5.2 Numerical Engineering: The Search for Practicality

Estimation theory is comparatively straightforward in its goals, and in the methods of solution. When it comes to real problems, particularly those involving fluids, the main issues tend to be much less the principle of what one wants to do (it is usually reasonably clear), and more the problems of practicality. Even linear three dimensional fluid models, particularly those arising in the geophysical world, readily overwhelm the largest available computers, storage devices, and investigators, even when one exploits the special structure of the simultaneous equations represented by time-evolving models. The major issues are primarily those of "numerical engineering"—finding practical methods adequate for a particular goal, while keeping the underlying theory in mind as a guideline. Engineering involves all aspects of the problem, including the forward model, the algorithms for doing minimization, representation and computation of weight matrices, finding adequate estimates of model, and overall system errors. Because of the diversity of the problems that arise, only some very general description of various applications and remedies can be described here.

5.2.1 Meteorological Assimilation

"Data assimilation " is a term widely used in numerical weather prediction (NWP) to describe the process of combining a forecast with current observations for the primary purpose of updating a dynamical model—usually in preparation for another forecast. In this book, we use the term "state estimation" for the more general problem of forming model/data combinations, and will reserve "assimilation" for the specific meteorological application. For fluid models, forecasting is probably more highly developed in meteorology than in any other field. Astronomers forecasting planetary or cometary positions have a longer history, and ballistic engineers are greatly experienced with a range of trajectory and impact prediction problems. But the meteorological problem is of much greater dimension than any of these, and the economic stakes are so high, that many person-years have been devoted to making and improving weather forecasts. The field is thus a highly developed one, and worth examination. A correspondingly large literature on meteorological assimilation exists.[159]

Much data assimilation involves simplified forms of objective mapping, in which the model dynamics are used in a primitive fashion to help choose covariances in both time and space for interpolation as in Chapter 2.[160]. The formal uncertainties of the forecast are not usually computed—the forecaster learns empirically, and very quickly, whether and which aspects of his forecast are any good. If something works, then one keeps on doing it; if it doesn't work, one changes it. Because of the short time scale, feedback from the public, the military, farmers, the aviation industry, etc., is fast and vehement. Theory often takes a backseat to practical experience. It is important to note that, despite the dense fog of jargon that has come to surround meteorological practice, that the methods in actual use remain, almost universally, attempts at the approximate least-squares fitting of a time-evolving atmospheric model to the oncoming observations. The primary goal is forecasting, rather than smoothing.

5.2.2 Nudging and Objective Mapping

A number of meteorological schemes can be understood by referring back to the Kalman filter averaging step,

{67001} $$\tilde{\mathbf{x}}(t) = \tilde{\mathbf{x}}(t,-) + \mathbf{K}(t)\left[\mathbf{y}(t) - \mathbf{E}\tilde{\mathbf{x}}(t,-)\right]. \tag{5.3}$$

This equation has the form of a predictor-corrector—the dynamical forecast of $\tilde{\mathbf{x}}(t,-)$ is compared to the observations and corrected on the basis of the discrepancies. Some assimilation schemes represent guesses for $\mathbf{K}$ rather than the computation of the optimum choice, which we know—for a linear model—is given by the Kalman gain, replacing (5.3) with,

{67002} $$\tilde{\mathbf{x}}(t) = \tilde{\mathbf{x}}(t,-) + \mathbf{K}_m\left[\mathbf{y}(t) - \mathbf{E}\tilde{\mathbf{x}}(t,-)\right] \tag{5.4}$$

where $\mathbf{K}_m$ is a modified gain matrix. Thus, in "nudging", $\mathbf{K}_m$ is diagonal or nearly so, with elements which are weights that the forecaster assigns to the individual observations.[161] To the extent that the measurements have uncorrelated noise, as might be true of pointwise meteorological instruments like anemometers, and the forecast error is also nearly spatially uncorrelated, pushing the model values pointwise to the data may be very effective. If, in (5.4), the observations $\mathbf{y}(t)$ are direct measurements of state vector elements (e.g., if the state vector includes the density and $\mathbf{y}(t)$ represents observed densities), then $\mathbf{E}(t)$ is very simple—but only if the measurement point coincides with one of the model grid points. If, as is often true, the measurements occur between model grid points, $\mathbf{E}$ is an interpolation operator from the model grid to the data location. In the most usual situation, there are many more model grid points than data points, and this direction for the interpolation is the most reasonable and accurate. With more data points than model grid points, one might better interchange the direction of the interpolation. Formally, this interchange is readily accomplished by rewriting (5.3) as,

$$\tilde{\mathbf{x}}(t) = \tilde{\mathbf{x}}(t,-) + \mathbf{K}_m \mathbf{E} \left[\mathbf{E}^{+} \mathbf{y}(t) - \mathbf{x}(t,-) \right] \tag{5.5}$$ {67003}

where $\mathbf{E}^{+}$ is any right inverse of $\mathbf{E}$ in the sense of Chapter 2, for example, the Gauss-Markov interpolator or some plausible approximation to it.

There are potential pitfalls of nudging, however. If the data have spatially correlated errors, as is true of many real observation systems, then the model is being driven toward spatial structures that are erroneous. More generally, the expected great variation in time and space of the relative errors of model forecast and observations cannot be accounted for with a fixed diagonal gain matrix. A great burden is placed upon the insights and skill of the investigator who must choose the weights. Finally, one can calculate the uncertainty of the weighted average (5.4), using this suboptimal gain, but it requires that one specify the true covariances. As noted, however, in NWP formal uncertainty estimates are not of much interest. User feedback is, however, rarely available when the goal is understanding—the estimation problem—rather than forecasting the system for public consumption. When forecasts are made in many contexts, e.g., for decadal climate change, the time scale is often so long as to preclude direct test of the result.

As with the full Kalman filter, the "analysis step" where the model forecast is averaged with the observations, there is a jump in the state vector as the model is pulled (usually) toward the observations. Because the goal is usually forecasting, this state vector discontinuity is not usually of any concern, except to someone instead interested in understanding the time evolution of the atmosphere.

Another more flexible, approximate form of time-dependent estimation can also be understood in terms of the Kalman filter equations. In the filter update equation (4.51), all elements of the state vector are modified to some degree, given any difference between the measurements and the model-prediction of those measurements. The uncertainty of the statevector is *always* modified whenever data become available, even if the model should perfectly predict the observations. As time evolves, information from measurements in one part of the model domain is distributed by the model dynamics over the entire domain, leading to correlations in the uncertainties of all the elements.

One might suppose that some models propagate information in such a way that the error correlations diminish rapidly with increasing spatial and temporal separation. Supposing this to be true (and one must be aware that fluid models are capable of propagating information, be it accurate or erroneous, over long distances and times), static approximations can be found in which the problem is reduced back to the objective mapping methods employed in Chapter 2. The model is used to make an estimate of the field at time t, $\tilde{\mathbf{x}}(t,-)$, and one then finds the prediction error $\Delta\mathbf{y}(t) = \mathbf{y}(t) - \mathbf{E}\tilde{\mathbf{x}}(t,-)$. A best estimate of $\Delta\mathbf{x}(t)$ is sought based upon the covariances of $\Delta\mathbf{y}(t)$, $\Delta\mathbf{x}(t)$, etc.—that is, objective mapping—and the improved estimate is,

{67006}

$$\tilde{\mathbf{x}}(t) = \tilde{\mathbf{x}}(t,-) + \Delta\tilde{\mathbf{x}}(t) = \tilde{\mathbf{x}}(t,-) + \mathbf{R}_{xx}\mathbf{E}^T(\mathbf{E}\mathbf{R}_{xx}\mathbf{E}^T + \mathbf{R}_{nn})^{-1}\Delta\mathbf{y}\,, \tag{5.6}$$

which has the form of a Kalman filter update, but in which the state uncertainty matrix, $\mathbf{P}$, is replaced in the gain matrix, $\mathbf{K}$, by $\mathbf{R}_{xx}$ representing the prior covariance of $\Delta\mathbf{x}$. $\mathbf{R}_{xx}$ is fixed, with no dynamical evolution of the gain matrix permitted. Viewed as a generalization of nudging, this appoach, it permits one to specify spatial structure in the noise covariance through choice of a nondiagonal $\mathbf{R}_{nn}$. The weighting of the $\Delta\mathbf{y}$ and the modification for $\tilde{\mathbf{x}}$ is then much more complex than in pure nudging.

The major issues are the specification of $\mathbf{R}_{xx}$, $\mathbf{R}_{nn}$. Most attempts to use these methods have been simulations by modelers who were content to ignore the problem of determining $\mathbf{R}_{nn}$ or to assume that the noise was purely white. In principle, estimates of $\mathbf{R}_{xx}$ can be found either from observations or from the model itself.

Methods that permit data to be employed from finite-time durations, weighting them inversely with their deviation from some nominal central time, are localized approximations to smoothing algorithms of the Wiener type. Many variations on these methods are possible, including the replacement of $\mathbf{R}_{xx}$ by its eigenvectors (the singular vectors or EOFs), which again can be computed either from the model or from data. Improvements could be made by comparison of the covariance matrices used against the estimates emerging from the calculations of $\tilde{\mathbf{x}}(t)$, $\tilde{\mathbf{n}}(t)$.

All practical linearized assimilation methods are a weighted average of a model estimate of the oceanic state with one inferred from the observations. If the model and the observations are physically inconsistent, the forced combination will be impossible to interpret. Thus, the first step in any assimilation procedure has to be to demonstrate that model physics and data represent the same fluid—with disagreement being within the error bounds of both. Following this confirmation of physical consistency, one recognizes that the weighted average of model and data will be useful only if the weights make sense—chosen to at least well-approximate the relative uncertainties of these two. Otherwise, the result of the combination is an average of "apples and oranges."