Chapter 3

Optimal Interpolation

Optimal Interpolation or OI is a commonly used and fairly simple but powerful method of data assimilation. Most weather centers around the world used OI for operational numerical weather forecasts throughout the 1970s and 80s. In fact, Canada was the first to implement an OI for operational weather forecasting in the 1970's with the others following suit. Only recently, in the last decade, has there been a shift away from OI and toward variational methods.

OI is also a good place to start when studying data assimilation since only the spatial dimensions are used. Later on we will introduce the time dimension as well.

OI is so named because it is a minimum variance estimator, as we shall see. However, this is only in theory. In practice, it will never be optimal, so it is often referred to as "Statistical Interpolation" or SI.

This chapter is primarily based on Daley (1991) chapter 4. For more details, the interested reader is referred to Daley (1991), chapters 3, 4 and 5.

3.1 A two city example

Let us return to our simple example in which we estimate temperature or ozone levels at Toronto and Montreal but now based on observations and background estimates at both places. We also assume that the observation and background errors are unbiased and that we have access to their variances. The instrument type is assumed the same for both locations and equals $(\sigma^r)^2$. The correlation of observation error between Toronto and Montreal is 0. The background error variance is $(\sigma^b_T)^2$ at Toronto and $(\sigma^b_M)^2$ at Montreal. The correlation of background error between Toronto and Montreal is given by ρ. Recall that we introduced an analysis equation:

$$\mathbf{x}^a = \mathbf{x}^b + \mathbf{W}(\mathbf{x}^o - \mathbf{x}^b). \tag{3.1}$$

where

$$\begin{aligned} \mathbf{x}^b &= (\mathrm{x}^b_T, \mathrm{x}^b_M)^{\mathrm{T}}, \\ \mathbf{x}^o &= (\mathrm{x}^o_T, \mathrm{x}^o_M)^{\mathrm{T}}, \\ \mathbf{x}^a &= (\mathrm{x}^a_T, \mathrm{x}^a_M)^{\mathrm{T}}, \end{aligned}$$

where the subscripts T and M refer to Toronto and Montreal and the superscripts b, a and o refer to the background, analyses and observations, respectively. The weight matrix is 2 x 2 with

components:

$$\mathbf{W} = \begin{bmatrix} w_{TT} & w_{TM} \\ w_{MT} & w_{MM} \end{bmatrix}. \tag{3.2}$$

What we'd like to do is determine the weight matrix using the knowledge we possess. To do this, let's follow the simple scalar example of section 1.5.

First, assume the existence of a *truth* although we don't need to know what this is. By doing so, we can rewrite the analysis equation in terms of errors. With the errors defined as

$$\begin{aligned} \mathbf{e}^a &= \mathbf{x}^a - \mathbf{x}^t \\ \mathbf{e}^b &= \mathbf{x}^b - \mathbf{x}^t \\ \mathbf{e}^r &= \mathbf{x}^o - \mathbf{x}^t \end{aligned}$$

and the definitions,

$$\mathbf{P}^a = < (\mathbf{e}^a)(\mathbf{e}^a)^{\mathrm{T}} > \qquad \mathbf{P}^b = < (\mathbf{e}^b)(\mathbf{e}^b)^{\mathrm{T}} > \qquad \mathbf{R} = < (\mathbf{e}^r)(\mathbf{e}^r)^{\mathrm{T}} >, \tag{3.3}$$

the analysis equation is

$$\mathbf{e}^a = \mathbf{e}^b + \mathbf{W}(\mathbf{e}^r - \mathbf{e}^b). \tag{3.4}$$

Now recall the assumption about unbiased errors. Because of this assumption we can easily see (by taking the expectation of (3.4)) that the analysis error will be unbiased also. This assumption is not critical however, because if biases exist we can define new analysis variables that subtract the biases as in problem 1.3. To solve for the weights, let's use the information we have, namely, the variances. Since the variances are the diagonals of the covariance matrix we shall form the latter by multiplying (3.4) on the right by the transpose of itself and then applying the expectation operator. The result is

$$\mathbf{P}^a = \mathbf{P}^b + \mathbf{W}(\mathbf{R} + \mathbf{P}^b)\mathbf{W}^{\mathrm{T}} - \mathbf{W}\mathbf{P}^b - \mathbf{P}^b\mathbf{W}^{\mathrm{T}} \tag{3.5}$$

if we assume no correlation between background and observation errors. If the background comes from a model forecast, then this is a completely independent information source from the measurement so this assumption is reasonable, to first order. In reality, both the forecast error and observation error may be functions of the true atmospheric state and are therefore not independent. However, the error of this assumption is much smaller than some approximations used in practice so it is safe to make the assumption. Now, let's take the derivative of the trace of (3.4) with respect to the weight matrix and set it to zero:

$$0 = 2\mathbf{W}(\mathbf{R} + \mathbf{P}^b) - 2\mathbf{P}^b. \tag{3.6}$$

At this point it should be clear that it is necessary to become familiar with vector algebra for this course. While the problem can be solved by writing out the components of the vector, this is onerous. It is easier to use the rules found in Todling (1999) ch. 4, problem 4:

$$\frac{d[\mathrm{Tr}(\mathbf{AB})]}{d\mathbf{A}} = \mathbf{B}^{\mathrm{T}}, \qquad \frac{d[\mathrm{Tr}(\mathbf{BA})^{\mathrm{T}}]}{d\mathbf{A}} = \mathbf{B}, \qquad \frac{d[\mathrm{Tr}(\mathbf{ACA}^{\mathrm{T}})]}{d\mathbf{A}} = 2\mathbf{AC} \tag{3.7}$$

where **AB** and **C** are symmetric. These rules are easy to confirm by writing out the terms of each side. As an example, the first rule is derived in Appendix B. Try to derive the other rules in the same manner. Now we can rearrange the above to solve for **W**:

$$\mathbf{W} = (\mathbf{P}^b)(\mathbf{R} + \mathbf{P}^b)^{-1}. \tag{3.8}$$

With this weight we can obtain our minimum variance analysis. We can also substitute this expression into (3.5) to obtain an error estimate of our analysis:

$$\begin{aligned} \mathbf{P}^a &= \mathbf{P}^b(\mathbf{I} - \mathbf{W}^{\mathrm{T}}) = (\mathbf{I} - \mathbf{W})\mathbf{P}^b \\ &= [\mathbf{I} - \mathbf{P}^b(\mathbf{R} + \mathbf{P}^b)^{-1}]\mathbf{P}^b \\ &= [(\mathbf{R} + \mathbf{P}^b)(\mathbf{R} + \mathbf{P}^b)^{-1} - \mathbf{P}^b(\mathbf{R} + \mathbf{P}^b)^{-1}]\mathbf{P}^b \\ &= \mathbf{R}(\mathbf{R} + \mathbf{P}^b)^{-1}\mathbf{P}^b. \end{aligned} \tag{3.9}$$

This can be inverted to get

$$\begin{aligned} (\mathbf{P}^a)^{-1} &= (\mathbf{P}^b)^{-1}(\mathbf{R} + \mathbf{P}^b)\mathbf{R}^{-1} \\ &= [(\mathbf{P}^b)^{-1}\mathbf{R} + \mathbf{I}]\mathbf{R}^{-1} \\ &= [(\mathbf{P}^b)^{-1} + \mathbf{R}^{-1}] \end{aligned}$$

Thus we have solved our problem. The analysis equation is (3.1) with weights given by (3.8) and analysis error, (3.9). Note the similarity between (3.8) and (1.6) and between (3.9) and (1.7). Of course the vector case is more general and includes the scalar case when the dimension of $\mathbf{x}$ and $\mathbf{x}^o$ are equal to 1. We can do a similar analysis of the impact of observations on the analysis, as in section 1.5. If the observations are perfect or the background error is very poor, then $(\sigma^r)^2 << (\sigma^b_{T,M})^2$ so $\mathbf{R} << \mathbf{P}^b$ and $\mathbf{W} = \mathbf{I}$. (Note that in this simple example, both $\mathbf{P}^b$ and $\mathbf{R}$ are 2x2 matrices so we can define $\mathbf{P}^b > \mathbf{R}$ as meaning $\mathbf{P}^b - \mathbf{R}$ is positive definite.) The observations are given full weight and the analysis at Toronto and Montreal equals the observations at Toronto and Montreal, respectively. In contrast, if the observations are of very poor quality or are not available, $\mathbf{P}^b << \mathbf{R}$ and $(\sigma^r)^2 \to \infty$ so $\mathbf{W}$=**0** and $\mathbf{x}^a = \mathbf{x}^b$. In the absence of observations, the analysis reverts to the background estimates. Finally, if $\mathbf{P}^b = \mathbf{R}$, then $\mathbf{W}$=0.5 **I** and observations and background are combined with equal weights at both locations.

We have discussed the solution to the data assimilation problem for some very special cases. What about the more general case? In order to consider the more general case, let us first write everything back in component form. From (3.1), we have

$$\mathbf{x}^a_T = \mathbf{x}^b_T + w_{TT}(\mathbf{x}^o_T - \mathbf{x}^b_T) + w_{TM}(\mathbf{x}^o_M - \mathbf{x}^b_M) \tag{3.10}$$

$$\mathbf{x}^a_M = \mathbf{x}^b_M + w_{MT}(\mathbf{x}^o_T - \mathbf{x}^b_T) + w_{MM}(\mathbf{x}^o_M - \mathbf{x}^b_M). \tag{3.11}$$

The terms in round brackets are components of the **innovation** vector or **observation increment**. The innovations are simply the differences between the observation and background values evaluated at observation locations. Now we see that the observation at Toronto affects the analysis at both Toronto and Montreal. What is the weight given to the observation increment at Toronto? To see this let's write the components of the weight matrix according to (3.8). First note that

$$\mathbf{P}^b = \begin{bmatrix} (\sigma^b_T)^2 & \rho\sigma^b_T\sigma^b_M \\ \rho\sigma^b_M\sigma^b_T & (\sigma^b_M)^2 \end{bmatrix}, \qquad \mathbf{R} = \begin{bmatrix} (\sigma^r)^2 & 0 \\ 0 & (\sigma^r)^2 \end{bmatrix} \tag{3.12}$$

so that

$$\mathbf{R}+\mathbf{P}^b = \begin{bmatrix} (\sigma^r)^2+(\sigma_T^b)^2 & \rho\sigma_T^b\sigma_M^b \\ \rho\sigma_M^b\sigma_T^b & (\sigma^r)^2+(\sigma_M^b)^2 \end{bmatrix} \tag{3.13}$$

and

$$(\mathbf{R}+\mathbf{P}^b)^{-1} = \frac{1}{|\mathbf{D}|}\begin{bmatrix} (\sigma^r)^2+(\sigma_M^b)^2 & -\rho\sigma_T^b\sigma_M^b \\ -\rho\sigma_M^b\sigma_T^b & (\sigma^r)^2+(\sigma_T^b)^2 \end{bmatrix} \tag{3.14}$$

with

$$|\mathbf{D}| = |\mathbf{R}+\mathbf{P}^b| = ((\sigma^r)^2+(\sigma_M^b)^2)((\sigma^r)^2+(\sigma_T^b)^2)+\rho^2(\sigma_T^b)^2(\sigma_M^b)^2. \tag{3.15}$$

Combining (3.12), (3.14) and (3.15) according to (3.8) yields

$$\mathbf{W} = \frac{1}{(1+\alpha_M)(1+\alpha_T)+\rho^2}\begin{bmatrix} (1+\alpha_M)-\rho^2 & \rho\sqrt{\alpha_M}\sqrt{\alpha_T} \\ \rho\sqrt{\alpha_M}\sqrt{\alpha_T} & (1+\alpha_T)-\rho^2 \end{bmatrix} \tag{3.16}$$

with $\alpha_M = (\sigma^r)^2/(\sigma_M^b)^2$ and $\alpha_T = (\sigma^r)^2/(\sigma_T^b)^2$. If the correlation between the background (forecast) error at Toronto and Montreal is zero, then the weight simplifies to:

$$\mathbf{W} = \begin{bmatrix} (1+\alpha_T)^{-1} & 0 \\ 0 & (1+\alpha_M)^{-1} \end{bmatrix}. \tag{3.17}$$

Thus only the observation at Toronto impacts the analysis at Toronto and similarly for Montreal. As in the scalar case, the observation and background are combined according to their *relative* accuracies. As the correlation increases from zero, the off-diagonal terms in (3.16) increase in magnitude. Thus the weight of an observation at Montreal on the analysis at Toronto increases and vice versa. The maximum weight that can be given to an observation at Montreal for an analysis at Toronto is when ρ=1 and

$$w_{TM}^{\max} = \frac{\sqrt{\alpha_M}\sqrt{\alpha_T}}{(1+\alpha_M)(1+\alpha_T)+1}.$$

When ρ=1, the observation at Toronto receives a weight of

$$w_{TT} = \frac{\alpha_M}{(1+\alpha_M)(1+\alpha_T)+1}$$

and $w_{TM}^{\max}/w_{TT} = \sqrt{\alpha_T}/\sqrt{\alpha_M}$. If the background error variance is the same at both locations, then $\alpha_M = \alpha_T$, and the weight given to both is the same. Recall that both observation error variances were assumed to be the same. If the background error at Montreal is lower, then $\alpha_M > \alpha_T$ and its observation is given less weight than the one in Toronto since the background (forecast) is quite good in Montreal. Similarly, if the background error variance were lower in Toronto, the observation at Montreal would have more weight.

Clearly, observations at one location can influence the analysis at other locations. The key to determining the influence is the correlation, ρ, which is the correlation of background errors between the observation location and the analysis location. Since this correlation is derived from the background error covariance matrix, it is clear that we need to know more about this matrix and how to compute it. This will be postponed until later in this chapter. In the next section, we will see the impact of two observations on a single analysis gridpoint.

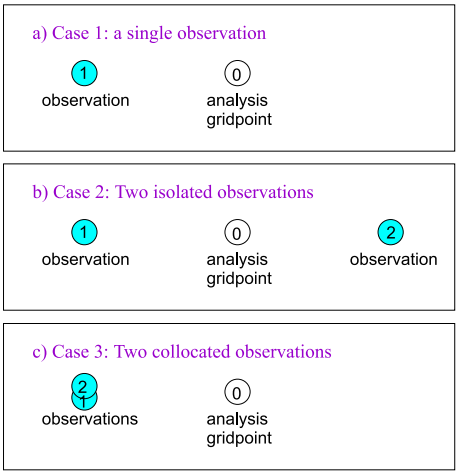

Figure 3.1: The influence of two observations on the analysis at gridpoint 0. The various cases are discussed in the text.

3.2 Two observations on a 1-D grid

Now consider the case of an analysis grid point influenced by two observations. This example comes from Daley (1991) chapter 4.6 and is illustrated in Fig. 3.1.

Both observations are of the same type and so have the same error variance of $(\sigma^r)^2$. Similarly, the background error variance at both obs stations is assumed the same:

$$< (\epsilon_1^b)^2 >=< (\epsilon_2^b)^2 >= (\sigma^b)^2,$$

Subscripts 1 and 2 refer to observation locations 1 and 2. Subscript 0 refers to the location of the grid point where the analysis is desired. The observation error is assumed horizontally uncorrelated, i.e.

$$< (\epsilon_1^r)(\epsilon_2^r) >= 0.$$

The observation and background errors are uncorrelated:

$$< (\epsilon_c^b)(\epsilon_d^r) >= 0$$

where $c, d \in \{0, 1, 2\}$. The analysis equation is

$$\mathrm{x}_0^a = \mathrm{x}_0^b + w_1(\mathrm{x}_1^r - \mathrm{x}_1^b) + w_2(\mathrm{x}_2^r - \mathrm{x}_2^b). \tag{3.18}$$

To determine the weights applied to each observation according to a minimum variance principle, first form the analysis error variance from (3.18) and apply the expectation operator.

$$< (\epsilon_0^a)^2 >= (\sigma^b)^2 + (w_1^2 + w_2^2)[(\sigma^r)^2 + (\sigma^b)^2] - 2w_1\rho_{10}(\sigma^b)^2 - 2w_2\rho_{20}(\sigma^b)^2 + 2w_1w_2\rho_{12}(\sigma^b)^2 \tag{3.19}$$

where we have defined

$$< \epsilon_0^b\epsilon_1^b >= \rho_{10}(\sigma^b)^2, \qquad < \epsilon_0^b\epsilon_2^b >= \rho_{20}(\sigma^b)^2, \qquad < \epsilon_1^b\epsilon_2^b >= \rho_{12}(\sigma^b)^2$$

and where all terms involving correlations of observation and background errors have been dropped. Now minimize $< (\epsilon_0^a)^2 >$ w.r.t. w_1 and w_2:

$$\begin{aligned} w_1(1+\alpha) + w_2\rho_{12} &= \rho_{10} \\ w_1\rho_{12} + w_2(1+\alpha) &= \rho_{20} \end{aligned}$$

where

$$\alpha = (\sigma^r)^2/(\sigma^b)^2.$$

Solving for w_1 and w_2 yields:

$$w_1 = \frac{\rho_{10}(1+\alpha) - \rho_{12}\rho_{20}}{(1+\alpha)^2 - \rho_{12}^2} \tag{3.20}$$

$$w_2 = \frac{\rho_{20}(1+\alpha) - \rho_{12}\rho_{10}}{(1+\alpha)^2 - \rho_{12}^2}. \tag{3.21}$$

With these optimal weights, (3.19) becomes

$$< (\epsilon_0^a)^2 >= (\sigma^b)^2 \left\{ 1 - \frac{(1+\alpha)(\rho_{10}^2 + \rho_{20}^2) - 2\rho_{10}\rho_{20}\rho_{12}}{(1+\alpha)^2 - \rho_{12}^2} \right\}. \tag{3.22}$$

3.2.1 Case One: A single observation

What is the analysis at gridpoint 0 if only the observation at gridpoint 1 is available? In this case, the analysis equation, (3.18) reduces to

$$\mathrm{x}_0^a = \mathrm{x}_0^b + w_1(\mathrm{x}_1^r - \mathrm{x}_1^b),$$

the weight, (3.20) becomes

$$w_1 = \frac{\rho_{10}}{1+\alpha} \tag{3.23}$$

and the analysis error variance in (3.22) becomes

$$< (\epsilon_0^a)^2 >= (\sigma^b)^2 \left\{ 1 - \frac{\rho_{10}^2}{1+\alpha} \right\}. \tag{3.24}$$

If the observation is coincident with the analysis gridpoint, then ρ_{10}=1 and we reproduce our simple example of section 1.5. As the observation becomes further away from our analysis, we expect the correlation between background errors at the two locations to drop in magnitude. Finally, when the observations is so far away that the correlation is 0, the weight is 0 and the analysis error variance reverts to the background error variance. Thus the weight given to an observation depends on the distance between it and the analysis grid point and the way the background error correlation varies with distance. Clearly, we need to know more about this background error correlation and how it varies in space. This will be examined in section 3.5 .

3.2.2 Case Two: Two isolated observations

Now consider the case where the observations are located on either side of the analysis grid point. Let us assume that $\rho_{12} \approx 0$, i.e., that the two observations are so far from each other that the background error correlation between the two locations is zero. Let's also assume that $\rho_{10} = \rho_{20} = \rho$, i.e. that the two observations are equidistant from the analysis gridpoint, and their background error correlations with that at the analysis location is identical. In this case, (3.20) and (3.21) reduce to

$$w_1 = w_2 \approx \frac{\rho}{1+\alpha}, \tag{3.25}$$

and (3.22) reduces to

$$< (\epsilon_0^a)^2 >= (\sigma^b)^2 \left\{ 1 - \frac{2\rho^2}{1+\alpha} \right\}. \tag{3.26}$$

Comparing (3.24) and (3.26) reveals that having 2 observations results in a lower analysis error than having only 1 observation.

3.2.3 Case Three: Two collocated observations

What if, instead of being located on either side of the analysis gridpoint, the two observations are collocated? In this case, $\rho_{12} = 1$ and $\rho_{10} = \rho_{20} = \rho$ so that

$$w_1 = w_2 = \frac{\rho}{2+\alpha}, \tag{3.27}$$

and

$$< (\epsilon_0^a)^2 >= (\sigma^b)^2 \left\{ 1 - \frac{2\rho^2}{2+\alpha} \right\}. \tag{3.28}$$

The weight given to the collocated observations is less than that for the isolated observations. The analysis error is also smaller when there are isolated observations. Why? More information is obtained for *independent* observations. Two collocated observations do not provide independent information so they each contribute less than if they had been independent.

3.2.4 Case Four: Two observations on a 1-D network

We've seen that a two observations are better than one and that two isolated observations are better than two collocated ones. If you have two observations, where is the best place to put them, presuming one can choose this? To find out, consider holding one the observations fixed at x/L=-2. The analysis gridpoint is at x/L=0. The second observation's location will vary with x from -∞ to +∞. Daley's Fig. 3.2 illustrates the result of this experiment for $\alpha_1 = \alpha_2 = 0.25$ and $\rho_{10} = 0.406$. To determine the remaining correlations, ρ_{20}, and ρ_{12}, a model of the variation of the correlation with distance is adopted:

$$\rho^b(\Delta x) = \left(1 + \frac{|\Delta x|}{L}\right) \exp\left(-\frac{|\Delta x|}{L}\right).$$

Thus the background error correlation between two points depends only on the distance between the points. Daley's Fig. 3.2 depicts the weights, w_1 and w_2, and the normalized analysis error

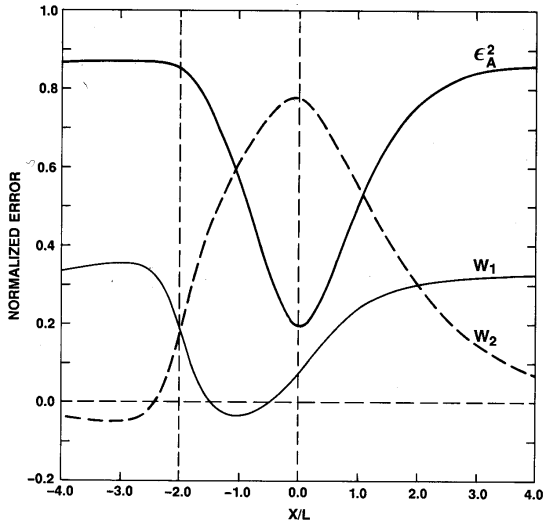

Figure 3.2: A posteriori weights w_1 and w_2 and normalized expected analysis error variance $< (\epsilon_0^a)^2 > /(\sigma^b)^2$ for the analysis gridpoint at x=0, observation 1 at x=-2.0 and the position of observation 2 varying between $x = -\infty$. (From Daley 1991, Fig. 4.7).

variance, $< (\epsilon_0^a)^2 > /(\sigma^b)^2$. When observation 2 is to the left of observation 1, the weight given to it is very small. The weight given to observation 1 is close to the single observation value, (3.23). When it is coincident with observation 1, the weights given to both observations are the same and given by (3.27). As observation 2 moves closer to the analysis location, its weight increases while the weight given to observation 1 decreases. Overall though, the analysis error begins to decrease until observation 2 coincides with the analysis location. At this point, the weight for obs 2 is maximized. As obs 2 moves further to the right, its weight begins to decrease and the total analysis error increases. When obs 2 is at x/L=2, the same distance from the analysis gridpoint as obs 1, the weights are again equal but larger than when the observations were collocated. Finally as obs 2 moves further to the right, its weight drops off and its impact on the analysis becomes diminished. Note that the weight for obs 1 can be negative when obs 2 is closer to the analysis point. Similarly, the weight for obs 2 can be negative when obs 1 is closer to the analysis location. This is the effect of observation screening, when the weight given to a more distant observation can actually be negative due to the presence of a closer observation.

3.3 Spatial Interpolation

The simple example of section 3.1 almost describes the general Statistical Interpolation (SI or OI) algorithm. To get to the general case, we need only redefine our vectors and allow for observations not coincident with analysis gridpoints. First, for the general case, consider a model state vector,

$$\mathbf{x}^{\mathrm{T}} = (\mathrm{x}_1, \mathrm{x}_2, \ldots, \mathrm{x}_n)^{\mathrm{T}}.$$

The background, $\mathbf{x}^b$, and analysis, $\mathbf{x}^a$ are both on this grid. (For a spectral model, the state may consist of spectral coefficients rather than gridpoints, but for this discussion let's assume they are gridpoint values since this is easier to envision.) Note that $\mathbf{x}^b$ and $\mathbf{x}^a$ are n-vectors. Let us assume that there is an observation network of m measurements. Let us define the observation vector as

$$\mathbf{z}^{\mathrm{T}} = (\mathrm{z}_1, \mathrm{z}_2, \ldots, \mathrm{z}_m)^{\mathrm{T}}.$$

Since the observations are not necessarily at analysis grid points, we need a spatial interpolation from the observation locations to the model grid. Let's call this operator, H. This operator, also called the *forward model*, maps the model state to the observed variables and locations. Thus, if the observation is a radiance from a satellite instrument, the forward model operator involves integration of the temperature (and perhaps water vapour) over a column in the model atmosphere and using the precise location of the satellite at the time of measurement. Clearly this operator can be a nonlinear function of the model variables. We indicate this with the notation, H, being NOT bold. (If our model state had been spectral coefficients, this operator also includes an inverse spectral transform back to the physical space used for measurements.) Our analysis equation is then

$$\mathbf{x}^a = \mathbf{x}^b + \mathbf{K}(\mathbf{z} - H(\mathbf{x}^b)). \tag{3.29}$$

Note that we renamed the weight matrix, $\mathbf{K}$ instead of $\mathbf{W}$ in anticipation of the development of the Kalman filter in later chapters. Note the similarity between (3.1) and (3.29). We can now introduce our stochastic measurement equation:

$$\mathbf{z} = \mathbf{z}^t + \boldsymbol{\nu} \tag{3.30}$$

$\boldsymbol{\nu}$ is the measurement error (see section 1.4). $\mathbf{z}^t$ is the "true" atmospheric quantity being sensed. However, we are more interested in the truth projected onto our imperfect model basis. So, let's introduce our imperfect forward model operator into the above equation. The result is

$$\begin{aligned} \mathbf{z} &= H(\mathbf{x}^t) + \mathbf{z}^t - H(\mathbf{x}^t) + \boldsymbol{\nu} \\ &= H(\mathbf{x}^t) + \mathbf{v} \end{aligned} \tag{3.31}$$

where

$$\mathbf{v} = [\mathbf{z}^t - H(\mathbf{x}^t)] + \boldsymbol{\nu}. \tag{3.32}$$

The term is square brackets is called the *representativeness* error and reflects the fact that our forward model, H, is not perfect. Recall that H includes a mapping of model variables to observed variables and a spatial interpolation from the model grid (or state) to the observed locations. The sum of the measurement and representativeness errors form the *observation* error, $\mathbf{v}$. The observation error bias is given by

$$< \mathbf{v} >= \bar{\mathbf{v}}$$

and the observation error covariance matrix is:

$$\mathbf{R} =< (\mathbf{v} - \bar{\mathbf{v}})(\mathbf{v} - \bar{\mathbf{v}})^{\mathrm{T}} > .$$

Our errors can then be defined as before, with a new addition:

$$\begin{aligned} \mathbf{e}^a &= \mathbf{x}^a - \mathbf{x}^t \\ \mathbf{e}^b &= \mathbf{x}^b - \mathbf{x}^t \\ \mathbf{v} &= \mathbf{z} - H(\mathbf{x}^t). \end{aligned}$$

The analysis equation in terms of errors is then

$$\begin{aligned}
\mathbf{e}^a &= \mathbf{e}^b + \mathbf{K}[\mathbf{z} - H(\mathbf{x}^t) + H(\mathbf{x}^t) - H(\mathbf{x}^b)] \\
&= \mathbf{e}^b + \mathbf{K}[\mathbf{v} + H(\mathbf{x}^b + \mathbf{x}^t - \mathbf{x}^b) - H(\mathbf{x}^b)] \\
&\approx \mathbf{e}^b + \mathbf{K}[\mathbf{v} + H(\mathbf{x}^b) + \mathbf{H}(\mathbf{x}^t - \mathbf{x}^b) - H(\mathbf{x}^b)] \\
&= \mathbf{e}^b + \mathbf{K}[\mathbf{v} - \mathbf{H}(\mathbf{e}^b)].
\end{aligned} \tag{3.33}$$

To get the 3rd line, we approximated the second term in square brackets as a Taylor series truncated after the linear term. Thus we have introduced a new operator, the *Tangent Linear Forward Model* operator, which is defined as

$$\mathbf{H} = \left.\frac{dH}{d\mathbf{x}}\right|_{\mathbf{x}^b}. \tag{3.34}$$

H is the derivative of the forward model operator with respect to the model state vector and evaluated at the model background state. Thus we have performed a linearization of the nonlinear observation operator around the background state, implicitly assuming that the truth is not too far from the background. To form the analysis error covariance, multiply (3.33) by the transpose of itself and apply the expectation operator:

$$\mathbf{P}^a = \mathbf{P}^b + \mathbf{K}(\mathbf{R} + \mathbf{H}\mathbf{P}^b\mathbf{H}^{\mathrm{T}})\mathbf{K}^{\mathrm{T}} - \mathbf{K}\mathbf{H}\mathbf{P}^b - \mathbf{P}^b\mathbf{H}^{\mathrm{T}}\mathbf{K}^{\mathrm{T}}. \tag{3.35}$$

We now minimize the analysis error variance or trace of $\mathbf{P}^a$ with respect to the weight **K**. Thus

$$0 = \frac{d\mathrm{Tr}(\mathbf{P}^a)}{d\mathbf{K}} = 2\mathbf{K}(\mathbf{R} + \mathbf{H}\mathbf{P}^b\mathbf{H}^{\mathrm{T}}) - 2\mathbf{P}^b\mathbf{H}^{\mathrm{T}} \tag{3.36}$$

or, on solving for **K**:

$$\mathbf{K} = \mathbf{P}^b\mathbf{H}^{\mathrm{T}}(\mathbf{H}\mathbf{P}^b\mathbf{H}^{\mathrm{T}} + \mathbf{R})^{-1}. \tag{3.37}$$

This is the choice of weight that gives the minimum variance of the estimate. Substituting (3.37) into (3.35) reveals the analysis error covariance for this optimal weight:

$$\mathbf{P}^a = (\mathbf{I} - \mathbf{K}\mathbf{H})\mathbf{P}^b. \tag{3.38}$$

In summary, the OI algorithm includes the analysis equation, (3.29), the weight, (3.37), and the analysis error covariance matrix, (3.35).

$$\boxed{\begin{array}{lcl}
\mathbf{x}^a &=& \mathbf{x}^b + \mathbf{K}[\mathbf{z} - H(\mathbf{x}^b)] \\
\mathbf{K} &=& \mathbf{P}^b\mathbf{H}^{\mathrm{T}}(\mathbf{H}\mathbf{P}^b\mathbf{H}^{\mathrm{T}} + \mathbf{R})^{-1} \\
\mathbf{P}^a &=& (\mathbf{I} - \mathbf{K}\mathbf{H})\mathbf{P}^b
\end{array}}$$

In OI, we linearly combine two sources of information, an observation vector and a background vector, according to their relative accuracies. The weight assigned to the observation increment (or innovation) is optimally determined to minimize the analysis error variance. However, the method is "optimal" only if we really know the error covariances involved. In reality, we will never know these. Although we can estimate the covariances, they are still only estimates and could be incorrect. Thus, in practice, the method is not optimal and for that reason is often called *Statistical Interpolation.*

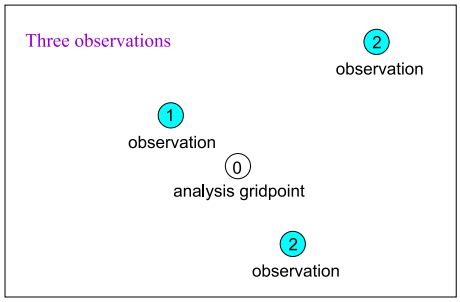

Figure 3.3: An example where three observations influence the analysis at gridpoint 0.

3.4 Example: 3 observations

Assume three observations of a variable, s, are distributed near an analysis grid point (denoted with subscript a). The observations have subscripts 1, 2 and 3. The analysis equation is

$$\mathrm{x}_0^a = \mathrm{x}_0^b + \left[\begin{array}{ccc} K_1 & K_2 & K_3 \end{array} \right] \left[\begin{array}{c} < z_1 - H(\mathbf{x}) > \\ < z_2 - H(\mathbf{x}) > \\ < z_3 - H(\mathbf{x}) > \end{array} \right]. \tag{3.39}$$

The weights assigned to the observations are given by:

$$\boxed{\mathbf{K}(\mathbf{H}\mathbf{P}^b\mathbf{H}^{\mathrm{T}} + \mathbf{R}) = \mathbf{P}^b\mathbf{H}^{\mathrm{T}}}$$

The observation and background errors are assumed to be unbiased. Let us assume that the instrument type is the same for each measurement and is uncorrelated in space, i.e. $\mathbf{R}$ is $(\sigma^r)^2\mathbf{I}$. (Because $\mathbf{R}$ includes representativeness error, we are also assuming that it is uncorrelated in space and has uniform variance.) Then the above becomes:

$$\left[\begin{array}{c} K_1 \\ K_2 \\ K_3 \end{array} \right]^{\mathrm{T}} \left\{ \left[\begin{array}{ccc} < \epsilon_1^b \epsilon_1^b > & < \epsilon_1^b \epsilon_2^b > & < \epsilon_1^b \epsilon_3^b > \\ < \epsilon_2^b \epsilon_1^b > & < \epsilon_2^b \epsilon_2^b > & < \epsilon_2^b \epsilon_3^b > \\ < \epsilon_3^b \epsilon_1^b > & < \epsilon_3^b \epsilon_2^b > & < \epsilon_3^b \epsilon_3^b > \end{array} \right] + \mathbf{I}(\sigma^r)^2 \right\} = \left[\begin{array}{c} < \epsilon_1^b \epsilon_0^b > \\ < \epsilon_2^b \epsilon_0^b > \\ < \epsilon_3^b \epsilon_0^b > \end{array} \right]$$

where superscripts b, r, a refer to background, observed and analysis variables. The subscript 0 denotes the analysis location. The first term in curly brackets is the background error covariance matrix evaluated at observation locations 1, 2 and 3. The diagonal terms are variances and the off-diagonals are covariances. Note that if any observation locations coincide, this matrix becomes singular. Then, if the observation error is small (observations are very accurate), $\mathbf{H}\mathbf{P}^b\mathbf{H}^{\mathrm{T}} + \mathbf{R}$ is very nearly singular and difficult to invert. Therefore, to avoid this problem, one can choose observations that are not (nearly) collocated, or combine co-located observations into "superobs" using the same analysis equation and ending up with a reduced observation error for the "superob".

In this example, we are starting to get an idea of how the weights are solved for, in practice. A matrix of size $m \times m$ must be inverted. The components of the matrix to invert include $\mathbf{H}\mathbf{P}^b\mathbf{H}^{\mathrm{T}}$ which is the background error covariance matrix evaluated at observation locations, and $\mathbf{R}$, the observation error covariance matrix. Since the inception of numerical weather prediction, the

backbone of the observing network has been the radiosonde network. The sondes are launched every 6 or 12 hours at primarily land stations using a helium filled balloon. The instrument package is lost after launch. Each station launches one of a few kinds of sondes, whose error characteristics are known. Thus, the observation error at two different locations, because the instruments are different, is not correlated. On the other hand, satellite observations are often averages of atmospheric variables over a footprint or line of sight. Thus, horizontal observation error correlations are possible. Nevertheless, because radiosondes were the basis of the observing network for weather forecasting, it is often assumed that all observations are horizontally uncorrelated, and thus the $\mathbf{R}$ matrix is diagonal. If observation errors are horizontally correlated, they must be thinned or averaged over the correlation length scale to produce uncorrelated observations. If $\mathbf{P}^b$ were also diagonal, then it would be easy to invert $\mathbf{H}\mathbf{P}^b\mathbf{H}^{\mathrm{T}} + \mathbf{R}$. However, this is not the case. $\mathbf{P}^b$ is in general a full matrix. What does this matrix look like, and how is it estimated? We consider these questions in the next section.

3.5 Background Error Covariance Matrix

The background error covariance is by definition:

$$\mathbf{P}^b =< (\mathbf{x}^b - \mathbf{x}^t)(\mathbf{x}^b - \mathbf{x}^t)^{\mathrm{T}} > . \tag{3.40}$$

Note that we are assuming that the background errors are unbiased. If they are biased, we can redefine a new variable with this bias subtracted out. For a state $\mathbf{x}$ of dimension $n = 10^7$, this matrix is $10^7 \times 10^7$. If we typically have an observation vector of size $m = 10^5$, it is clear that we cannot estimate the elements of $\mathbf{P}^b$ based on the observations (as a proxy for the truth). Moreover, to estimate even 1 element statistically requires more than 1 observation so it is really impossible to determine each of the 10^{14} elements of $\mathbf{P}^b$. Thus, the best that we can hope for is to *model* the covariance and then estimate the parameters of this model. The more simplistic the model, the fewer the parameters to estimate, and the better we can estimate them with observations. The problem is that a simple model is unlikely to be valid. Yet the data assimilation algorithm crucially depends on these statistics for its optimality.

We have no choice but to model the background error covariance matrix. Let's look at how this was done in the past with NWP (Numerical Weather Prediction) OI schemes and how valid the simplifying assumptions were.

3.5.1 Horizontal correlations

The covariance matrix for a 3-dimensional meteorological field discretized to gridpoints can be simplified if the vertical and horizontal structures are separable. That means that

$$C^b(x_i, y_i, z_i, x_j, y_j, z_j) = C^b_H(x_i, y_i, x_j, y_j)C^b_V(z_i, z_j)$$

where C^b_H and C^b_V are the horizontal and vertical covariances. Let us first consider only the horizontal covariances for a variable $\mathbf{x}$. The horizontal covariances depend upon the location of two points, i and j, on a 2D surface. Thus the covariances depend upon 4 parameters. To simplify this, we can assume that the horizontal covariance is **homogeneous**. This means that the covariance depends only upon the distance between the two points i and j. In 2 dimensions, this distance

vector can be written as a function of only two parameters, distance and angle. Thus we could write,

$$C_H^b(x_i, y_i, x_j, y_j) \approx C_x^b(r, \theta)$$

where $r^2 = (x_i - x_j)^2 + (y_i - y_j)^2$ and $\tan\theta = (y_i - y_j)/((x_i - x_j)$. Note that we are modelling a discrete matrix by evaluating a continuous function at discrete locations. Thus r and θ are actually continuous variables (not discrete ones). Recall that the variances are the diagonals of the covariance matrix, i.e. when i=j. When $i = j$, $r = 0$ so that $C_x^b(0, \theta)$ determines the variance. Thus for homogeneous covariances, the variances must be *independent* of location.

An additional assumption that we could make is that the covariance is **isotropic**. This means that the background error covariance does not depend upon direction either. Then,

$$C_H^b(x_i, y_i, x_j, y_j) \approx C_x^b(r).$$

As in the homogeneous case, the variance is given by $C_x^b(0)$ and is independent of location.

Are these reasonable assumptions? Fig. 4.13 of Daley (1991) shows that the standard deviation (square root of the variance) of the 250 mb geopotential background error is not constant over North America. Then the horizontal covariance for geopotential background error is not homogeneous. A less restrictive assumption is to assume that only the *correlations* are homogeneous or isotropic. The correlation matrix is defined by:

$$\boldsymbol{\rho}^b = \frac{C(x_i, y_i, x_j, y_j)}{\sigma(x_i, y_i)\sigma(x_j, y_j)} \tag{3.41}$$

or

$$\mathbf{P}^b = \mathbf{D}\boldsymbol{\rho}^b\mathbf{D}$$

where $\mathbf{D}$ is a diagonal matrix of standard deviations. The standard deviations exist for each point on the grid and each variable. Now are the correlations reasonably homogeneous and isotropic? Fig. 4.2 of Daley (1991) shows that the 500 mb geopotential background error correlations are not isotropic. The contours are skewed to look more like ellipses than circles. Nevertheless, the field is reasonably isotropic.

Now if we make these simplifying assumptions of homogeneity and isotropy, we can model the background error covariance matrix. Let us assume that we have observations which are linearly related to model variables (such as radiosonde measurements of temperature and wind). Then the observation operator, H, is linear and $[\mathbf{z} - H(\mathbf{x}^b)] = [\mathbf{v} - \mathbf{H}\mathbf{e}^b]$. Recall that $\mathbf{z} - H(\mathbf{x}^b)$ is called the *innovation* vector and represents the difference between observed and model variables at observation locations. Now consider the covariance of the innovations assuming no correlations between background and observation errors.

$$\begin{aligned} < (\mathbf{v} - \mathbf{H}\mathbf{e}^b)(\mathbf{v} - \mathbf{H}\mathbf{e}^b)^{\mathrm{T}} > &= < (\mathbf{v})(\mathbf{v})^{\mathrm{T}} > + \mathbf{H} < \mathbf{e}^b\mathbf{e}^{b^{\mathrm{T}}} > \mathbf{H}^{\mathrm{T}} \\ &- \mathbf{H} < \mathbf{e}^b(\mathbf{v})^{\mathrm{T}} > - < \mathbf{v}(\mathbf{e}^b)^{\mathrm{T}} > \mathbf{H}^{\mathrm{T}} \\ &= \mathbf{R} + \mathbf{H}\mathbf{P}^b\mathbf{H}^{\mathrm{T}} \end{aligned}$$

Now assume that all observations are of the same type, (i.e. radiosondes) so that the instrument and representativeness errors are the same at all observation locations. Thus the diagonal elements of $\mathbf{R}$ are identical. If the observation error is also horizontally uncorrelated (as in the case of

radiosondes), then $\mathbf{R}$ is a diagonal matrix and we can write $\mathbf{R} = (\sigma^r)^2\mathbf{I}$. The above can then be simplified to:

$$< (\mathbf{v} - \mathbf{He}^b)(\mathbf{v} - \mathbf{He}^b)^{\mathrm{T}} > \quad = \quad (\sigma^r)^2\mathbf{I} + \mathbf{HP}^b\mathbf{H}^{\mathrm{T}}. \tag{3.42}$$

Now with the homogeneity and isotropy assumptions, we can gather statistics of innovations from data assimilation cycles. For example we can accumulate innovations for all radiosonde station pairs over North America and bin them according to separation distance. Fig. 4.3 of Daley (1991) shows an example of such an exercise. We would like to determine a continuous correlation function that fits the points. Clearly there is a lot of scatter so it won't be easy to uniquely fit a function to the data. What kind of function should we choose? We want a function that will be a valid correlation function, so it must have certain properties.

Consider a correlation function in 1D and assume homogeneity. Then the spectrum of the correlation function, ρ is given by the Fourier transform:

$$g(m) = \frac{1}{\pi L}\int_0^\infty \rho(x)\cos(mx)dx \tag{3.43}$$

with

$$\rho(x) = 2L\int_0^\infty g(m)\cos(mx)dm.$$

$g(m)$ is the **spectral density function**. L is a distance (correlation length scale) to be defined later and m is a wavenumber. Multiplying the above by the variance (where $cov(x) = \sigma^2\rho(x)$) and evaluating at $x = 0$ gives

$$\sigma^2 = \int_0^\infty 2L\sigma^2 g(m)dm.$$

Thus $2L\sigma^2 g(m)dm$ represents the variance in the spectral interval between m and $m + dm$. Also, because of the homogeneity assumption, $\rho(x) = \rho(-x)$, so ρ is symmetric about 0. Then $g(m)$ is symmetric about $m = 0$. A very important theorem in stochastic processes or random field theory is the Wiener- Khinchine theorem (see Todling (1999) ch. 2.6). This theorem states that the spectral density function of an autocorrelation function must be real, continuous and positive. Additionally, the correlation matrix formed by evaluating the correlation function at discrete locations is strictly positive definite and has real eigenvalues providing the locations are distinct. In the 1-D case, we can easily show that positivity of the spectrum corresponds to an autocorrelation function that decreases as x increases.

$$\begin{aligned} |\rho(x)| &= \left|2L\int_0^\infty g(m)\cos(mx)dm\right| \le 2L\int_0^\infty |g(m)||\cos(mx)|dm \\ &= 2L\int_0^\infty |g(m)|dm = \rho(0). \end{aligned} \tag{3.44}$$

Fig. 3.4 of Daley (1991) illustrates the following correlation model:

$$\rho(r) = \left[\cos(cr) + \frac{\sin(cr)}{Lc}\right]e^{-r/L} \tag{3.45}$$

where c and L are specified constants. This correlation was defined for climatological data. Note that the correlation can be negative for some values. However the spectral density must be strictly

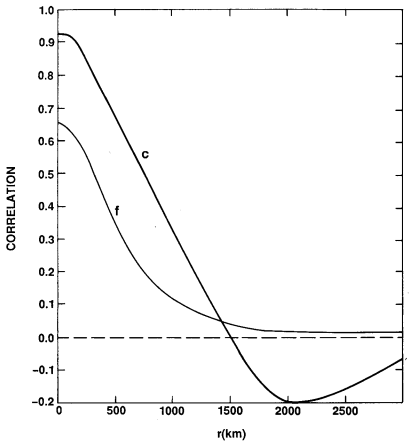

Figure 3.4: Observation-minus-background correlation for the 500 mb geopotential as a function of distance between stations. Curve c is for a climatological background and curve f is for a forecast background. Adapted from Schlatter, *Mon. Wea. Rev.* **103**: 246, 1975. The American Meteorological Society. (From Daley 1991, Fig. 4.4).

positive. For background states coming from short term NWP forecasts, a more appropriate model is the above in the limit that c goes to zero:

$$\rho(r) = \left[1 + \frac{r}{L}\right] e^{-r/L}. \tag{3.46}$$

This is curve f of Daley's Fig. 3.4. This curve remains positive. Another correlation model that has been employed is a Gaussian function:

$$\rho(r) = \exp\left(\frac{-r^2}{2L^2}\right). \tag{3.47}$$

Now let us define the length scale L of (3.43). In the 1D homogeneous case, it is

$$L^2 = -\left.\frac{\rho(x)}{d^2\rho/dx^2}\right|_{x=0} = \frac{\int_0^\infty g(m)dm}{\int_0^\infty g(m)m^2dm}. \tag{3.48}$$

Note that $g(m)$ is positive for an autocorrelation function. Then $m^2g(m)$ is also positive so L^2 must be positive. Also, the second derivative of $\rho(x)$ must be negative at $x = 0$. Recall that $\rho(x)$ is symmetric about $x = 0$ so that the first derivative should be zero. Thus the autocorrelation function is approximately parabolic near x=0. L roughly gives an idea of the inverse curvature of $\rho(x)$ at the origin. For a sharp function, the curvature is high and the length scale would be small. Similarly, if the curvature is low, the correlation function is wide and the length scale would be large. This correlation length scale gives us an idea of the distance over which the influence of the observation extends. Variables such as temperature tend to have larger correlation length scales than those such as wind. This means that background (forecast) errors of temperature have more

energy at larger scales than wind background errors. In two dimensions, the characteristic length is given by

$$L^2 = -\frac{2\rho(x)}{\nabla^2\rho}\bigg|_{\mathbf{r}=0}. \tag{3.49}$$

The factor 2 appears in the numerator because L corresponds to a 2D length scale. To see this note that under isotropic conditions, $L^2 = L_x^2 + L_y^2 = 2L_x^2$.

3.5.2 Vertical correlations

Thus far we've only considered the horizontal part. Assuming separability, we can estimate the vertical part, $C_V^b(z_i, z_j)$, by computing sample covariances between vertical levels averaged over all stations and all times (for a season). Vertical correlation functions computed for geopotential background error at ECMWF are shown in Daley's Fig. 4.8. Note that a curve such as that in the lower left panel shows the correlation between the background error at 400 mb and that at other levels. What this tells us is that an observation at 400 mb will receive a good weight for the analysis of geopotential at 400 mb. However, the observation will have a slightly smaller weight for the analysis at 500 or 300 mb. The correlation becomes 0.4 with a level near the surface so the observation would influence this analysis only a little. The bottom right panel shows the influence of an observation at the surface. The main influence is for levels below 500 mb. Surface observations on average have little influence in the upper troposphere. This is because the planetary boundary layer confines the flow and the influence of the observations to lower levels. Consequently, although the vast majority of in situ observations are taken at the surface, it is difficult to use these observations in an assimilation.

Comparing the curves in Fig. 4.8 of Daley (1991) reveals that they are not the same (apart from a vertical displacement). Thus, the homogeneity assumption is not valid for vertical correlations, and was not made operationally. Instead, correlation functions for each level were computed.

3.5.3 Optimality of the scheme

Remember that Optimal Interpolation is only optimal if we really know the error statistics. We have seen that we've made a number of assumptions about the error statistics such as:

1. no correlation between background and observation errors,
2. no horizontal correlation of observation (measurement and representativeness) errors,
3. homogeneous and isotropic horizontal correlations,
4. separability of vertical and horizontal correlations.

Are these assumptions reasonable? If not, then our statistics are not correct and the OI scheme will not be optimal. Moreover, the analysis error estimate based on (3.38) will be wrong.

For satellite observations, horizontal error correlations can occur.

Daley's Fig. 4.5 shows that the isotropic component of the horizontal correlation function for geopotential background error at 500 mb changes with location. Thus, the horizontal correlations are not homogeneous. The most important trend is the flattening of the curves (meaning longer correlation lengths) in the tropics. The correlation length scale increases as the equator is approached.

Daley's Fig. 4.2 shows that the 500 mb background error correlation for geopotential is not isotropic.

If the correlations are really separable, then we should be able to use a single horizontal correlation function (and length scale) at each vertical level. Daley's Fig. 4.12 plots the vertical variation of correlation length scale and it is not constant. In fact it is roughly constant until 300 mb, increasing with height thereafter.

Thus, none of our assumptions are really correct. Nevertheless, we had to make some assumptions in order to reduce the number of parameters used to define a background error covariance model to a number small enough to be estimated from the observations. By modelling the background error covariance matrix by a continuous function in the horizontal, we need to estimate only the correlation length scale L and perhaps one or two more constants. The vertical correlations, being non-homogeneous will need to be estimated for each pair of vertical levels. Thus, the total number of parameters is small enough to be estimated from innovation statistics. Since we are forced to make some assumptions which are not really correct, the OI can never be optimal. Thus, because of such approximations, in practice, the algorithm is called Statistical interpolation. The analysis error will not be optimal but can be calculated using (3.35) or, on rewriting (3.35):

$$\mathbf{P}^a = \mathbf{KRK}^{\mathrm{T}} + (\mathbf{I} - \mathbf{KH})\mathbf{P}^b(\mathbf{I} - \mathbf{KH})^{\mathrm{T}}. \tag{3.50}$$

3.6 Multivariate Analyses

Thus far, we have only considered the background error covariances for a single variable such as temperature or wind (u or v) or ozone. By doing a multivariate analysis, we can allow observations of one variable to influence the analysis of another variable. Thus observations of temperature could be used to improve the wind analysis if we had a linear relationship between temperature and wind increments. The balances must be linear because they are applied through the modelling of the background error covariance matrix.

Let us examine how the matrices are configured for the multivariate problem. First, for simplicity, assume that the analysis and observation locations coincide ($\mathbf{H} = \mathbf{I}$) and define

$$X_i = (p_i, u_i, v_i)^{\mathrm{T}}$$

For one observation location the analysis equation is:

$$\begin{bmatrix} p_i^a \\ u_i^a \\ v_i^a \end{bmatrix} = \begin{bmatrix} p_i^b \\ u_i^b \\ v_i^b \end{bmatrix} + \begin{bmatrix} w_{pp} & w_{pu} & w_{pv} \\ w_{up} & w_{uu} & w_{uv} \\ w_{vp} & w_{vu} & w_{vv} \end{bmatrix} \begin{bmatrix} p^o - p^b \\ u^o - u^b \\ v^o - v^b \end{bmatrix}.$$

For K observation locations, define:

$$\begin{aligned} \mathbf{x}^{\mathrm{T}} &= [X_1^{\mathrm{T}}, X_2^{\mathrm{T}}, \ldots, X_K^{\mathrm{T}}] \\ &= [p_1, u_1, v_1, p_2, u_2, v_2, \ldots, p_K, u_K, v_K]. \end{aligned}$$

Then the analysis equation is

$$X_i^a = X_i^b + \mathbf{K}_i[\mathbf{x}^o - \mathbf{x}^b] \tag{3.51}$$

where

$$\mathbf{K}_i = [W_{i1}, W_{i2}, \ldots, W_{iK}]$$

and where

$$W_{ik} = \begin{bmatrix} w_{pp}(\mathbf{r}_i, \mathbf{r}_k) & w_{pu}(\mathbf{r}_i, \mathbf{r}_k) & w_{pv}(\mathbf{r}_i, \mathbf{r}_k) \\ w_{up}(\mathbf{r}_i, \mathbf{r}_k) & w_{uu}(\mathbf{r}_i, \mathbf{r}_k) & w_{uv}(\mathbf{r}_i, \mathbf{r}_k) \\ w_{vp}(\mathbf{r}_i, \mathbf{r}_k) & w_{vu}(\mathbf{r}_i, \mathbf{r}_k) & w_{vv}(\mathbf{r}_i, \mathbf{r}_k) \end{bmatrix}.$$

Note that the size of the model state and observed vectors in this example is $n = 3K$. The weights are solved using:

$$(\mathbf{P}^b + \mathbf{R})\mathbf{K}_i = B_i$$

with

$$\mathbf{P}^b = \begin{bmatrix} b_{11} & b_{12} & \dots & b_{1K} \\ b_{21} & b_{22} & \dots & b_{2K} \\ \vdots & \vdots & \dots & \vdots \\ b_{K1} & b_{K2} & \dots & b_{KK} \end{bmatrix}, b_{ij} = \begin{bmatrix} C_{pp}(\mathbf{r}_j, \mathbf{r}_k) & C_{pu}(\mathbf{r}_j, \mathbf{r}_k) & C_{pv}(\mathbf{r}_j, \mathbf{r}_k) \\ C_{up}(\mathbf{r}_j, \mathbf{r}_k) & C_{uu}(\mathbf{r}_j, \mathbf{r}_k) & C_{uv}(\mathbf{r}_j, \mathbf{r}_k) \\ C_{vp}(\mathbf{r}_j, \mathbf{r}_k) & C_{vu}(\mathbf{r}_j, \mathbf{r}_k) & C_{vv}(\mathbf{r}_j, \mathbf{r}_k) \end{bmatrix}$$

B_i is the ith column of $\mathbf{P}^b$. The analysis equation shows that all $3K$ observations can impact on the analysis of all 3 variables at gridpoint i. Thus the OI equation is as before only we've expanded our definitions of the matrices. The elements of the background error covariance matrix, $\mathbf{P}^b$, are now 3x3 submatrices. The submatrix, b_{ij} has 9 elements which describe the auto- or cross-covariances of the prognostic variables at that gridpoint. In this example, there are only 3 prognostic variables, a mass variable (pressure, temperature, geopotential, etc.) and two wind components.

What do these auto- and cross-correlation functions look like? Let us consider the covariance matrices involving all grid points. Consider, for example, $C_{pu}(\mathbf{r}_i, \mathbf{r}_j)$. Let us define two points on the sphere,

$$\mathbf{r}_i = (x_i, y_i), \qquad \mathbf{r}_j = (x_j, y_j).$$

It will be useful to also define

$$u_j = u(x_j, y_j) \qquad v_j = v(x_j, y_j)$$

in order to simplify the notation. Thus we can write

$$C_{pu}(\mathbf{r}_i, \mathbf{r}_j) =< p_i, u_j > .$$

Now we need to introduce some linear relationships. Until the 1990's simple balances such as geostrophy were used, so let's assume a geostrophic balance of innovations. Let's also assume our mass variable subscript, p, refers to geopotential. Then we are assuming:

$$fu = -\frac{\partial \phi}{\partial y}, \qquad fv = \frac{\partial \phi}{\partial x}. \tag{3.52}$$

We can write

$$\begin{aligned} C_{pu}(x_i, y_i, x_j, y_j) &= < p_i, u_j >= -\frac{1}{f} < \phi_i \frac{\partial \phi_j}{\partial y_j} > \\ &= -\frac{1}{f}\frac{\partial}{\partial y_j} < \phi_i \phi_j > . \end{aligned} \tag{3.53}$$

Why can we simply exchange the derivative and ensemble operators? To see this consider the definition of a derivative:

$$\begin{aligned} < \phi_i \frac{\partial \phi_j}{\partial y} > &= \left\langle \phi_i \lim_{\Delta y_i \to 0} \left[\frac{\phi(x_j, y_j + \Delta y_j) - \phi(x_j, y_j)}{\Delta y_j} \right] \right\rangle \\ &= \lim_{\Delta y_i \to 0} \left[\frac{< \phi_i \phi(x_j, y_j + \Delta y_j) > - < \phi_i \phi(x_j, y_j) >}{\Delta y_j} \right] \\ &= \frac{\partial}{\partial y_j} < \phi_i \phi_j > . \end{aligned} \tag{3.54}$$

Now let's introduce an assumption of homogeneity for further simplification. Define,

$$\tilde{\mathbf{r}}^2 = \tilde{x}^2 + \tilde{y}^2, \qquad \tilde{x} = x_i - x_j, \qquad \tilde{y} = y_i - y_j.$$

Then,

$$\frac{\partial}{\partial y_i} = \frac{\partial}{\partial \tilde{y}}, \qquad \frac{\partial}{\partial y_j} = -\frac{\partial}{\partial \tilde{y}}$$

and similarly for x derivatives. Now we can write our 9 covariances as

$$\begin{aligned} C_{pp}(\mathbf{r}_i, \mathbf{r}_j) &= < \phi_i \phi_j > \\ C_{pu}(\mathbf{r}_i, \mathbf{r}_j) &= -C_{up}(\mathbf{r}_i, \mathbf{r}_j) = \frac{1}{f} \frac{\partial}{\partial \tilde{y}} < \phi_i \phi_j > \\ C_{pv}(\mathbf{r}_i, \mathbf{r}_j) &= -C_{vp}(\mathbf{r}_i, \mathbf{r}_j) = -\frac{1}{f} \frac{\partial}{\partial \tilde{x}} < \phi_i \phi_j > \\ C_{uu}(\mathbf{r}_i, \mathbf{r}_j) &= -\frac{1}{f^2} \frac{\partial^2}{\partial \tilde{y}^2} < \phi_i \phi_j > \\ C_{vv}(\mathbf{r}_i, \mathbf{r}_j) &= -\frac{1}{f^2} \frac{\partial^2}{\partial \tilde{x}^2} < \phi_i \phi_j > \\ C_{uv}(\mathbf{r}_i, \mathbf{r}_j) &= C_{vu}(\mathbf{r}_i, \mathbf{r}_j) = \frac{1}{f^2} \frac{\partial^2}{\partial \tilde{x} \partial \tilde{y}} < \phi_i \phi_j > \end{aligned}$$

If we had modelled our autocorrelation for ϕ by $F(r)$, we could write

$$C_{pp}(\mathbf{r}_i, \mathbf{r}_j) =< \phi_i \phi_j >= E_p^2 F(r),$$

where E_p^2 is the variance of ϕ background error. Then, on introducing some shorthand notation for the derivatives as in the appendix of Mitchell et al. (1990):

$$\begin{aligned} \Gamma &= -[R + \tilde{y}^2 R^2] \\ \Delta &= -[R + \tilde{x}^2 R^2] \\ \Theta &= [\tilde{x} \tilde{y} R^2] \\ \Xi &= [\tilde{y} R] \\ \Pi &= [\tilde{x} R] \\ R &= \frac{1}{r} \frac{\partial}{\partial r}. \end{aligned} \tag{3.55}$$

On defining the relationship between streamfunction and geopotential variance,

$$E_p = fE_\psi,$$

we can write

$$\begin{aligned}
C_{pp}(\mathbf{r}_i, \mathbf{r}_j) &= E_p^2 F(r), \\
C_{pu}(\mathbf{r}_i, \mathbf{r}_j) &= -C_{up}(\mathbf{r}_i, \mathbf{r}_j) = E_p E_\psi \Xi[F(r)] \\
C_{pv}(\mathbf{r}_i, \mathbf{r}_j) &= -C_{vp}(\mathbf{r}_i, \mathbf{r}_j) = -E_p E_\psi \Pi[F(r)] \\
C_{uu}(\mathbf{r}_i, \mathbf{r}_j) &= E_\psi^2 \Gamma[F(r)] \\
C_{vv}(\mathbf{r}_i, \mathbf{r}_j) &= E_\psi^2 \Delta[F(r)] \\
C_{uv}(\mathbf{r}_i, \mathbf{r}_j) &= C_{vu}(\mathbf{r}_i, \mathbf{r}_j) = E_\psi^2 \Theta[F(r)].
\end{aligned}$$

These correlations are plotted in Mitchell et al. (1990) Fig. 17. The x and y axes of each plot is simply distance in the x and y direction from an observation located at the origin. The total domain of each of the 9 panels is 2500 km x 2500 km. In the top left corner, one can recognize the concentric circles for contours as being due to the homogeneous, isotropic modelling of the geopotential autocorrelation function. Thus, the maximum impact of an observation at the origin is on an analysis at the origin. The impact on other grid points radially distant from the observation decreases with increasing distance. For an analysis gridpoint 1250 km away from the observation, the background error correlation has dropped to less than 0.2.

The top middle panel shows $C_{pu}(\mathbf{r}_i, \mathbf{r}_j)$. One can obtain this picture by taking a derivative of the top left figure with respect to y. The way to interpret this is as follows. Suppose one has a wind observation at the origin. The biggest impact of the observation will be on the geopotential at grid points about 300 km directly to the north and south of the observation. The impact on the geopotential analysis at the observation location is nil. This is a direct consequence of the spatial derivatives used in the geostrophic assumption since wind is related to geopotential derivatives in the north-south direction. Note that the middle left panel is the negative of the top middle panel, as expected from our relationships developed above. A similar discussion ensues for the geopotential-v cross-correlations except that derivatives in the x-direction are involved.

The u-u autocorrelation is found in the middle panel. In the x-direction, the correlation decreases away from the plot's origin so that the impact of a u observation at the origin decreases with distance from the observation. On the other hand, as one proceeds north from the u observation at the origin, the correlation decreases and becomes negative. A secondary extrema is found about 700 km to the north of the observation. Thus the biggest impact of u observation on a u analysis is felt at the location of the observation. 700 km to the north and south, a smaller impact is felt. Again, this pattern is due to taking a second derivative in y of the pattern in the top left panel. Empirically, the small negative correlations are seen in data (e.g. Mitchell et al. 1990, Fig. 1b) and indicate an approximate geostrophic balance of innovations.

In this section we have seen that observations of 1 variable can impact the analysis of another variable, if the multivariate covariances indicate a cross-correlation between the two variables. In this section, a geostrophic balance was imposed in the horizontal. More complex balances can be imposed, but for 3-dimensional data assimilation schemes, these balances must be linear because they appear in a matrix ($\mathbf{P}^b$). Daley (1991) discusses the more general case where the divergent component of the wind is also analysed. The interested reader is referred to Daley's chapter 5 for a much more extensive discussion of multivariate analyses.

3.7 Statistical Interpolation in practice

OI for Numerical Weather Prediction is an intermittent data assimilation scheme. This means it is run at every synoptic hour (00, 06, 12, 18Z), four times a day, rather that inserting data continuously. Observations are binned into 6 hr intervals centered on the analysis time. Because the same background error covariance is used regardless of the analysis time step, all realizations of the background state at all times are assumed to have the same statistics. That means we are assuming that the background errors are stationary. The background error statistics are computed using innovations collected over 1-3 months and are thus stationary on this time scale. However, the statistics are computed for at least the 4 seasons , and sometimes monthly. Although the background error covariance matrix is assumed stationary, it changes from month to month. The product of the assimilation at any given time is a full analysis of each model prognostic variable on the model grid (or spectral coefficients). This is then used as an initial state for the integration of the forecast model.

The OI equations are given by (3.29), (3.37) and (3.35). As we shall see later on, the OI corresponds to the analysis step of the Kalman filter. (The Kalman filter does not assume stationary statistics, but rather, uses the model's own dynamics to propagate the forecast error covariance matrix in time.)

To solve for the weights, using (3.37), a matrix inversion is required. For a state vector of dimension n=10^7 and an observation vector of m=10^5, $\mathbf{HBH}^{\mathrm{T}} + \mathbf{R}$ is $m \times m$. Clearly, this matrix inversion is too expensive. So, some approximations have to be made to solve this problem, in practice. These are given below.

1. Assume the generalized interpolation, H, is linear. Then $\mathrm{H}(\mathbf{x}) = \mathbf{H}\ \mathbf{x}$. This means that observed variables must be linearly related to model variables. For indirect data such as radiances, a separate inversion processes is required.

2. $\mathbf{P}^b$ is continuous. $\mathbf{HP}^b\mathbf{H}^{\mathrm{T}}$ can then be evaluated at observation sites as an $m \times m$ matrix without ever needing to know $\mathbf{P}^b$ on the model grid, which is $N \times N$. Linear dynamical constraints can then be applied through modelling of $\mathbf{P}^b$. Typically, background errors are assumed to be geostrophic and hydrostatic.

3. Data selection is used so that the analysis equation is solved n times. Each equation is then solved for a scalar x^a. By limiting the number of observations that influence a given analysis point to p (<100), we can further reduce the size of $\mathbf{HP}^b\mathbf{H}^{\mathrm{T}}$ to $p \times p$. Thus the inversion of an $m \times m$ matrix has been replaced by n inversions of $p \times p$ matrices. If the background error autocorrelation drops to zero in finite distance (decorrelation length scale), then the correlation between background errors at two points separated by a length larger than the decorrelation length scale is zero. Since these correlations do in fact decrease with separation distance, it is reasonable to impose a cutoff radius beyond which observations are unimportant. However, a consequence of data selection and applying a cutoff radius, is that exact satisfaction of constraints applied through $\mathbf{P}^b$ is prevented because each analysis point can use a different set of observations. The constraints in $\mathbf{P}^b$ apply globally. Thus, smoothness of the analysis is not guaranteed even if a geostrophic, and hydrostatic assumption is made. As a result, the analysis will have to be filtered to prevent initial state imbalances from exciting spurious gravity wave activity and destroying the forecast.

3.8 Filtering Properties

In this final but rather important section, we consider the filtering properties of the OI scheme. We make the usual assumption of no horizontal correlations of observation errors. In this case, the spatial structure of the background error covariance matrix determines the filtering properties of the OI algorithm.

This discussion follows Daley (1991)'s section 4.5. The analysis equation is

$$\mathbf{x}^a = \mathbf{x}^b + \mathbf{P}\mathbf{H}^{\mathrm{T}}(\mathbf{H}\mathbf{P}\mathbf{H}^{\mathrm{T}} + \mathbf{R})^{-1}[\mathbf{z} - \mathbf{H}(\mathbf{x}^b)] \tag{3.56}$$

For simplicity of notation, the superscript, b, on the background error covariance matrix, $\mathbf{P}^b$ has been dropped. To focus on the filtering properties, let us eliminate the interpolation aspects of the analysis by assuming $\mathbf{H} = \mathbf{I}$. Then define $\mathbf{y} = [\mathbf{z} - \mathbf{H}(\mathbf{x}^b)]$. Then (3.56) can be written as:

$$\mathbf{x}^a - \mathbf{x}^b = \mathbf{d} = \mathbf{P}(\mathbf{P} + \mathbf{R})^{-1}\mathbf{y}$$

or

$$\mathbf{d} = \mathbf{A}\mathbf{y}$$

where

$$\mathbf{A} = \mathbf{P}(\mathbf{P} + \mathbf{R})^{-1} = (\mathbf{I} + \mathbf{R}\mathbf{P}^{-1})^{-1}.$$

Now let us simplify $\mathbf{A}$ further. Let $\mathbf{P} = (\sigma^b)^2\mathbf{C}$. Also assume that all observations are from the same instrument (and have the same representativeness error statistics). Then, $\mathbf{R} = (\sigma^r)^2\mathbf{I}$. Finally, define the eigenvalues and eigenvectors of $\mathbf{C}$ as λ and $\mathbf{e}$, i.e.,

$$\mathbf{C}\mathbf{e} = \lambda\mathbf{e}.$$

Then,

$$\mathbf{P}\mathbf{e} = (\sigma^b)^2\mathbf{C}\mathbf{e} = (\sigma^b)^2\lambda\mathbf{e}$$

and

$$\mathbf{R}\mathbf{e} = (\sigma^r)^2\mathbf{I}\mathbf{e} = (\sigma^r)^2\mathbf{e}$$

so that

$$\mathbf{R}\mathbf{P}^{-1}\mathbf{e} = \frac{(\sigma^r)^2}{(\sigma^b)^2\lambda}\mathbf{e}$$

and

$$(\mathbf{I} + \mathbf{R}\mathbf{P}^{-1})\mathbf{e} = (1 + \frac{(\sigma^r)^2}{(\sigma^b)^2\lambda})\mathbf{e}.$$

Finally, we see that,

$$\mathbf{A}\mathbf{e} = (\mathbf{I} + \mathbf{R}\mathbf{P}^{-1})^{-1}\mathbf{e} = \frac{1}{1 + \frac{\alpha}{\lambda}}\mathbf{e}$$

where

$$\alpha = \frac{(\sigma^r)^2}{(\sigma^b)^2}.$$

If the observation increment can be written as a superposition of eigenvectors of $\mathbf{C}$, i.e.,

$$y = \sum_{i=1}^{N} c_i\mathbf{e}_i$$

then

$$\mathbf{d} = \mathbf{A}\mathbf{y} = \sum_{i=1}^{N} c_i \mathbf{A}\mathbf{e}_i = \sum_{i=1}^{N} c_i \left(\frac{1}{1 + \frac{\alpha}{\lambda_i}} \right) \mathbf{e}_i$$

Now, if the eigenvalues of $\mathbf{C}$ are large, then $\lambda_i \gg \alpha$, and the term in round brackets goes to 1. Large eigenvalues typically correspond to large spatial scales. Thus large scales are not damped much. However, if the eigenvalues of $\mathbf{C}$ are small, then $\lambda_i \ll \alpha$, and the term in round brackets goes to zero. Thus small small spatial scales (small eigenvalues) are damped. The spectral structure of the correlation matrix for background errors determines the filtering properties of the analysis. Thus, if the background error correlation function has most energy at large scales, the OI will act as a low-pass filter. If the correlation function is dominated by energy at small scales, OI will act as a high-pass filter.

Appendix A: The Sherman-Morrison-Woodbury formula

The Sherman-Morrison-Woodbury formula is

$$(\mathbf{P}^{-1} + \mathbf{H}^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{H})^{-1}\mathbf{H}^{\mathrm{T}}\mathbf{R}^{-1} = \mathbf{P}\mathbf{H}^{\mathrm{T}}(\mathbf{H}\mathbf{P}\mathbf{H}^{\mathrm{T}} + \mathbf{R})^{-1} \tag{3.57}$$

There are many ways to prove this, all involving matrix multiplications. Here is one possible proof. Expand the left hand side:

$$\begin{aligned}
(\mathbf{P}^{-1} + \mathbf{H}^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{H})^{-1}\mathbf{H}^{\mathrm{T}}\mathbf{R}^{-1} &= (\mathbf{P}^{-1} + \mathbf{H}^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{H})^{-1}[\mathbf{R}(\mathbf{H}^{\mathrm{T}})^{-1}]^{-1} \\
&= [(\mathbf{R}(\mathbf{H}^{\mathrm{T}})^{-1})(\mathbf{P}^{-1} + \mathbf{H}^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{H})]^{-1} \\
&= [(\mathbf{R}(\mathbf{H}^{\mathrm{T}})^{-1}\mathbf{P}^{-1} + \mathbf{H}]^{-1} \\
&= [(\mathbf{R}(\mathbf{H}^{\mathrm{T}})^{-1} + \mathbf{H}\mathbf{P})\mathbf{P}^{-1}]^{-1} \\
&= [(\mathbf{R} + \mathbf{H}\mathbf{P}\mathbf{H}^{\mathrm{T}})(\mathbf{H}^{\mathrm{T}})^{-1}\mathbf{P}^{-1}]^{-1} \\
&= \mathbf{P}\mathbf{H}^{\mathrm{T}}(\mathbf{R} + \mathbf{H}\mathbf{P}\mathbf{H}^{\mathrm{T}})^{-1}
\end{aligned}$$

Appendix B: Proof of some derivative formula

Verify that

$$\frac{d\mathrm{Tr}(\mathbf{A}\mathbf{B})}{d\mathbf{A}} = \mathbf{B}^{\mathrm{T}}$$

Let $\mathbf{A}$ be $n \times r$ and $\mathbf{B}$ be $r \times n$ since $\mathbf{A}\mathbf{B}$ is symmetric. We can write the individual matrix elements as

$$(\mathbf{A}\mathbf{B})_{ij} = \sum_{k=1}^{r} a_{ik}b_{kj}.$$

Thus we can write the trace of this matrix as

$$\mathrm{Tr}(\mathbf{A}\mathbf{B}) = \sum_{i=1}^{n}(\mathbf{A}\mathbf{B})_{ii} = \sum_{i=1}^{n}\sum_{k=1}^{r} a_{ik}b_{ki}.$$

Now we can write that

$$\frac{d\mathrm{Tr}(\mathbf{A}\mathbf{B})}{d\mathbf{A}} = \frac{d}{da_{lm}}\left[\sum_{i=1}^{n}\sum_{k=1}^{r} a_{ik}b_{ki}\right] = b_{ml} = \mathbf{B}^{\mathrm{T}}.$$

Appendix C: Eigenvalues of covariance matrices

A covariance matrix is real, symmetric and positive definite. Since it is real, symmetric, its eigenvalue decomposition may be written as

$$\mathbf{A} = \mathbf{E}\mathbf{D}\mathbf{E}^{\mathrm{T}} \tag{3.58}$$

where $\mathbf{D}$ is a diagonal matrix of eigenvalues and $\mathbf{E}$ is a unitary matrix of eigenvectors. That is, $\mathbf{E}^{\mathrm{T}} = \mathbf{E}^{-1}$. Now if $\mathbf{A}$ is positive definite, then for all vectors, $\boldsymbol{x}$, we have that

$$\boldsymbol{x}^{\mathrm{T}}\mathbf{A}\boldsymbol{x} > 0.$$

We can substitute for $\mathbf{A}$ using (3.58):

$$\boldsymbol{x}^{\mathrm{T}}\mathbf{E}\mathbf{D}\mathbf{E}^{\mathrm{T}}\boldsymbol{x} = \boldsymbol{y}^{\mathrm{T}}\mathbf{D}\boldsymbol{y} > 0$$

where $\boldsymbol{y} = \mathbf{E}^{\mathrm{T}}\boldsymbol{x}$. Expanding this we have that

$$\sum_{i=1}^{n}(y_i)^2\lambda_i > 0. \tag{3.59}$$

But this must be true for all $\boldsymbol{x}$ and therefore for all $\boldsymbol{y}$. The only way to ensure this is when all eigenvalues, λ_i are positive. Another way to see this is to choose a particular $\boldsymbol{y}$ since (3.59) must hold for all $\boldsymbol{y}$. Choose $\boldsymbol{y}$ = (0,0,...,0,1,0,...,0), i.e. the vector with all 0 elements except for the ith element (which is a 1). For this choice of $\boldsymbol{y}$, (3.59) becomes

$$\lambda_i > 0.$$

This can be repeated for all i, $1 \leq i \leq n$.

The eigenvalues of a real, symmetric, positive definite matrix are real and positive. Thus covariance matrices have real and positive eigenvalues.

An excellent reference on eigenvalue problems is Wilkinson (1965).

REFERENCES

1. Daley, R., 1991: *Atmospheric Data Analysis.* Cambridge University Press. 457 pp.

2. Mitchell, H. L., C. Charette, C. Chouinard and B. Brasnett, 1990: Revised interpolation statistics for the Canadian data assimilation procedure: Their derivation and application. *Mon. Wea. Rev.*, **118**, 1591-1614.

3. Todling, R., 1999: Estimation Theory and Foundations of Atmospheric Data Assimilation, DAO Office Note 1999-01.

4. Wilkinson, J. H., 1965: *The Algebraic Eigenvalue Problem.* Oxford Clarendon, 662 pp.

3.9 Problem Set 3

1. Suppose you are lost at sea during the night and have no idea of your location. You take a star sighting to establish your position using two completely different instruments. For simplicity, let's assume that the position is a 1D variable. The observation from the first instrument is z_1 with error variance σ_1^2. The observation from the second instrument is z_2 with error variance σ_2^2. The observation errors from the two instruments are uncorrelated and each is unbiased. Use this information to obtain an optimal estimate of your position, x, where x is a scalar. Note that there is no background information available.

   (a) First form the scalar analysis equation. Then write this in terms of errors or departures from the truth. For an unbiased analysis, what constraint must be applied to the two weights? Note that
   $$\mathrm{e}_1^r = \mathrm{z}_1 - \mathrm{x}^t, \qquad \mathrm{e}_2^r = \mathrm{z}_2 - \mathrm{x}^t.$$

   (b) Form the equation for the analysis error variance, using the constraint derived above. Find the weights that minimize the analysis error variance. Write the analysis equation again, now with these weights. What is the analysis error variance with the optimal weights? How does this results relate to the scalar example in Ch. 1 section 1.5?

   (c) Now we will redo the problem using vector notation. The measurement equation is $\mathbf{z} = \mathbf{H}\mathrm{x}^t + \mathbf{v}$ where
   $$\mathbf{z} = \begin{pmatrix} \mathrm{z}_1 \\ \mathrm{z}_2 \end{pmatrix}, \mathbf{v} = \begin{pmatrix} \mathrm{e}_1^r \\ \mathrm{e}_1^r \end{pmatrix}, \mathbf{H} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$$
   What are the elements of $\mathbf{R}$, the observation error covariance matrix? What is the analysis equation in vector form? Form the analysis error equation by subtracting the truth from both sides. What is the vector form of the constraint that arises for an unbiased analysis error? Write the equation for the analysis error variance. Find the weights that minimize this equation. Note that because of the constraint, the weight matrix has only 1 independent variable. Therefore, if we write k_2 in terms of k_1, we only want to take the derivative of of $(\sigma^a)^2$ with respect to k_1. To use vector notation, then we need the rule (e.g. CRC Standard Mathematical Tables 10.2.1):
   $$\frac{\partial y}{\partial x} = Tr[\frac{\partial y}{\partial \mathbf{K}} \frac{\partial \mathbf{K}^{\mathrm{T}}}{\partial x}]$$

where y and x are scalars and where each element of $\mathbf{K}$ is a function of x.

2. Let us examine the case of two observations on a 1-D grid from section 3.2. As in the text, obs 1 is located at x/L=-2.0, the analysis is done at x/L=0 and obs 2 varies over x/L=[-4,4]. Also, $\alpha = (\sigma^r/\sigma^b)^2$=0.25 and

$$\rho^b(\Delta x) = \left(1 + \frac{|\Delta x|}{L}\right) \exp\left(-\frac{|\Delta x|}{L}\right).$$

Obtain MATLAB script **prob3p2.m**. This script plots the weights w_1 and w_2 from (3.20) and (3.20) as well as the normalized analysis error for (3.22). Running **prob3p2.m** will reproduce Fig. 4.7 of Daley (1991). In this problem, we consider how changes to the free parameters of this system, the position of obs 1 and α affect the results.

i) Consider the impact of changing α on the results. Try for example, α=0.1,0.25,0.5,1.0 .

(a) What happens to the analysis error as α increases? Why?

(b) What happens to the phenomenon of observation screening?

ii) Consider how well the results vary when obs 1 is moved. Plot the results when α=0.25 and obs 1 is located at x/L=-3.0, -2.0, -2.0, 0.0 .

(a) When obs 1 is located at x1, for what positions of obs 2 are the weights equal?

(b) What locations of obs 2 have no effect on the analysis?

(c) What is the best location for obs 1? For this location, over what length scale does obs 2 influence the analysis? Why?

3. In this problem we generate a correlation matrix for a specific grid and examine its properties. Consider the interval $(-L_x, L_x]$ and divide it into J gridpoints. Consider the homogeneous and isotropic, Gaussian correlation function in 1D:

$$\rho(x, y) = \rho(r = |x - y|) = \exp(-\frac{1}{2}(x - y)^2/L_d^2). \tag{3.60}$$

r is the distance between two points in the domain and L_d is the decorrelation length. Points in the discrete domain are defined by

$$x_j = j\Delta x$$

where $\Delta x = 2L_x/J$ for $j \in \{-J/2 + 1, J/2\}$, and the elements of the homogeneous, isotropic correlation matrix, $\mathbf{B}$ are given by

$$B_{ij} = \rho(x_i, y_j).$$

a) Construct a MATLAB function that returns the correlation matrix $\mathbf{B}$, given the half-length of the domain, L_x, the number of gridpoints J, and the decorrelation length, L_d. For $(L_x, L_d, J) = (1, 0.2, 32)$, compute $\mathbf{B}$ using this function. Make a contour plot of the correlation array. (Note: use meshgrid to generate this matrix. Don't forget to use ".*" or ".hat 2" for matrix multiplys. "hat" is the symbol on the 6 key. I couldn't get LaTeX to print out a "hat" symbol.)

b) For the parameters of part (a), plot the correlation functions at the following two specific locations: $x_j \in \{0, L_x\}$. There is a MATLAB function plot.

c) Is the $\mathbf{B}$ obtained above an *acceptable* correlation matrix? Why or why not? Hint: check its eigenvalues. MATLAB has a function called eig.

d) From the figures constructed in parts (a) and (b), we see that the correlations decrease very quickly toward zero. What if we actually set some of these small value to zero in order to save some storage space? Without actually changing the storage of the matrix, define a new matrix $\mathbf{B}_c$ by setting elements corresponding to $|r| > L_c$ to zero. Use the same parameters as in (a) and a cutoff value of $L_c = 3L_d$. Make a contour plot, and plot the correlation function at two locations as in part (b). Is $\mathbf{B}_c$ an acceptable correlation matrix?

e) Repeat parts (a) to (c) for the Triangular correlation function,

$$T(x,y) = \begin{cases} 1 - |x-y|/L_c, & \text{for} |x-y| \le L_c \\ 0 & \text{otherwise} \end{cases}$$

f) Construct a matrix $\mathbf{Q}$ as the Hadamard product of the matrices $\mathbf{B}$ and $\mathbf{T}$ by multiplying the matrices element by element:

$$\mathbf{Q} = \mathbf{B} \circ \mathbf{T} = [B_{ij} T_{ij}].$$

MATLAB can do this product trivially ($Q = B. * T$). Make a contour plot of $\mathbf{Q}$. Is $\mathbf{Q}$ an acceptable correlation matrix? Plot the correlation functions from $\mathbf{Q}$, $\mathbf{B}$ and $\mathbf{T}$ on the same graph, for x=0.

4. Let us now consider how we can generate an ensemble of vectors that have a specified covariance matrix. That is, we want to find vectors, $\mathbf{w}$ such that

$$E(\mathbf{w}\mathbf{w}^{\mathrm{T}}) = \mathbf{Q}$$

where $\mathbf{Q}$ is specified and can be a full matrix.

First assume we can generate N samples, ϵ_i, $i = 1, 2, \ldots, N$ from a normal distribution $\mathcal{N}(0, \sigma^2)$. Put these into a vector called $\mathbf{v} = (\epsilon_1, \epsilon_2, \ldots, \epsilon_N)^{\mathrm{T}}$. Note that the sample covariance matrix for the vectors so formed should be:

$$\begin{aligned} < \mathbf{v}\mathbf{v}^{\mathrm{T}} > &= < \begin{pmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_N \end{pmatrix} (\epsilon_1 \quad \epsilon_2 \quad \ldots \quad \epsilon_N) > = \begin{bmatrix} < \epsilon_1\epsilon_1 > & < \epsilon_1\epsilon_2 > & \ldots & < \epsilon_1\epsilon_N > \\ < \epsilon_2\epsilon_1 > & < \epsilon_2\epsilon_2 > & \ldots & < \epsilon_2\epsilon_N > \\ \vdots & \vdots & \ddots & \vdots \\ < \epsilon_N\epsilon_1 > & < \epsilon_N\epsilon_2 > & \ldots & < \epsilon_N\epsilon_N > \end{bmatrix} \\ &= \begin{pmatrix} \sigma^2 & 0 & \ldots & 0 \\ 0 & \sigma^2 & \ldots & 0 \\ 0 & 0 & \ddots & 0 \\ 0 & 0 & \ldots & \sigma^2 \end{pmatrix} = \sigma^2 \mathbf{I}. \end{aligned} \tag{3.61}$$

a) What if each element was drawn from a different Normal distribution, $\mathcal{N}(0, \sigma_i^2)$? What form does the sample covariance matrix, $\mathbf{R} = < \mathbf{v}\mathbf{v}^{\mathrm{T}} >$ take now?

b) Now consider an arbitrary covariance matrix, $\mathbf{Q}$, where $\mathbf{Q} = \mathbf{L}\mathbf{L}^{\mathrm{T}}$. What is the covariance matrix for $\mathbf{L}\mathbf{v}$ if $< \mathbf{v}\mathbf{v}^{\mathrm{T}} >= \mathbf{I}$? i.e. elements of $\mathbf{v}$ are drawn from a $\mathcal{N}(0,1)$ distribution.

c) Finally consider

$$\mathbf{Q} = \mathbf{E}\mathbf{D}\mathbf{E}^{\mathrm{T}}.$$

How would you generate vectors, $\mathbf{w}$, having the covariance matrix $\mathbf{Q}$, given an ensemble of vectors, $\mathbf{v}$, where $\mathbf{v}$ are drawn from a $\mathcal{N}(0,1)$ distribution?

d) Using MATLAB, contour plot sample correlation matrices for $\mathbf{Q}$ where the correlation function is

$$\rho(r = |x - y|) = \exp(-\frac{1}{2}(x - y)^2/L_d^2)$$

for $L_d = 0.2$. Use the same parameters as in prob. 3 for the mesh and other constants. Assume variances are all equal to 1. Use 100, 1000 and 1000 samples in the ensemble. Now contour plot the Identity matrix for the same 32-dimensional space using 1000 samples. How does this plot differ from that for $\mathbf{Q}$? What assumption is really being made when an Identity covariance matrix is used?

5. Filtering properties. In this problem, we are going to examine the structure of the weight matrix and verify the assertion of section 3.8 that the filtering of observation increments is controlled by the background error covariance matrix. Start by defining a grid as in prob. 4.3. Consider the interval $(-L_x, L_x]$ and divide it into J gridpoints. We will use the homogeneous and isotropic, Gaussian correlation function in 1D:

$$\rho(|x - y|) = \exp(-\frac{1}{2}(x - y)^2/L_d^2). \tag{3.62}$$

r is the distance between two points in the domain and L_d is the decorrelation length. Points in the discrete domain are defined by

$$x_j = j\Delta x$$

where $\Delta x = 2L_x/J$ for $j \in \{-J/2+1, J/2\}$. Form the the homogeneous, isotropic correlation matrix, $\mathbf{Q}$ using

$$\mathbf{Q}_{ij} = \rho(x_i, y_j).$$

(a) Find the eigenvalues and eigenvectors of $\mathbf{Q}$. Plot the eigenvalues as a function of index. Plot the eigenvectors corresponding to the 6 largest eigenvalues, as a function of x. Draw a zero-line and count the number of zero crossings for each eigenvector plotted. (Type **help eig** for help on how to use MATLAB's eigenvalue function.)

(b) Now compute

$$\mathbf{A} = \mathbf{B}(\mathbf{B} + \mathbf{R})^{-1}$$

where $\mathbf{R} = (\sigma^r)^2\mathbf{I}$, $\mathbf{B} = \mathbf{D}\mathbf{Q}\mathbf{D}$ and $\mathbf{D}$ is a diagonal matrix with each diagonal element equal to σ^b. Let σ^r=1 and σ^b=2. Compute the eigenvalues and eigenvectors of $\mathbf{A}$ making sure to sort by magnitude. Compare the eigenvalues of $\mathbf{A}$ to $(1 + \alpha/\lambda_q)^{-1}$ when $\alpha = (\sigma^r/\sigma^b)^2$ as before and λ_q is an eigenvalue of $\mathbf{Q}$. Plot the eigenvectors of $\mathbf{A}$. Are they the same as those for $\mathbf{Q}$?

(c) Now let's create some observation increments. To avoid interpolation, place an obs at every grid point. To see the filtering aspects, define an obs increment by

$$y = \cos(c\pi x)$$

where the x-grid values are from (-1,1]. Create an obs increment vector, and compute the analysis increment using

$$\mathbf{d} = \mathbf{A}\mathbf{y}.$$

Plot $\mathbf{y}$, $\mathbf{d}$ versus x for various values of c. Which waves are filtered the least? For which waves is the amplitude dropped by 80% or more?

(d) How are filtering properties affected by a change in L_d?

6. Linear advection equation model. You must obtain the following MATLAB scripts: {oi, gauss, gcorr, getpsi, sqrwv, obspat, upwind}.m . In this problem, we will run an optimal interpolation scheme for a forecast model which is basically a passive tracer advection in a 1D periodic domain. The forecast model is simply

$$\frac{\partial u}{\partial t} + U\frac{\partial u}{\partial x} = 0. \tag{3.63}$$

The x domain is [-2,2] and is periodic. The initial condition is a rectangular wave of the form:

$$u(x, t = 0) = \begin{cases} 1, & 1 \le x \le 0 \\ 0, & \text{otherwise} \end{cases} \tag{3.64}$$

Using an upwind finite difference scheme, we can write the solution as

$$u_j = Cu_{j-1} + (1 - C)u_j$$

where $C = U\Delta t/\Delta x$ is the Courant number and v_j is the numerical solution for $u(x = j\Delta x)$. Δx and Δt are the gridspacing and time step, respectively.

a) Simulation experiments. Let us first examine the forecast model. Run the model alone (no data assimilation) by typing `oi(0,1,1,1)`. This provides a Courant number of 1 and integrations from $T_o = 0$ to T_{final}=1. Obtain a plot of the initial and final states. Repeat this exercise for different Courant numbers of 0.95, 0.9 and 0.25. What is happening to the solution as the Courant number decreases?

b) Now we will run an OI. We simulate the truth by running the forecast model. To simulate the observations, we perturb the truth by the observation error variance:

$$\mathbf{z} = \mathbf{H}\mathbf{x} + \mathbf{v}.$$

The observation network will be simple. Observations are available every kth gridpoint on either the left half of the domain, or the whole domain. In time, observations are available every n timesteps. The analysis is generated whenever there is data, every n timesteps. Otherwise, the analysis reverts to the background state. Once a new analysis is obtained, the model is integrated forward, by another time step. Thus, we have an intermittent assimilation scheme.

(i) Run the OI scheme by typing `oi(0,1,0.95,0)`. The Courant number will now be fixed at 0.95 and T_{final}=1 again. Three questions will be asked. Hitting "return" will give the default. First enter observation frequency of 5 (obs every 5 time steps), an observation sparsity of 1 (obs at every gridpoint), and "return" for the obs error standard deviation. This will give the default value of 0.02. How does the analysis compare with the truth? Does the error estimate make sense?

(ii) Now let's make the problem a little harder. Again type `oi(0,1,0.95,0)`, but provide the an obs error of 0.2. Keep the obs frequency of 5 and the obs sparsity of 1. What happened to the analysis and the error estimate?

(iii) Now let's see what happens when there are data gaps. Type `oi(0,1,0.95,0)`, but answer "return" to all questions. This gives an obs every time step, over the left half of the domain with a std deviation of 0.02. How the analysis fare? Now decrease the observation frequency by typing first 2 then 5 and keeping the obs pattern and error std dev the same as before. Now what happens to the solution? Why?