Appendix A: 4DVAR - an elementary introduction

Assume that we have two reports, +10° and +15° C, of the temperature for a certain location. The first has an assumed error of 2 K, the second 3 K. The true value is therefore probably closer to +10° than to +15°. A traditional statistical least-square technique weights together the two observations, with weights proportional to the "precision" or accuracy of the measurements defined as the inverse of the variances of the assumed errors. In our case these variances are 4 and 9, the weights become 9/(4+9) ≈0.7 and 4/(4+9)≈0.3,which yields an "analysed" value of

10 * 0.7 + 15 * 0.3 = 11.5 C.

Let us now assume that of the two values one is an observation (O), the other a background field value (F) with accuracies σ_O and σ_B respectively. In accordance with what stated above the weighting formula may then be written

$$A = O\frac{\sigma_B^2}{\sigma^2_B + \sigma^2_O} + F\frac{\sigma_O^2}{\sigma^2_B + \sigma^2_O}$$

Most meteorological objective analysis techniques are further developments of this simple least-square approach. They have, however, certain weaknesses. So for example, the method is local: only observations within a limited area could be considered at each time to influence a given grid point. They could not cope well with non-conventional data. That is why a variational technique was introduced.

This technique starts by introducing a so called "cost function" J(S) which measures the sum of the squares of the distances (or misfit) of different atmospheric states S to the observation O and the background F.

$$J(S) = \frac{1}{2}\left[\frac{(F-S)^2}{\sigma_B^2} + \frac{(O-S)^2}{\sigma_O^2}\right]$$

The observation O is known, and we are looking for a value of A that will make the cost function J(S) as small as possible. We do this by differentiating J with respect to S. This will result in the same value of A as for the least square approach above, which shows the internal consistency of the two approaches.

One may therefore ask the motivation for defining this cost function? The reason is that it opens up a door to a mathematical formalism which will provide a powerful tool, in particular when the scalars in the right hand equation are replaced by vector and matrices.

Let us start by trivially re-arranging the equation above

$$J(S) = \frac{1}{2}\left[(F-S)\frac{1}{\sigma_B^2}(F-S) + (O-S)\frac{1}{\sigma_O^2}(O-S)\right]$$

We now introduce an operator H which can be anything from a simple interpolation of the analysis to the exact position of an observation, to a complicated equation that converts temperature and moisture information into radiance values. Instead of using the difference between analysis and observation, we use the difference between the converted value H(S) and the observation Y:

$$J(S) = \frac{1}{2}\left[(F-S)\frac{1}{\sigma_F^2}(F-S) + (Y-H(S))\frac{1}{\sigma_O^2}(Y-H(S))\right]$$

The next step is to consider not one, not many, but *all* observations made simultaneously over the globe. The scalars A, F and Y now become vectors of enormous dimensions since all the global observations enter as elements. So for example when there is "only" one million observations globally available at a certain time interval the vector **Y**=(obs_1, obs_2, obs_3,................obs_{999999}, $obs_{1000000}$).

The error $\sigma_O{}^2$ and $\sigma_B{}^2$ variances turn from scalars into the error covariance matrices **R** for the observations. To match the vectors they are formally of the size 1000000 x 1000000. The matrice.**B** for the background or first guess is approximately of the same dimension.

$$J(\mathbf{S}) = \frac{1}{2}[(\mathbf{F}-\mathbf{S})^T\mathbf{B}^{-1}(\mathbf{F}-\mathbf{S}) + (\mathbf{Y}-\mathbf{H}(\mathbf{S}))^T\mathbf{R}^{-1}(\mathbf{Y}-\mathbf{H}(\mathbf{S}))]$$

This is the basic equation for the 3DVAR assimilation system. In spite of the enormity of the numerical calculations, it still preserves the structure of our first formula above.

What characterizes 4DVAR is that *time* enters as an additional element. The aim is to find the A_0 in the beginning (t=0) of a 12-hour time interval, which minimizes J. The minimization will not only be dependent on the conditions at t=0, but to the whole dynamic evolution during the 12-hour interval.

So while the first term on the right hand side is kept alone, the second term will be evaluated for every single time step (t=1 -> N). The operator **H** is as before the operator which converts model states into observation states. But we now also introduce operator $\mathbf{M_n}$, which is the forecast model at time step n.

By integrating the atmosphere forward by $\mathbf{M_n}$ and “backward” by the so called adjoint in the 12-hour window, 4DVAR process ensures that the initial conditions at 09 and 21 UTC are worked out in a way that provides optimum fit to the observations throughout the 12 hour window.

So mathematically we are looking for an atmospheric state S_0=A at time n=0 which minimizes the cost function J(S).

$$J(\mathbf{S}) = \frac{1}{2}[(\mathbf{F_0} - \mathbf{S_0})^T \mathbf{B_0^{-1}} (\mathbf{F_0} - \mathbf{S_0}) +$$

$$\sum_{n=1}^{N} ((\mathbf{Y}_n - \mathbf{HM}_n(\mathbf{S}_n))^T) \mathbf{R}_n^{-1} (\mathbf{Y}_n - \mathbf{HM}_n(\mathbf{S}_n))$$

The first term: the difference between the first guess and the initial state determines only partly the size of J.

The second term: the Σ–term, sums up all the differences between the evolving forecast and the *n* number of observations of varying kind.