# DATA ASSIMILATION:
# OPTIMUM INTERPOLATION APPROACH/VARIATIONAL APPROACH

**Jean Pailleux**
**European Centre for Medium-Range Weather Forecasts**
**Reading, U.K.**

## 1. INTRODUCTION

1989 is mainly the 200th anniversary of the French revolution. However, for people who are more interested in global data assimilation than in history, 1989 is the 10th anniversary of two important meteorological events. One is the First GARP Global Experiment (FGGE) which provided in 1979 the best global observation set for numerical models. The second one is the beginning of operational forecasting at ECMWF with a system based on a global forecast model and a global analysis. At the beginning of the eighties, the production of the FGGE analyses (III b data sets) by the ECMWF and the GFDL assimilation systems gave a unique opportunity to assess the performance of global assimilations on a large number of cases involving new observing systems such as buoys, automatic aircraft reports (ASDAR) and TOVS data from the NOAA satellites.

The first operational assimilation system at ECMWF was based on a 3 dimensional multivariate optimal interpolation analysis (3D multivariate OI) and a non-linear Normal Mode Initialization (NMI). The 3D OI and NMI principles are still the basis of the 1989 operational assimilation, so a short glance at the scientific principles gives the impression that nothing new has happened during the last ten years. However, all the objective and subjective evaluations show a continuous improvement of the quality of the ECMWF assimilation, which must be explained by many developments on the implementation aspects of the 3D OI and NMI (although the basic principles remained the same). The purpose of this paper is to explain some of these developments and to show that they are organised along three main general tendencies.

The first general tendency for assimilation systems is to be more and more built around a forecast model. This aspect is developed in section 2: the end of section 2 is a comparison between the 3D OI and the 3 dimensional variational analysis (3D VAR). It is shown that the development of variational assimilation schemes is an extra step in the logic of making the analysis system more and more dependent on the forecast model.

The second general tendency for assimilation systems is that they use observed parameters which are closer and closer to the genuine observed quantity. This point is illustrated in section 3, the best examples being found with the use of satellite data. The development of variational analysis schemes for operational purposes at the beginning of the nineties should confirm this tendency during the next decade.

The third and last tendency is that assimilation systems are more and more demanding regarding the quality of observations. This is illustrated (in section 4) by several retunings of the ECMWF operational analysis, mainly during the last five years (retunings of the quality control tests and of the use of data).

These three main ideas are in the general logic of the evolution of most of the operational assimilation systems, not only the ECMWF one. They are also true for the limited area models with a high resolution (used for short range forecasts). They are also consistent with two other main trends in Numerical Weather Prediction (NWP):

- The range of observed parameters becomes larger and larger as new observing systems are being developed;

- With the continuous increase of the computer power, the resolution and the accuracy of the forecast models (both dynamical and physical processes) is improved.

## 2. ASSIMILATION SYSTEMS MORE AND MORE BUILT AROUND A FORECAST MODEL

In the sixties and the early seventies the first operational analysis schemes were mainly designed and described as the "mathematical interpolation schemes computing the values of meteorological fields on a regular grid from the values at observation points". The analysed fields were used for a series of applications such as production of documents for forecasters and initialisation of forecast models. The notion of guess field (or first-guess) was already present in these schemes, but is was not the "core" of the analysis schemes which were tuned in order to "draw reasonably to the observed data" and to produce "realistic features on a map". For practical reasons (e.g., computer limitations) these analysis schemes were 2 dimensional, performed level by level on a set of standard pressure levels; see for example Cressman (1959).

In the period 1975-79, when the ECMWF analysis scheme was designed and developed, more attention was progressively given in the analysis to methods for using more information from the forecast model, and to the balance properties required by the model equations. Such an evolution was stimulated by the tendency of the operational models to be "primitive equation" (rather than "quasi-geostrophic filtered" before) and to cover a larger and larger area. The balance requirements for primitive equation models (balance between mass and wind fields) led to the important development of NMI, see Machenhauer (1977). The progressive extension of the model domain to the whole globe led to the necessity of an accurate first guess in areas such as the Southern Hemisphere where the conventional data were (and still are) extremely sparse.

The assimilation scheme implemented operationally at ECMWF in 1979 was a 6 hour intermittent scheme based on a global 3D multivariate OI, Lorenc (1981), and on adiabatic NMI. Although a 3D multivariate OI was already operational in Canada, see Rutherford (1976), the ECMWF scheme was at this time a major step toward more consistency with the model dynamics (and the atmosphere dynamics!).

- The 3D OI analysis of the departures "observations - 6h forecast" was using the 3D statistical structure of the forecast errors, and was improving considerably the vertical coherence of the analysis;

- A good dynamical balance between mass and wind fields was achieved by the NMI and the multivariate aspect of the OI;

- The "box technique" consisting in setting up one OI system for a volume of atmosphere led to consistent increments (increments = "Analysis-guess") at the scale of the boxes.

Although the 1979 ECMWF analysis system was largely built "around" a global forecast model, the analysis did not treat the model variables directly. The analysis was performed on 15 standard pressure levels (from 1000 to 10 hPa) rather than on the model levels (sigma coordinate). The statistical interpolation scheme was performed on the height and wind components $\Phi$, u and v, rather than on the true historical variables, of the forecast model
(temperature T, u and v components then replaced by vorticity and divergence when the forecast model became spectral). Also the humidity was treated by a simple 2D relative humidity analysis, carried out level by level, through a "weighted-average" technique, whereas the model prognostic variable is q in sigma coordinate.

The following developments of the ECMWF assimilation system are all consistent with the general idea of using more information from the forecast model or making the analysis more consistent with the model.

a) Vertical interpolation of increments rather than interpolation of the full fields, to go for the standard analysis levels to the model levels (1980). This was particularly important in the boundary layer where the model resolution was much higher than the analysis resolution (4 model levels between 1000 and 850 hPa).

b) Diabatic heating introduced in the initialization (1982). The diabatic term is computed by integrating the model on 10 time steps).

c) 3D multivariate OI performed directly on the model levels, abandoning then the pressure standard levels as "artificial interface" between the observation levels and the model levels (1984).

d) Tides introduced in the initialization (1986). The treatment of the tidal signal is a development aiming at making the assimilation consistent with a particular phenomenon which exists in the atmosphere as well as the forecast model (since the forecast model had a diurnal cycle).

e) 3D multivariate OI performed directly on the gaussian grid of the model (1986). This was achieved with a complete recoding of the operational analysis. Before, the analysis was carried out on a fixed grid (regular latitude-longitude grid; $\Delta\Phi = \Delta\lambda = 1.875°$) which was the one of the first ECMWF forecast model; an extra horizontal interpolation was needed to go from the analysis grid to the model gaussian grid.

f) Divergence allowed in the increment fields (1988). Before, the increments were constrained to be non-divergent by the geostrophic assumption made on the structure functions. This meant that a divergent feature in the atmosphere, seen by some observations but not the the first guess (6h forecast) was filtered out by the analysis.

Nevertheless, the ECMWF system as it is in 1989, is constrained by a major limitation which is intrinsic to all the intermittent OI systems: in the assimilation the analysis step is performed independently of the initialization step (NMI) and of the 6 hour forecast step. This means that when the analysis is run, it forgets completely about the model equations. The only tool transferring some information from the forecast model to the analysis is the 3D statistical structure of the forecast errors (used as input to the OI system). Many attempts have been made in different NWP centres to improve the OI analysis schemes by using "dynamical" structure functions. "Dynamical" means "flow-dependent" in this context: by using dynamical structure functions, one tries to take into account in the OI the specific properties of a specific area, a specific air mass or a specific meteorological situation. Several statistical studies show that fixed structure functions are an important limitation to OI, see Pailleux (1982). At ECMWF, the OI structure functions are dynamical, but only to a small extent. Two important retunings have been made between 1979 and 1989, one in 1984 (see Shaw et al., 1987) and one in 1988.

At the UK Met. Office, a constant feature in the operational assimilation scheme during the last decade has been the concept of "repeated insertion technique". At each time step the model is relaxed toward the observed values, using weights which are calculated in different ways, see for example Lorenc (1984). This relaxation technique (called sometimes "nudging" technique) is illustrated in Fig. 1.

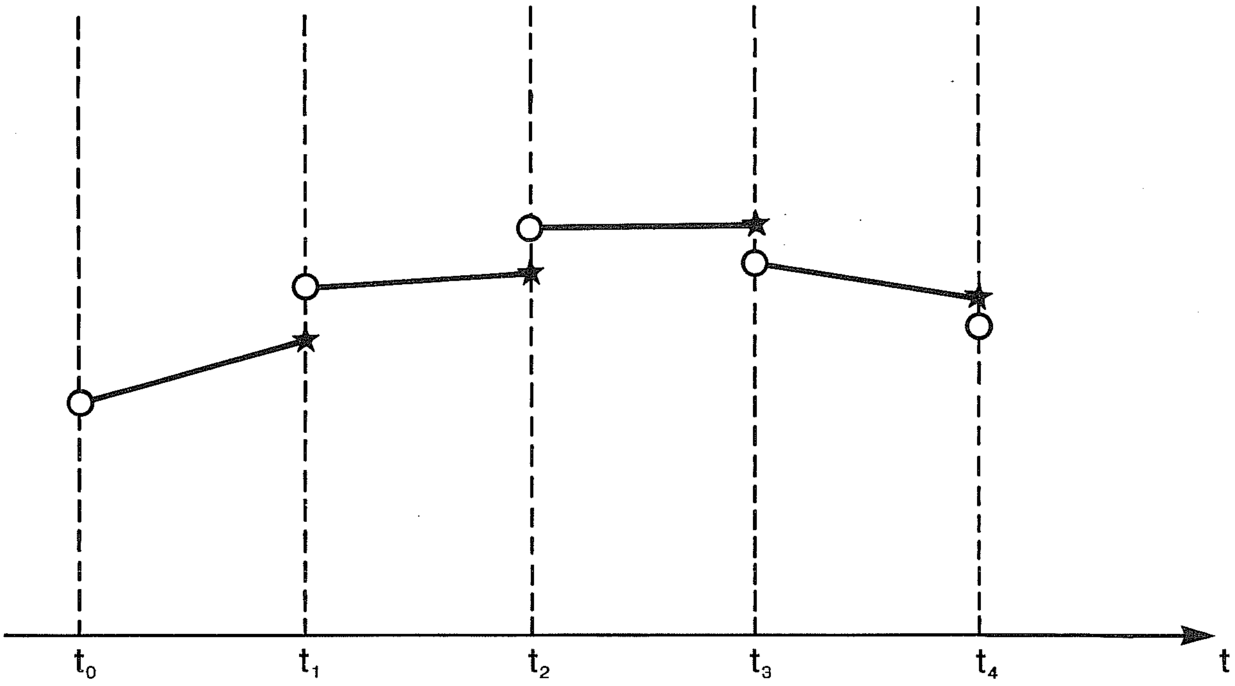

Fig. 1 Schematic representation of the "repeated insertion technique". Each segment represents the integration of the forecast model on one time step from the "analysis points" (Θ) to the model points (*).

By using such a technique, the UK Met. Office scheme tries to achieve the double goal of satisfying both the model equations and the observations. Maybe more than in the ECMWF scheme, the forecast model is the core of the assimilation system. However, neither the analysis points (Θ) nor the forecast points (*) respect the model equations in this relaxation technique. To solve the problem in a completely clean way, one would need to find a model trajectory (i.e. an ensemble of successive model states obtained by direct integration of the model equations) which fits also the available observations with a "reasonable" accuracy. "Reasonable" means that the fit to the observed parameters must depend on the quality of the observations. This minimization problem is illustrated in Fig. 2.

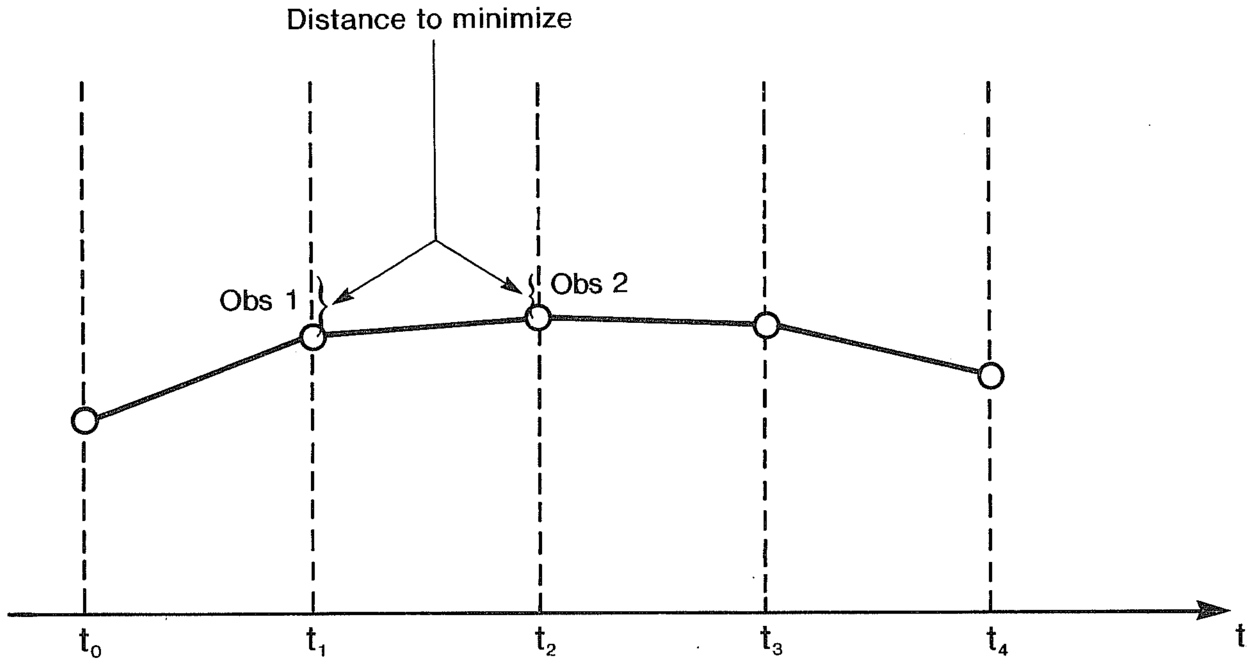

Fig. 2 Schematic representation of the 4D assimilation problem.

It was found in the eighties that the natural mathematical tool needed to solve this minimization problem in a correct way is the notion of adjoint operation. Although it is not used in operations anywhere in 1989, the adjoint of the forecast model provides a practical way to minimize the distance between a model trajectory and the observations available on a time period (i.e. to solve the 4D variational assimilation problem - 4D VAR). For the general concept of the adjoint model, one can read Le Dimet and Talagrand (1986) or Lewis and Derber (1985). The theory and practical aspects related to the design of a variational analysis are fully documented in the Proceedings of the 1988 ECMWF seminar. One can see also Talagrand and Courtier (1987) and Courtier and Talagrand (1987).

A variational analysis does not need to be necessarily four dimensional to be useful. Performed at one time only, the 3D VAR presents several aspects which are more flexible than the 3D multivariate OI. These aspects are fully discussed in the 1988 Proceedings. They are also summarized in table 1 where a parallel is made between a general 3D OI scheme and a 3D VAR scheme. The 4D VAR concept is a variational problem posed in the space of the model variables. It uses the model equations directly and explicitly. It is clear that in this context the analysis is entirely built around the forecast model.

| | 3D OI | 3D VAR |
|---|---|---|
| Principle | Minimize variance of interpolation error (in a stat-sense) | Minimize distance between model state and available information<br>$J = J_{obs} + J_{guess}$<br>$= (HX-d)^t O^{-1} (HX-d) +$<br>$(X-X_g)^t P^{-1} (X-X_g)$ |
| Method | Solve a set of linear equations for a particular ensemble of analysis variables | Use std. minim. scheme<br>- steepest gradient<br>- conjugate gradient<br>- quasi-newton |
| Data Selection | "Grid-point" or "Box" technique | Not needed |
| Quality Control | - Check against FG<br>- Check against an at obs point | idem |
| Time Coordinate (generalization to 4D) | Difficult | Straightforward using forecast model and adjoint |
| Use of observations | Observed quantities linked to the model variables by a <u>linear</u> operator | Observed quantities linked to the model variables by a <u>differentiable</u> operator |

Table 1 Comparison of optimum interpolation analysis with variational analysis

## 3. ASSIMILATION SYSTEMS USE VARIABLES CLOSER AND CLOSER TO GENUINE OBSERVED QUANTITIES

The bottom part of table 1 shows that there is a much wider range of observations which can potentially be used by a variational analysis than by an OI analysis. The restriction to linear operators between analysed and observed variables in OI is a serious constraint. As an example, an observed wind direction (without observed speed) cannot be used directly in the multivariate OI because of its non-linear link with the wind components (involving some sine and cosine functions). It can be used in a natural way in a variational analysis. In other words, the variational analysis is a step towards using more observations, compared to the OI analysis.

The best example is the one of meteorological satellites which observe radiances. The radiances are linked to the model variables (temperature and humidity profiles, surface parameters) through the radiative transfer operator (which is non-linear). The mathematical framework of the variational analysis gives the opportunity to combine the inversion step and the analysis step in a consistent scheme which uses directly the clear radiances. By "inversion step" we refer to the process going from the radiances to the temperature/humidity profiles at each observation point.

The direct use of one radiance in a variational analysis system requires the following chain or operators:

a) M: Model integration from $t_i$ (initial time) to $t_{ob}$ (observation) if the assimilation is 4D.

b) $I_s$: Inverse spectral transforms (if the model is spectral) to provide the model fields on the gaussian grid.

c) $I_h$: Horizontal interpolation from the gaussian grid to the satellite observation point.

d) $I_v$: Vertical interpolation from the model levels to the set of pressure levels required by the radiative transfer model (performed at the observation point).

e) $T_r$: Radiative transfer operator computing (from the model) the analogue $R_m$ of the observed radiance $R_o$.

At this stage $R_o$ and $R_m$ can be compared, and a distance function can be computed between the model and the observed radiance. The obvious distance function is

$J = ((R_0-R_m)/\sigma)^2$, $\sigma$ being the standard deviation of the radiance observation error. The gradient of J with respect to $R_m$ is obvious to compute : $2\ (R_m - R_0)/\sigma$. Then to get the gradient with respect to the original control variable, we have to apply the chain of adjoint operators $T_r^*$, $I_v^*$, $I_h^*$, $I_s^*$, and $M^*$ (note the opposite order). This evolution towards a direct use of clear radiances in the assimilation could be pursued in the nineties by a direct use of raw radiances (genuine observed quantity in the case of TOVS data). "Clear" radiances means here decontaminated from the effect of clouds. The direct use of raw radiances requires the insertion of cloud parameters in the control variable of the minimization problem, see Eyre (1987).

In 1989 most of the operational assimilation systems use the TOVS data as inverted profiles, (called also retrievals or SATEMs). The use of SATEMs in operational analysis was affected by an evolution which is illustrated by the following examples:

a) Before the 3D multivariate OI, the SATEMs were used by adding the "observed" thickness (reported in the SATEM message) to a surface analysis, in order to produce a pseudo geopotential observation. These pseudo observations were then used as radiosonde geopotential observations.

b) The first operational scheme at ECMWF used directly the 14 thickness layers reported by the SATEM code between 1000 and 10 hPa. It is obvious that this ensemble of 14 layers is far from the real observed information which cannot describe the vertical resolution with such an accuracy.

c) From 1985 to 1987, the ECMWF analysis system was changed progressively in order to use 7 layers rather than 14 from the SATEM data. 7 layers is a better representation of the real information content of the TOVS radiances.

This slow evolution, not as drastic as the direct use of radiances, is clearly a series of steps towards feeding the assimilation by something closer to the genuine observed quantity. Another example which is likely to be treated operationally in a few years is the one of the scatterometer data as they will be produced by the satellite ERS1. The backscatter signal $\sigma_0$ can be used to compute surface winds over the ocean with an ambiguity on the wind direction. Some "dealiasing" techniques are used to guess the wind direction, and then the wind data can be used in an analysis, see Andersson et al., (1987). The trend for the future is (like for the TOVS) to assimilate directly the $\sigma_0$ rather than "dealiased" winds far away from the observation content.

## 4. ASSIMILATION SYSTEMS MORE AND MORE DEMANDING REGARDING THE QUALITY OF THE OBSERVATIONS

Ten years ago, when the first global models were implemented operationally, the meteorologists were probably more concerned about the lack of data (especially in the Southern Hemisphere) than about the poor quality of data. In 1989, we are in the situation where TOVS data are available in operations at ECMWF with a 80km horizontal resolution, and it is clear from Kelly (1989), that the satellite data quality has become one of the main priorities. This is a natural evolution, as the quality of any observing system has always to be judged in the analysis by comparison with the forecast quality (first-guess). In ten years, the 6 hour forecast quality has been improved dramatically by more efficient numerical techniques, more sophisticated physical parametrizations and increase of the computer power. In the mean time the TOVS instruments have not changed.

During the period 1979-1989 the quality control tests had to be tuned several times in the ECMWF analysis system. One example is the recent "stability check" implemented on the TOVS data in January 1989, see Kelly (1989). Another one is the general retuning of the wind first-guess check limits made by Lönnberg (pers. comm.) in July 1988. Before the retuning, the first-guess check was rejecting an observation wind when it was about 40 m/s away from the first-guess. After the retuning this limit had to go down to about 20 m/s for AIREPs, 14 m/s for cloud-winds, and even 7m/s for cloud-winds when the "asymmetric check" is on (described later). These limits are actually variable, depending on the forecast error standard deviation, rather than fix numbers. This tightening of the quality control appeared to be necessary (from experimentation) when we tried to use the observations at higher resolution through a retuning of the horizontal correlation functions in the OI.

One choice was made at ECMWF in 1979, consisting in monitoring carefully the availability and quality of the observations over the whole globe. This work is still carried out by the "MET OPS" section. Ten years after the beginning of operations, it is clear that this choice was good, and that the monitoring work is very beneficial for the analysis quality.

a) The direct effect of the monitoring work is the maintenance of exclusion lists for different types of observations, preventing too bad stations from entering the analysis.

b) The indirect effect (perhaps more beneficial) is the accumulation of some empirical knowledge on the quality and the weaknesses for each observation type. This knowledge (archived as "statistical files") can then be exploited:

   - To tune the OI statistics in the analysis (mainly the observation error standard deviations);

- To tune the quality control checks in the analysis;
- To design and implement new quality control tests aiming at tackling some specific weaknesses of some specific observing systems.

One particular illustration of the "learning" process is the use of cloud winds and its evolution during the ten year period at ECMWF.

a) In 1979, all the cloud winds were used in the analysis.

b) In 1982, they were used over sea only. It was clear that the quality was far below the radiosonde wind quality, and the idea was to let the analysis be driven by the conventional wind data over land.

c) In 1985, they were reintroduced over land inside the tropics (from 20°N to 20°S), as it was clear from the monitoring results that the main weakness in cloud winds was underestimation of strong winds in jet streams (which are normally outside the tropics).

d) In 1987, an "asymmetric" check was implemented after a careful study of this underestimation problem.

   The idea was to be much more severe in the first guess checks for high level cloud winds outside the tropics, when the observed wind speed is below the first guess wind speed. The rejection limits were again tightened in 1988, as mentioned before.

The very recent monitoring results (accumulated statistics in July and August 1989) show that the use of cloud winds is still not optimal regarding the quality control. They still have a tendency to slow down the jets in some areas, and further quality control developments have to be considered.

## 5. CONCLUSION

We have illustrated three general trends of the meteorological data assimilation systems. These trends have been observed during the past ten years (or even more), and are likely to be observed again during the next decade. It is easy to predict that the assimilation will draw benefit from model developments and it will be more and more difficult to separate the respective merits of "assimilation" and "model". It is also clear that the quality control developments have to follow the other developments (otherwise the quality of data may become the blocking point in NWP). It is likely that in the nineties, variational analysis schemes will use an increasing variety of observations. However, the details of implementation are difficult to predict, as they are dependent on model developments and computer evolutions.

References

Anderson, D., A. Hollingsworth, S. Uppala, and P. Woiceshyn, 1987: A study of the feasibility of using sea and wind information from the ERS-1 satellite. ECMWF contract report to ESA. Part 1: Wind scatterometer data. 121 p.p.

Courtier, P. and O. Talagrand, 1987: Variational assimilation of meteorological observations with the adjoint vorticity equation. Part 2: Numerical results. Quart.J.R.Met.Soc., 113, 1329-1368.

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

Eyre, J.R., 1987: Inversion of cloudy satellite sounding radiances by non-linear optimal estimation: theory and simulation for TOVS. Quart.J.R.Met.Soc., Dec 87.

Kelly, G.A., 1989: Use of satellite data in global modelling: present and future (these Proceedings).

Le Dimet, F.X. and O. Talagrand, 1986: Variational algorithms for analysis and assimilation of meteorological observations: theoretical aspects. Tellus, 38A, 97-110.

Lewis, J.M. and J.C. Derber, 1985: The use of adjoint equations to solve a variational adjustment problem with advective constraints. Tellus, 37A, 309-322.

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

Lorenc, A.C., 1984: Data assimilation by repeated insertion into a forecast model - Principles, practice and plans. Proceedings of the ECMWF seminar/workshops on Data assimilation systems and Observing system experiments with particular emphasis on FGGE. Reading, September 1984. 191-214.

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

Pailleux, J., 1982: Mesoscale analysis at "Direction de la Météorologie". Proceedings of the ECMWF workshop on current problems in data assimilation. Reading, November 1982. 241-252.

Rutherford, I.D., 1976: An operational three-dimensional multivariate statistical objective analysis scheme. GARP Programme on Numerical Experimentation. Proceedings of the JOC study group conference on four-dimensional data assimilation, Paris, November 1975. Vol 11, pp 98-121.

Shaw, D.B., P. Lönnberg, A. Hollingsworth and P. Unden, 1987: Data assimilation: The 1984/85 revisions of the ECMWF mass and wind analysis. Quart.J.R.Met.Soc., Vol. 113, pp 533-566.

Talagrand, O. and P. Courtier, 1987: Variational assimilation of meteorological observations with the adjoint vorticity equation. Part 1: Theory. Quart.J.R.Met.Soc., 113, 1313-1328.