PROBLEMS IN FOUR-DIMENSIONAL DATA ASSIMILATION

BY

L. BENGTSSON

EUROPEAN CENTRE FOR MEDIUM-RANGE WEATHER FORECASTS

1. Introduction

With the introduction of new observing systems based on asynoptic observations, the analysis problem has changed in character. In the near future we may expect that a considerable part of meteorological observations will be unevenly distributed in four dimensions, i.e. three dimensions in space and one dimension in time.

The term analysis, or objective analysis in meteorology, means the process of interpolating observed meteorological observations from unevenly distributed locations to a network of regularly spaced grid points. Necessitated by the requirement of numerical weather prediction models to solve the governing finite difference equations on such a grid lattice, the objective analysis is a three-dimensional (or mostly two-dimensional) interpolation technique.

As a consequence of the structure of the conventional synoptic network with separated data-sparse and data-dense areas, four-dimensional analysis has in fact been intensively used for many years. Weather services have thus based their analysis not only on synoptic data at the time of the analysis and climatology, but also on the fields predicted from the previous observation hour and valid at the time of the analysis. The inclusion of the time dimension in objective analysis will be called four-dimensional data assimilation. From one point of view it seems possible to apply the conventional technique on the new data sources by simply reducing the time interval in the analysis-forecasting cycle. This could in fact be justified also for the conventional observations. We have a fairly good coverage of surface observations 8 times a day and several upper air stations are making radiosonde and radiowind observations 4 times a day. If we have a 3-hour step in the analysis-forecasting cycle instead of 12 hours, which is applied most often, we may without any difficulties treat all observations as synoptic. No observation would thus be more than 90 minutes off time and the observations even during strong transient motion would fall within a horizontal mesh of 500 km x 500 km.

From another point of view it may be difficult to reduce the interval of the analysis-forecasting cycle. Before we can start a numerical integration from an analysis of a given time, the analysis has to be adjusted to attain an initial state which is compatible with the prediction model. This process if called initialization. The initialization so far applied in practice is only using the analysed quantities which are valid initially. During the initialization the atmospheric parameters are forced to fulfil certain dynamical constraints, the most important being a diagnostic relation between the mass field and the wind field (geostrophic equation, balance equation). Since this constraint is valid approximately only in the atmosphere, where the wind is varying around the geostrophic wind, it will take some time for the model to establish a wind field which is adjusted to the time history of the mass field. It will also take some time for the model to generate its own structure patterns (frontal structure, small-scale areas of vertical motion and humidity, etc.), which cannot be inferred satisfactorily from the observations. The model-consistent mechanisms for transporting heat, moisture and momentum and generation of precipitation mechanisms take considerable time to establish as well, especially in areas where the dynamic activity is low.

It is evident from this that the four-dimensional data assimilation not only means a blending of observations with predicted values but also implies that we take care of the necessary adjustment between new observations and predicted values. The objective of the four-dimensional assimilation is thus to determine a sequence of fields of the dependent variables, consistent with the observations as well as with the physical prediction equations governing the evolutions of the field in time.

The observed variables of primary importance for definition of the large-scale motion are the surface pressure (or pressure at any level), temperature, the horizontal wind components and humidity. For the hydrostatic system this information is sufficient to close the system of equations since the vertical motion is a derived quantity and need not be observed.

Recently general summary articles have been published. For a more extensive discussion on different aspects on data-assimilation, the reader is referred to Kasahara (1972), Bengtsson (1975) and McPherson (1975).

2. Historical Review and Background

A sophisticated model needs a more complex initialisation procedure than a simple model. Diagnostic relations as the balance and ω- equations are rather cumbersome to calculate, in particular for a global domain, nor can they be used in the equatorial region. By the aid of calculus of variation more complete constraints can be used but this leads to extremely complicated numerical problems.

During the end of 1960:s the idea appeared to use the numerical model itself in an heuristic way by successively inserting data into the model during a given time sequence. Since we, for the linearized equation, can replace initial conditions on the windfield by the massfield and its time derivatives there was reason to believe that it also would be possible in the non-linear case.

This hypotesis was tried using different models and it was found that the insertions of time sequences of temperature and pressure data generated a correct windfield and similarly the insertions of time sequences of wind data generated a correct massfield. The only drawback was that the process was very slow and of the order of weeks. The experiments also had the limitations that the models were updated with data generated by the model itself in an earlier experiment.

Unfortunately when the experiments were extended to real data or to data generated by other models the results got much worse and in some cases it was not possible to finish the experiments due to numerical instability.

The reason for this is now well understood. (See e.g. Morel et al. (1974).)

If the generated observations are free of errors the forecasting model is able to reproduce exactly the reference flow for an indefinite period of time and thus the predictability of the model with respect to the reference flow is in principle infinite. The model dependent experiments are equivalent to experiments performed by a perfect model using real data.

However, if a difference exists between the model and the data set of observations at any given time, the difference between two numerical integrations initialized from these two states will grow to eventually reach the variability of the model.

Since the reference flow is never perfectly reconstituted in the course of a data assimilation there will always be an error growth of the forecast with respect to the reference flow. It is evident that the time for the error growth must be larger than the data acquisition time, that is the time to assimilate a non-redundant data set for the model.

3. Observational Limitations

In the following we will examine different ways to improve the data acquisition capability of the forecasting model. Before we do that, however, we will study some problems related to the limitations in the basic observations, that is in the level II data base. Table 1 summarizes all sources of information which are available to us. As can be seen, we have also included a valid forecast as well as climatology as relevant information.

NO.	INFORMATION SOURCES	00Z 12Z	06Z 18Z	03Z, 09Z 15Z, 21Z
1	Climatology	X	X	X
2	Forecast	X	X	X
3	Raobs	X	X1)	
4	Dropsondes	X		
5	SIRS	X2)	X2)	X2)
6	SYNOP	X	X	X
7	Buoys	X	X	
8	Constant Levels Balloons (CLB)	X2)	X2)	X2)
9	Cloud Winds	X	X	X
10	Aireps	X	X	X

Table 1

Remarks 1) mainly at NH

2) Information is available at every time interval along bands following the satellite tracks.

We will assume as a reasonable hypothesis that the best analysis is an analysis where we have minimized the initial error. In order to do that, we have to combine or blend the different sources of information in an optimum way. That could most simply be done by giving a normalized weight to the data sources inversely proportional to the RMS error of the corresponding information.

Due to the variation in climate variance, forecasting capability data coverage etc. between different regions, we will have large variations in the dominating data sources from area to area. In table 2 and 3 we have made an attempt to estimate the relative importance of observations from different sources. As can be seen from the tables,we have studied the relative value of the observations in 11 different areas. For simplicity these areas are only separated by latitudes and by land or sea. Such a subdivision is necessary, since the data control- and analysis procedures usually are related to the most reliable sources of information. When checking observations, they must be compared with some other independent information. This can either be observations, a forecast or climatology. Checking a radiosonde against a group of SIRS-soundings giving non-consistent information can easily lead to deletion of the radiosonde, since the SIRS data often are strongly space-correlated. Thus the control system may frequently eliminate correct observations if a general routine is used without paying attention to the data source.

Table 2 presents the situation at the observational hours 00Z and 12Z. At these hours observations from Synop (or Ship) and the upper air network dominates and in areas where these observations have a relatively dense and uniform coverage, they constitute the essential data source.

The usefulness of temperature observations from SIRS is mainly a function of latitude. These observations are less useful at low latitudes, since the error of the observations if of the same order of magnitude as the variance of the climatological monthly means.

The forecast information consists of short range forecasts from 6 to 24 hours. They will of course be most useful in areas where the forecasts have a high skill (middle and high latitudes) and in areas and downstream areas where we have small errors in the initial state. Climatology, finally, is of course more useful the smaller the variance is. The relative value of climatology also depends on whether we have available information from other sources.

Table 3 presents the situation at the observational hours 06Z and 18Z. At this time we have very little upper air observations and thus we have to rely upon surface and satellite observing systems. The relative merits of the forecast increase, compared with the situation at 00Z and 12Z.

It is obvious that we are challenged with a very difficult data problem. The data problem is further complicated by the fact that the error structure of observations from satellite systems shows that the errors are systematic; they depend on the weather situation and they are correlated in space.

	Lat. extent	Land/ sea	Dominating data sources in order					
			1	2	3	4	5	
1.	90N-30N	Land	Raobs	Forecast	Sirs	Climatology		Vert. profile
			Synop			Airep		1 vert. value
2.	90N-30N	Sea	Raobs*	Sirs	Forecast	Cloud winds	Climatology	
			Synop	Airep				
3.	30N-10N	Land	Raobs	Forecast	Sirs	Climatology		
			Synop					
4.	30N-10N	Sea	Raobs*	Cloud winds	Forecast	Sirs	Climatology	
			Synop*	Airep				
5.	10N-10S	Land	Drop sondes	Raobs	Forecast	Climatology	Sirs	
			Synop		Aireps			
6.	10N-10S	Sea	Drop sondes	Cloud winds	Raobs*	Forecast	Climatology	
			Synop*		Aireps			
7.	10S-30S	Land	Raobs	Forecast	Climatology	Sirs		
			Synop	CLB	Aireps			
8.	10S-30S	Sea	Raobs*	Cloud winds	Sirs	Climatology	Forecast	
			Synop*	CLB				
9.	30S-65S	Land	Raobs	Sirs		Forecast	Climatology	
			Synop	CLB	Aireps			
10.	30S-65S	Sea	Raobs*	Sirs	Forecast	Climatology		
			Buoys	Synop*	CLB			
11.	65S-90S	Land	Raobs	Sirs	Forecast	Climatology		
			Synop	CLB				

Table 2

The relative value of different meteorological information in different regions at 00Z and 12Z.

* indicates good quality information, but very limited in space.

	Lat. extent	Land/ sea	Dominating data sources in order					
			1	2	3	4	5	
1.	90N-30N	Land	Raobs*	Forecast	Sirs	Climatology		
			Synop	Aireps				
2.	90N-30N	Sea		Sirs	Forecast	Cloud Winds	Climatology	
			Synop	Aireps				
3.	30N-10N	Land		Forecast	Sirs	Climatology		
			Synop	Aireps				
4.	30N-10N	Sea	Cloud winds	Forecast	Sirs			
			Synop*	Synop	Aireps			
5.	10N-10S	Land		Forecast	Climatology	Sirs		
			Synop	Aireps				
6.	10N-10S	Sea	Cloud winds	Forecast	Climatology	Sirs		
			Synop*	CLB	Aireps			
7.	10S-30S	Land		Forecast	Sirs	Climatology		
			Synop	CLB				
8.	10S-30S	Sea	Cloud winds	Forecast	Climatology	Sirs		
			Synop*	CLB				
9.	30S-65S	Land	Sirs	Forecast		Climatology		
			Synop	CLB				
10.	30S-65S	Sea	Sirs		Forecast	Climatology		
			Buoys	Synop*	CLB			
11.	65S-90S	Land		Sirs	Forecast	Climatology		
			Synop	CLB				

Table 3

The relative value of different meteorological information in different regions at 06Z and 18Z.

4. Data-assimilation by simple Insertion

The first and very unsuccessful data assimilation experiments were carried out in a very simple way. The observations simply replaced the forecasted value in the grid point closest to the observation. This is incorrect in two different ways.

Firstly, from the informational point of view. An observation in a particular place gives in fact information for a whole area or a whole volume in space. One atmospheric parameter is also giving information about other parameters,e.g. wind will give information about geopotential and vice versa. The size of the information area can be derived from the actual structure function. As we have discussed above, also the forecast and in the tropical area, climatology, carries an information content which can have errors comparable to the errors of the observations themselves. 12 hour forecast errors from reliable short range predictions are in fact comparable in quality to good satellite temperature soundings. A multivariate analysis of the deviation between the forecast and the observations which is described elsewhere in this report is a way of blending this information in an optimum way.

Secondly, simple or direct insertion procedures are incorrect from the initialization point of view. The direct insertion also creates high amplitude waves which give a shock to the system. These non-meteorological waves must be damped during the course of the data assimilation procedure.The damping could either be accomplished by a damping time-integration scheme (Euler-backward) or by a time filter. However, this damping must be rather strong and can have harmful effects on the meteorological modes. Since the structure functions basically are derived from observed data, they will more or less represent the structure of the quasi-geostrophic flow. We may therefore expect that the shock effect will be reduced when we insert locally analysed data instead of observations only. In particular it seems important to apply a local multivariate analysis or correspondingly relate wind and geopotential by the geostrophic relation and insert both winds and height at the same time. As can be seen from Fig. 1, this will speed up the updating considerably.

5. Data-assimilation as an intermittent Process

Assimilation of observations in four dimensions can in principle be carried out in two different ways. The first of these will be called an intermittent scheme. In this scheme the data assimilation is separated into two well defined parts, analysis and initialization, and is more or less similar to procedures now used in operational weather prediction.

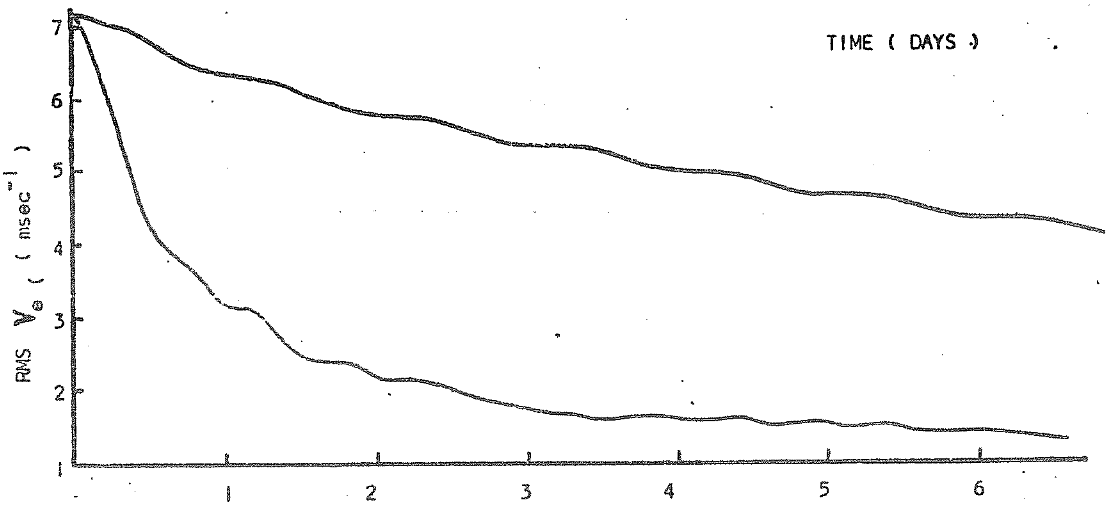

Fig. 1

The upper curve shows the successive reduction of the wind-error using a direct insertion of height data in a barotropic model. The lower curve shows corresponding results, using height data as well as from these derived geostrophic winds. (After McPherson)

In the second scheme, which will be discussed in the next section, the observations will be inserted as they become available. Such a scheme is called a continuous scheme. Insertion means here that the observations are assimilated by a proper analysis procedure applied in the actual influence area.

The first step of the intermittent scheme consists of a 3-dimensional consistent objective analysis, obtained by using an optimal combination of climatology, forecasted values and available information. Statistical as well as dynamical constraints are applied. Such 3-dimensional data sets can in principle be generated for any time-step, but every 12th or possibly every 6th hour may be sufficient.

In the second step the analysed fields are adjusted to the forecasting model by different kinds of initialization procedures (static initialization, dynamic initialization etc.) (See Bengtsson (1975) for terminology and further information).
It is necessary to use sophisticated interpolation procedures during the first step in order to reduce the amount of computations in the second step.

Below we will outline a procedure which basically is an extension of systems which are in operational use at the weather services. We will not go into any details about the space inter-p-lation procedures, but only assume that it should be any form of a multivariate method.

As was mentioned above, it is important to perform the analysis as carefully as possible in order to speed up the initialization procedure. Some studies have indicated that an initialization using several time levels of observations will be more accurate than if (as usually is done) only one time level is used. This is true in particular at middle and high latitudes with strong transient flow. An analysis scheme which will give analyses at t_o, t_o - 6hr and t_o + 6hr is outlined below. The initialization procedure, which follows will yield an initialized state at the time $t = t_o$. Consequently the initialization will be carried out with respect to all data in a 12 hr interval.

During the first step in the analysis procedure the mass field is analysed for three time-levels. At this step as in the other steps below, forecast- and climatological information is included. A three-dimensional consistent analysis of the mass field is carried out. This can be achieved by using the hydrostatic equation as a constraint during the analysis. In the <u>second</u> step the mass field is re-analysed with respect to a modified and improved observational control. <u>Thirdly</u>, the wind field is analysed using all available mass field information at the different time levels as a first guess. In the <u>fourth</u> step the wind field is re-analysed with respect to a modified and improved observational control. <u>Finally</u>, the humidity field is analysed using the wind field to compute non-isotropic structure functions. Each analysis will be processed by making use of all available information in a 12hr interval (± 6hr). This means that the model will be <u>adjusted to observations available in a certain time interval.</u>

Asynoptic information (observations which are not made, say, ± 90 min. within every 6 hr) must be transferred to the actual time intervals. This could preferably be done before the analysis procedure starts by using the valid forecast field. Another possibility is to reduce the time interval to three hours. In such a case all observations can be regarded as "synoptic" but we have to produce five time levels of analyses for each 12-hour interval.

The control procedure is an integrated part of the analysis system. During the <u>first control step</u> the mass field information (p_s,Z and T) is checked and the mass field is checked a <u>second</u> time making use of the available mass field analyses and the wind information. In the <u>third control step</u> the humidity information is checked and the wind field is checked a second time. During the observational control, different checks are made of the observations; hydrostatic control, internal consistency control between the different variables (e.g. geostrophic balance), check against available forecasts and checks against continuity in space and time.Erroneous observations are corrected if possible and other erroneous observations are eliminated from the analyses. The analyses are carried out in principle in the same way at the different time levels.

The objective analysis scheme thus presents information of all parameters at all mandatory levels for three different time-levels. Due to the preliminary fields and the dynamical constraints which are implicitly imposed in the analysis scheme we will assume that the mass field, wind field and humidity field are relatively well adjusted.

The analysis is assumed to be performed in the p-system but the forecast model will most likely use the σ-system. The interpolation from p- to σ- surfaces can obviously again create an imbalance. It seems, however, reasonable that this will not be so serious if a high vertical resolution is used and if mountain areas are treated carefully.

In order to initialize a model with respect to the observations available in a time-interval, dynamical initialization is the most efficient procedure.

During the dynamic initialization the forecasting model will be integrated forward and backward around the time t_o. The details in this performance can be designed in many different ways and several experiments have to be carried out to find the best procedure. One simple procedure would be to integrate backward and forward in time over the 12hr interval and then get the initial state by a weighted mean . (See Fig.2) Another approach would be to integrate forward and backward between -6hr and +6hr, blending analysed fields and predictions at the times where analyses are available.

There are several disadvantages with the intermittent data assimilation. The data handling is very complicated and necessary dynamical constraints are difficult to apply in particular on a global domain. Dynamical initialization is very time-consuming and the forward-backward integration can correspond to several days' computation.

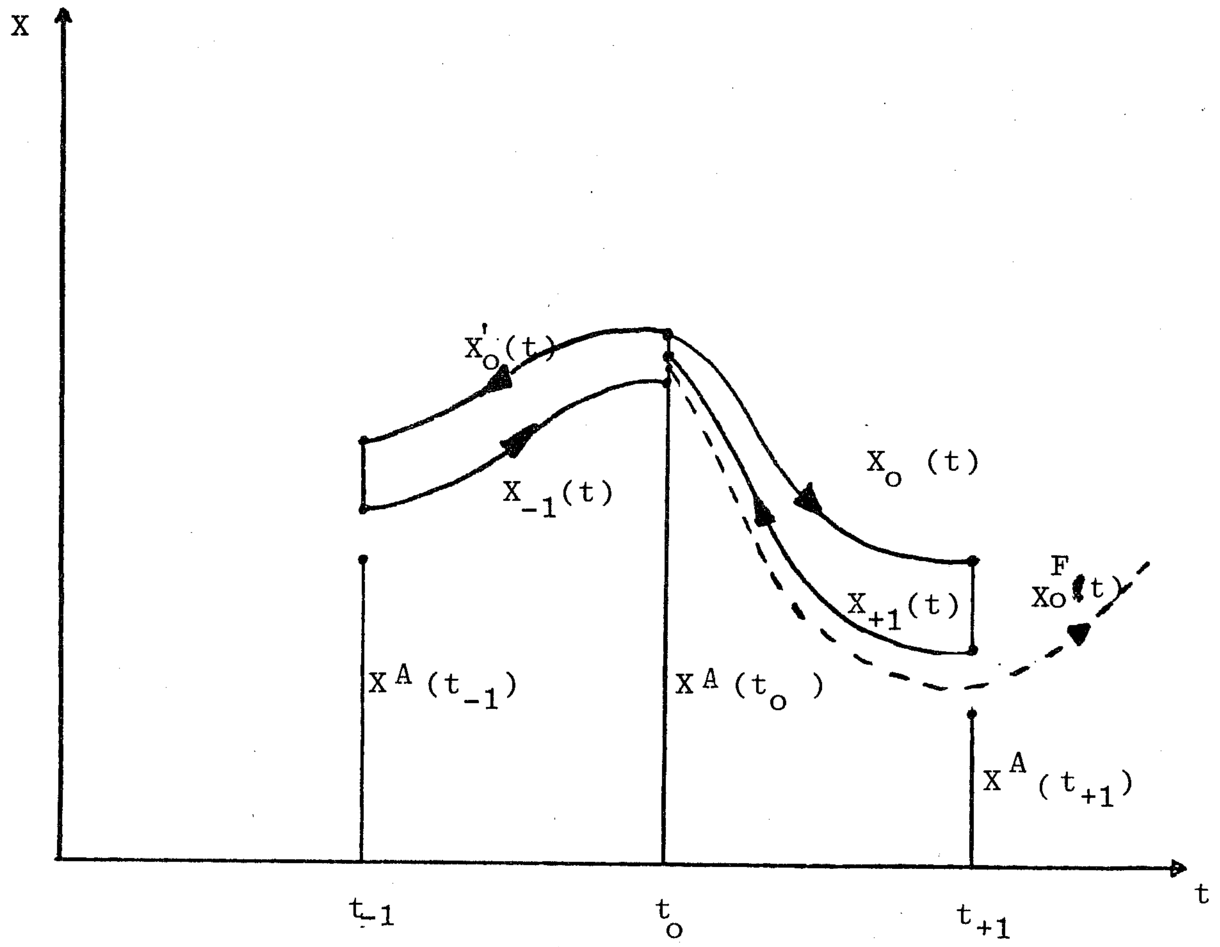

Fig. 2

Schematic illustration of a forward and backward initialization method. The X-axis represents the 3-dimensional phase space. In this particular example we assume that we integrate from $t = t_o$ backward in time 6 hr and forward in time 6 hr. $X^A(t_{-1})$, $X^A(t_o)$ and $X^A(t_{+1})$ represent the analysed values at the 3 different times t_{-1}, t_o and t_{+1}. The forward integration, $X_o(t)$ starts at t_o and goes to t_{+1}. The new initial state at t_{+1} is integrated backwards to t_o (called $X_{+1}(t)$). A corresponding procedure is done backward in time. Finally the two integrations $X_{+1}(t)$ and $X_{-1}(t)$ are combined with $X^A(t_o)$ to give the final state $X_o^F(t)$.

Another severe disadvantage can be seen in Fig. 3, which illustrates the problem of initializing precipitation. A continuous assimilation is in that case superior. The reason is that it will take some time for the model to generate its own structure patterns (frontal structures, small-scale areas of vertical motion and humidity etc.) which cannot be inferred satisfactorily from the observations. The model-consistent mechanisms for transporting heat, moisture and momentum and generation of precipitation mechanisms take considerable time to establish as well, especially in areas where the dynamic activity is low.

6. Data assimilation as a continuous Process

Contrary to the intermittent scheme the continuous data assimilation is straightforward and from the formal point of view very simple.

A major problem, however, and a difference from conventional methods is that we are blending observations with a forecast with a different time span at the different grid points.

For any model the forecast error varies in time and space. The amplitude expressed by the forecast error variance has an exponential increase at least for small errors,but the horizontal scale of the error which can be expressed by the autocorrelation increases very slowly.

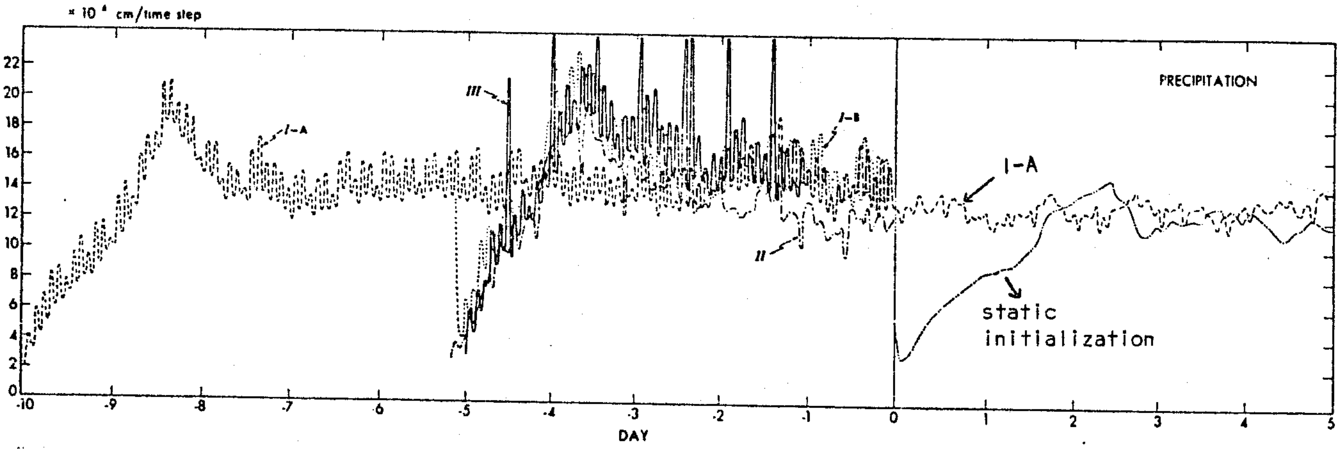

Fig. 3

Predicted averaged precipitation based on (a) static initialization and, (b) 4-dimensional data assimilation (I-A). Observe the long time for the model to generate its own precipitation mechanism in case (a). (After Miyakoda)

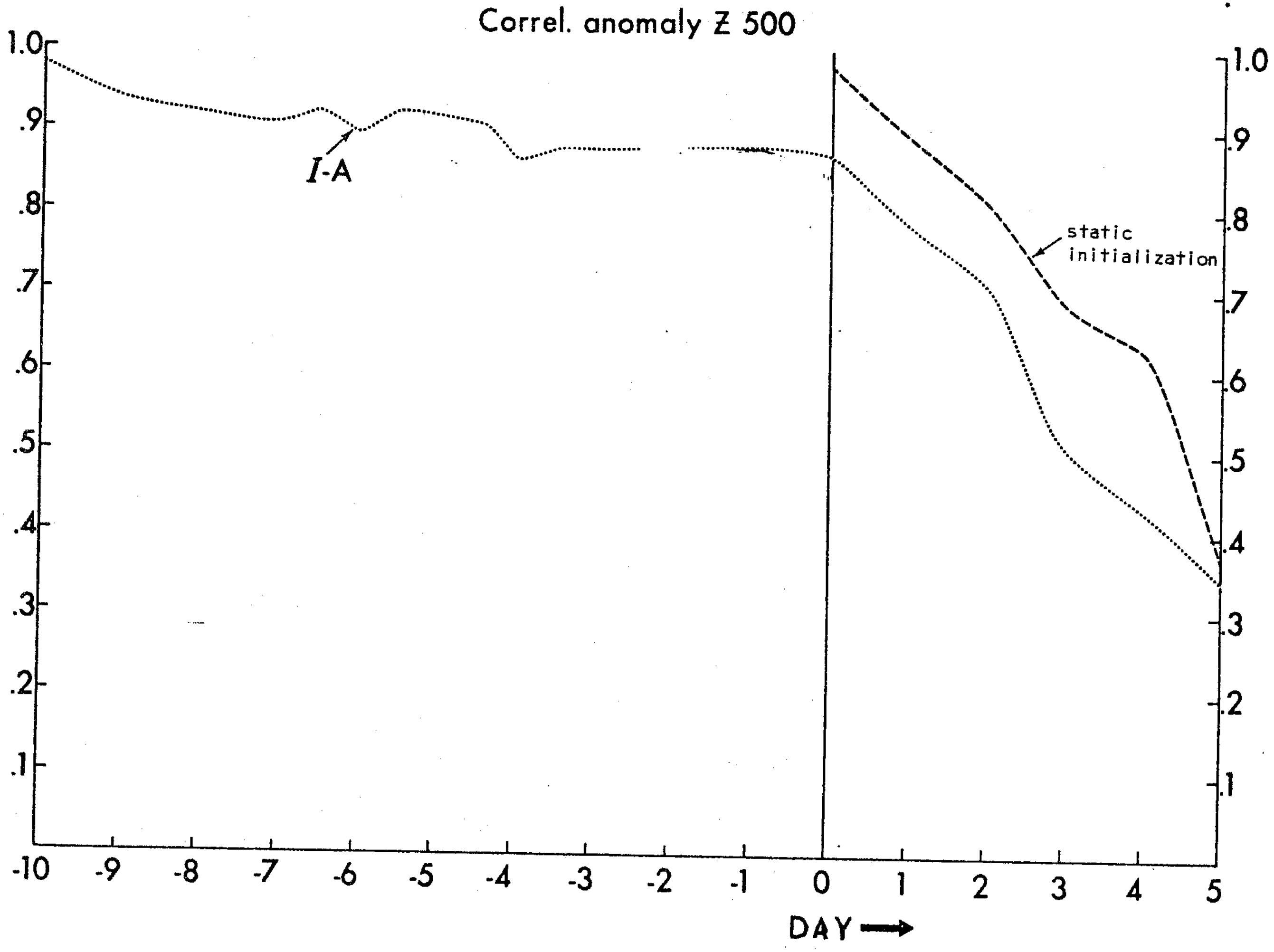

Fig. 4

Forecast experiments based on (a) static initialization (dashed line) (b) continuous 4-dimensional data assimilation (I-A) dotted line).

As a first approximation we may assume that the structure of the forecast error is roughly unchanged in the forecast interval 3- 12 hr. This means that we can assume the autocorrelation of the forecast error to be the same and independent of the forecast length. The amplitude of the forecast error will, of course, vary strongly in space and time. In order to optimize the information, we would like to know how the variance changes. This is extremely time-consuming to do in a statistically correct way. However, as a first approximation we may apply the following simple procedure.

By the aid of optimum interpolation we compute the mean square interpolation error, E, in a grid point, i at the time, t. This reads:

$$E = m_{iit} - \sum_{k=1}^{N} m_{ikt} p_k \qquad (1)$$

where $m_{iit} = \overline{(\alpha_{it} - \alpha_{it}^{F})^2}$ denotes the variance of the forecast error in the grid point i. α is an arbitrary parameter and superscript F denotes the forecasted value.

m_{ikt} is the corresponding value for the difference between the observed and forecasted value and p_k is the weight computed by optimum interpolation.

Expression (1) can be used to compute the new variance estimated after the analysis has been completed in the influence area. As a very simple first approximation we can assume that the variance grows linearly in time to a maximal value $(m_{ii})_{max}$ during a time period T. Thus

$$m_{ii}\,(t+\Delta t) = m_{iit} + \frac{\Delta t}{T}\,(m_{ii})max. \qquad (2)$$

The maximum value of the forecast error variance can be put equal to the total variance (deviation from climatology).

As was discussed in section 2 the insertion of new information during the course of the integration will create shock effects in a primitive model. Different methods have been tried to overcome this problem . Miyakoda et al (1975) have carried out a global data assimilation experiment during the whole GATE experiment. During that experiment the observations were first interpolated in time to 2 hour intervals. These data were then inserted repeatedly every time step for each 2-hour interval, beginning one hour before the valid time . The repeated insertion speeded up the assimilation considerably. In order to damp gravity waves, an Euler-backward or Matsuno scheme was applied during the whole experiment . Unfortunately the application of the Euler-backward scheme also has a negative effect on the meteorological modes and in particular small scale systems will be smoothed. Fig. 4 shows the result from another similar experiment carried out by Miyakoda. The correlation anomaly for the geopotential at 500 mb indicates that the forecast based on continuous data assimilation is inferior to a forecast based on a conventional static initialization. As can be seen from the figure this is due to the deviation (error) in the initial state.

On the other hand, for initialization of precipitation, the continuous data assimilation is superior to the static initialization (Fig. 3).

Evidently there are still remaining problems in continuous assimilation of real observations. The basic problem is still to find a way to put the information into the meteorological modes. From that point a filtered model would be a way to solve this problem. A spectral model, which uses stream function and velocity potential as historical parameters instead of the wind components can easily be modified in order to suppress non-meteorological modes.

REFERENCES

Bengtsson, L. 1975:

4-dimensional assimilation of meteorological observations. GARP Publications Series No. 15. Global Atmospheric Research Programme, WMO-ICSU Joint Organizing Committee, 76 pp.

Kasahara, A., 1972:

Simulation experiments for meteorological observing systems for GARP. Bull. Amer. Meteor.Soc., 53, 3, pp 252-264

McPherson, R.D., 1975:

Progress, problems and prospects in meteorological data assimilation. Office Note 110, NMC Development Division, NWS, NOAA. 23pp.

Miyakoda, K. Umscheid, L., Lee, D.H. Sirutis, J., Lusen, R. and Pratte, F., 1975:

The near-real time global four-dimensional analysis experiment during the GATE period, Part I. Geophysical Fluid Dynamics Laboratory/NOAA.

Morel, P. and Talagrand, O., 1974:

Dynamic approach to meteorological data assimilation, Tellus 26 pp 334-343.