ASSIMILATION OF OBSERVATIONS INTO NUMERICAL MODELS

Olivier Talagrand
Laboratoire de M t orologie Dynamique, IPSL, ENS, Paris, France

Summary *Assimilation* of meteorological or oceanographic observations is the process by which observations are combined together with a numerical dynamical model of the flow in order to produce as accurate as possible a description of the state of the atmosphere or the ocean.. The basic problems raised by the development and implementation of assimilation, as well as the methods and numerical algorithms that are used in practice, are presented and discussed.

Numerical modeling has now become a major component of the study of the dynamics of the atmosphere and the oceans. A large variety of numerical models, which compute the temporal evolution of the atmospheric or oceanic flow for given initial and lateral boundary conditions, have been developed, and cover a very broad range of physical processes, as well as of spatial and temporal scales. Some models are intended at studying specific fundamental dynamical phenomena, such as convection or baroclinic instability. At the other extreme, models have been developed for simulating, as accurately as allowed by the present state of knowledge and the available computing resources, the whole range of complex interactions which determine and govern the climate°: mechanical and thermal exchanges between the atmosphere and the oceans, atmospheric water cycle and oceanic salinity cycle, biogeochemical cycles, atmospheric and oceanic chemistry, etc. Other models have been developed for a very specific, directly utilitarian purpose, such as for instance weather prediction.
It is numerical weather prediction that is at the origin of the development of *assimilation of observations*. In order to perform a numerical weather forecast, it is necessary to have at oneÕs disposal, not only an appropriate numerical model, but also as accurate as possible a description of the corresponding initial conditions. The available observations do not provide however the temporally synchronous and spatially homogeneous description of the atmosphere that is required by a numerical model. In addition, many observations, such as satellite observations, which have now become the major source of information on the state of the atmosphere, are ÔindirectÕ in that they bear, not on the physical quantities in terms of which the state of the flow is normally described (temperature, pressure, wind components, É), but on more or less complicated functions (often, one-dimensional space integrals) of those quantities. Assimilation of observations is the process by which observations are combined together with a numerical model of the flow in order to produce as accurate as possible a spatially homogeneous description of the state of the flow.
Any information will always be affected with some uncertainty. Uncertainty is most easily described by probability distributions, and this leads to consider assimilation as a problem in *bayesian estimation*, *i. e.*, to look for the conditional probability distribution of the state of the flow, given the available information and the probability distribution of the associated uncertainty. In that respect, assimilation of observations is one of many *inverse problems* that are encountered in many fields of science of technology. In spite of the variety of applications, most (if not all) inverse problems are solved using the same basic mathematical methods. Difficulties that are specific to meteorological and oceanographical applications are the large numerical dimensions of the problems to be solved (the number of individual scalar meteorological observations is now in the range 10^6-10^7 per 24-hour period**,** with a similar value for the state vector dimension of the assimilating models), and the nonlinearity of the underlying dynamics. In the case of numerical weather prediction, one must add to those two basic difficulties the need for timely, reliable and robust production of the forecast.
Because of various reasons, and particularly of the large numerical dimensions involved, bayesian estimation cannot be implemented in practice, in its full generality, in meteorological and oceanographical applications. One must restrict oneÕs ambitions to much more modest goals. Two basic approaches have been used so far.
Statistical linear estimation [1], which looks for the estimated state as the linear combination of the data (observations and model equations) that minimizes the statistical variance of the corresponding estimation error. Linear estimation can be useful for what is basically a nonlinear problem to the extent the so-called *tangent linear approximation* is valid. This means that a prior estimate, or *background*, is available for the state of the flow, which is close enough to the real state so that the estimation problem can be linearized in terms of deviations from that prior estimate. Statistical linear estimation produces the so-called *Best Linear Unbiased Estimate* (*BLUE*) of the state of the flow, and requires the *a priori* knowledge of only the expectation and covariance matrix of the data errors. In addition to an estimate of the state of the flow, statistical linear estimation also provides in theory an estimate of the covariance matrix of the corresponding estimation error. Moreover, if the data errors are gaussian, the *BLUE* achieves bayesian estimation in the sense that if entirely defines the conditional probability distribution of the state of the flow, given the data.
Ensemble assimilation [2], which produces an ensemble of possible states of the flow, whose distribution is meant to describe the desired conditional probability distribution. Ensemble assimilation is of particular interest in circumstances where a tangent linear approximation is not valid.
Most assimilation methods have so far been based on statistical linear estimation, with innumerable variants as to the specification of the expectation and covariance matrix of the data error, and as to the algorithmic implementation. When the temporal dynamics of the flow is explicitly taken into account, two classes of algorithms are possible. *Sequential assimilation* constantly updates the most recent estimate of the state of the flow with new observations. The exact *BLUE* form of sequential assimilation is *Kalman filtering* [3] which, because of the need for explicitly computing the temporal evolution of the covariance matrix of the estimation error, goes however well beyond available numerical

resources. *Variational assimilation* globally adjusts the assimilating model to observations distributed over a period of time. It becomes numerically feasible through the use of the *adjoint* of the assimilating model. One advantage of variational assimilation is that it provides a very efficient way to carry information both forward and backward in time. In addition, and contrary to Kalman filter, whose optimality requires that the errors affecting the data must be uncorrelated in time, it can relatively easily cope with time-correlated errors. On the other hand, and contrary to Kalman filter, it does not provide an explicit estimate of the estimation error.

Ensemble assimilation, which, as said, is particularly appropriate for situations when a local linear approximation is not valid, is most often sequential in nature. The ensemble of possible states is evolved in time, and constantly updated as new observations become available. It is a natural extension of Kalman filtering to nonlinear dynamics.

All assimilation algorithms require an *a priori* estimate of, at least, part of the probability distribution of the errors affecting the data. That information cannot be obtained from the data themselves, even through appropriate statistical processing. It entirely depends on independent hypotheses that cannot be objectively validated from the data themselves. This introduces an additional difficulty in the development of assimilation algorithms.

As said, assimilation of observations originated from the needs of numerical weather prediction. Together with improvements in numerical models and (to a lesser extent) improvements in the observing system, improvements in assimilation techniques have led to the steady progress, over the last three decades, of the quality of meteorological forecasts. Assimilation has also extended to many various applications. It first extended to observations of the oceanic circulation, then to reassimilation of past observations, and to other domains, such as atmospheric and oceanic chemistry and biochemistry. Assimilation has also led to the development of powerful numerical tools (for instance, adjoint models, or algorithms for computing the dominant unstable modes of the flow) which have in turn been applied to a large variety of diverse uses. One can mention in particular sensitivity studies, and the definition and optimization of observing systems. One specific example is *observation targeting*, intended at identifying, for a particular meteorological situation, critical areas in which observations can be expected to be particularly useful. Assimilation has also played a basic role in the development of ensemble and probabilistic prediction, intended at determining, not one possible future state of the flow, but a whole probability distribution, reflecting the various sources of uncertainty on the future evolution of the atmosphere or the ocean.

References

[1] Talagrand, O.: Assimilation of Observations, an Introduction. *J. Meteor. Soc. Japan* **75**:191-209, 1997.

[2] Evensen, G.: The Ensemble Kalman Filter: Theoretical Formulation and Practical Implementation. *Ocean Dynamics* **53**:343-367, 2003.

[3] Ghil, M.: Advances in Sequential Estimation for Atmospheric and Oceanic Flows. *J. Meteor. Soc. Japan* **75**:289-304, 1997.