CURRENT PROBLEMS IN FOUR-DIMENSIONAL DATA ASSIMILATION

L. Bengtsson

ECMWF

ABSTRACT

The purpose of this lecture is to review recent development in data analysis, initialization and data assimilation. The development of 3-dimensional multivariate schemes has been very timely because of its suitability to handle the many different types of observations during FGGE.

Great progress has taken place in the initialization of global models by the aid of the non-linear normal mode technique.

However, in spite of great progress, several fundamental problems are still unsatisfactorily solved. Of particular importance is the question of the initialization of the divergent wind field in the Tropics and to find proper ways to initialize weather systems driven by non-adiabatic processes. The unsatisfactory ways in which such processes are being initialized are leading to excessively long spin-up times.

1. INTRODUCTION

In the paper "Das Problem der Wettervorhersage betrachtet vom Standpunkt der Mechanik und der Physik" published in Meteorologische Zeitschrift in 1904, V. Bjerknes outlines for the first time a consistent hydrodynamical theory to the problem of weather prediction. The necessary and sufficient conditions for a rational solution of the forecasting problem are the following:

i) A sufficiently accurate knowledge of the state of the atmosphere at the initial time.

ii) A sufficiently accurate knowledge of the laws according to which one state of the atmosphere develops from another.

To meet the first of these conditions which is of relevance in the context of the present lecture, Bjerknes gave the following general plan:

"The main task of observational meteorology will be to obtain regular simultaneous observations of all parts of the atmosphere, at the surface of the earth and aloft, over land and over sea.

The first task of the theoretical meteorologist will be to work out, on the basis of these observations, the best possible overall picture of the physical and dynamical state of the atmosphere at the time of these observations. And this representation must have such a form that will enable it to serve as the starting point for weather prediction by rational dynamical-physical methods.

Even this first preliminary task is a sizeable one. For it is of course much more difficult to represent the state of the atmosphere at all elevations than only at sea level, as it is now done. In addition, our direct observations of the higher layers of air will always be very limited. One must therefore use each observation from the higher levels to the utmost. From the directly observable quantities one has to compute to the greatest extent all accessible data about the non-observable ones. In doing this, one has to utilize the physical relationships between the quantities. Even to construct a coherent picture of the total state of the atmosphere out of scattered observations, one has to use, to a large extent, dynamical-physical methods".

It is interesting to note that in essence V. Bjerknes had identified the fundamental problem in data analysis and data assimilation. Although the observing system during FGGE in 1978/79 is beyond what could be expected in 1904, the same methods as Bjerknes indicated must still be adhered to. In spite of all upper data, both in the form of radiosondes and satellite data, all effort has to be made to use observations from higher levels to the utmost.

Moreover, it is equally necessary to use the observable quantities in order to obtain data about the non-observable ones and in doing this, utilize the physical relationship between the quantities. Finally, in order to obtain a coherent picture of the total state of the atmosphere from the scattered observation, dynamical-physical methods must be used to a large extent.

In this lecture I will take the opportunity to review recent developments in data analysis and data assimilation and show how closely we are still following the main strategy outlined by V. Bjerknes. I will also try to identify remaining fundamental problems in 4-dimensional data assimilation and associated problems in numerical weather prediction.

2. DATA ANALYSIS

The meteorological observing systems have gone through a rapid development in recent years, culminating in FGGE.

Fig. 1 shows the data distribution for 00Z 9 February 1979 and is typical for the data coverage during SOP I.

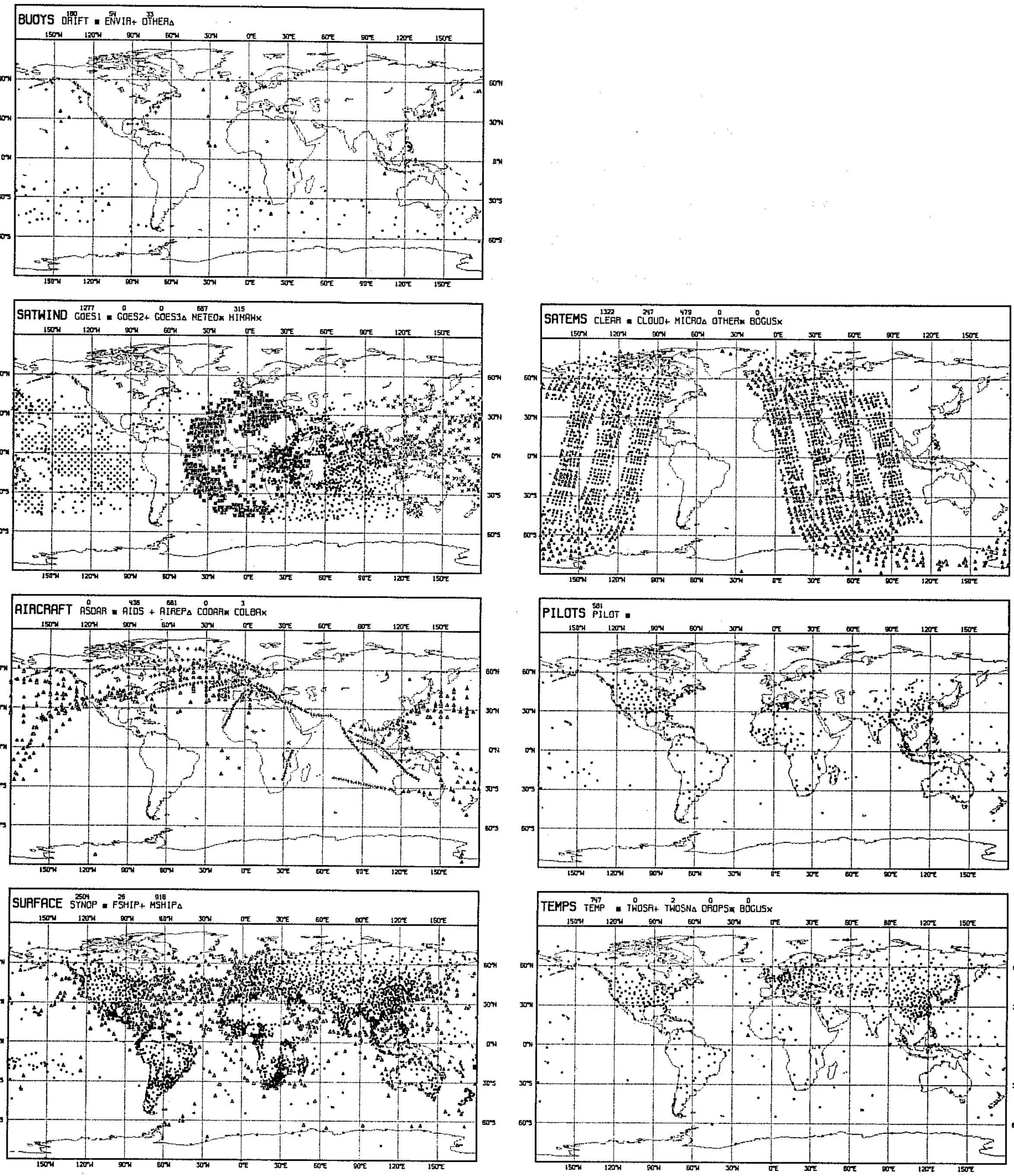


Fig. 1 Distribution of FGGE level II-b data, 12 GMT 9 January 1979. The number of reports from each observing system is given in individual maps.

Table 1 summarises the estimated observational accuracy and to what extent these observations can be used to represent the large scale flow of the atmosphere. In spite of the very good coverage, the data distribution is very scattered, in particular if we want to provide initial data for surface pressure and geopotential (potential temperature) wind and humidity for a global high resolution model. As can be seen, there are large areas where there are hardly any wind observations (e.g. high latitudes at the Southern Hemisphere). The observation of humidity is unsatisfactory almost everywhere and temperature and wind observations are inaccurate in the boundary layer and around the tropopause, except where high resolution radiosonde data are available. Clearly, the utmost must be done by a combination of statistical, dynamical and physical methods to extract an initial state for the numerical prediction from these observations.

TABLE 1 Observational errors used for different observing systems

	RADIOSONDES		SATELLITE TEMP.		SATELLITE WIND			AIREPS	
			TIROS-N clear/ partly cloudy	TIROS-N micro-wave	NESS Wisc-onsin	ESA/ LMD	Hima-wari	ASDAR AIDS	Convent. Aireps
Pressure level (mb)	V (ms^{-1})	T (^oC)	T (^oC)	T (^oC)	V (ms^{-1})	V (ms^{-1})	V (ms^{-1})	V (ms^{-1})	V (ms^{-1})
10	6	4.5	2.8	2.8	8	8	13	6	8
20	6	3.8	2.6	2.7	8	8	13	6	8
30	6	3.2	2.5	2.6	8	8	13	6	8
50	6	2.7	2.4	2.5	8	8	13	6	8
70	6	2.3	2.2	1.4	8	8	13	6	8
100	6	2.1	2.0	1.6	8	8	13	6	8
150	6	2.1	2.0	1.7	8	8	13	6	8
200	6	2.0	1.9	1.8	8	8	13	6	8
250	6	1.8	1.9	1.9	8	8	13	6	8
300	6	1.6	1.8	2.0	8	8	13	6	8
400	5	1.5	1.8	2.2	7	8	10	5	8
500	4	1.2	1.7	2.2	6	8	10	4	8
700	3	1.1	1.8	2.5	5	8	6	3	8
850	2	1.1	2.0	3.9	4	7	6	2	7
1000	2				4	7	6	2	7

CONSTANT LEVEL BALLOON, DROPSONDES, TWOS-NAVAID: Calculated as a function of position. In good condition the wind errors are similar to those of the ASDAR system.

SURFACE PRESSURE: SYNOP/SHIP 1.0 mb, Drifting Buoys 2.0 mb

Considerable progress has taken place towards the development of consistent statistical interpolation procedures. These schemes are generalized procedures derived from the so-called optimum interpolation or uni-variate analysis first proposed by Eliassen (1954) and further developed by Gandin (1963). In the first experiment with this method the analysis of a quantity α_i^A was done by expressing it as a linear combination of the normal value α_i^N (climatology) and the deviation of the observation from the normal value $(\alpha_k^O - \alpha_k^N)$ in the vicinity of the grid point (subscript k refers to the observations).

$$\alpha_i^A = \alpha_i^N + \sum_{k=1}^{K} p_k \, (\alpha_k^o - \alpha_k^N) \qquad (1)$$

The weights p_k are determined by minimizing the mean square error of interpolation. The mean is representing ensemble averages and the calculation of p_k implies that we must know the necessary covariance function $\langle\alpha_k \; \alpha_\ell\rangle$. The simple application of optimum interpolation can be generalized in the following ways:

i) Instead of using climatology as a reference field, a short range forecast (0-12 hours) is used as a normal field (background field). In doing so the information from previous time steps is projected into the future through the forecasting model. The same algorithm (1) can be used but new covariance functions must be calculated. The usage of a short range forecast is important for FGGE data which are highly asynoptic (SATTEMP and aircraft winds).

ii) Optimum interpolation can be extended to include also the vertical dimension. This is in particular very important for FGGE since many observations are only available at a single or at a few vertical levels (SYNOP, BUOYS, aircraft winds, constant level balloons and SATWIND).

iii) The interpolation can be extended to other parameters (e.g. we use wind observations to analyse the height field or vice versa). This implies the computation of a number of cross-correlation functions. This method is usually called multi-variate analysis.

iv) Interpolation can be extended to include time. This is in particular important if we do not use i) or if the time-interval for up-dating is too short.

We will here consider the extension of the optimum interpolation given in points (i) - (iii).

Rutherford (1973) and Schlatter (1975) have described a system for multi-variate analysis of height and wind data. Rutherford (1976) has further described an extension to 3-dimensions using an economical, split vertical and horizontal scheme capable of using height, temperature and wind-data without fully taking into account their 3-dimensional structure. A fully 3-dimensional method has been developed by Lorenc et al (1977) and Bergman (1979). The analysis system used at ECMWF to produce FGGE level III-b data sets consists of a fully 3-dimensional multi-variate scheme analysing wind, height and thicknesses. A detailed description of this system has been given by Lorenc (1980). The application of a 3-dimensional multi-variate scheme is straightforward and can as in the case of the ECMWF system be developed into a fully automatic system. A 3-dimensional multi-variate scheme is by and large a practical necessity if we require global analysis for a multi-level model.

The observing system has, as has been pointed out above, a very incomplete coverage both horizontally and vertically. It consists of different kinds of observations with different accuracy and representativeness. It is necessary to have an analysis system which can draw from all kinds of observations and weight them according to their accuracy. Moreover, a model using the primitive equations must be updated with both mass- and wind field information. Daley and Puri (1980) have recently shown that a primitive model mainly assimilates wind field information. As is shown in Daley's contribution to these Proceedings when mass field (height) information only is inserted, having a scale distribution which is typical for updating of prediction increments (analysis-first guess), the mass field information is projected onto the gravity wave modes and consequently rapidly dissipated by the model. It can also be shown that even the application of advanced initialization procedures, as the non-linear normal mode, does not improve the assimilation. It therefore seems necessary, in the case where only increments in the mass field are available, to calculate the corresponding increment in the wind. These conditions, the irregular distribution of observation, the variable accuracy of the observation as well as the properties of the dynamical adjustment, are the fundamental reasons why a 3-dimensional multi-variate scheme is so useful for analysing the FGGE data.

A very important additional advantage in using a 3-dimensional multi-variate scheme is the provision for a consistent way of checking the data.

In spite of rapid progress in recent years, there are several problems which are still unsolved and poorly understood. Before we can start to analyse these problems, we must define the fundamental purpose of an analysis scheme. It is certainly straightforward that the analysis should satisfy the criteria that we fit the observed parameters to the analysis in a way consistent with their accuracy. However, this is neither our main criteria for a satisfactory analysis, nor is it particularly difficult to satisfy. What is difficult to satisfy on the other hand is to have an analysis which can provide good forecasts. An evaluation of the analysis scheme becomes therefore very difficult, since there is no straightforward way to separate analysis error from deficiencies in the initialization, the data assimilation and the prediction model itself.

In the analysis procedure itself there are 3 questions which need attention, namely the structure functions of the parameters to be analysed (covariances and cross-covariances), the estimation of the accuracy of the first guess and the data selection (selection of observations in space to influence a given grid point). The evaluation of the weights p_k in the multi-variate scheme implies the solution of a linear system of equations through a matrix inversion process. To avoid ill-conditioning of the matrix the covariance function must be positive definite, and it is therefore necessary to fit a given function to the atmosphere structure function. In most cases the covariance function for the height field increment z' is given by an expression of the form:

$$\langle z'_k \, z'_\ell \rangle = \langle z'_i \rangle^2 \, e^{-br^2} \tag{2}$$

where subscripts k, ℓ and i indicate positions, b is an empirical constant and r the distance between k and ℓ.

The cross-covariance between winds and heights cannot be calculated with sufficient accuracy from observed data, and they are therefore obtained from the geostrophic relationship. It is also possible to calculate the wind covariances from the geostrophic relation and this is also done in most applications. Very little is known about the importance of the value of b and of the covariance function for the forecast. Experiments carried out at ECMWF show very little sensitivity during the first 4 days of the forecast, while a considerable sensitivity can be seen thereafter. This may suggest greater sensitivity in data sparse areas, mainly perhaps in the Tropics and therefore some time is needed to influence middle and high latitudes. There is clearly a great need to calculate the value of b and its variation by altitude and latitude as well as the covariance functions given by the model. Such studies are presently being started at ECMWF.

The second problem area is to find simple ways of estimating the accuracy of the first guess. This is in particular of great importance when the prediction model is accurate and where the accuracy of the first guess is similar or even higher than some observations. A simple approach has been implemented at ECMWF based upon an idea from Bengtsson and Gustafsson (1972). This is as follows: the error of the first guess (short range forecast) is first obtained from the interpolation error in the previous analysis step. This error is then assumed to increase in the forecast with a certain percentage of the climate variance. At present it is assumed that the increase in the 6 hour forecast cycle is 15% of the climatic variance. This increase is rather rapid and cannot be extrapolated more than about 2 cycles. However, since almost every area of the globe, because of the satellite systems, receives data in every 12 hour cycle and due to the fact that the forecast error normally grows relatively fast in the very beginning, this has been found to be a realistic, although a somewhat heuristic assumption.

Fig. 2 shows the estimated error of the first guess for the 500 mb geopotential. It should be stressed that the assumption of observational and forecast errors is mainly of importance for the checking of the observations. In case of a satisfactory data coverage, the analysis system usually draws from the observations if they are consistent and support each other.

The problem of data selection also needs careful consideration. In particular, for a 3-dimensional multi-variate scheme it is important that the influencing observations are selected from a space volume of reasonable dimensions in order to obtain signals from all significant components of the spectrum and also to have a mix of wind, height and thickness observations. Experiments at ECMWF have shown sensitivity to the data selection algorithm from forecasts beyond 3 days.

Finally, the most serious limitation in objective analysis is due to the fact that we are using <u>statistical structure functions</u> instead of <u>dynamical structure functions.</u>

This is the technique used in the synoptic analysis of weather maps developed by the Bergen School when the observations are fitted to the actual structure of the individual weather systems. Bengtsson (1980) has shown by a simple analytic and numerical experiment that a time sequence of surface pressure observations can initialize a baroclinic wave if the observations are being interpolated in the vertical along the sloping axis of the baroclinic perturbation (mode update). If the actual interpolation is done using statistical weighting functions, where no attention is paid to the structure of the weather system being analysed (standard update), the result is quite unsatisfactory. In the example shown in Fig. 3 surface pressure observations have been inserted during the first ten time steps. One time step is equal to 1 hour. So far there is no obvious way of applying dynamical structure functions under fully realistic conditions, but possibly the system could be implemented with the aid of computer/man interactive systems.

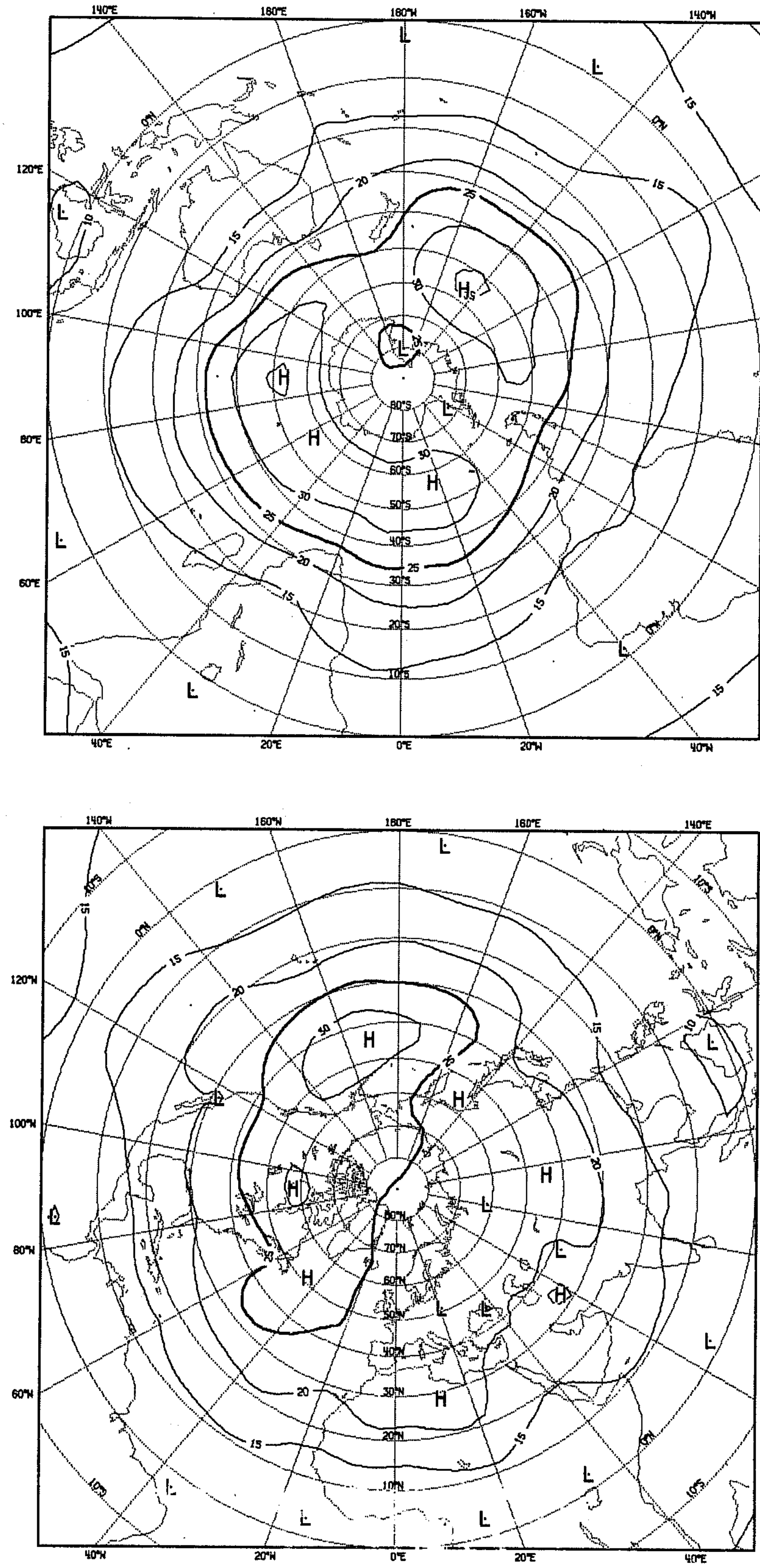


Fig. 2 **Estimated first guess errors 12 GMT 16 February 1979.**
<u>Top</u>: Southern hemisphere, <u>Bottom</u>: Northern hemisphere.

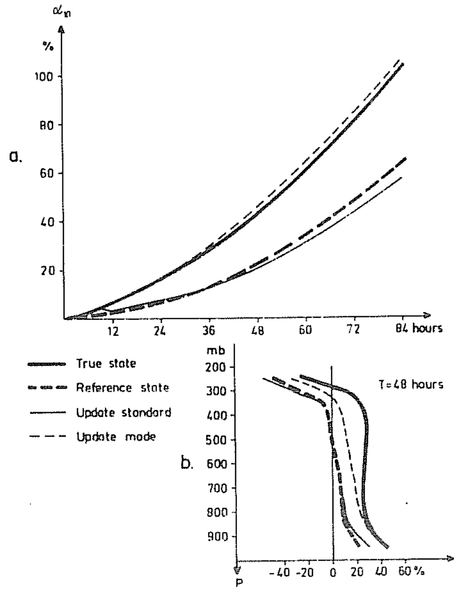


Fig. 3 (a) Percentage change in the amplitude of a perturbation at 950 mb in a 10-level model. Full heavy lines give amplitude as a function of time for the true state, dashed heavy line gives the reference state. Standard updating is illustrated by a full thin line and mode updating by a dashed thin line. (b) Vertical amplitude structure for the same cases as t = 48 h. Wavelength of the perturbation 4000 km in zonal and meridional direction. (From Bengtsson, 1980).

3. INITIALIZATION

A breakthrough has taken place in recent years in the initialization of primitive equation models by the application of normal modes. As shown by Longuet-Higgins (1968) the atmospheric flow can under normal conditions (low Rossby and Froude numbers) be separated into three classes based on their frequencies. This is done by linearizing the model around a simple basic state and then solving the resulting eigenvalue problem. The solution of this problem gives the normal modes and their associated eigenfrequencies. The classes of flow are the high-frequency eastward and westward propagating gravity waves and the low-frequency westward propagating Rossby modes. In a primitive model with m degrees of freedom (grid points or spectral coefficients) there would be 3 m normal modes divided between the three classes. Following Daley and Puri (1980) we may call the set of m Rossby modes $\underline{Y}$ and the set of 2 m gravity modes $\underline{Z}$. If we now project the equations of the assimilating model onto its normal modes, we will obtain a set of ordinary differential equations, which we write symbolically as

$$\dot{Z} = - i\underline{\Lambda}_z \underline{Z} + N_z (\underline{Z},\underline{Y}) \tag{3}$$

$$\dot{Y} = - i\underline{\Lambda}_y \underline{Y} + N_y (\underline{Z},\underline{Y}) \tag{4}$$

where Z and Y are column vectors of gravity mode and Rossby mode expansion coefficients, respectively, $\underline{\Lambda}_z$, $\underline{\Lambda}_y$ are diagonal matrices whose elements are the individual eigenfrequencies λ_z, λ_y of the normal mode; N_z N_y are the projections of the non-linear and forcing terms of the model onto the normal modes $\underline{Z}$ and $\underline{Y}$ ($i = \sqrt{-1}$). The terms $-i\underline{\Delta}_z\underline{Z}$ and $-i\underline{\Delta}_y\underline{Y}$ come from the linear terms of the model equation, which appear in this diagonalized form because the normal modes are eigenfunctions of the linearized equations. N is a non-linear operator and a function of all the normal mode expansion coefficients $\underline{Z}$ and $\underline{Y}$. The linear version of equations 3 and 4 has a strict literal interpretation and we will not deal with this here.

For the non-linear case the situation is more complicated. Machenhauer (1977) showed that for the shallow water equation the gravity modes $\underline{Z}$ contained a low-frequency component as well as the expected high-frequency component. This low-frequency component is essentially due to a quasi-balance between the two terms on the right-hand side of equation 3. The low-frequency balancing component $\underline{Z}_B$ can be obtained by dropping the time-derivative $\underline{Z}$ which gives

$$\underline{Z}_B = (i\underline{\Lambda}_z)^{-1} N_z (\underline{Z},\underline{Y}) \tag{5}$$

Machenhauer (1977) demonstrated that equation (5) is very effective in initializing a shallow water equation model. For an f-plane Leith (1980) has identified the vector $\underline{Z}_B$ with the low frequency ageostrophic flow. The set of Rossby modes $\underline{Y}$ plus the low frequency balancing component $\underline{Z}_B$ constitutes the slow manifold of Leith (1980). Fig. 4 which is taken from Daley and Puri (1980) illustrates the slow manifold concept. The Rossby mode amplitude $\underline{Y}$ is the abscisse and the ordinate is the gravity mode amplitude $\underline{Z}$. The curved line M represents the locus of low-frequency model states which Leith (1980) calls the <u>slow manifold</u>. The projection M on $\underline{Z}$ gives the balancing component $\underline{Z}_B$. When $\underline{Z}_B$ tends to zero the slow manifold M becomes identical with the Rossby manifold. Provided the non-linear terms of the model are not too large, the balancing component $\underline{Z}_B$ is relatively small compared to Y, so that the slow manifold is not far from the Y-axis.

It must be realised that this is a very simplified description. In reality $\underline{Z}$ and $\underline{Y}$ are multi-dimensional and M moves in a multi-dimensional space. Moreover, the question of higher order corrections and the approach to the slow manifold by iterations, which is done in practical applications, has not been considered.

The objective of initialization is to project the data manifold D onto the slow manifold M. This is indicated by I in Fig. 4.

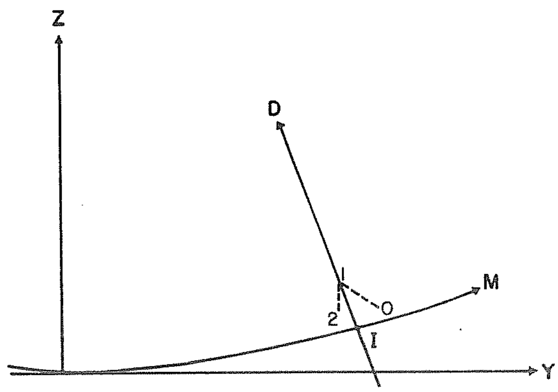


Fig. 4 The slow manifold diagram. After Leith, 1980.

In the case where no data is inserted into the model, any state vector on M will continue to move slowly on M, while any state vector not on M will oscillate about it. It is presumed that the amplitude of the high-frequency transient gravity modes is proportional to the distance of the state vector from the slow manifold M.

The non-linear normal mode scheme according to Machenhauer (1977) can now easily be illustrated. We first perform the insertion step moving the state vector from point 0 (first guess) to point 1 (insertion). We then calculate an adjustment to the state vector, demanding that the new state vectors $\underline{Y}(2)$ and $\underline{Z}(2)$ satisfy

$$Z(2) = (i\underline{\Lambda}_Z)^{-1} \; N_Z[Z(2), \underline{Y}(2)] \qquad (6)$$

where Z(2) will be recognised as the balance component $\underline{Z}_B$ calculated in (5). The solution of (6) requires an iterative step because $\underline{Z}(2)$ appears on both sides. However, the convergence is very fast and only a few iterations are required. In terms of Fig. 4 point 2 would be on a vertical straight line passing through point 1 and would also be on or near the slow manifold M. However, point 2 would not necessarily be close to the ideal (initialization) point I. This scheme would give very little excitation of gravity waves but the observation would not necessarily be fitted in an optimal way.

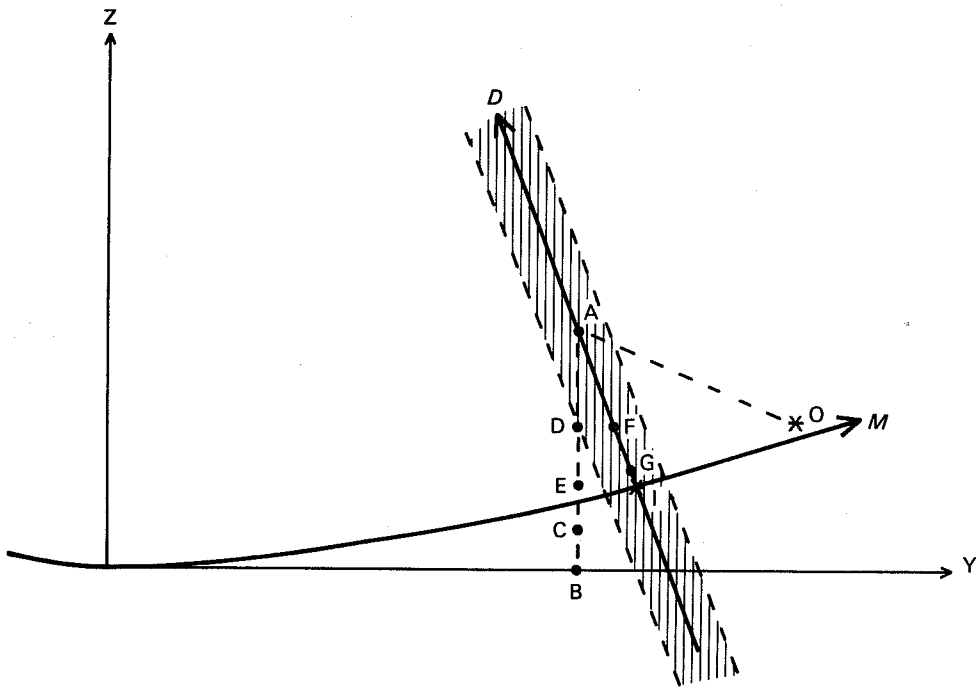


Fig. 5 **Schematic slow manifold diagram to illustrate different initialization procedures. O indicates model state vector before insertion, and I ideal initialization point, being the intersection between the data manifold *D* and the slow manifold *M*. Position A-G indicates different initialization procedures. For further information see text.**

Let us use Fig. 5 to illustrate seven different kinds of initialization, where we have assumed that we only observe geopotential using model simulated observations which are assimilated by a spectral shallow water equation model. These initializations are as follows:

A	Direct insertion (equivalent to 1 in Fig. 4)
B	Geostrophic insertion
C	Geostrophic insertion + quasi-geostrophic divergence
D	Linear normal mode
E	Non-linear normal mode
F	Linear normal mode (forced)
G	Non-linear normal mode (forced)

In contrast to Fig. 4 we will here symbolize the data manifold with a band, where the width of the band is inversely proportional to the accuracy of the observation. A correct observation will fall on the centre line. The initialization examples indicated by "forced", within brackets, incorporate a variational principle according to Daley (1978) where specific weighting functions, inversely proportional to the accuracy of the observations, are being used as a constraint on the initialization.

As in Fig. 4, a direct insertion will move the state vector from 0 to A. If we now assume that we will initialize the inserted geopotential by making use of the non-divergent geostrophic relation, this can be illustrated by moving the state vector to point B (on the Y-axis), since all the gravity modes will be removed.

If this initialization will be further improved by calculating the divergent wind from the quasi-geostrophic ω-equation, which has been a common procedure in the past, this will create ageostrophic divergent winds and consequently move the state vector closer to the slow manifold (point C).

The four different normal mode schemes can be exemplified in this very simple diagram in a similar way. The linear normal mode scheme would in the case when no winds are inserted mean, that the difference field (observed minus forecast) is to be projected entirely onto the Rossby modes. This would give point D in the diagram. The non-linear normal mode initialization will take us to point E, since it will end up on the vertical axis through A but closer to the slow manifold. Finally, the forced normal mode schemes, F and G, will just bring the state vector towards the centre of the data manifold as indicated in the figure. Both theoretical evaluations and numerical experiments given below, indicate that G comes closer to I, the ideal insertion point, with F and E as second alternatives.

Examples of these different kinds of initialization can be found in Daley and Puri (1980). They show the initial data and results from scheme A, D and F. It is clearly demonstrated that direct insertion of geopotential or mass field information is very badly assimilated by the shallow water equations.

In the case both analysed height and winds are used, the data manifold is changed to an area. Experiments show that direct insertion of winds assimilates very well, which indicates the wind data are closer to the slow manifold. This follows from the geostrophical adjustment theory as has recently been demonstrated by Daley (1980).

Initialization in particular using the normal mode technique can readily be extended to a baroclinic model. Andersen (1977) and Daley (1979) have demonstrated this. The non-linear normal mode initialization of the ECMWF 15-level global model has been described by Temperton and Williamson (1979).

Talagrand has calculated the correlation between observed and initialized surface pressure tendencies (calculated over 3 hours). They are given in Table 2.

TABLE 2 Correlation between observed surface pressure tendencies and tendencies calculated from the model initialized by the non-linear normal mode technique for one particular day. Surface pressure tendency observations have been stratified in four different regions and also grouped according to the magnitude of the tendency. Without initialization there is no correlation between observed and calculated tendencies (number of observations within brackets).

SURFACE PRESSURE TENDENCIES

Correlation observed/initialized 00Z 16 January 1979

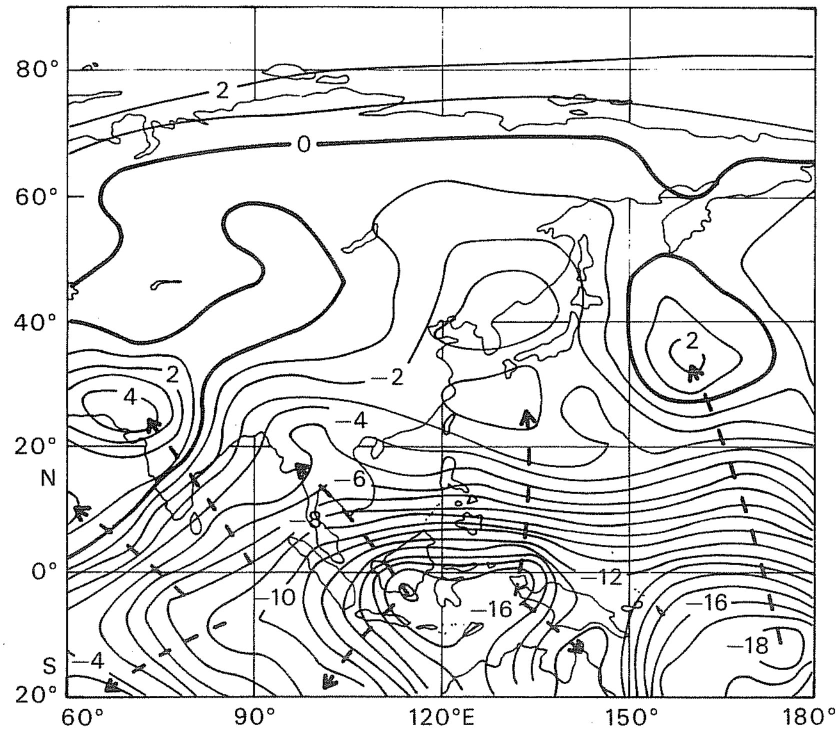
Global	(2277)	0.70
N.H.	(1520)	0.76
Tropics	(627)	0.26
S.H.	(130)	0.39

	<1 mb	>1 mb	>2 mb	>3 mb	>4 mb	>5 mb
Global	0.28	0.82	0.86	0.87	0.89	0.89
N.H.	0.32	0.86	0.89	0.90	0.91	0.91

In spite of many advantages the non-linear normal mode initialization has drawbacks. In particular the method is presently not very successful to initialize the divergence in the Tropics. Fig. 6 shows the analysed and initialized divergence in a tropical area around Indonesia, where the divergent wind has a maximum. It is found that the initialization reduces the amplitude of the divergent wind by at least a factor of 2, in particular at the upper level of the troposphere (200 mb).

The initialization is in particular harmful for the large-scale components of the divergence (Fig. 7), which depicts the mean meridional circulation from the south- to the north pole. Initialization reduces in particular the outflowing branch of the Hadley circulation.

The Hadley circulation recovers after about a day (Fig. 8), while it takes several days for wavenumber 1 to build up to full strength. Consequently the FGGE data base at ECMWF will not include initialized data. It seems clear that the assumption underlying the concept of the non-linear mode e.g. (5) is not directly valid in the Tropics and some form of forced adjustment of observed winds is necessary.



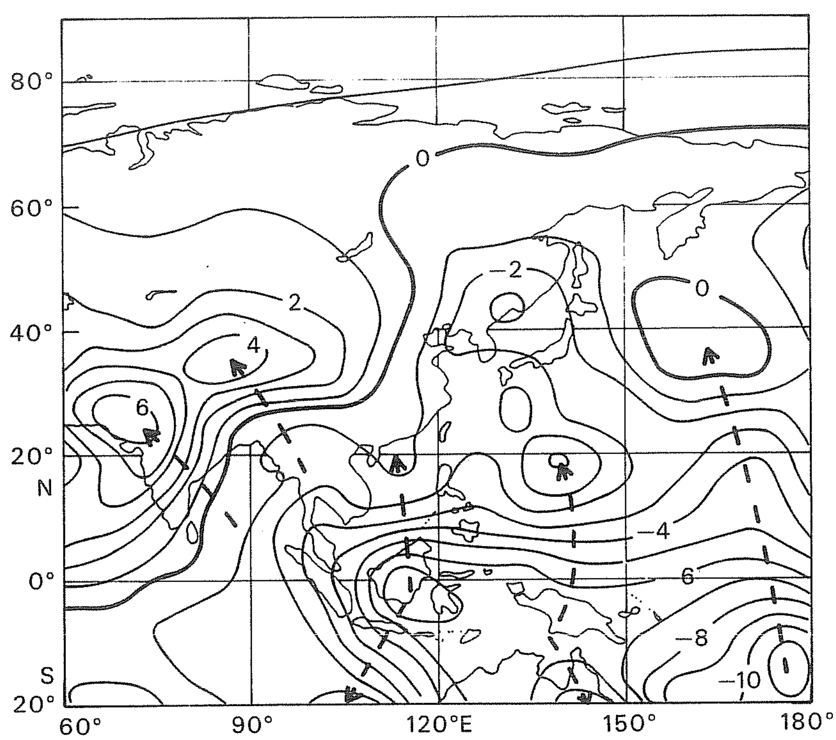


Fig. 6 Velocity potential at 200 mb over eastern Asia at 00 GMT 27 December 1978. Analysed values (top) and initialized values (bottom). Units $10^6\ m^2\ sec^{-1}$.

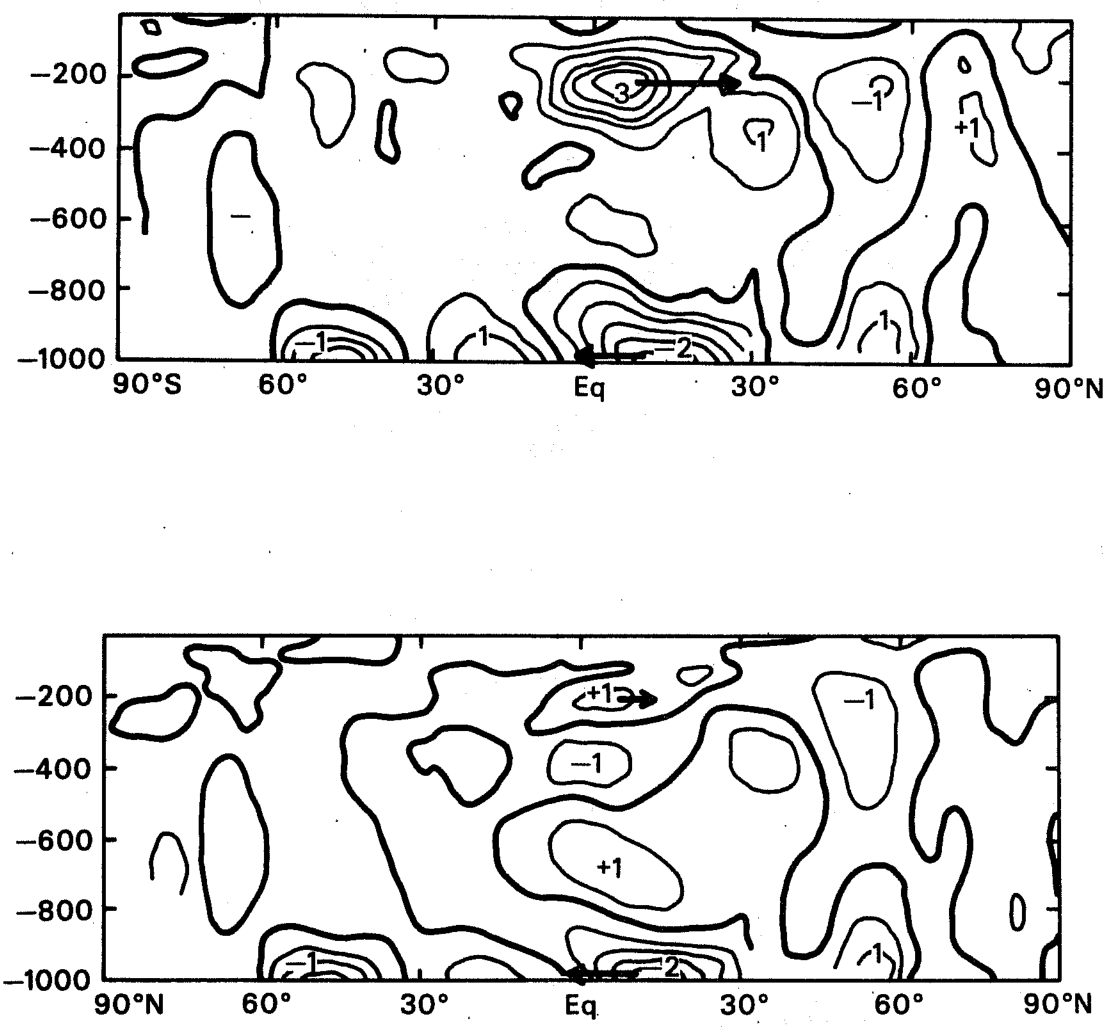


Fig. 7 Mean meridional circulation for 00 GMT 16 January 1979 in m sec^{-1}. Uninitialized data (top) and initialized (bottom). Positive values northerly winds. The heavy arrows indicate maximum inflow and outflow in the Hadley circulation. The length of the arrow is proportional to the magnitude of the meridional wind.

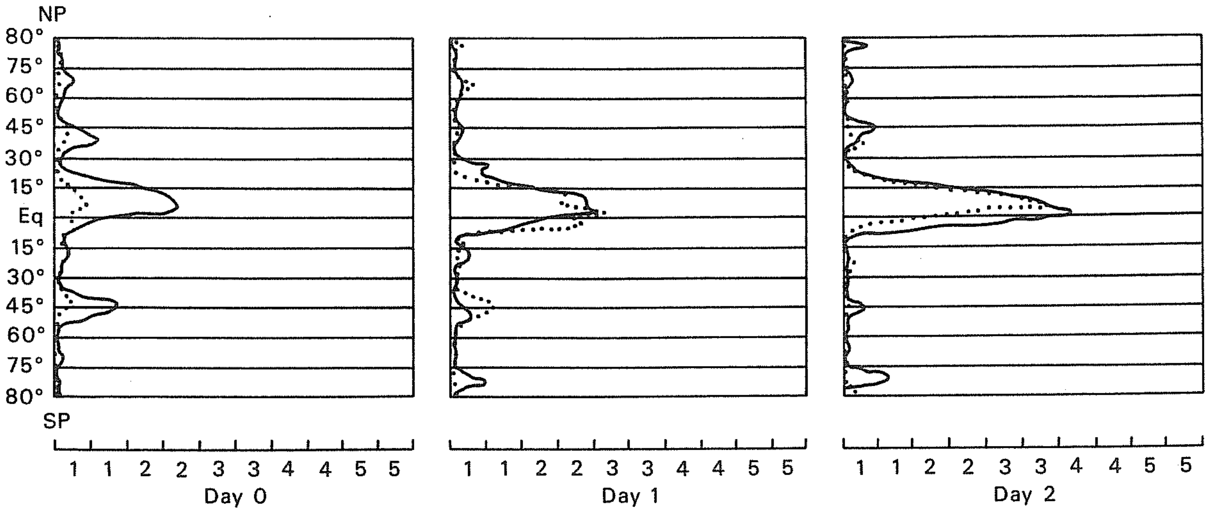


Fig. 8 Latitudinal distribution of divergent kinetic energy at 200 mb for the analysed flow (full lines) and uninitialized (day 0) and predicted (day 1 and day 2) (dotted lines). Units: $m^2\ sec^{-2}$.

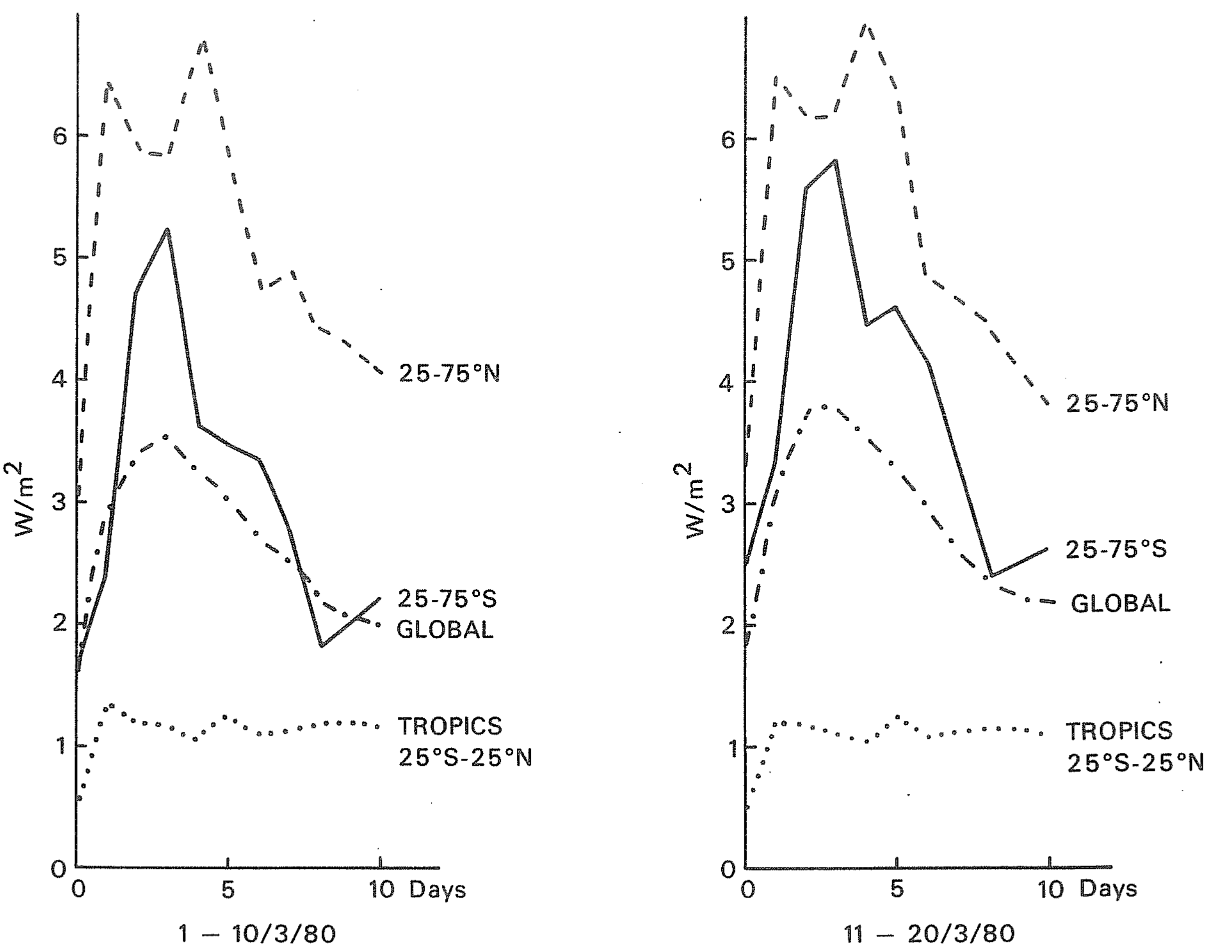


Fig. 9 Spin-up time for total dissipation in Watt/m^2 for different regions. These figures which are ensemble values have been calculated for two different periods in March 1980.

4. THE SPIN-UP PROBLEM

A fundamental problem in numerical forecasting and in data assimilation is the problem of spin-up. The spin-up time is the time necessary for the model to fully reach its own circulation regime. By this we mean the establishment of boundary layer fluxes, fronts, large-scale precipitation system, convection processes etc. For a complete general circulation type of model, the spin-up time is very long and can last for several days. For the ECMWF model the average spin-up time is about three days.

Fig. 9 shows the spin-up for the total dissipation. The long spin-up time means that each analysis produces a "shock" which influences the model forecast over several days. The spin-up effect points at fundamental problems in data assimilation and in numerical forecasting. The reason is of course that the coverage and accuracy of observations are unsatisfactory to correctly analyse intense small-scale features which at least to some extent can be described by a high resolution model with realistic and accurate parameterization.

The spin-up problem can be called the adjustment problem of the second kind, which means the adjustment between the observation and the dynamical/physical structure of the weather systems.

Some improvement can be obtained by minimising the interpolation errors in the data assimilation cycle. Talagrand at ECMWF has carried out a study where only the increments (observed - first guess) are interpolated between the sigma and p-surfaces. (The analyses are carried out at pressure surfaces). Some improvement is obtained as can be seen from Fig. 10. It is also clear that the initialization itself affects the spin-up and as can be seen from Fig. 11 the removal of initialization reduces the spin-up effect in particular for the large-scale convection. Unfortunately the forecast also becomes slightly worse.

I believe that in order to solve the spin-up problem or at least reduce the spin-up time there is very much to be learned from subjective or manual analysis of the weather maps. The Bergen School fostered meteorologists with masterly skill with the weather maps. I think it is very important that a few interesting weather situations could be selected and carefully analysed by skilled meteorologists using the automatically produced FGGE analyses as a first guess. Such a study could teach us very much in finding meteorologically better ways in analysing the weather, because, as V. Bjerknes correctly said, we must do the utmost in using the available observations.

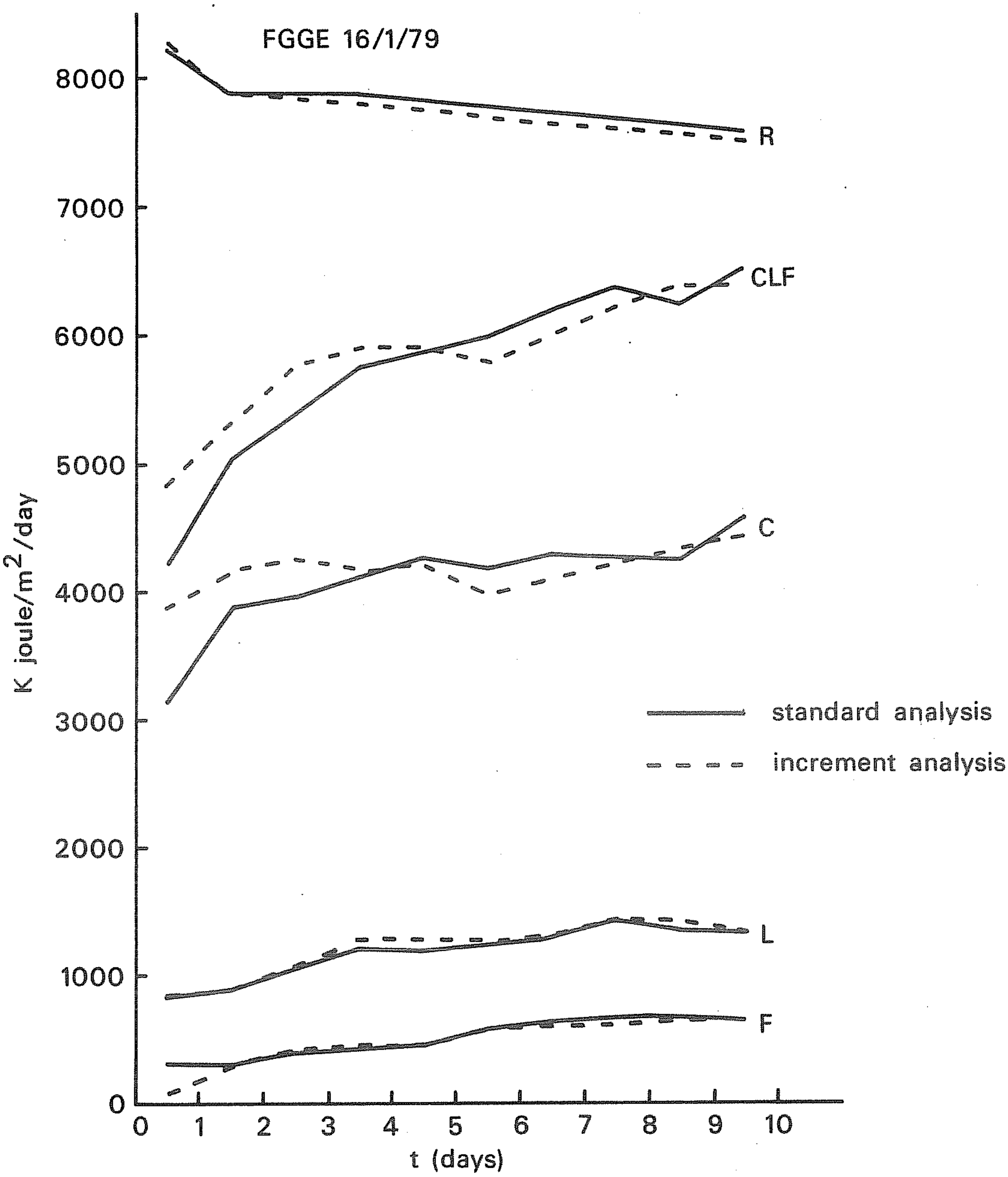


Fig. 10 Spin-up effect on the heat balance for the model from 3 February 1979 in KJoule/m^2/day. Ideally the heating processes from surface fluxes (F) convection (C) and large-scale release of latent heat (L) should balance the cooling from radiation (R). CLF is the sum of C, L and F. Full lines show the spin-up in the standard analyses when the whole fields are interpolated between σ- and p-surfaces. Dashed lines show the spin-up effect when only the analysed deviation from the first guess (increments) is interpolated. It can be seen that the interpolation of the complete field affects the spin-up time noticeably.

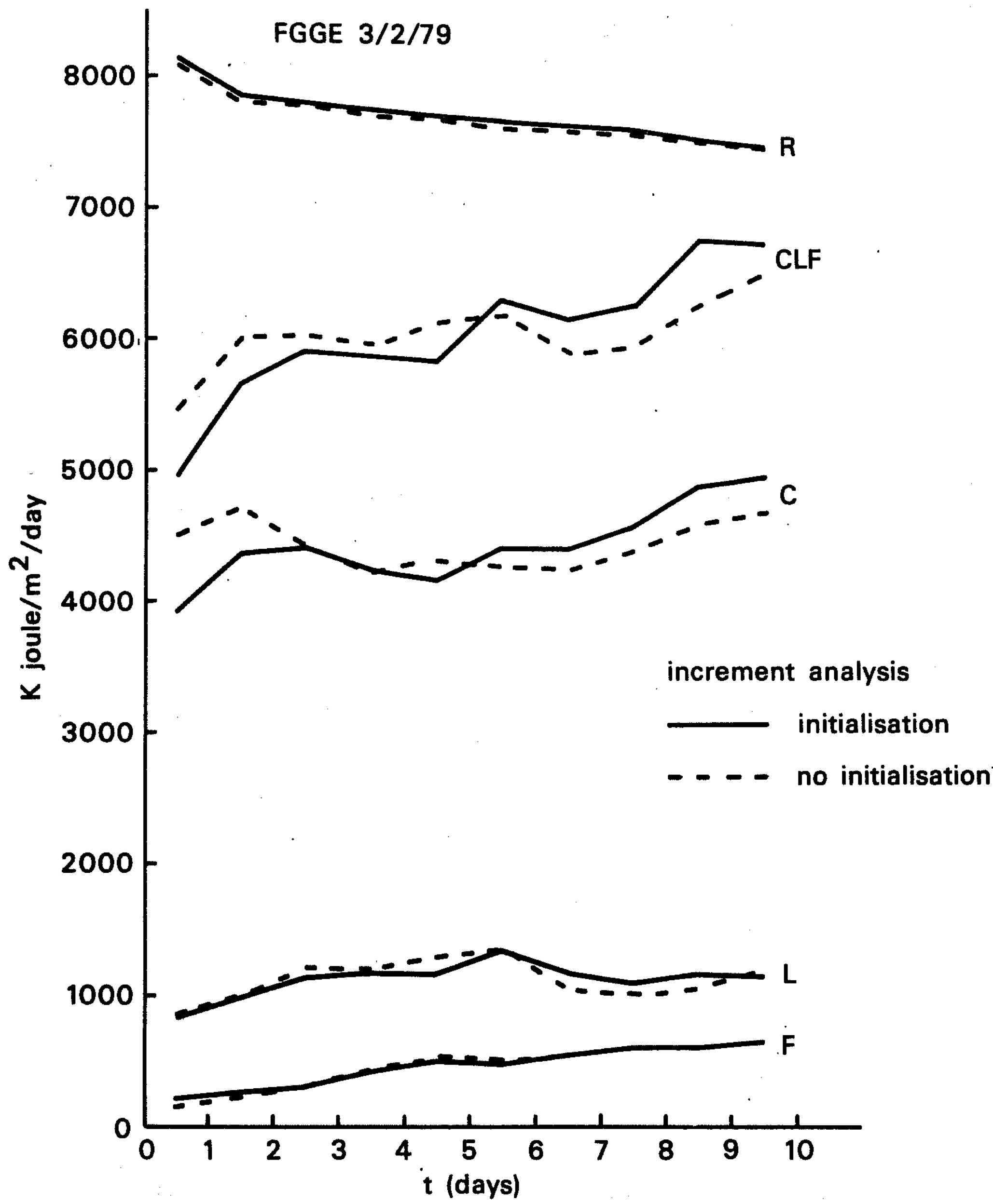


Fig. 11 The same as Fig. 10, but comparing initialized (full lines) and non-initialized (dashed lines). Initialization also affects the spin-up time in particular for the convection.

REFERENCES

Andersen, J., 1977 A routine for normal mode initialization with non-linear correction for a multi-level spectral model with triangular truncation. ECMWF Internal Report 15, 41 pp (Available from ECMWF)

Bengtsson, L., and Gustafsson, N., 1972 Assimilation of non-synoptic observations. Tellus, 24, 383-399.

Bengtsson, L., 1980 On the use of a time sequence of surface pressures in four-dimensional data assimilation. Tellus, 32, 189-197.

Bergman, K., H., 1979 Multivariate analysis of temperatures and winds using optimum interpolation. Mon.Wea.Rev., 107, 1423-1444.

Bjerknes, V., 1904 Das Problem der Wettervorhersage betrachtet vom Standpunkte der Mechanik und Physik. Met.Z., 21, 1-6.

Daley, R., 1978 Variational non-linear normal mode initialization. Tellus, 30, 201-218.

Daley, R., 1979 The application of non-linear normal mode initialization to an operational forecast model. Atmosphere - Ocean, 17, 98-123.

Daley, R., and Puri, K., 1980 Four-dimensional data assimilation and the slow manifold. Mon.Wea.Rev., 108, 88-99.

Daley, R., 1980 The development of efficient time integration schemes using normal modes. Mon.Wea.Rev., 108, 100-115.

Eliassen, A., 1954 Provisional report on calculation of spatial covariance and auto-correlation of the pressure field. Videnskaps Akademiets Institutt for Vaer-og Klimaforskning, Report No.5, 11 pp.

Leith, C.E., 1980 Non-linear normal mode initialization and quasi-geostrophic theory. J.Atmos.Sci., 37, 958-968.

Longuet-Higgins, M., 1968 The eigenfunctions of Laplace's tidal equation over a sphere. Phil.Trans.Royal Soc., A262, No.1132.

Lorenc, A., Rutherford, I. and Larsen, G., 1977 The ECMWF analysis and data assimilation scheme. Analysis of mass and wind fields. ECMWF Tech. Report No. 6, (available from ECMWF)

Lorenc, A., 1980 A global three-dimensional multi-variate statistical interpolation scheme. Submitted to Mon.Wea.Rev.

Machenhauer, B., 1977 On the dynamics of gravity oscillations in a shallow-water model, with applications to normal mode initialization. Beitr. Phys. Atmos. 50, 253-271.

Rutherford, I., 1973 Experiments on the updating of PE forecasts with real wind and geopotential data. 3rd Conf. on Prob. and Stats. in Atmos.Sci. Preprint Am.Met.Soc.

Rutherford, I., 1976 An operational three-dimensional multi-variate statistical objective analysis scheme. Proceedings, JOC Study Group Conference on four-dimensional data assimilation. GARP Program on numerical experimentation. Report No.11, 98-121.

Schlatter, T., 1975 Some experiments with a multi-variate statistical objective analysis scheme. Mon.Wea.Rev., 193, 246-257.

Temperton, C. and Williamson, D., 1979 Normal mode initialization for a multi-level grid point model. ECMWF Tech. Report No. 11, 99 pp (available from ECMWF).