Optimizing Data Assimilation for Re-Analysis

Mike Fisher

ECMWF
Shinfield Park, Reading, Berks RG2 9AX, UK

1. Introduction

Development of a state-of-the-art data assimilation system is a major undertaking, requiring large resources. For this reason, most re-analysis projects have used existing numerical-weather-prediction (NWP) analysis systems. This is likely to continue to be the case. However, the requirements of NWP and re-analysis are different. An analysis system optimized for NWP is unlikely to be optimal for retrospective analysis. We must therefore seek to adapt NWP analysis systems to make them more suitable for re-analysis purposes, but without requiring radical and costly alteration.

The most fundamental difference between analysis for NWP and retrospective analysis is that retrospective analysis can make use of observations that lie in the future with respect to the analysis time. That is, retrospective analysis is a “smoothing” problem, whereas forecasting is a “filtering” problem. To date, however, little account has been taken of this difference, except in the pioneering work of Cohn *et al*. (1994).

2. Choice of Analysis Time in 4dVar

Four-dimensional variational (4dVar) assimilation defines an analysis that is a trajectory of the assimilating model (or close to a trajectory) over the analysis “window”. We are free to choose the analysis window in such a way as to put the nominal analysis time at the beginning, end or in the middle of the window. Surprisingly little investigation has been carried out on the effects of this choice on the quality of the analysis.

It is clear that, for a given length of analysis window, maximum use of observations that lie in the future with respect to the analysis time requires that the analysis time coincide with the start of the analysis window. In this case, observations in the past with respect to the analysis time affect the analysis through their contribution in determining the background state. This transfer of information from earlier observations via the background is sub-optimal since the specified background error covariance matrix is at best a crude approximation to the true covariance matrix of background error. Typically, for example, the specified covariances are averaged over time, and are largely isotropic and homogeneous. Hence, it may be preferable to sacrifice some future information in order to make better use of past information. The best choice of analysis time is clearly a matter for experiment.

3. Longer Assimilation Windows

One possibility for reducing the loss of information incurred in cycling the analysis system is illustrated schematically in Figure 1. In Figure 1a, two cycles of 4dVar with 12-hour assimilation windows are shown. The nominal analysis time is placed at the beginning of the window. In this case, observations up to 12 hours in the future contribute to the analysis. Past observations also contribute, but all past information is transferred sub-optimally to the analysis via the background, $\mathbf{x}_b$.

By increasing the analysis window to 24 hours (Figure 1b), and placing the nominal analysis time at the centre of the window, the use of future observations remains the same, whereas observations up to 12 hours in the past can contribute directly to the analysis, without being sub-optimally “filtered” by the background covariance matrix.

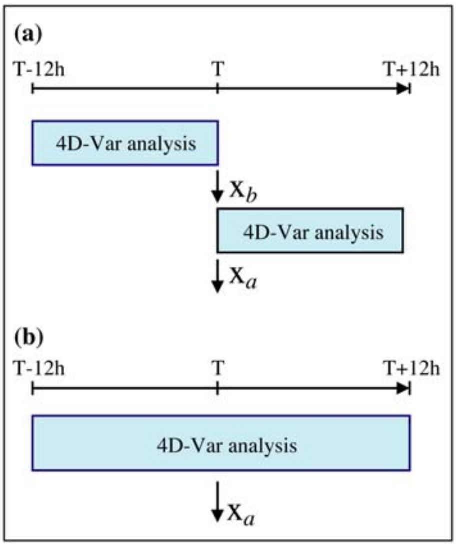

Figure 1: (a) Schematic showing a pair of consecutive 4dVar analyses with 12-hour analysis windows, and with the analysis time defined to coincide with the start of the analysis window. (b) An alternative configuration with a 24-hour window.

4. Very Long Assimilation Windows

To take the idea presented above to its extreme, we may envisage a 4dVar analysis window that is sufficiently long to encompass *all* observations that have an influence on a given analysis. To assess the length of window that would be required, Fisher *et al*. (2005) considered a data-denial experiment whereby all satellite data were withdrawn from a version of the ECMWF analysis-forecast system. The analysis error increased steadily for each cycle during the week after the data were withdrawn, showing that the analyses retained a memory of the satellite data for approximately seven days. Figure 2 shows the reverse experiment in which satellite data were re-introduced into the analysis system at 00z 15 August 2005 after a long period in which no satellite data were assimilated.

The short-term forecast error (shown by the black curve) decreased steadily for about three days after the satellite data were introduced. The red curve in Figure 2 shows short-term forecast error for a second run of the analysis-forecast system that had included satellite data for a long period prior to 15 August. Note that the two experiments’ forecast scores become nearly indistinguishable from about three days after the re-introduction of the satellite data, showing that observations older than three days have no influence on the quality of the analysis.

It is interesting that Figure 2 suggests a significantly shorter memory than the experiment considered by Fisher *et al*. (*op. cit.*). A likely explanation is that memory for earlier observations depends on the number of observations assimilated in the intervening period. In this case, the shorter memory suggested by Figure 2 may be more representative of current assimilation systems than the seven days suggested by Fisher *et al*. (*op*. *cit*.).

Swanson et al. (1998) successfully used strong-constraint 4dVar with analysis windows of several days in an idealised quasi-geostrophic system. They showed that analysis quality improved significantly as the length of the analysis window was increased. In a more realistic system, it is unlikely that model error can be neglected over such time intervals. For this reason, Fisher *et al*. (*op*. *cit*.) considered weak-constraint 4dVar.

They showed, for a simple nonlinear system, that analysis quality at the end of the assimilation window improved monotonically as the length of the analysis window was increased, eventually reaching an asymptotic level equivalent to that of the extended Kalman filter once the length of the analysis window exceeded the time period over which memory for earlier observations was lost. Moreover, the solution in the middle of the assimilation window was more accurate than that at the end of the window, since it took into account future observations as well as observations in the past.

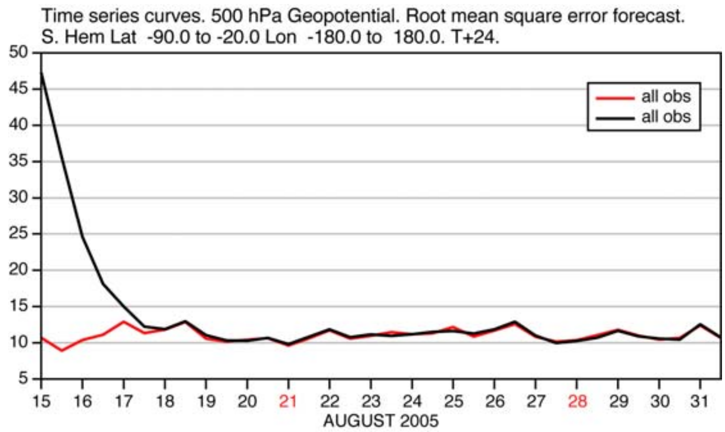

Figure 2: RMS 24-hour forecast scores for southern-hemisphere geopotential at 500hPa for two runs of the analysis-forecast system, differing only in the initial background state.

Finally, we note that if the analysis window is longer than the interval at which analyses are required to be produced, then we must either issue several analyses per cycle, at different times during the analysis window, or we must allow analysis windows to overlap. The former solution is not ideal since analyses will vary in quality and in character depending on when during the window they are issued. The alternative approach was used by Fisher et al. (*op. cit.*), and is illustrated in Figure 3. This solution ensures considerable consistency between consecutive analyses. (Note that the "first guess" indicated in Figure 3 is used as the starting point and linearization state for each minimization. This ensures that each minimization starts from a very good

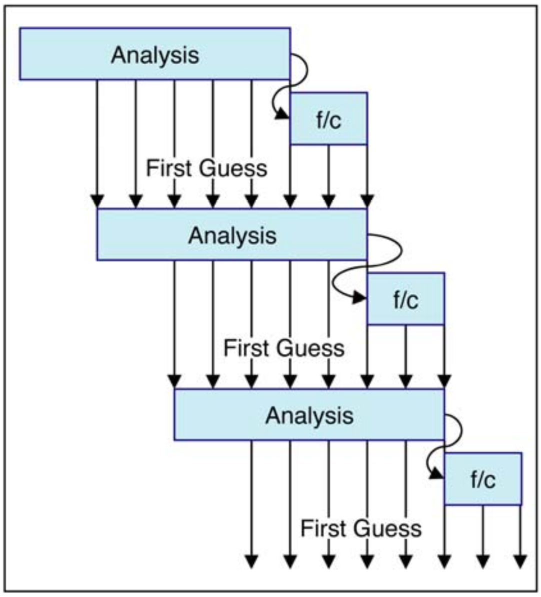

Figure 3: Schematic showing the cycling of overlapping analysis windows. "f/c" indicates a short forecast initialized at the end of the preceding analysis window.

initial state, and may only require a few iterations to converge. However, there is no statistical constraint that any analysis stay close to its first guess. Fisher *et al*. (*op. cit*.) did not incorporate a background term into their analysis system. However, it clear that statistical consistency would require any background state to be taken from a non-overlapping cycle that had not "seen" any of the current observations.)

5. Conclusion

Reanalysis differs fundamentally from analysis for weather prediction. This fact should be recognized in the design of assimilation algorithms. A range of possibilities exists, from a simple tuning of the boundaries of the assimilation window, to the more exotic possibilities suggested by overlapping, long-window, weak-constraint 4dVar. In principle, the latter allows time-consistent analyses equal in quality to those of a full, un-approximated Kalman smoother. The possibilities of such an approach are the subject of active research.

6. References

Cohn S.E., Sivakumaran N.S., and Todling R., 1994: A Fixed-Lag Kalman Smoother for Retrospective Data Assimilation. *Monthly Weather Review*: **122**, No. 12, 2838-2867

Fisher M., Leutbecher M. and Kelly G.A., 2005: On the equivalence between Kalman smoothing and weak-constraint four-dimensional variational data assimilation. *Q.J.R.Meteor.Soc*., **131**, 3235-3246.

Swanson K., Vautard R. and Pirés C., 1998: Four-dimensional variational assimilation and predictability in a quasi-geostrophic model. *Tellus*, **50A**, 369-390.