Data assimilation principles

Angela Benedetti
(European Centre for Medium-Range Weather Forecast)

Based on a lecture by:

Mike Fisher (ECMWF)

Outline

- General assimilation concepts and historical perspective
- One-dimensional example
- Extension to multi-dimensions
- Introduction to 3D and 4D-VAR
- Introduction to Tangent Linear and Adjoint operators

Introduction

Analysis: The process of approximating the true state of a (geo)physical system at a given time.

- For example:

  - Hand analysis of synoptic observations (1850 LeVerrier, Fitzroy).

  - Polynomial Interpolation (1950s Panofsky)

- An important step forward was made by Gilchrist and Cressman (1954), who introduced the idea of using a previous numerical forecast to provide a preliminary estimate of the analysis.
- This prior estimate was called the background.

Introduction (cont.)

- Bergthorsson and Doos (1955) took the idea of using a background field a step further by casting the analysis problem in terms of increments which were added to the background.
- The increments were weighted linear combinations of nearby observation increments (observation minus background), with the weights determined statistically.
- This idea of statistical combination of background and synoptic observations led ultimately to Optimal Interpolation.
- The use of statistics to blend model fields with observations is fundamental to all current methods of analysis.

Data Assimilation

- An important change of emphasis happened in the early 1970s with the introduction of primitive-equation models.
- Primitive equation models support inertia-gravity waves. This makes them much more fussy about their initial conditions than the filtered models that had been used previously.
- The analysis procedure became much more intimately linked with the model. The analysis had to produce an initial state that respected the model's dynamical balances.
- Unbalanced increments from the analysis procedure would be rejected as a result of geostrophic adjustment.
- Initialization techniques which suppress inertia-gravity waves became important.

Data Assimilation (cont.)

- The idea that the analysis procedure must present observational information to the model in a way in which it can be absorbed (i.e. not rejected by geostrophic adjustment) led to the coining of the term data assimilation.
- A final impetus towards the modern concept of data assimilation came from the increasing availability of asynoptic observations from satellite instruments.
- It was no longer sufficient to think of the analysis purely in terms of spatial interpolation of contemporaneous observations.
- The time dimension became important, and the model dynamics assumed the role of propagating observational information in time to allow a synoptic view of the state of the system to be generated from asynoptic data.

One dimensional analysis

- Suppose we want to estimate the temperature of this room, given:
- A prior estimate: Tb.

E.g., we measured the temperature an hour ago, and we have some idea (i.e. a model) of how the temperature varies as a function of time, the number of people in the room, whether the windows are open, etc.

- A thermometer: To.

Denote the true temperature of the room by Tt.

- The errors in Tb and To are:

$\varepsilon b = Tb - Tt$

$\varepsilon o = To - Tt$

- We will assume that the error statistics of Tb and To are known, and that Tb and To have been adjusted (bias corrected) so that their mean errors are zero: $\varepsilon b = \varepsilon o = 0$

One dimensional analysis (cont.)

- We estimate the temperature of the room as a linear combination of Tb and To:

$$Ta = \alpha To + \beta Tb + \gamma$$

- Denote the error of our estimate as $\varepsilon a = Ta - Tt$
- We want the estimate to be unbiased: $\varepsilon a = 0$.
- We have:

$$Ta = Tt + \varepsilon a = \alpha(Tt + \varepsilon o) + \beta\,(Tt + \varepsilon b) + \gamma$$

- Taking the mean and rearranging gives:

$$\varepsilon a = (\alpha + \beta - 1)Tt + \gamma$$

Since this holds for any Tt, we must have

$$\gamma = 0 \quad \text{and} \quad \alpha + \beta - 1 = 0$$

hence: $Ta = \alpha\, To + (1 - \alpha)Tb$

One dimensional analysis (cont.)

- The general Linear Unbiased Estimate is:

$$Ta = \alpha\, To + (1 - \alpha) Tb$$

- Now let's consider the error of this estimate.
- Subtracting Tt from both sides of the equation gives

$$\varepsilon a = \alpha\, \varepsilon o + (1 - \alpha)\, \varepsilon b$$

- The variance of the estimate is:

$$\overline{\varepsilon_a^2} = \alpha^2 \overline{\varepsilon_o^2} + 2\alpha(1-\alpha)\overline{\varepsilon_o \varepsilon_b} + (1-\alpha)^2 \overline{\varepsilon_b^2}$$

- The quantity $\overline{\varepsilon_o \varepsilon_b}$ represents the covariance between the error of our prior estimate and the error of our thermometer measurement.
- There is no reason for these errors to be connected in any way.
- We will assume that $\overline{\varepsilon_o \varepsilon_b} = 0$.

One dimensional analysis (cont.)

- The minimum-variance estimate occurs when

$$\frac{d\overline{\varepsilon_a^2}}{d\alpha} = 2\,\alpha\overline{\varepsilon_o^2} - 2\,(1-\alpha)\,\overline{\varepsilon_b^2} = 0$$

which gives an estimate for α:

$$\alpha = \frac{\overline{\varepsilon_b^2}}{\overline{\varepsilon_b^2} + \overline{\varepsilon_o^2}}$$

- It can be shown that the error variance of this minimum-variance estimate is:

$$\overline{\varepsilon_a^2} = \left(\frac{1}{\overline{\varepsilon_o^2}} + \frac{1}{\overline{\varepsilon_b^2}}\right)^{-1}$$

Extension to multiple dimensions

- The major difference between the simple scalar example and the multi-dimensional case is that there is no longer a one-to-one correspondence between the elements of the observation vector and those of the background vector.
- It is no longer trivial to compare observations and background.
- Observations are not necessarily located at model grid-points
- The observed variables (e.g. radiances) may not correspond directly with any of the variables of the model.
- To overcome this problem, we must assume that our model is a more-or-less complete representation of reality, so that we can always determine “model equivalents” of the observations.

Extension to multiple dimensions

- We formalize this by assuming the existence of an observation operator, H.
- Given a model-space vector, x, the vector H(x) can be compared directly with the observation vector y and represents the “model equivalent” of y.
- For now, we will assume that H is perfect. I.e. it does not introduce

any error, so that H(xt) = yt, where xt is the true state, and yt contains the true values of the observed quantities.

- The analysis equation in the multi-dimensional case becomes:

$$x_a = x_b + K(y - H(x_b))$$

K plays the same role played by the constant α in the scalar example.

- K is called the gain matrix and determines the weight given to observations
- It also handles the transformation of information defined in “observation space” to the space of the model variables (i.e. from radiance to temperature).

Extension to multiple dimensions

- Th expression for the analysis error is

$$\overline{\boldsymbol{\varepsilon_a}\,\boldsymbol{\varepsilon_a^T}} = (\mathrm{I} - \mathrm{KH})\overline{\boldsymbol{\varepsilon_b}\,\boldsymbol{\varepsilon_b^T}}\,(I - KH)^T + K\overline{\boldsymbol{\varepsilon_o}\,\boldsymbol{\varepsilon_o^T}}\,K^T$$

which is equivalent to the scalar case $\overline{\varepsilon_a^2} = (1-\alpha)^2\overline{\varepsilon_b^2} + \alpha^2\overline{\varepsilon_o^2}$

but $\boldsymbol{\varepsilon_a}\,\boldsymbol{\varepsilon_a^T}, \boldsymbol{\varepsilon_b}\,\boldsymbol{\varepsilon_b^T}, \boldsymbol{\varepsilon_o}\,\boldsymbol{\varepsilon_o^T}$ are covariance matrices.

Again we see that K plays essentially the same role in the multi-dimensional analysis as α plays in the scalar case.

In the scalar case we chose α so that the variance of the analysis error was at a minimum.

In the multi-dimensional case a similar thing can be done but since we are dealing with matrices we have to chose K so that the trace (sum of the diagonal elements) of the analysis error covariance matrix is zero. This is mathematically equivalent to the scalar case although the formalism is more complex.

Extension to multiple dimensions

It can be shown than

$$\mathrm{K} = \overline{\boldsymbol{\varepsilon_b}\,\boldsymbol{\varepsilon_b^T}}\,H^T[H\overline{\boldsymbol{\varepsilon_b}\,\boldsymbol{\varepsilon_b^T}}\,H^T + \overline{\boldsymbol{\varepsilon_o}\,\boldsymbol{\varepsilon_o^T}}]^{-1}$$

This optimal gain matrix is called the Kalman Gain Matrix.

Note the similarity with the optimal gain we derived for the scalar analysis:

$$\alpha = \frac{\overline{\varepsilon_b^2}}{\overline{\varepsilon_b^2} + \overline{\varepsilon_o^2}}$$

The variance of analysis error for the optimal scalar problem was:

$$\overline{\varepsilon_a^2} = \left(\frac{1}{\overline{\varepsilon_b^2}} + \frac{1}{\overline{\varepsilon_o^2}}\right)^{-1}$$

The equivalent expression for the multi-dimensional case is:

$$\overline{\boldsymbol{\varepsilon_b}\,\boldsymbol{\varepsilon_b^T}} = [\left(\overline{\boldsymbol{\varepsilon_b}\,\boldsymbol{\varepsilon_b^T}}\right)^{-1} + H^T\left(\overline{\boldsymbol{\varepsilon_o}\,\boldsymbol{\varepsilon_o^T}}\right)^{-1}H]^{-1}$$

Notation

- The standard notation for the covariance matrices is:

$$\overline{\boldsymbol{\varepsilon_b}\,\boldsymbol{\varepsilon_b^T}} = P_b$$
$$\overline{\boldsymbol{\varepsilon_a}\,\boldsymbol{\varepsilon_a^T}} = P_a$$
$$\overline{\boldsymbol{\varepsilon_o}\,\boldsymbol{\varepsilon_o^T}} = R$$

- In many analysis schemes, the true covariance matrix of background error is not known or is too large to be used
- In this case, we use an approximate background error covariance matrix, indicated with the symbol B.

Assimilation Methods: Optimal Interpolation

- Optimal interpolation is a statistical data assimilation method based on the multi-dimensional analysis equations we have just derived. This method was used operationally at ECMWF from 1979 until 1996, when it was replaced by 3D-Var.
- The basic idea is to split the global analysis into a number of boxes which can be analysed independently

$$x_a^{(i)} = x_b^{(i)} + K^{(i)} (y^{(i)} - H^{(i)} (x_b))$$

- In principle, one should use all available observations to calculate the analysis for each box. However, this is too expensive.
- For computational reasons, OI restricts the observations used for each box to those observations which lie in a surrounding selection area.
- Also, because the analysis solution is computed with direct methods and the matrices have to be specified explicitly, OI can only be used with simple interpolation operators. This is why it was replaced by 3D-Var.

Assimilation Methods: 3D-Var

- Iterative methods to solve the analysis equations are more efficient than the direct methods used in OI.
- They can be applied to much larger problems than direct techniques and do not require access to matrix elements.
- Linear 3D-Var analysis can be seen as an application of iterative solution methods to the linear analysis equation.
- Historically, 3D-Var was not developed this way.
- We will now consider this alternative derivation.
- We will need to apply Bayes' theorem:
- $p(A \; given \; B) = \frac{p(\, B \; given \; A)\, p(A)}{p(B)}$

Maximum likelihood

- We developed the linear analysis equation by searching for a linear combination of observation and background that minimized the variance of the error.
- An alternative approach is to look for the most probable solution, given the background and observations:

$$x_a \; such \; as \; \text{p(x given y and } x_b\text{)) is at its max}$$

- It is convenient to define a cost function

$$J = -\log(\text{p(x given y and } x_b) \; + \text{const}$$

- Since log is a monotonic function:

$$x_a \; such \; as \; J(x) \; is \; at \; its \; minimum$$

Maximum likelihood (cont.)

- The maximum likelihood approach is applicable to any probability density functions.
- Considering the special case of Gaussian probability distributions

$$p(x_b) = const \ * \exp\left[-\frac{1}{2}\,(x_b - x)^T (B)^{-1}\,(x_b - x)\right]$$

$$\mathrm{p}(y) = const \ * \exp\left[-\tfrac{1}{2}\,\big(y - H(x)\big)^T (R)^{-1}\,(y - H(x))\right]$$

- With an appropriate choice of constants:

$$J(x) = \frac{1}{2}\,(x_b - x)^T (B)^{-1}\,(x_b - x) + \frac{1}{2}\,\big(y - H(x)\big)^T (R)^{-1}\,(y - H(x))$$

- This is the 3D-Var cost function
- At the minimum, the gradient of the cost function is zero and the probability of a certain state (analysis) happening is maximum

Maximum likelihood (cont.)

- The maximum likelihood approach can be naturally expressed in terms of a cost function representing minus the log of the probability of the analysis state.
- The minimum of the cost function corresponds to the maximum likelihood (probability) solution.
- For Gaussian errors and linear observation operators, the maximum likelihood analysis coincides with the minimum variance solution.
- This is not the case in general.
- In the nonlinear case, the minimum variance approach is difficult to apply.
- The maximum-likelihood approach is much more generally applicable.
- As long as we know the probability distributions, we can define the cost function.
- However, finding the global minimum may not be easy for highly non-Gaussian probability distribution functions.
- In practice, background errors are usually assumed to be Gaussian.
- This makes the background-error term of the cost function quadratic.

Strong Constraint 4D-Var

- So far, we have tacitly assumed that the observations, analysis and background are all valid at the same time, so that H includes spatial, but not temporal, interpolation.
- In 4D-Var, this assumption is relaxed.
- Let's use G to denote a generalised observation operator that:

  - Propagates model fields defined at some time t0 to the (various) times at which the observations were taken.

  - Spatially interpolates these propagated fields

  - Converts model variables to observed quantities
- We will use a numerical forecast model to perform the first step.
- Note that, since models integrate forward in time and we do not have an inverse of the forecast model, the observations must be available for times tk greater or equal to t0.

Strong Constraint 4D-Var

- Formally, the 4D-Var cost function is identical to the 3D-Var cost function. We simply replace H by G:

$$J(x) = \frac{1}{2}\,(x_b - x)^T (B)^{-1}\,(x_b - x) + \frac{1}{2}\,\big(y - G(x)\big)^T (R)^{-1}\,(y - G(x))$$

- However, it makes sense to group observations into sub-vectors of observations, yi , that are valid at the same time, ti .
- It is reasonable to assume that observation errors are uncorrelated in time. Then, R is block diagonal, with blocks Ri corresponding to the sub-vectors yo .
- The cost function is then expressed as follows:

$$J(x) = \frac{1}{2}\,(x_b - x)^T (B)^{-1}\,(x_b - x) + \frac{1}{2}\sum_{i=0}^{N}(y_i - G_i\,(x))^T\,R^{-1}\,(y_i - G_i\,(x))$$

Now, each generalised observation operator can be written as $G_i = H_i\,M(t_o\,, t_i\,)$ where:

- M(t0,ti) represents an integration of the forecast model from time t0 to time ti .
- Hi represents a spatial interpolation and transformation from model variables to observed variables (observation operator)

Weak Constraint 4D-Var

- Note that by introducing the vectors xi we have converted an unconstrained minimization problem into a problem with strong constraints as she solution xi has to satisfy the model equation:

$$x_i = M(x_{i-1})$$

- For this reason, this form of 4D-Var is called strong constraint 4D-Var.
- The generalised observation operators Gi are assumed to be perfect, i.e. error-free.
- This is called perfect model assumption.
- The perfect model assumption limits the length of analysis window that can be used to roughly 12 hours (for an NWP system).
- Relaxing this assumption (to account for model deficiencies) and allowing for the model to have an associated error leads to an alternative formulation of the 4D-Var problem which is called weak-constraint 4D-Var.

Summary

- Current assimilation methods for Numerical Weather Prediction or Chemical Data Assimilation are based on a linear or linearized analysis equation which can be solved in different ways.
- Direct methods of solution such as OI are costly for large problems, while iterative methods such as 3D-Var can be used with a variety of observations operators and for different applications
- 4D-Var is used at many NWP centres and it has been proven successful for NWP and CDA applications
- Ensemble Kalman methods are very attractive as they provide a way to estimate flow-dependent error covariance matrices, and do not require the development of tangent linear and adjoint models.
- Many centres are turning to hybrid techniques in which the background error covariance matrix is calculated with an ensemble, and the analysis is then performed with the 4D-Var approach.