THE ATMOSPHERIC DATA ASSIMILATION PROBLEM - PAST, PRESENT AND FUTURE

Roger Daley,
Canadian Climate Centre,
Toronto, Ontario, Canada.

In the history of atmospheric data assimilation, two events occurred in the year 1979 which were of particular significance. Firstly, the Global Weather Experiment (GWE) collected the most comprehensive global set of atmospheric observations ever assembled. Secondly, the ECMWF data assimilation system commenced operations by analyzing the GWE data. These two coupled events have stimulated a decade of rapid progress in atmospheric data assimilation.

This ECMWF seminar, commemorating ten years of progress, provides a propitious moment to examine the past, present and future of atmospheric data assimilation. Examination of the past is not done for nostalgic reasons, but rather to give a better perspective with which to predict future developments in a rapidly changing environment. Consequently, the view taken here will be very broad.

It is only in the last ten years, that the atmospheric data assimilation problem has been soundly formulated using the appropriate concepts of estimation theory. This has been an important step forward. By planting the atmospheric problem firmly in the mathematical mainstream, concepts and techniques can be readily exchanged with our sister sciences. Because the connection has been made so recently, a historical analysis of atmospheric data assimilation must really examine two separate roots - the meteorolgical root and the estimation theory root.

Before examining these roots, however, it is first necessary to consider some very basic issues. The object of data assimilation is to produce a regular, physically consistent four dimensional representation of the state of the atmosphere from a heterogeneous array of in situ and remote instruments which sample imperfectly and irregularly in space and time. In designing data assimilation systems, one must first consider the goal - weather prediction, physical understanding, signal detection etc. The characteristics of the atmosphere and the phenomena of interest must be taken into account. This requires knowledge of the spatial and temporal spectra of the atmosphere and its underlying physical laws.

The characteristics of the observing system are very important and must be well understood. The global observing system is a multipurpose network which is not now and never will be optimally configured to sample the atmosphere according to sound physical principles. The reason for this was given many years ago by L. F. Richardson,

> "An existing meteorological station in the British Isles has been either an outgrowth from an astronomical or magnetic observatory, or it has adjoined the house of an enthusiast who lived there for reasons unconnected with meteorology, or it has been pushed out to the confines of the islands to grasp as much weather as possible, or it has been placed in charge of coastguards because they are on duty at night, or it has been set on a mountain to test the upper air. Excellent practical reasons all these, but it is remarkable that the properties of the atmosphere, which are expressed by its dynamical equations and its equation of continuity, appear to have no influence on its selection."
>
> Weather Prediction by Numerical Process - 1922.

The global observing system is a compromise, based on international agreements amongst the nations of the world and there is wide variation in coverage and quality. The observations are imperfect and may be biased as well as being contaminated by spatially or temporally correlated errors. There are, in addition, gross errors caused by communication problems, human error etc.

Now to get back to the roots of atmospheric data assimilation itself. Consider first the meteorological root. Atmospheric data assimilation has developed because it is a fundamental part of weather prediction. As discussed by Bjerknes in 1911, deterministic weather prediction can be broken into three components: the observation component (see above), the diagnostic or analysis component and the prognostic component. Atmospheric data assimilation is the diagnostic component and, historically, cannot be easily separated from the prognostic component. Thus, the first true subjective analyses, by LeVerrier in France and Fitzroy in England in the 1850's, were to provide the diagnostic foundation for the first true weather forecasts. Subjective analysists grow more and more skilful until by the late 1940's techniques existed for the graphical calculation of gradient and thermal winds, horizontal advection, deformation, divergence, vorticity, vertical motion etc. Unfortunately, this was very manpower intensive and the Weather

Bureau - Air Force - Navy Analysis Center of 1949 employed 126 people - mostly analysts and plotters.

In addition to being manpower intensive, weather forecasts produced by these subjective techniques were not very accurate. It has been noted that, apart from the introduction of the Norwegian Frontal Model at the time of th First World War, there was little real improvement in weather forecasts between 1860 and 1950.

This all changed with the invention of the practical, multipurpose electronic computer in the 1940's and the successful numerical forecasts of the Princeton group. Numerical weather prediction also has a diagnostic component which amounts to the specification of the initial atmospheric state from which the forecast is integrated or marched forward in time. The initial state has to to specified in terms of some uniform, regular grid. Unfortunately, the observations, as noted by Richardson above, are not so neatly arrayed. Richardson had suggested facetiously, that the observation stations should be re-located to the gridpoints of the forecast model. This solution being impractical, the only solution was to interpolate in some fashion, the regular gridpoint values from the irregularly distributed observations. At first this was done manually, by subective analysis, followed by digitization. This solution was also impractical, because the manual extraction took much longer than the numerical forecast.

What was needed was an automatic procedure to estimate the atmospheric dependent variables on a regular three dimensional grid using the data from the irregular observation network. The procedure had to be robust enough to work without human intervention and without consuming an inordinate amount of computer time. Such procedures are called "objective" because they do not rely on the judgement of a human analyst.

The first serious attempt at objective analysis was that of Panofsky in 1949. His prcedure used a polynomial fit to all the observations in a small area of the analysis domain that included several analysis points. The coefficients of the expansion were determined by a least squares fit and the analyzed values obtained by evaluating the series at the analysis points. Objective analysis algorithm are applications (not always well-founded) of a branch of statistics called estimation theory - the second root of atmospheric data assimilation.

Estimation theory, like so many other concepts useful to physicists, owes its origins to Karl Friedrich Gauss. At the end of the eighteenth century, Gauss was concerned with estimating the orbits of comets from fragmentary telescopic sightings. To this end, he invented the principle of least squares and most (but not all) data assimilation algorithms are based explicitly or implicitly on this prnciple. One practical quotation from Gauss's treatise - Theoria Motus Corporum Coelestium - 1809 is particularly interesting.

> "If the astronomical observations and other quantities on which
> the computation of orbits is based were absolutely correct, the
> elements also, whether deduced from three or four observations
> 1 would be strictly accurate **(so far indeed as the motion**
> **is supposed to take place exactly according to the laws** of
> Kepler), and therefore, if other observations were used, they
> 2 might be confirmed, but not corrected. But since our **measure-**
> **ments and observations are nothing more than approximations to**
> **truth,** the same must be true of all calculations resting upon
> 3 them, and **the highest aim of all computation made concerning**
> **concrete phenomena must be to approximate, as nearly as practi-**
> **cable to the truth.** But this can be accomplished in no other
> 4 way than by **a suitable combination** of **more observations**
> 5 **than the number absolutely requisite for the determination of**
> **the unknown quantities.** This problem can only be properly
> 6 undertaken when an **approximate knowledge of the orbit has**
> **been already attained,** which is afterward to be corrected,
> 7 so as **to satisfy all of the observations in the most accurate**
> **manner possible."**

This quotation is worthy of some comment, because there are many parallels with today's data assimilation problem.

(1) There exists a dynamic model of the system, which is not exact.

(2) All observations have errors.

(3) The final analysis or estimate will only approximate the truth.

(4) The observations must be weighted in some optimal fashion.

(5) The problem should be overdetermined, that is, the number of observations should be larger than the number of degrees of freedom of the system that one wishes to determine.

(6) A background or prior estimate (obtained from the dynamic model) improves the estimate.

(7) The final estimate or analysis should fit all the observations to within their presumed accuracies.

Stripped to its bare essentials, atmospheric data assimilation is no different than applications of estimation theory in a number of related disciplines. As examples, one might note the determination of the structure of the mantle from seismic observations, determining the internal structure of human organs from transmitted or scattered radiation, determination of the internal structure of the ocean, obtaining orbital parameters of artificial satellites and estimating the present and controlling future states of electromechanical devices.

Viewed from the perspective of 1989, much of the early work on atmospheric data assimilation seems somewhat ad hoc and poorly anchored to the underlying principles of estimation theory. This can be explained, in part, by the absence of sufficient computer resources. It is only in the last decade, that computer resources have been adequate to make a really sound formulation of the problem worthwhile.

Nonetheless, despite their somewhat erratic course toward a mathematically satisfying formulation of the data assimilation problem, meteorologists can take considerable pride in their achievements. Firstly, producing near real time coherent analyses from such a large, heterogeneous data base is a real achievement. Moreover, when we realize that this has been going on for over 30 years, it is clear that meteorologists were really pioneers. Secondly, many of the elements of modern estimation theory have been employed (albeit heuristically) by meteorologists for many years. Thus, the use of background or prior estimates from climatology or model forecasts goes back to Bergthorsson, Doos and Thompson in the late 1950's. The use of observation and background error statistics, optimal weighting and consistent multivariate coupling goes back to Gandin in the early 1960's. In the spatial domain, analysis of the signal, noise and filtering properties of objective analysis algorithms began with Stephens in the late 1960's. In the temporal domain, theory and techniques for separating low frequency signal from high frequency noise (initialization) were discussed by Charney as long ago as 1947. With all this background, it was a relatively minor step to place the atmospheric data assimilation problem firmly within the mainstream of estimation theory as has been done in the last decade by Rodgers, Lorenc, Ghil and others.

And so we arrive at the present - 1989 and the tenth anniversary of operational forecasting at ECMWF. The atmospheric data assimilation cycle has become a major end in itself, with an enormous suite of complex softwear for data quality control, integration of the assimilation model, objective analysis, initialization and the generation of diagnostics. Statistical interpolation and continuous forward assimilation are the preferred methods of objective analysis, while normal mode techniques are preferred for initialization. Work is well in hand on more advanced methods based on estimation theory - three and four dimensional variational procedures, in particular.

But, what of the future? New and improved techniques will continue to evolve, but other speakers will discuss these developments more convincingly. Here, we will be concerned with two other aspects of the future - the nature of data assimilation itself and the future observing system. These two aspects are strongly interconnected.

Prediction of the future is always dangerous, as any meteorologist knows full well. However, there are some fairly strong signs that the nature of the data assimilation problem may change radically over the next two decades. Consider a few scientific and socio-political facts.

(1) There is increasing public concern with such environmental problems as climate change (increasing CO2 and the ozone hole). Pressures are mounting on governments to act, and intelligent policy and legislation requires good scientific input. Sound scientific judgement needs accurate monitoring of the present situation and modelling to anticipate the future.

(2) In partial response to (1), ambitious new remote sensing programs such as EOS (Earth Observing System) are in pre-implementation phase to monitor many new environmental variables. These include external forcing variables, concentrations of radiatively and chemically important race species, land surface and oceanic variables.

(3) Objective analyses are increasingly used for a variety of new purposes including climate change signal detection.

From these facts, a possible vision of the future of data assimilation follows. Firstly, data assimilation will no longer be entirely or even primarily concerned with short (or medium) range weather forecasting. Secondly, the data base will become incredibly diverse in both variables measured and type of observing system. Thirdly, assimilating models will be much more comprehensive, involving ocean, land surface and stratospheric components. Finally, there will be considrably more emphasis on the long term stability of algorithms to facilitate climate change signal detection.

Atmospheric scientists have a long experience in the real time analysis of a vast heterogeneous data base and our sister environmental sciences look to us for leadership in this regard. The present human and computer resources are clearly inadequate for the coming task. It is clear, then, that the profile of data assimilation has to be raised in the scientific community and that more gifted young scientists must be attracted into the field.

1 REFERENCE

Daley, R., 1990: Atmospheric data analysis. Cambridge University Press, New York, 400pp.