Introduction to Data Assimilation

Fundamental Principles of Data Assimilation

As defined previously, **data assimilation** is the process by which observations are assimilated in order to update a 'first guess' for the initial conditions that is typically drawn from a short-range (1-6 h) numerical model forecast. Data assimilation methods seek to obtain the best-possible estimate of the initial atmospheric state while, hopefully but not necessarily intrinsically, satisfying an appropriate balance condition. Using a short-range numerical model forecast as the first guess enables atmospheric phenomena on scales of those resolved by the model grid and larger to be represented, at least in part, within the initial conditions (e.g., "warm start"). It also allows for approximate initial conditions to be available in locations where few or no observations exist.

Data assimilation involves modifying the first guess based upon observation characteristics and physical principles. Consider the assimilation of a single wind observation taken in the core of an upper tropospheric jet streak. This observation should not only update the kinematic fields at the location of the observation but should also update kinematic fields at nearby locations (given the scale of and typical variation in wind across a jet streak) as well as closely related model fields such as geopotential height and temperature (due to geostrophic and thermal wind balance) at and near the location of the observation. Specifying how an assimilated observation should affect changes in space and across variables is a major challenge associated with data assimilation, particularly given discrepancies between the scales resolved by the observation and the first guess, inaccuracies in converting from observed quantities to model variables, and our limited ability to explicitly physically specify how updating one model variable should also update another variable.

Furthermore, data assimilation methods should update the first guess more in locations of higher data density than in locations of lower data density. In the former, the observations provide a tighter constraint upon the 'true' values of the relevant model variables, whereas in the latter the 'true' values of the relevant model variables are less certain. In other words, the first guess forms the basis for the initial conditions in the absence of observations to the contrary.

Sequential versus Continuous Data Assimilation

There exist two categories of data assimilation: **sequential** and **continuous**. With sequential data assimilation methods (Figs. 1 and 2), observations are assimilated in batches at a single analysis time. Consider a data assimilation cycle at 0000 UTC. In sequential data assimilation, all observations valid at a time deemed sufficiently close to 0000 UTC are assimilated as if taken at 0000 UTC. Typically, an interval of +/- 30 to 60 minutes centered on the assimilation time is used. Given a reasonable estimate for the resolved-scale kinematic fields, *time-space conversion* can be used to identify where a given observation would have been taken if it were taken at 0000 UTC, thus enabling for observations to be assimilated at their correct locations. Sequential data

assimilation methods are typically known as *cycled* methods, with the cycling interval reflecting how frequently observations are assimilated to update the first guess for the initial conditions.

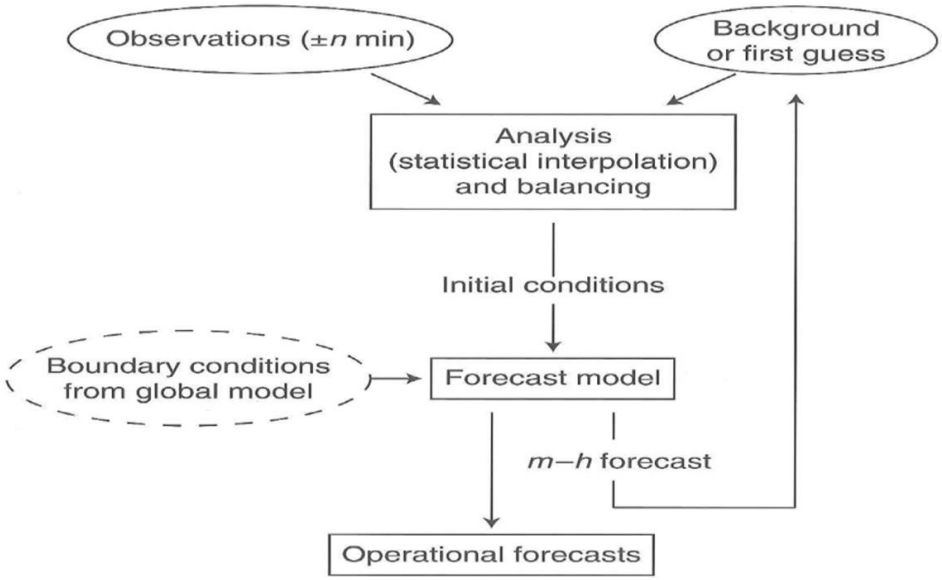

Figure 1. Flowchart of sequential data assimilation. Observations centered on a given analysis time are used to update the first guess for the initial conditions obtained from a previous model forecast valid at the analysis time. Figure reproduced from Warner (2011), their Fig. 6.5.

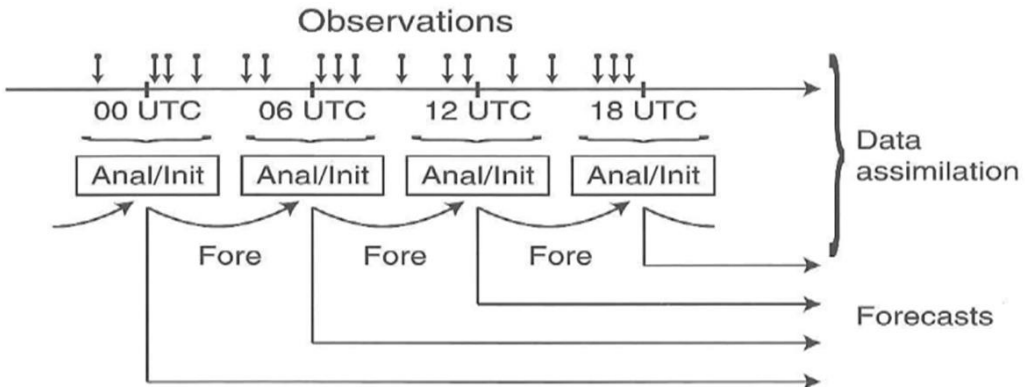

Figure 2. Sequential data assimilation as manifest in operational numerical weather prediction. Observations centered on a given analysis time are binned and assimilated at a fixed analysis time (e.g., 0000 UTC, 0600 UTC, 1200 UTC, or 1800 UTC) to obtain updated initial conditions. Figure reproduced from Warner (2011), their Fig. 6.6b.

By contrast, continuous data assimilation methods (Fig. 3) assimilate observations at the times at which they are valid. Sequential methods update the first guess in light of many observations at a single time, whereas continuous methods update the first guess in light of fewer observations at a larger number of times. Forecasts launched from initial conditions updated using continuous methods can theoretically be launched at any time; however, in practice, forecasts are typically launched only at synoptic times (i.e., when the most observations are available for assimilation). Continuous methods are more computationally intensive than are sequential methods, and as a result most (but not all) operational data assimilation systems are sequential in nature.

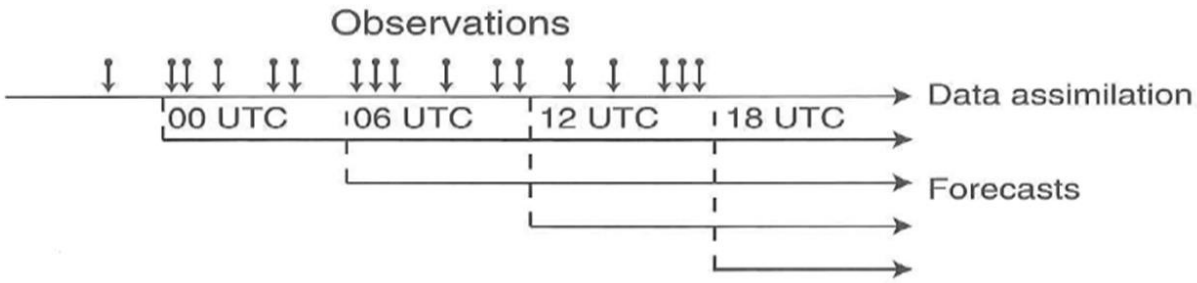

Figure 3. Continuous data assimilation as manifest in operational numerical weather prediction. Observations are assimilated as they are received. In this example, forecasts are launched every 6 h, though they may be launched more or less frequently. Figure reproduced from Warner (2011), their Fig. 6.6a.

Definitions

Before we consider the data assimilation problem, whether in one or many dimensions, we wish to first introduce a litany of common terms encountered within the data assimilation literature. These include the following:

- **State vector ($\vec{x}$)**: the vector that defines the simulated atmospheric state. It and the other vectors noted below contain all model variables at all model grid points at a given time.
- **True state vector ($\overrightarrow{x_t}$)**: the best-possible representation of the atmospheric state on the model grid. Discretizing the atmosphere onto a finite model grid and errors associated with doing so keep the true state vector from exactly matching reality.
- **Perfect state vector ($\overrightarrow{x_p}$)**: reality.
- **Background ($\overrightarrow{x_b}$)**: the 'first guess' estimate of the initial atmospheric state.
- **Analysis ($\overrightarrow{x_a}$)**: the post-assimilation estimate of the simulated atmospheric state.

Each of the aforementioned vectors are of dimension *n*, where *n* is equal to the product of the number of grid points and the number of model variables. Note that both boldface and vector notation are both used to refer to vector quantities throughout the remainder of this document.

Ideally, $\mathbf{x_a} = \mathbf{x_t}$. This is generally not feasible, however, due to limits in the observing network (e.g., observation density, observation precision, representativeness errors, etc.). Instead, we seek to *minimize* the error in $\mathbf{x_a}$, i.e.,

$$\vec{x_t} - \vec{x_a} \approx 0$$

The analysis and background are related to each other by an *analysis increment* δx, i.e.,

$$\vec{x_a} = \vec{x_b} + \delta x \approx \vec{x_t}$$

Note that the analysis increment is dependent upon characteristics of the observations, namely their specific values and error characteristics, as we shall see shortly.

- **Observation vector ($\vec{y}$)**: a vector of dimension *p*, where *p* = number of observations, of all observations to be assimilated.

Data assimilation starts by comparing the background $\mathbf{x_b}$ to the observations $\mathbf{y}$. This is done in observation space, meaning that it is done at the locations of the observations for the variables represented by the observations. This requires the use of a…

- **Forward/transform operator ($\vec{H}(\vec{x})$)**: operator to transform a model field to observation space. In its simplest form, where the observed variable is a model variable, the forward operator is a spatial interpolation operator. In more complex forms, it also invokes retrieval algorithms to convert a model variable to the observed variable.

- **Innovation ($\vec{y} - \vec{H}(\vec{x_b})$)**: the difference between the observations and the background. It is an estimate of the needed correction to the background state. This is equal to δx, except in observation (rather than model) space.

- **Analysis residual ($\vec{y} - \vec{H}(\vec{x_a})$)**: the difference between the observations and the analysis (transformed into observation space).

In the lack of perfect observations of every variable at every location on the model grid, the analysis will depart from the observations. Instead, we seek to minimize the analysis residual,

$$\vec{y} - \vec{H}(\vec{x_a}) \approx 0$$

This is the observation space equivalent to minimizing the error in the analysis $\mathbf{x_a}$ relative to the true state $\mathbf{x_t}$.

One-Dimensional Statistical Framework for Data Assimilation

We now wish to develop a one-dimensional statistical framework for data assimilation. To do so, we use least squares estimation, representing a form of linear regression. Consider the temperature in Milwaukee with a true value of T_t. There exist two estimates of this temperature: the background or first guess value T_b and the observed value T_o. Both background and observed values are imperfect measurements of T_t, with a background error ε_b and observational error ε_o. Background error may be drawn from climatology, while observational error is typically specified empirically based upon instrument characteristics and expected observation representativeness. To obtain the analysis temperature T_a, we wish to optimally combine T_b and T_o based on their individual error characteristics.

We start by defining T_b and T_o relative to T_t:

$$T_b = T_t + \varepsilon_b \qquad\qquad T_o = T_t + \varepsilon_o$$

In other words, each estimate for T is equal to T_t plus the error in the estimate.

We assume that errors in T_b and T_o are random, assuming unbiased methods to obtain T_b and T_o. We also assume that we have knowledge of the background and observational errors ε_b and ε_o.

Recall that variance – here taken to be the *sample* variance – can be defined generically as:

$$\sigma^2 = \frac{\sum (x-\mu)^2}{N-1}$$

Here, x is some estimate of the variable being considered. In this application, x can be viewed as analogous to either T_b or T_o. In statistical terms, μ is the mean of the variable being considered. In this application, μ can be viewed as analogous to T_t. N defines the population size, or number of estimates x. There can be one or many such estimates. The variance defines the mean squared error and, for unbiased methods for x and μ, can be expressed as:

$$\sigma^2 = E\left((x-\mu)^2\right)$$

where the expected value $E(\ )$ is analogous to the mean of an infinitely sampled discrete random variable. This holds here, except that our estimates x are finite in number.

With this information, we can write:

$$\sigma_b^2 = E\left((T_b - T_t)^2\right) = E\left(\varepsilon_b^2\right)$$

$$\sigma_o^2 = E\left((T_o - T_t)^2\right) = E\left(\varepsilon_o^2\right)$$

where we have used the definitions for ε_b and ε_o given above to obtain the final expressions. Note that errors in the background and observations are said to be uncorrelated, such that:

$$E\left(\varepsilon_o \varepsilon_b\right) = 0$$

The least-squares best fit of T_b and T_o, defining the analysis T_a, is given by:

$$T_a = a_b T_b + a_o T_o \qquad \text{where } a_b + a_o = 1$$

This represents the optimal linear combination of T_b and T_o. The a_b and a_o represent the weights applied to T_b and T_o, respectively. We choose these weights so as to *minimize* the mean squared error of T_a, defined by σ_a^2, i.e.,

$$\sigma_a^2 = E\left((T_a - T_t)^2\right)$$

This expression can be expanded by substituting for both T_a and T_t, noting that $T_t = a_b T_t + a_o T_t$ (a decomposition of T_t solely for algebraic purposes):

$$\begin{aligned} E\left((T_a - T_t)^2\right) &= E\left(\left((a_b T_b + a_o T_o) - (a_b T_t + a_o T_t)\right)^2\right) \\ &= E\left(\left(a_o (T_o - T_t) + a_b (T_b - T_t)\right)^2\right) \end{aligned}$$

The second step of the above simply involves grouping like a_o and a_b terms. Noting that $T_o - T_t$ is equal to ε_o and that $T_b - T_t$ is equal to ε_b, we obtain:

$$\sigma_a^2 = E\left((a_o \varepsilon_o + a_b \varepsilon_b)^2\right)$$

If we expand this, we obtain:

$$\sigma_a^2 = E\left((a_o \varepsilon_o)^2 + 2 a_o a_b \varepsilon_o \varepsilon_b + (a_b \varepsilon_b)^2\right)$$

However, we stated earlier that errors in the observations and background are uncorrelated. Thus, the $2 a_o a_b \varepsilon_o \varepsilon_b$ term is zero. Noting that $E(a+b) = E(a) + e(b)$, we obtain:

$$\sigma_a^2 = E\left((a_o \varepsilon_o)^2 + (a_b \varepsilon_b)^2\right) = E\left((a_o \varepsilon_o)^2\right) + E\left((a_b \varepsilon_b)^2\right) = a_o^2 \sigma_o^2 + a_b^2 \sigma_b^2$$

In obtaining the final expression, we applied the definitions of both σ_o^2 and σ_b^2. Consequently, this expression indicates that the analysis error variance is given by a linear combination of the observation and background error variances.

Recall that we are trying to find the a_o and a_b that minimize the mean squared error in T_a, given by σ_a^2. Let $a_o = k$, such that $a_b = 1 - k$. We define k as the optimal weighting factor. Substituting, we obtain:

$$\sigma_a^2 = k^2\sigma_o^2 + (1-k)^2\sigma_b^2$$

Calculus gives us a means of obtaining an expression that minimizes σ_a^2. We wish to minimize σ_a^2 relative to the optimal weighting factor k. Thus, by definition, σ_a^2 is minimized when its first partial derivative with respect to k is equal to zero, i.e.,

$$\frac{\partial\sigma_a^2}{\partial k} = 0$$

Plugging in to this expression, we obtain:

$$\frac{\partial\sigma_a^2}{\partial k} = \frac{\partial}{\partial k}\left(k^2\sigma_o^2 + (1-k)^2\sigma_b^2\right) = 0$$

The expansion of $(1 - k)^2$ is given by $k^2 - 2k + 1$. Thus, we obtain:

$$\frac{\partial}{\partial k}\left(k^2\sigma_o^2 + k^2\sigma_b^2 - 2k\sigma_b^2 + \sigma_b^2\right) = 2k\sigma_o^2 + 2k\sigma_b^2 - 2\sigma_b^2 = 0$$

Solving for k, we obtain:

$$k = \frac{\sigma_b^2}{\sigma_o^2 + \sigma_b^2}$$

In other words, the optimal weighting factor k is equal to the background error variance divided by the total error variance, or the observational plus background error variances. Where uncertainty in the background (σ_b^2) is large relative to uncertainty in the observations (σ_o^2), k will be relatively large. Where uncertainty in the background is small relative to uncertainty in the observations, k will be relatively small.

Recall that k is the coefficient on σ_o^2 in the definition of σ_a^2. Thus, observations are given more weight when k is large whereas the background is given more weight when k is small. Because of the definitions for T_a and k given earlier, we can write:

$$T_a = a_bT_b + a_oT_o = (1-k)T_b + kT_o = T_b + k(T_o - T_b)$$

In other words, the analysis temperature T_a is equal to the background temperature T_b plus an optimally weighted innovation (reflecting the departure of the background from the observation,

noting that no transform is needed between model and observation space in this 1-D example). The weighted innovation $k(T_o - T_b)$ is equal to the analysis increment.

We can plug in for k in the above:

$$T_a = a_b T_b + a_o T_o = (1-k)T_b + kT_o = \left(1 - \frac{\sigma_b^2}{\sigma_o^2 + \sigma_b^2}\right)T_b + \frac{\sigma_b^2}{\sigma_o^2 + \sigma_b^2} T_o$$

However, because $a_o + a_b = 1$,

$$\left(1 - \frac{\sigma_b^2}{\sigma_b^2 + \sigma_o^2}\right) = \frac{\sigma_o^2}{\sigma_b^2 + \sigma_o^2}$$

Thus, we obtain:

$$T_a = \frac{\sigma_o^2}{\sigma_o^2 + \sigma_b^2} T_b + \frac{\sigma_b^2}{\sigma_o^2 + \sigma_b^2} T_o$$

The analysis temperature T_a depends upon the background error and observation error variances. Where the observation error variance is large, the background temperature T_b gets larger weight. Where the background error variance is large, the observation temperature T_o gets larger weight.

As the above expression demonstrates, accurate specification of the observation and background error variances is crucial to obtaining the best-possible estimate for the analysis temperature T_a. Observation error variance is typically specified relative to instrument error characteristics, typical magnitudes for representativeness error, and typical uncertainties associated with the use of a retrieval algorithm (if applicable). Specification of the background error variance is far more challenging and is discussed in more detail later.

Recall that $\sigma_a^2 = k^2\sigma_o^2 + (1-k)^2\sigma_b^2$. Plug in to this expression for k to obtain:

$$\sigma_a^2 = \left(\frac{\sigma_b^2}{\sigma_b^2 + \sigma_o^2}\right)^2 \sigma_o^2 + \left(\frac{\sigma_o^2}{\sigma_b^2 + \sigma_o^2}\right)^2 \sigma_b^2 = \frac{\left(\sigma_b^2\right)^2 \sigma_o^2 + \left(\sigma_o^2\right)^2 \sigma_b^2}{\left(\sigma_b^2 + \sigma_o^2\right)^2}$$

The numerator of this expression can be rewritten as $\left(\sigma_b^2 + \sigma_o^2\right)\sigma_b^2\sigma_o^2$. Thus,

$$\sigma_a^2 = \frac{\sigma_o^2\sigma_b^2}{\left(\sigma_b^2 + \sigma_o^2\right)}$$

Applying the definition for k, we can write:

$\sigma_a^2 = k\sigma_o^2$ or, equivalently, $\sigma_a^2 = (1-k)\sigma_b^2$

Because $k \leq 1$, this means that the analysis error variance is smaller than or equal to *either* the observation or background error variance! For a multi-dimensional problem, this implies that the spread across the ensemble analyses will be smaller than the spread in either the observations or the ensemble backgrounds.

We can also view this in light of what is known as the *precision*, or the inverse variance. Take the inverse of the relationship for σ_a^2 to obtain:

$$\frac{1}{\sigma_a^2} = \frac{\left(\sigma_b^2 + \sigma_o^2\right)}{\sigma_b^2 \sigma_o^2} = \frac{1}{\sigma_o^2} + \frac{1}{\sigma_b^2}$$

Here, the precision of the analysis is equal to the sum of the precisions of the background and observations. Estimates with less error in the background and observations have higher precision. Combining two good estimates results in a very good (in theory, at least) analysis estimate!

Least-squares minimization represents one common approach to data assimilation and is used in many Kalman filter implementations (e.g., the ensemble adjustment Kalman filter embedded within NCAR's Data Assimilation Research Toolkit). A closely related approach is given by *cost function minimization*, which is often used by variational data assimilation. Here, we wish to find the analysis that minimizes the combined squared errors in T_o and T_b, each as weighted by the precision of their measurements.

We define the combined squared error in terms of a cost function $J(T)$ as follows:

$$J(T) = J(T_o) + J(T_b) = \frac{1}{\sigma_o^2}(T - T_o)^2 + \frac{1}{\sigma_b^2}(T - T_b)^2$$

Note that this is expressed relative to a generic T and not to T_t or T_a. The cost in the observations or the background is relatively small when $T \approx T_o$ or T_b and/or when the precision of the observations or background is relatively large. These characteristics are illustrated in Fig. 4.

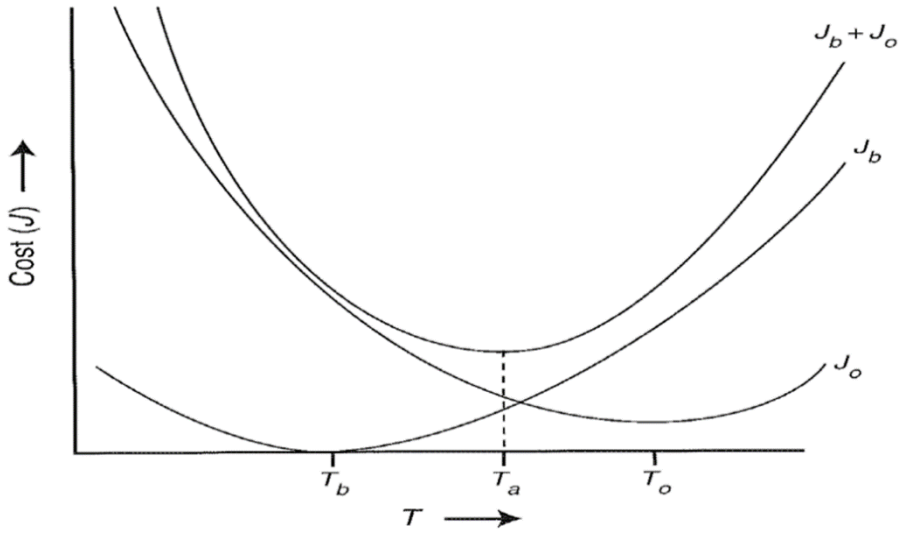

Figure 4. Schematic of $J(T_o)$, $J(T_b)$, and the total cost function $J(T)$ relative to T and J. The cost functions $J(T_o)$ and $J(T_b)$ are minimized where $T - T_o = 0$ and $T - T_b = 0$, respectively. (Note that the representation of $J(T_o)$ above is formally not accurate at $T = T_o$.) The total cost function $J(T)$ is simply equal to the sum of the two cost functions. Where this cost function is minimized with respect to temperature, as can be determined by finding where the first partial derivative equals zero, defines the analysis temperature T_a. Figure reproduced from Warner (2011), their Fig. 6.7.

We wish to minimize the cost function $J(T_a)$ with respect to the analysis temperature, i.e.,

$$\frac{\partial J(T_a)}{\partial T_a} = 0$$

This enables us to obtain an expression the analysis temperature T_a. For completeness, note that:

$$J(T_a) = \frac{1}{\sigma_o^2}(T_a - T_o)^2 + \frac{1}{\sigma_b^2}(T_a - T_b)^2 = \frac{1}{\sigma_o^2}\left(T_a^2 + T_o^2 - 2T_aT_o\right) + \frac{1}{\sigma_b^2}\left(T_a^2 + T_b^2 - 2T_aT_b\right)$$

Thus, we obtain:

$$\frac{\partial J(T_a)}{\partial T_a} = \frac{1}{\sigma_o^2}(2T_a - 2T_o) + \frac{1}{\sigma_b^2}(2T_a - 2T_b) = 0$$

Grouping T_a terms and dividing through by a common factor of 2, we obtain:

$$\frac{T_a}{\sigma_o^2} + \frac{T_a}{\sigma_b^2} = \frac{T_o}{\sigma_o^2} + \frac{T_b}{\sigma_b^2}$$

Operate on the left-hand side to obtain:

$$\frac{T_a\left(\sigma_b^2+\sigma_o^2\right)}{\sigma_o^2\sigma_b^2}=\frac{T_o}{\sigma_o^2}+\frac{T_b}{\sigma_b^2}$$

Isolate T_a and simplify to obtain:

$$T_a=\frac{\sigma_o^2\sigma_b^2}{\left(\sigma_b^2+\sigma_o^2\right)}\left(\frac{T_o}{\sigma_o^2}+\frac{T_b}{\sigma_b^2}\right)=\left(\frac{\sigma_b^2}{\sigma_b^2+\sigma_o^2}\right)T_o+\left(\frac{\sigma_o^2}{\sigma_b^2+\sigma_o^2}\right)T_b=kT_o+\left(1-k\right)T_b$$

This is identical to that obtained from the least squares minimization problem described above, albeit coming at the problem from a slightly different starting point. Differences in how each are implemented for multidimensional problems lead to non-identical results in practice.

Data Assimilation in Multiple Dimensions

In one dimension, the observation and background error variances determine the weighting given to the observation when assimilating it to update the background. In other words, they influence the magnitude of the analysis increment, i.e.,

$$T_a=T_b+k\left(T_o-T_b\right)=T_b+\frac{\sigma_b^2}{\sigma_o^2+\sigma_b^2}\left(T_o-T_b\right)$$

In the multidimensional problem, they also influence the *spread* of the analysis increment. This is manifest through (1) updates to the same variable at other locations, (2) updates to different variables at the same location, and (3) updates to different variables at other locations. To first order, the spread of the analysis increment is a function of the (mathematical, though also ideally physical) relationship, as manifest via their covariance, between the updated variable and other variables or locations. It is crucially dependent upon the accurate specification of the background error covariance matrix.

In one dimension, we dealt with the background error variance σ_b^2. The multidimensional analog is the *background error covariance matrix* **B**. The purpose of the background error covariance matrix is to translate information from an innovation vector $\vec{y}-\overrightarrow{H\left(x_b\right)}$ into a spatially varying analysis increment $\delta\vec{x}$ and apply it to the background to minimize the analysis error. This is the same conceptual framework as in one dimension, except with added dimensionality.

In the one-dimensional problem, the background error variance σ_b^2 is defined as:

$$\sigma_b^2=E\left(\varepsilon_b^2\right)=\overline{\left(x_b-x_t\right)^2}$$

Similarly, the multi-dimensional background error covariance matrix **B** is defined as:

$$\vec{B} = \overrightarrow{\left(\vec{x}_b - \vec{x}_t\right)\left(\vec{x}_b - \vec{x}_t\right)^T}$$

Equivalently, if the background errors are unbiased, the true state can be approximated by the mean of the background estimates, i.e., $x_t = \overline{x_b}$. Thus, the above expressions can be rewritten as:

$$\sigma_b^2 = E\left(\varepsilon_b^2\right) = \overline{\left(x_b - \overline{x_b}\right)^2}$$

$$\vec{B} = \overrightarrow{\left(\vec{x}_b - \overrightarrow{\overline{x}_b}\right)\left(\vec{x}_b - \overrightarrow{\overline{x}_b}\right)^T}$$

In the above, *T* denotes the transpose of the matrix. The background error covariance matrix is an *n* x *n* square matrix, recalling that *n* is the product of the number of grid points and the number of model variables. The diagonal of **B** is populated by the error variances for a given background estimate at a given location, similar to the one-dimensional problem. The off-diagonal terms of **B** are populated by the covariances between errors in two separate background estimates. For the case where *n* = 3, i.e., three variables at one grid point or three grid points for a single variable, **B** takes the form:

$$\vec{B} = \begin{bmatrix} \mathrm{var}(e_1) & \mathrm{cov}(e_2, e_1) & \mathrm{cov}(e_3, e_1) \\ \mathrm{cov}(e_1, e_2) & \mathrm{var}(e_2) & \mathrm{cov}(e_3, e_2) \\ \mathrm{cov}(e_1, e_3) & \mathrm{cov}(e_2, e_3) & \mathrm{var}(e_3) \end{bmatrix}$$

Note that $e_{\#} = x_{b\#}$. As stated before, the specification of the covariance terms defines both the spread of and the weight given to analysis increments. The variance terms define the weight given to analysis increments only for the variable being updated at the location that it is updated.

How is the background error covariance matrix specified? There exist two primary approaches that are used to do so:

- **Flow-dependent**: Consider an ensemble of short-range numerical forecasts valid at some analysis time. These comprise the background estimates. For unbiased background estimates, the true state can be approximated by the mean of the background estimates. Departures of the first guess from this approximate true state, defining background errors, and the variances of and covariances between the errors define **B**. This specification of **B** is implicitly flow-dependent (i.e., covariances are likely to be larger following the flow) and is explicitly temporally varying. See Fig. 5 for an idealized example.

- **Flow-independent**: Here, **B** is specified similar to the flow-dependent case, except single forecasts over a large number of cases (rather than many forecasts over a single case) are considered. As a result, any flow-dependence is washed out by the averaging over many cases, each of which has different background error characteristics. This can be viewed as

a climatological approach to specifying the background error covariance matrix, and a version of this method was used by NCEP for many years within their operational data assimilation routines. See Fig. 6 for an idealized example.

Whether flow-independent or flow-dependent methods are used to specify **B**, most data assimilation systems employ some means of *localization* to place an upper bound on the outward spread of the analysis increment to avoid spurious updates as may be associated with small non-physical covariances (e.g., 500 hPa temperature near Milwaukee to the 850 hPa zonal wind over Antarctica, as one hypothetical example). Specifying an innovation influence (or covariance) of zero beyond a certain radius in Figs. 5 and 6 is an illustrative example of localization.

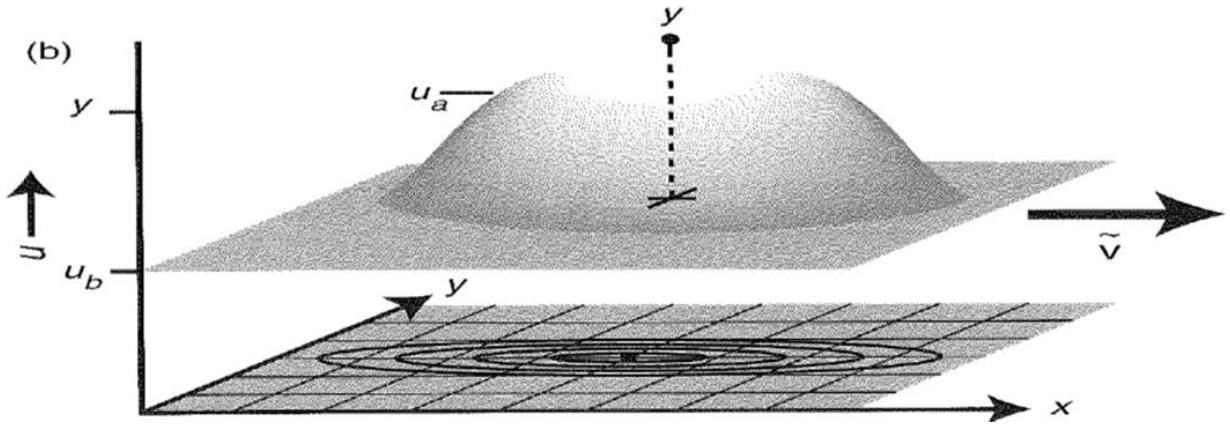

Figure 5. Hypothetical spread of the innovation, defining the analysis increment, in space for a flow-dependent specification of the background error covariance matrix. In this example, it was determined (using appropriate statistics) that the background errors are better-correlated following the large-scale westerly flow, and as a result the decay of the analysis increment away from the observation location is non-isotropic (i.e., less rapid following the flow, more rapid perpendicular to the flow). Figure reproduced from Warner (2011), their Fig. 6.9b.

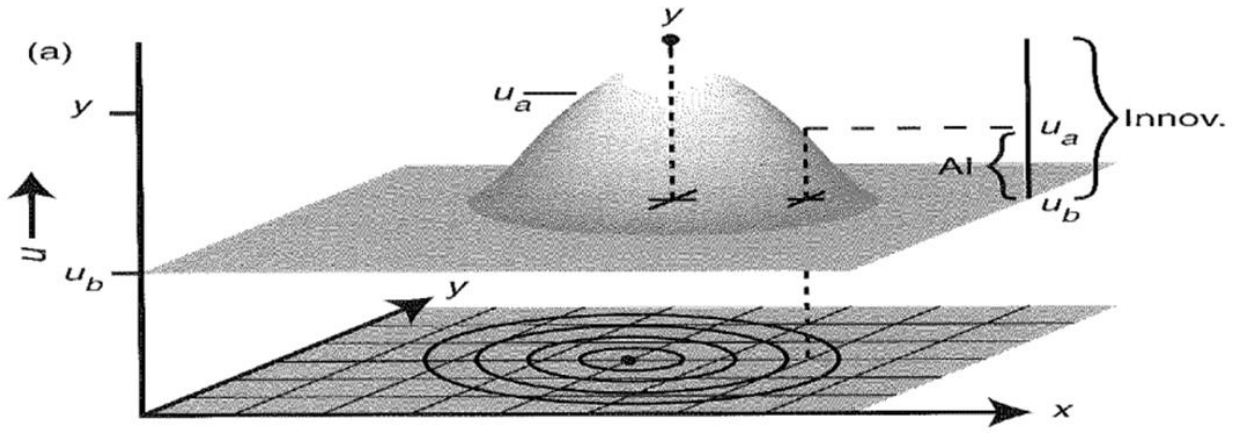

Figure 6. Hypothetical spread of the innovation, defining the analysis increment, in space for a flow-independent specification of the background error covariance matrix. The analysis increment is maximized at the location of the observation and decays isotropically (uniform in all directions) independent of the large-scale flow. Figure reproduced from Warner (2011), their Fig. 6.9a.

For completeness, we present the multidimensional analogs to the other variables defined for the one-dimensional data assimilation problem:

- **Observation error covariance matrix (R)**: the multidimensional analog to σ_o^2. Typically, it is assumed that any two observation errors are uncorrelated, such that the covariance terms of this matrix are all zero. Thus, **R** is a diagonal matrix of dimension *p* x *p*, where *p* is equal to the number of observations, and is comprised only of error variances.

- **Analysis error covariance matrix (A)**: the multidimensional analog to σ_a^2.

- **Forecast error covariance matrix (Q)**: the multidimensional analog to the forecast error variance $\sigma_f^2 = E(\varepsilon_f^2) = E((T_f - T_t)^2)$, where subscripts of *f* denote forecast quantities.

- **Weighting matrix (K)**: the multidimensional analog to *k*, defining both the weight and spread of analysis increments.

These definitions allow us to state, without formal derivation, the analogous expressions for the analysis and optimal weight in the multidimensional problem. Recall that the analysis for the one dimensional problem was expressed as:

$$T_a = T_b + k(T_o - T_b)$$

In the multidimensional problem, the analogous expression is:

$$\vec{x_a} = \vec{x_b} + \vec{K}\left(\vec{y} - \vec{H}\left(\vec{x_b}\right)\right)$$

Here, the analysis state is equal to the background state plus the weighted innovations, defined as the observations minus the transformed background estimates. Note that the transformation of the innovation back to model space is implicit to the above formulation.

Recall that the weight for the one-dimensional problem was expressed as:

$$k = \frac{\sigma_b^2}{\sigma_o^2 + \sigma_b^2}$$

In the multidimensional problem, the analogous expression is:

$$\vec{K} = \frac{\vec{B}\vec{H}^T}{\vec{H}\vec{B}\vec{H}^T + \vec{R}}$$

Here, the weighting matrix is equal to the ratio of the background error covariance matrix to the sum of the background and observation error covariance matrices. The **H** and $\mathbf{H^T}$ are transform operators between model and observation space. Data assimilation methods can vary with respect to how the background and, to some extent, observation error covariance matrices are defined, as was discussed above for the background error covariance matrix.