observed datum, the observation is typically expressed as a departure from the model first guess state which is also extrapolated.

A requirement to be borne in mind is that the analysis model should be the same conceptually as the prediction model. If an observation is made in the bottom of a valley not resolved by the model, the observation should not be used. Algorithms have been developed for the Centre's assimilation which meet this requirement in an approximate way.

2.2.10 The local nature of O.I.

All applications of O.I. are localised. Typically analyses are performed grid point by grid point. In the Centre's assimilation this is extented to analysis boxes 660 km square. The results presented by Cats suggest that these localised analyses may not be adequate and that one should seek to analyse the global modes separately from the localised O.I. It is not clear how real a problem this is. Further investigation is necessary.

2.2.11 Alternative sources of data and analysis variables

Most large scale analysis schemes exploit the same basic observational types.

SYNOP/SHIP, TEMP, PILOT, AIREPS, SATOBS, SATEMS, DRIBUS. There are alternatives, some examples of which were referred to in the workshop presentations - use of cloud top temperatures to infer divergence (Julian) ; and possible use of satellite radiance data to analyse humidity (Pailleux). Both are novel and merit further investigation. There is also the possibility of other analysis variables. It may be more important to analyse stability than geopotential, the latter requiring a double differentiation to infer the former. This may be particularly relevant in the tropics.

2.2.12 Diagnostic tools

Experience at ECMWF has shown that ability to identify weaknesses in the analysis scheme is very dependent on having appropriate diagnostic tools. Some facilities have proved very successful in highlighting errors - most notably the combined display of increment fields and suspect data ; rejection summary displays ; and the fit to observation statistics of the first guess, analysed and initialised fields. (The fit of observations to analysis can be a dubious measure of analysis quality, as has been found in FGGE intercomparison work). Direct insertion assimilation can yield a good fit but a bad use of data . Other centres also have diagnostic facilities to monitor the performance of their analyses. A pooling of ideas on this subject would be of mutual benefit.

2.2.13 Humidity analysis

There is no clear evidence that further effort on the analysis of humidity in mid-latitudes would bring significant improvements in the forecasts. The question is more important in the tropics and is discussed in section 4.

2.3 Recommendations

These recommendations are approximately ordered by importance.

2.3.1 Statistical functions and resolution

This is considered to be the overriding current requirement for the Centre's assimilation technique. It is noted that accumulation of statistics is now being done and that revisions of structure functions, first guess errors and observational errors (at least for radiosondes) will be feasible in the coming months. It is recognised that the effective resolution of the O.I. analysis system is controlled by these functions.

2.3.2 Quality control

All possible tools for quality control of data should be developed. In particular, quality control at the pre-analysis stage is emphasised. Time consistency checking is an obvious example. There should be a comprehensive comparison of the pre-analysis quality control procedures employed at different centres, including an exchange of information on flagging rates. This may best be achieved by a separate workshop on the subject, or by recruitment of a consultant. Some centres devote much greater resources to manual rejection and correction of data than does the Centre. A pilot study of the benefits of manual rejection and correction should be done by simply comparing such information with the Centre's automated quality control.

An intercomparison of decoded data base contents, particularyly Centre v. Bracknell and Centre v. Offenbach would be beneficial.

2.3.3 Increased vertical resolution

The current work on the 19-level analysis is regarded as desirable and likely to lead to some improvement in predictive skill. Consideration of further increases in vertical resolution is also recommended, noting the need to adequately resolve jet stream features. This should be considered in conjunction with plans for increased horizontal resolution, and would of course require a corresponding structure in the prediction model.

2.3.4 Increased temporal resolution

It is noted that a 3-hourly assimilation revision of the current scheme is currently being evaluated. This is seen as a logical and desirable extension of the current scheme which should exploit the asynoptic data much more effectively.

2.3.5 Elimination of unnecessary vertical interpolations

The current assimilation scheme is thought to be degraded by the double transformation of model and data before and after the analysis. There is a need to eliminate such interpolations where possible, in the manner described by Pailleux for example.

2.3.6 Data monitoring

As distinct from actual intervention, the process of monitoring data to identify particular problems in its use is regarded as important. The most comprehensive way of doing this is by accumulation of statistics on departures of data from first guess, flagging rates, etc. Such statistics are now being accumulated for radiosondes at the Centre, and need to be extended to other data types. The information provided from such a study will be essentially a post-mortem activity based on a monthly review. Consequently the need for a daily operational monitoring remains important.

2.3.7 Modification of first guess errors by synoptic situation

This is recommended for further consideration. Some modest studies could be undertaken in clear examples of rapid synoptic development and in jet stream analyses.

2.3.8 Analysis of additional fields

There is a general feeling that pressure tendency information should be a useful tool in data assimilation, apart from the obvious time-consistency aspect. What is unclear is the best way of incorporating it fully in the main data assimilation cycle. It is recommended that an analysis of $\frac{\partial p_*}{\partial t}$ be produced operationally, using the initialised state as a first guess, but that in the first instance this analysis should not interact with the main assimilation. An accumulation of such fields could provide a useful set of case histories which will reveal problems in the use of the data and perhaps show how it is best accommodated in the main assimilation.

Noting that the Centre also has a package for anlysing surface parameters which should be implemented in the near future, it is considered desirable to separate out the rainfall and evaporation analyses involved in the SWC analysis since these too could prove useful in future diagnostic studies. There may also be the need for an analysis of such fields as 2 m temperature, even if such fields are not an interactive part of the assimilation scheme at present.

2.3.9 Error structure in the initialised fields

The iterative procedures described by Daley raises the question of whether the Centre's assimilation has extracted all the usable information after an analysis/ initialisation cycle. This could be checked by looking at the spatial correlations (of observed minus initialised or analysed values). If there is appreciable structure it suggests the need for a further iteration of some kind. The computation is a minor extension of that involved in the derivation of the analysis structure functions and appropriate accumulation of statistics is already taking place.

2.3.10 Data fitting

As discussed in 2.2.4, the possibility of adjusting first guess error fields to improve the fit to observations by the analysis should be considered further, with perhaps a few test cases being investigated.

2.3.11 Use of alternative data types

Practical alternative forms of data not currently used in the analysis should be reviewed.

2.3.12 Extrapolation of surface pressure

Stations measuring surface pressure should be asked to include such information in their report, in addition to any extrapolated value such as mean sea level pressure, in those cases where the station height is significant.

2.3.13 Inappropriate data

Data which is not conceptually consistent with the prediction model, in the manner described in 2.2.9, should not be used in the analysis.

3. ANALYSIS FOR FINE MESH MODELS

3.1 Review

Forecasting on the meso-alpha scale is just becoming feasible. With improved observation platforms, better models and more powerful computers it is possible, in principle, to produce high resolution ($\simeq$ 50 km) forecasts for Europe. At this time, however, most components of such a forecast system - analysis, initialisation, model are in their infancy. Some experimentation has already been done with limited area models (LAM) which serve to spotlight the enormous complexities of the problem. However, the difficulties are no more insuperable than those facing our predecessors thirty years ago when they first contemplated large-scale numerical forecasting.

Fine mesh analysis is a much more difficult problem than large-scale analysis. Quasi-geostrophic theory, which is so helpful in the large-scale problem, is of less assistance here. In fact, what theory is available on these scales, suggests that there are few useful approximations which can be made. Thus, the divergent wind component is much more important than on the large-scale and it may even be necessary to analyse some gravity modes. On the other hand, a great deal of experience has been obtained in large-scale analysis which may be helpful in the design of fine-scale analysis systems.

3.2 Discussion

The discussion that follows will proceed logically through the observation-analysis-initialisation cycle considering each component separately.

3.2.1 Data

The first requirement for a good analysis is good data. The first priority is accurate, high vertical, horizontal and temporal resolution wind measurements. The observing system should be able to determine both vorticity and divergence. Temperature and humidity observations are also desirable, but are given a lower priority.

The current upper air radiosonde network is too sparse (especially for measuring the boundary layer structure), too infrequent and too inaccurate for good mesoscale analysis. More sophisticated methods of analysing the existing upper air information would be helpful. High resolution SATEMS will be helpful, but do not address the wind problem. Ground or space-based lidar measurements would be invaluable if sufficiently accurate and dense. Indirect information can be inferred from radar and satellite pictures but objective procedures for using this information are still quite primitive.

A frequency of between one and three hours would be desirable.

Observation system experiments (OSE) with both real and simulated data should be performed with mesoscale models to assess the value of existing and proposed mesoscale observing systems.

Known data sets of high quality such as the Co-operative Convective Precipitation Experiment (CCOPE) and ALPEX should be available and used to benchmark analysis procedures.

3.2.2 Data checking

Data checking is just as important on the mesoscale as on the large-scale. Relationships between variables which can be used for consistency checks are more complicated on the mesoscale than the geostrophic relationships generally used on the large-scale.

As in the early years of large-scale analysis, manual intervention may be necessary. Hopefully, the need for manual intervention should disappear eventually, as automatic procedures become more sophisticated.

Manual intervention could be used in two ways - first for direct data checking and secondly for improving the pattern of the analyses in data poor areas (the data-voids in a mesoscale forecast over Western Europe are relatively as important as the large Southern Hemisphere voids to a large-scale global forecast). With modern interactive terminal systems and careful training on a simulator, manual intervention could be made quite useful. It is again stressed, however, that manual intervention is a short-term solution and in the long run automatic methods are preferable.

Automatic data checking procedures should not be ad hoc, but should have a sound theroretical basis. Iterative and cross-validation methods might be used.

3.2.3 First guess

A data assimilation cycle which uses a model generated first guess field (as opposed to a climatological first guess) is preferable. This first guess field should be as accurate as possible on all scales which are important. This may require linear combinations of global analyses or forecasts of the large-scale with LAM forecasts of the smaller scales. Scales for which the predictability time scale is short may be filtered. As yet, however, there do not appear to have been any good predictability studies in the mesoscale.

3.2.4 Analysis procedures

At this early stage all analysis procedures should be considered, none should be rejected out of hand.Procedures with a sound theoretical background are preferrable as they usually prove more successful in the long run. However, more ad hoc methods, such as continuous four dimensional data assimilation, have a place in the shorter term because they are easy to implement and valuable experience can be gained by using them. Possible candidates (in no particular order) are: (1) optimal interpolation (2) successive corrections (3) Kalman filtering (4) four dimensional time variational analysis with dynamical constraints (in the style of P. Thompson and J. Lewis).

Useful linear relationships (such as geostrophic) do not abound on the meso-scale, but there may be usable non-linear relations. These can be incorporated in combined analysis-initialisation cycles, but iteration will almost certainly be necessary.

This may require that proportionally more emphasis will have to be placed on analysis in the operational suite, but the pay-off would be worth it.

Analysis in prediction model coordinates avoids interpolation problems but isentropic coordinates also offer significant advantages.

3.2.5 Initialisation

The role of initialisation in defining the model initial state cannot be over-emphasized. Unfortunately, the initialisation problem on the mesoscale is very difficult. The recent success of large-scale initialisation procedures is encouraging, but not particularly helpful for the mesoscale. The most important difficulties are convergence for shallow systems, complicated topography, convective parameterisations and boundary effects (for LAM models).

There is also considerable uncertainty over which gravity modes should be retained and which suppressed. It is possible that Lagrangian methods could help the convergence problem.

3.3 Recommendations for analysis systems with grids of 50 km or less

3.3.1 Data requirements

There is a great need for more high resolution, high accuracy, high frequency data (particularly wind data) on the mesoscale. Better use could be made of existing indirect data as well.

3.3.2 Data checking

Data checking is very important on the mesoscale. Manual intervention should be seriously considered in the initial stages at least.

3.3.3 Analysis procedures

Investigation should proceed on many types of analysis procedures. None should be excluded at this early stage. Iterative procedures should be considered.

3.3.4 Mesoscale initialisation

Work on mesoscale initialisation procedures is of paramount importance.

3.3.5 Observing systems experiments

Observation System experiments (OSE) with both simulated and real data should be performed with mesoscale models to assess the value of existing and proposed mesoscale observing systems.

3.3.6 Research effort

In view of the special difficulties and lack of experience in mesoscale analysis a major research effort is needed.

4. LARGE SCALE TROPICAL ANALYSES

4.1 Introduction

The papers presented in the tropical session considered a number of important problems in the analysis and forecasting of the tropical atmosphere. Some of these were recognised already in the discussions of a previous workshop (Tropical Meteorology 1981) but several new insights into the importance of the analysis of the tropical diabatic forcing and of the performance of the OI scheme in analysing large scale tropical flow have been gained since then.

Most of the important recommendations of the earlier workshop have been, or are being,acted upon - work on diabatic initialization and sea surface temperature is complete, work on the structure functions and soil parameters is well advanced, various changes in quality control have been made and we now have a feasible proposal for a direct interpolation of the analysis increments into sigma coordinates. Some of the additional recommendations of that workshop are still outstanding but are not felt to be of major importance. Here we shall concern ourselves with several areas where we believe new or additional work has to be undertaken.

4.2 The analysis of tropical diabatic forcing

Tropical motions are strongly influenced by diabatic processes. In the context of initialization, diabatic tendencies are required in order to support divergent circulations.

The contributions of Wergen and Heckley indicate that both on a global scale and on a continental scale, the initial specification of diabatic forcing is wrong and that this leads to very large errors in the tropical forecasts. The errors arise both from analysis and model deficiencies. The work of Julian and of Krishnamurti indicate that an additional data source, infra red (window channel) radiometry, can be used to get valuable information about the distribution of tropical convective heating and of the associated velocity fields.

The initial diabatic heating can generally be specified either:

(i) internally by using a forecast-model

or

(ii) from additional data (IR radiometry).

Method (i), which is being used at the Centre, depends on the quality of the parameterization scheme and on the initial distribution of the fields of mass, wind and moisture. In Method (ii), the intensity of condensational heating is estimated by means of a regression method using the IR data, while the vertical profile is estimated using a convection model. With Method (ii) it must be ensured that the diabatic heating during the early stages of the forecast stays close to the initial specification.

Estimates of the depth and intensity of convection, based on IR data, have been evaluated mostly over oceanic areas as in GATE (cf Griffith, et al, 1978) where calibration against radar data was possible. Alternative techniques make use of the temporal evolution of the cloud anvil in order to derive upward cloud mass fluxes and precipitation rates. The accuracy of the estimates of precipitation rates is thought to be within 30%. We understand that the Florida State Group have used this kind of approach with encouraging results.

4.3 Analysis of humidity in the tropics

A good estimate of boundary layer moisture would be valuable but cannot be expected from microwave sounders for several years. Current microwave sounders do not provide sufficiently accurate estimates of boundary layer humidity for our purposes. One may mention also the Meteosat water vapour channel, which gives qualitative estimates of the water vapour field above 700 mb. Contact should be maintained with the groups studying these data sources.

4.4 Large scale mass and wind analysis in the tropics

The theoretical work presented by Cats indicated that in our present analysis system the tropical mass field makes an important contribution in the analysis of those large scale slow modes of the model which have large amplitude in the tropics. Simpler studies indicate that OI analysis techniques using a rotational structure function have some difficulty in analysing very large scale wind fields. In order to estimate the significance of Cats's results we need to know how they depend on the wind analysis.

The measurement of the tropical mass field is difficult and complicated, because of the inhomogeneities of the observing system and the small amplitude of the variations. Idealised studies of the ability of SATEM and SATOB measurements, (when combined with surface pressure data) to define the tropical large scale mass field would be useful to complete this study.

4.5 Analysis of tropical divergent fields

The analysis of the large scale wind field has been discussed in relation to the mass analysis in the previous section. The present ECMWF OI analysis scheme does not provide an adequate definition of the irrotational wind component in the tropics. This is especially true when the scheme is presented with observations indicative of a strongly convergent/divergent situation. Although it is not completely clear how faithful to the tropical observations the analysed irrotational wind field should be, it seems reasonable to conclude that its convergent/divergent structure should be consistent with the analysis of diabatic heating available to the initialization procedure and the model physics. This conclusion means that a better job of specifying the irrotational component must be done in order to facilitate the work outlined in Section 1.

4.6 Recommendations

4.6.1 Analysis of diabatic forcing

Studies should be undertaken to exploit the hitherto unused data source of IR radiometry to specify the intensity, extent, and vertical distribution of deep convective heating in the tropical atmosphere. Theoretical and experimental studies will be necessary in order to learn how to modify the initial analyses and/or the parameterization schemes so that the model can support and maintain the analysed distribution of heating.

The Centre should contact other groups with an interest in obtaining real-time IR radiometry, so that a concerted approach can be made to the satellite agencies to provide the required data. Also, the importance of the boundary layer humidity measurements measurements should be emphasised to the data producers.

4.6.2 Analysis of the very largest scales

The performance of the OI scheme in analysing wind data on the largest scales should be further investigated.

Further studies of the analysis of the large scale mass field and of its coupling to the wind field should be carried out.

4.6.3 Analysis of the divergent wind component

Continued work in aggregating analysis increment (first guess error) statistics of the wind field is essential to establish how the rotational and irrotational components are to be modelled in three dimensions. These aggregations should come from the operational forecasts and GTS observations as well as from re-analysis of the FGGE Final IIb data. A major purpose of this work will be to clarify some of the problems of the large scale analysis of mass and wind. A further purpose of this work will be to allow some relaxation of the constraint of non-divergence within an analysis volume in order to be consistent with the analysis of diabatic heating.