PERSPECTIVES IN DATA ASSIMILATION

O. Talagrand
Laboratoire de Météorologie Dynamique du CNRS
Paris, France

Summary: All algorithms that are, or have been, used in operational data assimilation, as well as most algorithms that are, or have been, studied at the research level, can be described as more or less simplified forms of *statistical linear estimation*. Even though the numerical models used in assimilation are highly nonlinear, the initial uncertainty on the state of the flow at the start of the assimilation is small enough so that a local approximation, linear with respect to the difference between the estimated and real states, is generally legitimate. The validity of this so-called *tangent linear approximation* justifies the use of the present assimilation algorithms, which all produce approximations to the *BLUE* (Best Linear Unbiased Estimate) of the state of the flow in terms of the available observations. Present assimimilation algorithms are of two forms. In sequential algorithms (the only ones used so far in operational Numerical Weather Prediction), whose exact BLUE form is Kalman filtering, observations are introduced sequentially in the course of one integration of the assimilating model over the assimilation interval. In variational algorithms, a sequence of model states is globally adjusted to the observations over the assimilation interval. This is made possible through the use of the adjoint equations of the assimilating model. Active research is at present performed on both types of algorithms.

Present assimilation algorithms produce results which are generally considered as satisfactory, and progress in assimilation methods has substantially contributed to the gradual improvements of the quality of numerical weather forecasts. A number of problems still remain, among which one can mention the following ones.

- The determination of the errors affecting the various sources of information used in the assimilation, namely the errors affecting the observations and the errors affecting the assimilating model. *A priori* specification of these errors is required by the theory of statistical estimation. In spite of a number of instructive works, we still are far from having reached a satisfactory knowledge of these errors. They can in principle be determined, at least in part, from appropriate statistics performed on the sequence of observations, or on the sequence of various quantities obtained as by-products of assimilation. A fundamental question is which components of the various errors can be reliably estimated, and at which cost. *Adaptive filters* are algorithms which, in the context of sequential estimation, extract the statistical information contained in the innovation vector (difference between the observation vector and the corresponding first-guess). In variational assimilation, useful information is contained in the misfit between the observations and the corrersponding minimizing model solution, as well as in the adjoint solution.

Variational assimilation, as it has been implemented so far, ignores model errors. An estimate of the relative magnitude of model and observation errors would be extremely useful. Model errors, once they are approximately estimated, can be introduced in variational assimilation either by imposing the model as a "weak constraint", or by weighting observations with time-varying weights.

Associated with the problem of the determination of the various errors is the problem of the impact, on the result of the assimilation, of a misspecification of the errors.

- The determination of the limits of the tangent linear approximation. As long as one is concerned with forecasting the large scale atmospheric circulation, the tangent linear hypothesis is valid to a range of about 48-72 hours. Typical questions that arise are : What are the limitations imposed on assimilation by the nonlinear, chaotic, character of the flow ? Until what "age" is an observation useful ? Are specific algorithms more appropriate for extracting the useful information contained in "old" observations ? Is the tangent linear approximation valid for all observing instruments (*e. g.*, scatterometers) ?

One particular form of nonlinearity results from the presence in assimilating models of threshold processes, such as convection. Much concern has been expressed about the possible negative impact of these processes on the efficiency, and even feasibility, of variational assimilation. The magnitude of the impact of threshold peocesses on variational assimilation has not yet been precisely evaluated. In any case, the problem also exists in sequential assimilation, which is fundamentally also based on the tangent linear hypothesis.