Simple Examples

Let's look at a few simple examples of OI analysis.

Example 1: Consider a scalar problem. We have one observation y which is located at the analysis point. We also have a background estimate x_b. In addition, we assume that we know the error statistics of y and x_b:

$$< x_b - x_t >= 0$$
$$< (x_b - x_t)^2 >= \sigma_b^2$$
$$< (y - y_t)(x_b - x_t) >= 0$$
$$< y - y_t >= 0$$
$$< (y - y_t)^2 >= \sigma_o^2$$

Using the notion of OI solution, n =1, p=1, **H=**(1), **R**=(σ_o^2), **B**=(σ_b^2).

$$(x_a) = (x_b) + \frac{(1)(\sigma_o^2)^{-1}[(y) - (1)(x_b)]}{(\sigma_b^2)^{-1} + (1)(\sigma_o^2)^{-1}(1)}$$

$$x_a = x_b + \frac{\sigma_b^2}{\sigma_b^2 + \sigma_o^2}(y - x_b) = \frac{\sigma_o^2}{\sigma_b^2 + \sigma_o^2} x_b + \frac{\sigma_b^2}{\sigma_b^2 + \sigma_o^2} y$$

$$\frac{1}{\sigma_a^2} = \frac{1}{\sigma_b^2} + \frac{1}{\sigma_o^2}$$

The analysis equation can be rewritten as

$$x_a = \frac{\sigma_o^2 \sigma_b^2}{\sigma_b^2 + \sigma_o^2}\left[\frac{1}{\sigma_b^2} x_b + \frac{1}{\sigma_o^2} y\right],$$

then $$\left(\frac{1}{\sigma_b^2} + \frac{1}{\sigma_o^2}\right) x_a = \frac{x_b}{\sigma_b^2} + \frac{y}{\sigma_o^2},$$

therefore $$\frac{x_a}{\sigma_a^2} = \frac{x_b}{\sigma_b^2} + \frac{y}{\sigma_o^2}.$$

We see that the analysis is a linear combination of the background and observation, with the weighting coefficients proportional to the inverse of variances. In fact, if we have *N* samples of

observations, or more generally, N measured estimates of state variable x, called y_n, plus one background estimate x_b, then the final estimate combining all available pieces of information is

$$\frac{x_a}{\sigma_a^2} = \frac{x_b}{\sigma_b^2} + \sum_{n=1}^{N} \frac{y_n}{\sigma_{on}^2},$$

where

$$\frac{1}{\sigma_a^2} = \frac{1}{\sigma_b^2} + \sum_{n=1}^{N} \frac{1}{\sigma_{on}^2}.$$

The variance of the analysis is smaller than that of both background and observations, i.e., $\sigma_a^2 < \min[\sigma_b^2, \sigma_{o1}^2, \sigma_{o2}^2, ...\sigma_{oN}^2]$. This is because the varianace is always positive, $\frac{1}{\sigma_a^2}$ should be larger than the largest term of the right hand side of the analysis covariance equation above, therefore σ_a^2 should be smaller than the smallest σ^2 on the right hand side that corresponds to the largest term.

Now assume that the analysis is for temperature and

$$y = 22.0,\ \mathbf{R} = (\sigma_o)^2 = 1.0$$

$$x_b = 20.5,\ \mathbf{B} = (\sigma_b)^2 = 2.0$$

we obtain the analysis and its variance

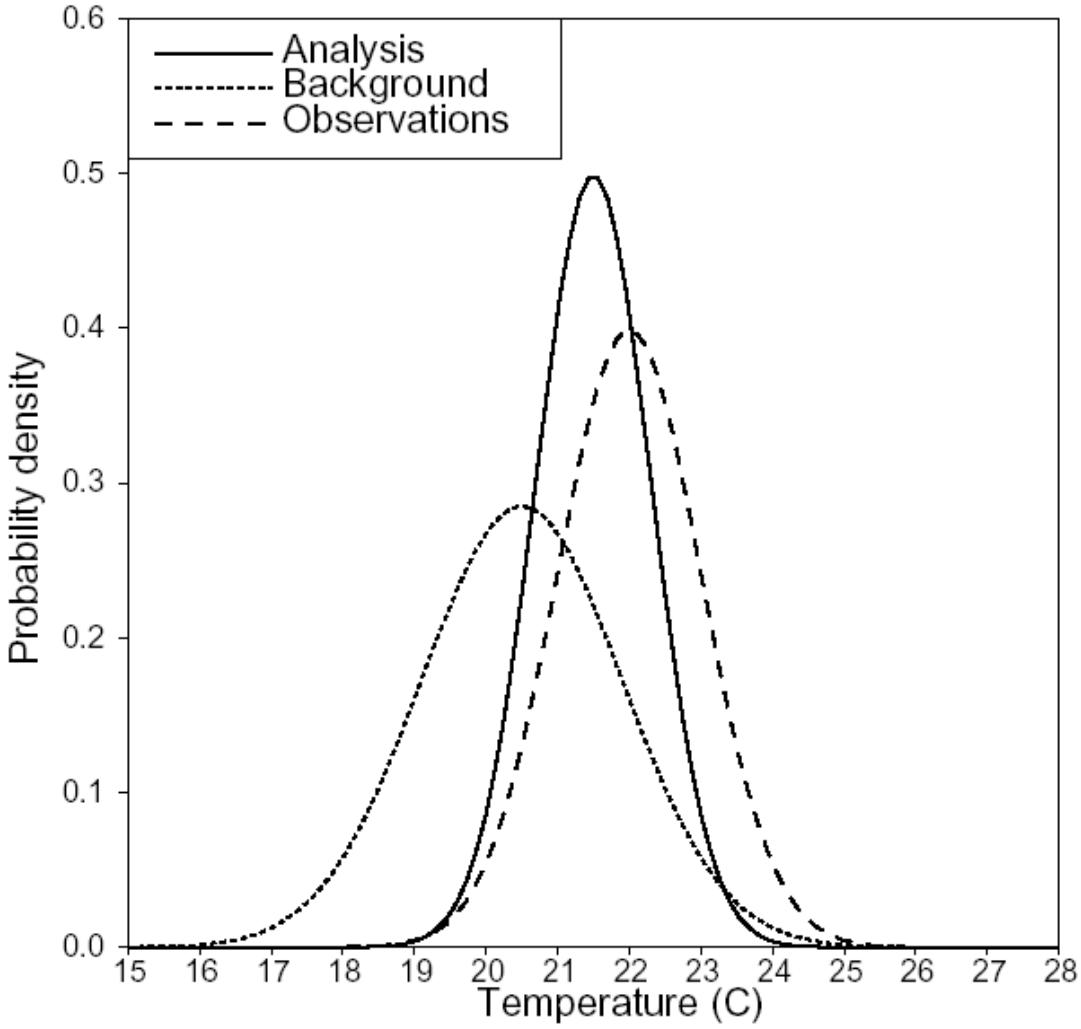
$$x_a = 21.5,\ \mathbf{A} \approx 0.7\ (\sigma_a \approx 0.8).$$

The following figure shows that background, observation and analysis probability distribution functions assuming the errors have Gaussian distributions.

It is clear that the analysis is between the background and observation, and is in this particular case closer to the observation because of its smaller expected errors.

It is also shown that the analysis has a higher probability and smaller expected errors compared to both the background and the observation.

Read also section 5.3.3 of Kalnay's book.



Probability distribution function for analysis, given observation and the background.

Example 2. Now, suppose we have one observation, *y*, located between two analysis points. We have background information on the two analysis points, denoted by x_{b1} and x_{b2}, and we can linearly interpolate the background information to the observation location as

$$\mathbf{H}\mathbf{x}_b = \begin{pmatrix} \alpha & 1-\alpha \end{pmatrix} \begin{pmatrix} x_{b1} \\ x_{b2} \end{pmatrix} = \alpha x_{b1} + (1-\alpha) x_{b2}$$

where $0 \leq \alpha \leq 1$.

For linear interpolation, $\alpha = \dfrac{z_2 - z}{z_2 - z_1}$, where *z* is the spatial coordinate.

The assumed observation error is as given before,

$$\mathbf{R} = (\sigma_o)^2 ,$$

while the background error covariance matrix now takes the form of

$$\mathbf{B} = \begin{pmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{pmatrix} = \sigma_b^2 \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}.$$

Here we have assumed that

$$b_{11} = b_{22} = \sigma_b^2$$
$$b_{21} = b_{12} = \sigma_b^2 \rho$$

with ρ being the background error correlation coefficient between the two grid points.

The OI solution is then

$$\begin{pmatrix} x_{a1} \\ x_{a2} \end{pmatrix} = \begin{pmatrix} x_{b1} \\ x_{b2} \end{pmatrix} + \sigma_b^2 \begin{pmatrix} \alpha + \rho(1-\alpha) \\ \rho\alpha + (1-\alpha) \end{pmatrix} \frac{y - [\alpha x_{b1} + (1-\alpha) x_{b2}]}{[\alpha^2 + 2\alpha(1-\alpha)\rho + (1-\alpha)^2]\sigma_b^2 + \sigma_o^2}. \qquad (5)$$

Let's consider three cases:

Case 1. The observation is collocated with analysis grid point 1 ($\alpha = 1$) and the background errors are not correlated between points 1 and 2 ($\rho = 0$). The above solution now reduces to

$$\begin{pmatrix} x_{a1} \\ x_{a2} \end{pmatrix} = \begin{pmatrix} x_{b1} \\ x_{b2} \end{pmatrix} + \sigma_b^2 \begin{pmatrix} 1 \\ 0 \end{pmatrix} \frac{y - x_{b1}}{\sigma_b^2 + \sigma_o^2}.$$

In this case, the solution at point 1 is identical to that in example 1. The solution at point 2 is equal to the background and no information from the observation is added there.

Case 2. The observation is collocated with analysis at grid 1 ($\alpha = 1$) as in case 1. However, the background errors are correlated between point 1 and point 2 ($\rho \neq 0$). The solution now reduces to

$$\begin{pmatrix} x_{a1} \\ x_{a2} \end{pmatrix} = \begin{pmatrix} x_{b1} \\ x_{b2} \end{pmatrix} + \sigma_b^2 \begin{pmatrix} 1 \\ \rho \end{pmatrix} \frac{y - x_{b1}}{\sigma_b^2 + \sigma_o^2}.$$

In this case, the solution at point 1 is unchanged from case 1, but the solution at point 2 is equal to the background plus ρ times the analysis increment added to point 1 – now we can see the role of background error correlation in spreading observational information (or more strictly the analysis increment).

Case 3. The observation is located inbetween of the analysis points ($\alpha \neq 1$) but the background errors are not correlated between points 1 and 2 ($\rho = 0$). Now the solution becomes

$$\begin{pmatrix} x_{a1} \\ x_{a2} \end{pmatrix} = \begin{pmatrix} x_{b1} \\ x_{b2} \end{pmatrix} + \sigma_b^2 \begin{pmatrix} \alpha \\ 1-\alpha \end{pmatrix} \frac{y - [\alpha x_{b1} + (1-\alpha) x_{b2}]}{[\alpha^2 + (1-\alpha)^2]\sigma_b^2 + \sigma_o^2}.$$

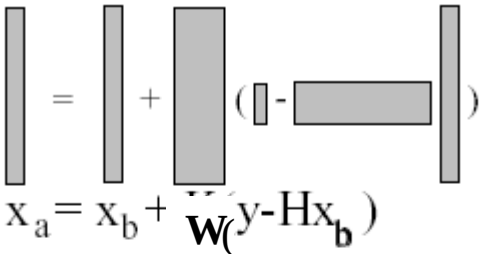
In this case, the analysis increments for point 1 and point 2 are proportional to α and 1 - α, respectively, i.e., the analysis result depends on the distance of the observation from the grid point.

This also says that even in the absence of background error correlation, the grid points involved in the forward observation operators are usually influenced directly by the observations, through the link in the observation operator.

Final Comments:

From the full solution (5), we can see that both the observation operator and error correlation have made contributions. However, when generalizing the solution from two analysis points to *n* points, the linear interpolation operator will only influence the analysis points around the observation, while the error correlations may spread information to all analysis points that have error correlation with the first guess observation $H(\mathbf{x}_b)$.

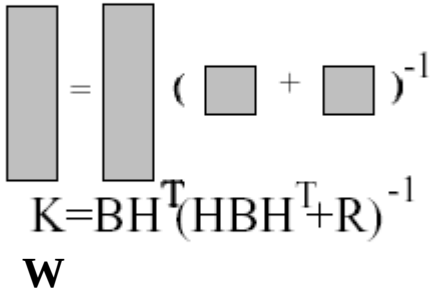
Approximations with Practical OI Implementation



- The OI analysis is given by

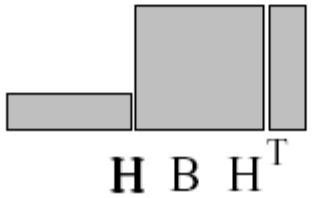
$$\mathbf{x}_a = \mathbf{x}_b + \mathbf{W}[\mathbf{y}_o - H(\mathbf{x}_b)]$$

$$\mathbf{W} = \mathbf{B}\mathbf{H}^T(\mathbf{R} + \mathbf{H}\mathbf{B}\mathbf{H}^T)^{-1}$$



- The actual implementation requires simplifications in the computation of the weight **W**.

- The equation for $\mathbf{x_a}$ can be regarded as a list of scalar analysis equations, one per model variable in the vector **x**.



- For each model variable the analysis increment is given by the corresponding line of **W** times the vector of background departures ($\mathbf{y} - H(\mathbf{x}_b)$).

Given

$$\mathbf{W} = \begin{bmatrix} \mathbf{w}_1^T \\ \mathbf{w}_2^T \\ \vdots \\ \mathbf{w}_n^T \end{bmatrix}, \text{ and } \begin{bmatrix} x_{a1} \\ x_{a2} \\ \vdots \\ x_{an} \end{bmatrix} = \begin{bmatrix} x_{b1} \\ x_{b2} \\ \vdots \\ x_{bn} \end{bmatrix} + \begin{bmatrix} \mathbf{w}_1^T \\ \mathbf{w}_2^T \\ \vdots \\ \mathbf{w}_n^T \end{bmatrix} [\mathbf{y}_o - H(\mathbf{x}_b)],$$

therefore, $x_{ai} = x_{bi} + \mathbf{w}_i^T [\mathbf{y}_o - H(\mathbf{x}_b)]$.

- The fundamental hypothesis in the typically implementation of OI is: *For each model variable, only a few observations are important in determining the analysis increment.*

Based on this assumption, the problem of matrix product and inversion is reduced by including only a smaller number of observations for the analysis at a given grid point. The following two figures show two data selection strategies.

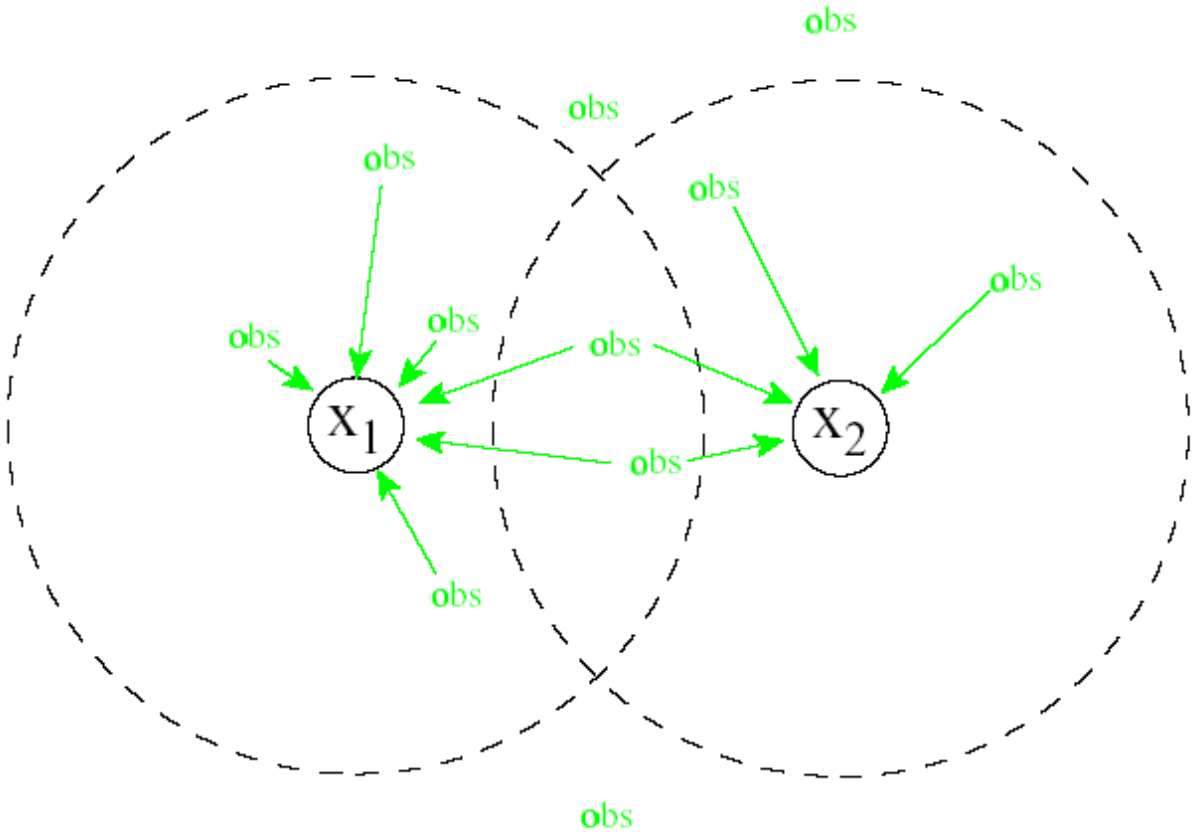


Figure 9. One OI data selection strategy is to assume that each analysis point is only sensitive to observations located in a small vicinity. Therefore, the observations used to perform the analysis at two neighbouring points x_1 or x_2 may be different, so that the analysis field will generally not be continuous in space. The cost of the analysis increases with the size of the selection domains.

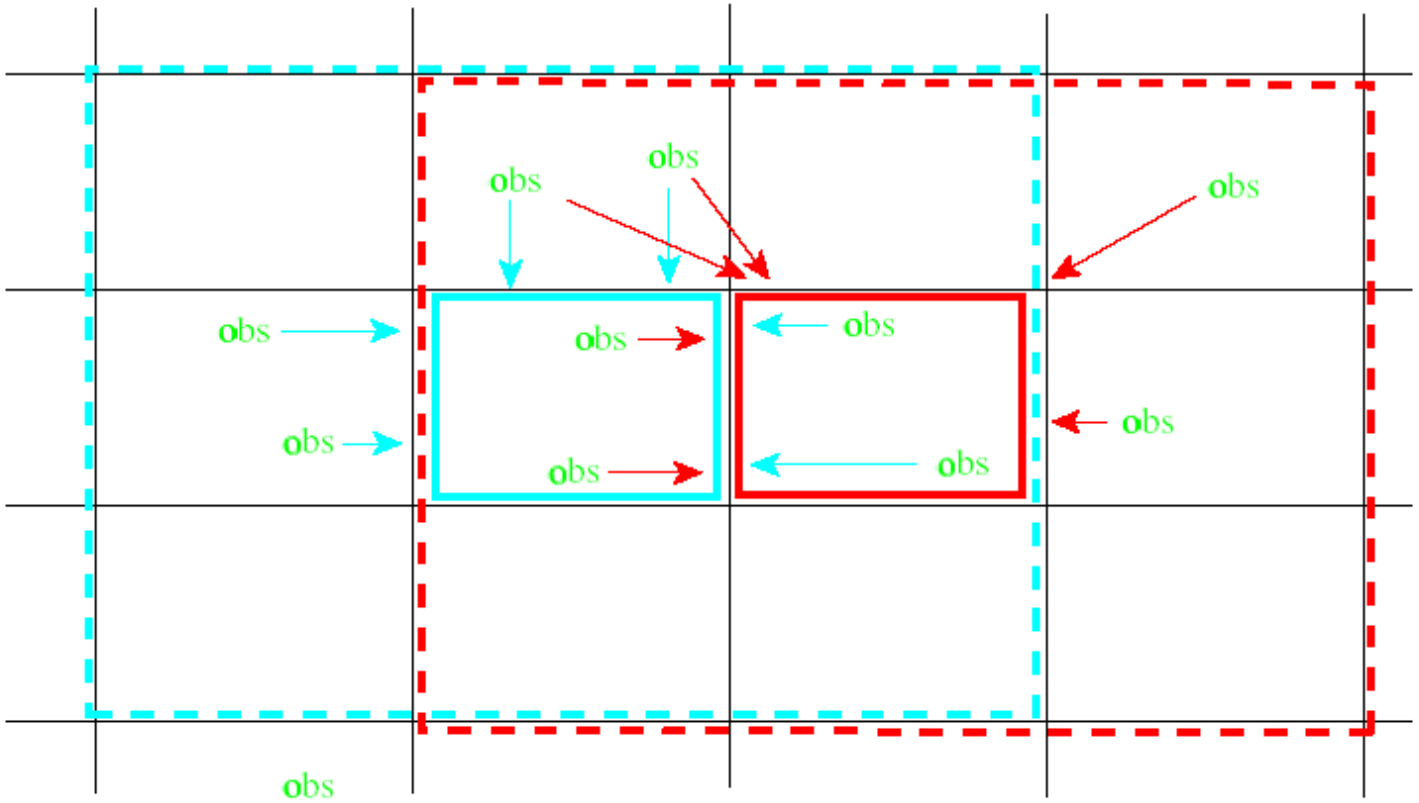


Figure 10. A slightly more sophisticated and more expensive OI data selection is to use, for all the points in an analysis box (black rectangle), all observations located in a bigger selection box (dashed rectangle), so that most of the observations selected in two neighbouring analysis boxes are identical.

- The actual implemented can be as follows:

  1) For each model variable x_i, select a small number of p_i observations using empirical selection criteria.

  2) Form the corresponding list of background departures $(\mathbf{y} - H(\mathbf{x}_b))_i$,

  3) Form the p_i background error covariances between the model variable x_i and the model state interpolated at the p_i observation points (i.e. the relevant p_i coefficients of the i-th line of $\mathbf{BH}^T$), and

  4) Form the $p_i \times p_i$ background and observation error covariance submatrices formed by the restrictions of $\mathbf{HBH}^T$ and $\mathbf{R}$ to the selected observations.

  5) Invert the positive definite matrix formed by the restriction of $(\mathbf{HBH}^T + \mathbf{R})$ to the selected observations,

  6) Multiply it by the i-th line of $\mathbf{BH}^T$ to get the necessary line of $\mathbf{W}$.

It is possible to save some computer time on the matrix inversion by solving directly a symmetric positive linear system, since we know in advance the vector of departures to which the inverse matrix will be applied. Also, if the same set of observations is used to analyze several model variables, then the same matrix inverse (or factorization) can be reused.

Models of Error Covariances

Correct specification of background and observation error covariances is crucial – they determine the relative weight of background and observations.

- Variances are essential for determining the magnitude of errors therefore the relative weight
- Covariance determines how observation information is spread in model space (when the model resolution does not match that of observation)

Observation error variances – include instrument errors and representativeness errors.

- Systematic observation biases should be removed before using the data

Observation error correlation/covariance – often assumed zero, i.e., measurements are assumed uncorrelated.

- Observation error correlation can show up when
    - Sets of observations are taken by the same platform, e.g., radar, rawinsonde, aircraft, satellite
    - Data preprocessing that introduce systematic errors
    - Representative errors of close-by observations
    - Error of the forward operator, e.g., interpolator that contains similar errors

- The presence of (positive) observation error correlation reduces the weight given to the average of the observations – reasonable because these observations are alike

- Observation error correlations are difficult to estimate and account for. In practice, efforts are made to minimize them through reducing bias, by

    - Avoiding unnecessary preprocessing
    - By thinning dense data (denser than grid resolution)
    - By improving model and observation operators (model plays the role of forward operator in the case of 4DVAR)

- After these are done, it is safer to assume the observation correlation to be zero, i.e., the observation error covariance matrix R is diagonal.

Background error variances – they are usually estimates of the error variances in the forecast used to produce $\mathbf{x_b}$.

- This is a difficult problem, because they are never observed directly – they can only be estimated in a statistical sense.

- If the analysis is of good quality (i.e. if there are a lot of observations) an estimate can be provided by the variance of the differences between the forecast and a verifying analysis.

- If the observations can be assumed to be uncorrelated, much better averaged background error variances can be obtained by using the observational method (or the Hollingworth-Lonnberg method – Tellus 1986).

- However, in a system like the atmosphere the actual background errors are expected to depend a lot on the weather situation, and ideally the background errors should be flow-dependent. This can be achieved by the Kalman filter, by 4D-Var to some extent, or by some empirical laws of error growth based on physical grounds.

- If background error variances are badly specified, it will lead to too large or too small analysis increments.

- In least-squares analysis algorithms, only the relative magnitude of the background and observation error variances is important.

Background error correlations – they are essential because

- **Information spreading**.
    - **In data-sparse areas**, the shape of the analysis increment is completely determined by the covariance structures.
    - The correlations in **B** will perform the spatial spreading of information from the observation points to a finite domain surrounding it.
- **Information smoothing**.

- **In data-dense areas**, the amount of smoothing of the observed information is governed by the correlations in **B**, which can be understood by noting that the left most term in **W** is **B**.

- The smoothing of the increments is important in ensuring that the analysis contains scales which are statistically compatible with the smoothness properties of the physical fields.

  - For instance, when analyzing stratospheric or anticyclonic air masses, it is desirable to smooth the increments a lot in the horizontal in order to average and spread efficiently the measurements.

  - When doing a low-level analysis in frontal, coastal or mountainous areas, or near temperature inversions, it is desirable on the contrary to limit the extent of the increments so as not to produce an unphysically smooth analysis. This has to be reflected in the specification of background error correlations.

- Both of the above address only the issue of correlations among same variables, or **auto-correlations**.

- **Balance properties** – correlation across variables, or **cross-correlations**

  - There are sometimes more degrees of freedom in a model than in reality (i.e., not all model variables are free from each other). For instance, the large-scale atmosphere is usually hydrostatic and is almost geostrophic – these relationships introduce ***balances* among the fields**

- These balance properties show up as correlations in the background errors

- Because of these correlations / balances, observations can be used more effectively, i.e., observing one model variable yields information about all variables that are in balance with it.

    - For example, a low-level wind observation allows one to correct the surface pressure field by assuming some amount of geostrophy.

    - When combined with the spatial smoothing of increments this can lead to a considerable impact on the quality of the analysis, e.g. a properly spread observation of geopotential height can produce a complete three-dimensional correction to the geostrophic wind field (see Figure).

    - The relative amplitude of the increments in terms of the various model fields will depend directly on the specified amount of correlation as well as on the assumed error variance in all the concerned parameters.

- Accurate estimation and use of background error (cross-) correlations can do the magic of ‘retrieving’ quantities not directly observed, a thing that ensemble Kalman filter attempts to do in, e.g., the assimilation of radar data.

- Accurate background errors are flow-dependent.

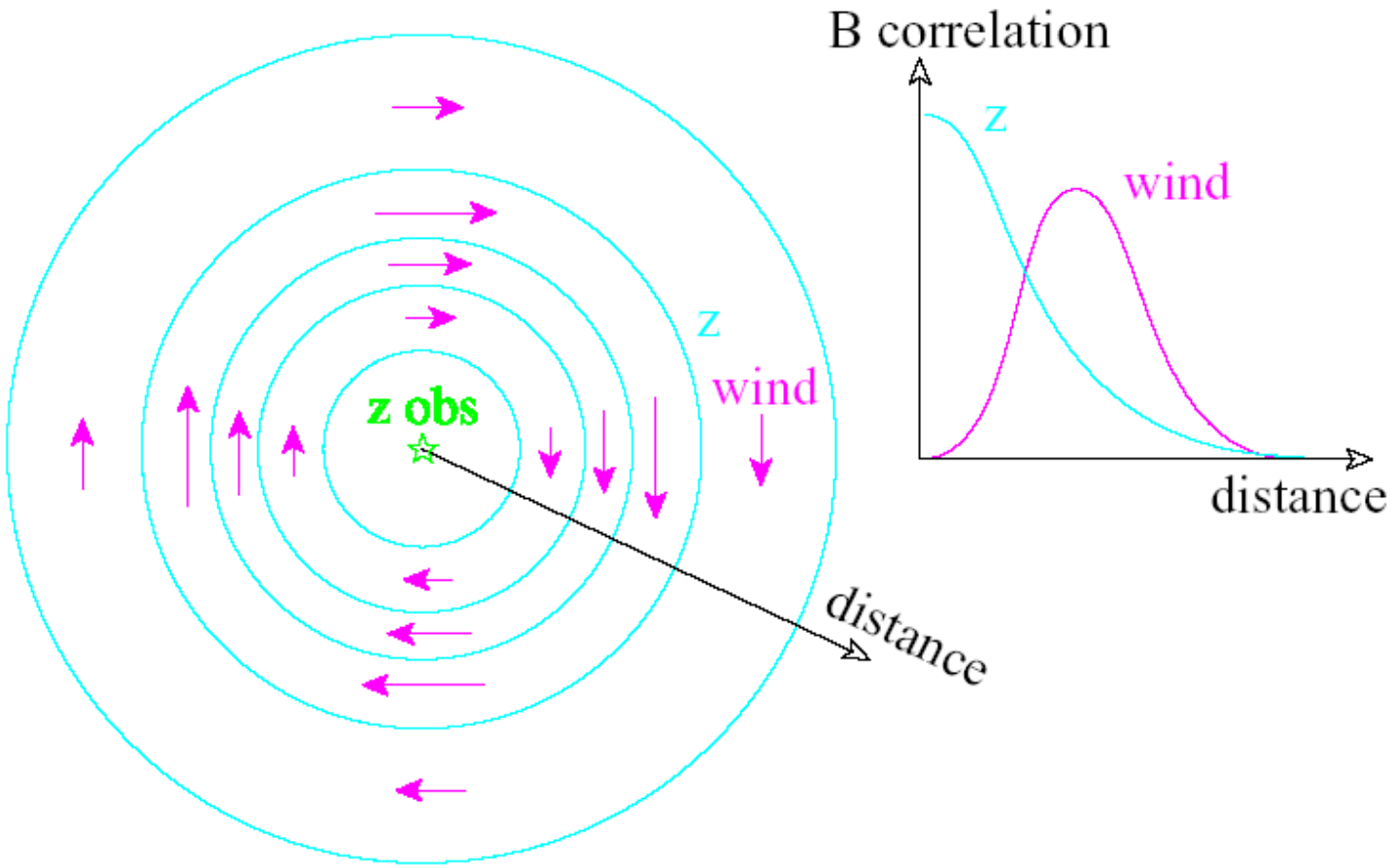


Figure 7. Example of horizontal structure functions commonly used in meteorology: the horizontal autocorrelation of height (or pressure) has an isotropic, gaussian-like shape as a function of distance (right panel). In turn, geostrophy implies that wind will be cross-correlated with height at distances where the gradient of height correlation is maximum. Hence, an isolated height observation will generate an isotropic height "bump" with a rotating wind increment in the shape of a ring.

The observational (or Hollingworth–Lonnberg) method for estimating background error

- This method relies on the use of background departures ($\mathbf{y} - H(\mathbf{x}_b)$) in an observing network that is dense and large enough to provide information on many scales, and that can be assumed to consist of uncorrelated and discrete observations.

- The principle (illustrated in Fig. 8) is to calculate an histogram of background departure covariances, stratified against separation (for instance).

- At zero separation the histogram provides averaged information about the background and observation errors, at nonzero separation it gives the averaged background error correlation.

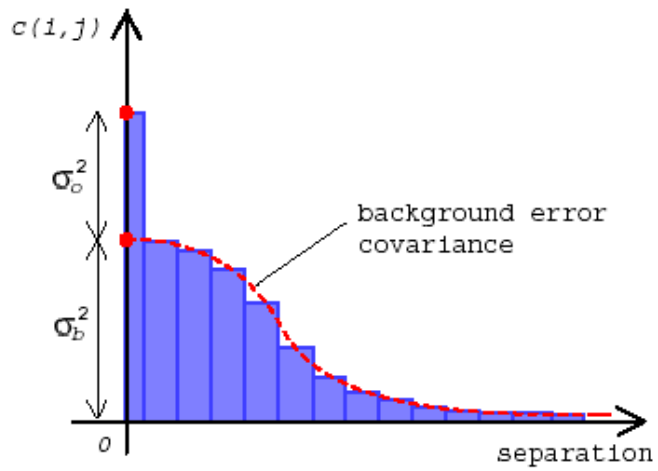


Figure 8. Schematic representation the observational method. The (observation – background) covariance statistics for a given assimilation system are stratified against distance, and the intercept at the origin of the histogram provides an estimate of the average background and observation error variances for these particular assimilation and observation systems.

- In most systems the background error covariances should go to zero for very large separations. If this is not the case, it is usually the sign of biases in the background and/or in the observations and the method may not work correctly (Hollingsworth and Lonnberg 1986.).

The formula is like this:

For the innovation covariance $\sigma_{ij,}$ between point *i* and *j*,

$$
\begin{aligned}
\sigma_{ij} &= E[(y_i - \mathbf{H}_i\mathbf{x}_b)(y_j - \mathbf{H}_j\mathbf{x}_b)^T] \\
&= E[\{(y_i - \mathbf{H}_i\mathbf{x}_t) + (\mathbf{H}_i\mathbf{x}_t - \mathbf{H}_i\mathbf{x}_b)\}\{(y_j - \mathbf{H}_j\mathbf{x}_t)^T + (\mathbf{H}_j\mathbf{x}_t - \mathbf{H}_j\mathbf{x}_b)^T\}] \\
&= E[(y_i - \mathbf{H}_i\mathbf{x}_t)(y_j - \mathbf{H}_j\mathbf{x}_t)^T] + \mathbf{H}_i E[(\mathbf{x}_t - \mathbf{x}_b)(\mathbf{x}_t - \mathbf{x}_b)^T]\mathbf{H}_j^T + 0 + 0 \\
&= \mathbf{R}_{ij} + \mathbf{H}_i\mathbf{B}\mathbf{H}_j^T
\end{aligned}
$$

Reference:

Hollingsworth, A. and P. Lonnberg, 1986: The statistical structure of short-range forecast errors as determined from radiosonde data. Part I: The wind field. *Tellus*, **38A,** 111-136.

The NMC method for estimating background errors

- The so-called "NMC method" (Parrish and Derber, 1992) estimates the forecast error covariance according to

$$\mathbf{B} \approx \alpha E\{[\mathbf{x}_f(48hr) - \mathbf{x}_f(24hr)][\mathbf{x}_f(48hr) - \mathbf{x}_f(24hr)]^T\},$$

  i.e., the structure of the forecast or background error covariance is estimated as the average over many (e.g., 50) differences between two short-range model forecasts verifying at the same time. The magnitude of the covariance is then appropriately scaled.

- In this approximation, rather than estimating the structure of the forecast error covariance from differences with observations, the model-forecast differences themselves provide a multivariate global forecast difference covariance.

- The forecast covariance strictly speaking is the covariance of the forecast differences and is only a proxy of the structure of forecast errors. Nevertheless, it has been shown to produce better results than previous estimates computed from forecast minus observation estimates. An important reason is that the rawinsonde observational network does not have enough density to allow a proper estimate of the global structures.

- The ‘NMC method’ has been in use at most operational centers because of its simplicity and comparative effectiveness.

- Being based on many past forecasts (over e.g., 1-2 months), the estimate is at best seasonally dependent, however.

Reference:

Parrish, D. E. and J. C. Derber, 1992: The National Meteorological Center's spectral statistical-interpolation analysis system. *Mon. Wea. Rev.*, **120,** 1747-1763.

The Modeling of background correlations

- The full **B** matrix is usually too big to be specified explicitly. The variances are just the *n* diagonal terms of **B**, which are usually specified completely.
  - The off-diagonal terms are more difficult to specify. They must generate a symmetric positive definite matrix.
  - Additionally **B** is often required to have some physical properties which are required to be reflected in the analysis:
  - the correlations must be smooth in physical space, on sensible scales,
  - the correlations should go to zero for very large separations if it is believed that observations should only have a local effect on the increments,
  - the correlations should not exhibit physically unjustifiable variations according to direction or location,

- the most fundamental balance properties, like geostrophy, must be reasonably well enforced.

- the correlations should not lead to unreasonable effective background error variances for any parameter that is observed, used in the subsequent model forecast, or output to the users as an analysis product.

- The complexity and subtlety of these requirements mean that the specification of background error covariances is a problem similar to physical parameterization. Some of the more popular techniques are listed below.

  - Correlation models can be specified independently from variance fields, under the condition that the scales of variation of the variances are much larger than the correlation scales,

  - Vertical autocorrelation matrices for each parameter are usually small enough to be specified explicitly.

  - Horizontal autocorrelations cannot be specified explicitly, but they can be reduced to sparse matrices by assuming that they are homogeneous and isotropic to some extent.

  - Three-dimensional multivariate correlation models can be built by carefully combining *separability*, *homogeneity* and *independency* hypotheses like: zero correlations in the vertical for distinct spectral wavenumbers, homogeneity of the vertical correlations in the horizontal and/or horizontal correlations in the vertical, property of the correlations being products of horizontal and vertical correlations. Numerically they imply that the correlation matrix is sparse because it is made of block matrices which are themselves block-diagonal

- Balance constraints can be enforced by transforming the model variables into suitably defined complementary spaces of *balanced* and *unbalanced* variables. The latter are supposed to have smaller background error variances than the former, meaning that they will contribute less to the increment structures.

- The geostrophic balance constraint can be enforced using the classical f-plane or β-plane balance equations, or projections onto subspaces spanned by so-called Rossby and Gravity normal modes.

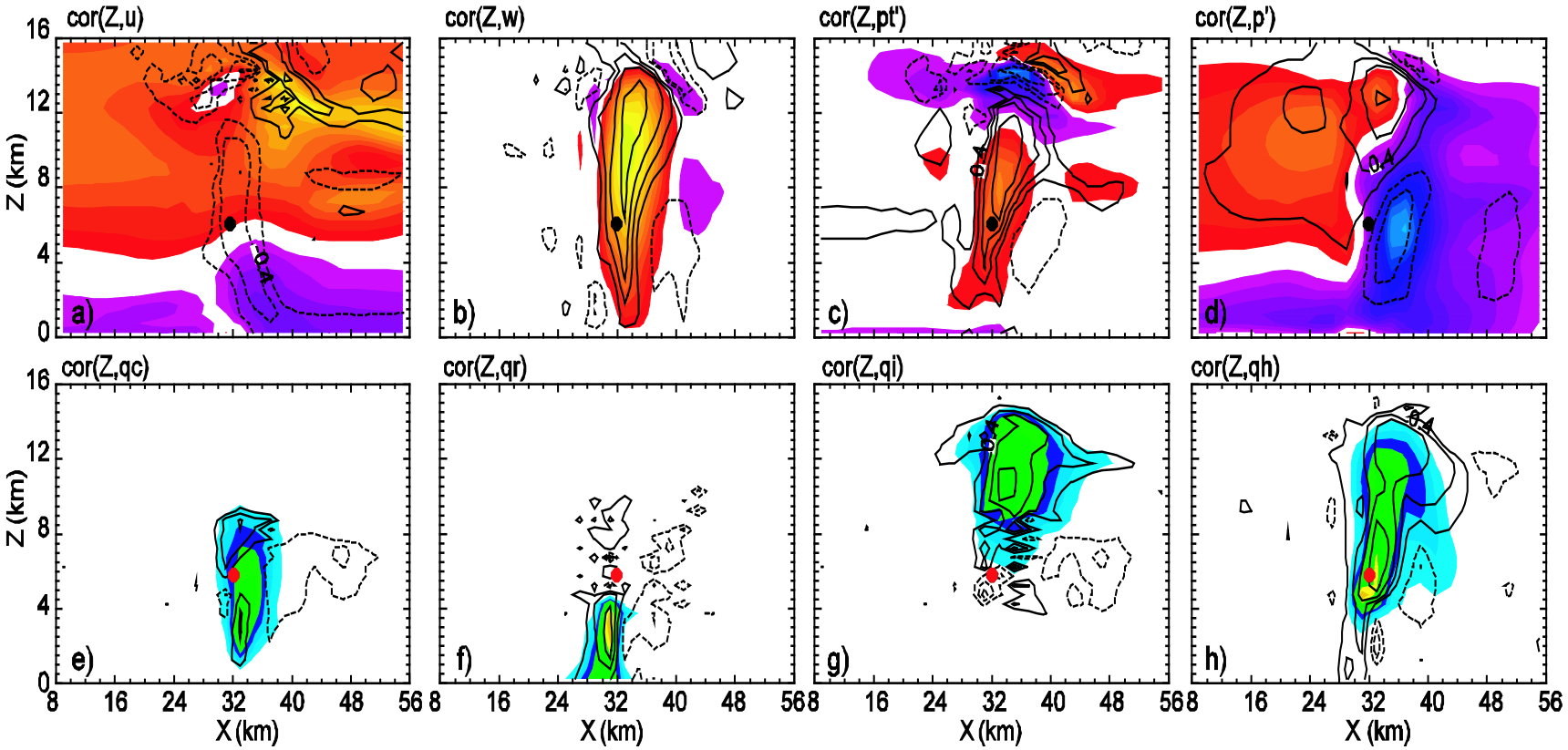
- More general kinds of balance properties can be expressed using linear regression operators calibrated on actual background error fields, if no analytical formulation is available.

Many of such treatments are done in the NCEP operational 3DVAR system. Good discussions can be found in:

Purser, R. J., W.-S. Wu, D. F. Parrish, and N. M. Roberts, 2003: Numerical aspects of the application of recursive filters to variational statistical analysis. Part II: Spatially inhomogeneous and anisotropic general covariances. *Mon. Wea. Rev.*, **131,** 1536-1548.

Purser, R. J., W.-S. Wu, D. F. Parrish, and N. M. Roberts, 2003: Numerical aspects of the application of recursive filters to variational statistical analysis. Part I: Spatially homogeneous and isotropic Gaussian covariances. *Mon. Wea. Rev.*, **131,** 1524-1535.

Correlation coefficients between Z at the black dot and model states at t=80min. Experiment assimilating Z(>10dBZ) only



(See Tong and Xue MWR 2005)

Comment on OI (and 3DVAR) versus other schemes

- **Perhaps the most important advantage of statistical interpolation schemes such as Optimal Interpolation and 3D-Var over empirical schemes such as successive correction method (SCM), is that the correlation between observational increments can be taken into account**.

- With SCM, the weights of the observational increments depend only on their distance to the grid point. Therefore, if a number of observations are "bunched up" in one quadrant, with just a single observation in a different quadrant, then all the observations will be given similar weight. In Optimal Interpolation (or 3D Var), by contrast, the isolated observational increment will be given more weight in the analysis than observations that are close together and therefore less independent.

- When several observations are too close together, then the OI solution becomes an ill-posed problem. In those cases, it is common to compute a "**super-observation**" combining the close individual observations. This has the advantage of removing the ill posedness, while at the same time reducing by averaging the random errors of the individual observations. The super-observation should be a weighted average that takes into account the relative observation errors of the original close observations.

The role of observation operator H

Earlier, we said that

- The **H** matrix (a $p \times n$ matraix) transforms vectors in model space (e.g., **x**, which is a vector of length *n*) into their corresponding values in observation space (vectors of length *p*).

- The transpose or adjoint of **H**, $\mathbf{H^T}$, (an $n \times p$ matrix) transforms vectors in observation space (e.g., **y**, a vector of length *p*) to vectors in model space (vectors of length *n*).

The observation operator and its adjoint can also operate on the error covariance matrices, **B** and **R**, and when they do so, they have similar effect as on vectors.

For example, we indicated earlier that $\mathbf{BH}^T$ is the background error covariances between the grid points and observation points, therefore $\mathbf{H^T}$ plays the role of 'taking background error from grid point to observational points, partially though because the product still represents correlations between grid and observation points'.

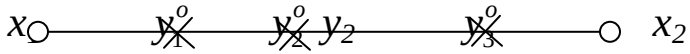
The background error is completely brought into the observation space by the **H** operator and its adjoint (transpose), so that $\mathbf{HBH}^T$ represents error covariances of the background in terms of the observed quantities. This can be seen from below:

$$\mathbf{y}_b = H(\mathbf{x}_b)$$
$$\mathbf{y}_b - \mathbf{y}_t = H(\mathbf{x}_t + \mathbf{x}_b - \mathbf{x}_t) - H(\mathbf{x}_t) \approx \mathbf{H}(\mathbf{x}_b - \mathbf{x}_t)$$
$$\left\{(\mathbf{y}_b - \mathbf{y}_t)(\mathbf{y}_b - \mathbf{y}_t)^{\mathrm{T}}\right\} \approx \left\{\mathbf{H}(\mathbf{x}_b - \mathbf{x}_t)(\mathbf{x}_b - \mathbf{x}_t)^T \mathbf{H}^T\right\} = \mathbf{HBH}^T$$

where approximation has been made to linearize the observation operator *H*.
For a linear observation operator, it is exact.

Further illustration of the point

Suppose we have three observations, y_1^o, y_2^o and y_3^o taken between two grid points with background values of x_1 and x_2:



The forward operator is simply the linear interpolator, so that

$$\mathbf{H}\mathbf{x}^b = \begin{bmatrix} \alpha_1 & 1-\alpha_1 \\ \alpha_2 & 1-\alpha_2 \\ \alpha_3 & 1-\alpha_3 \end{bmatrix} \begin{bmatrix} x_1^b \\ x_2^b \end{bmatrix} \equiv \begin{bmatrix} \alpha_1 & \beta_1 \\ \alpha_2 & \beta_2 \\ \alpha_3 & \beta_3 \end{bmatrix} \begin{bmatrix} x_1^b \\ x_2^b \end{bmatrix}.$$

The background error covariance matrix **B** is

$$\mathbf{B} = \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix}$$

therefore

$$\mathbf{BH}^T = \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix} \begin{bmatrix} \alpha_1 & \alpha_2 & \alpha_3 \\ \beta_1 & \beta_2 & \beta_3 \end{bmatrix}$$

$$= \begin{bmatrix} \alpha_1 b_{11} + \beta_1 b_{12} & \alpha_2 b_{11} + \beta_2 b_{12} & \alpha_3 b_{11} + \beta_3 b_{12} \\ \alpha_1 b_{21} + \beta_1 b_{22} & \alpha_2 b_{21} + \beta_2 b_{22} & \alpha_3 b_{21} + \beta_3 b_{22} \end{bmatrix}.$$

Indeed, the background error covariances have been 'interpolated' by the observation operator (interpolator in this case).

For example, the first element of $\mathbf{BH}^T$, $b_{11}\alpha_1 + b_{12}\beta_1$ represents the background error covariance between grid point one and observation 1, and is equal to the interpolation of the background error covariance between x_1 point with itself (b_{11}), plus the background error covariance between x_1 point and x_2 point (b_{12}) to the y_1 observation point, using interpolation coefficients α_1 and β_1. Let's denote the matrix

$$\mathbf{BH}^T = \begin{bmatrix} c_{11} & c_{12} & c_{13} \\ c_{21} & c_{22} & c_{23} \end{bmatrix}.$$

The first index of c denotes grid point location and second index indicates the observation point.

Applying **H** operator again to $\mathbf{BH}^T$ takes the 'other grid point end' of covariances also to the observation points, so that we are left with covariances between observation points, but still of background errors.

$$\mathbf{HBH}^T = \begin{bmatrix} \alpha_1 & \beta_1 \\ \alpha_2 & \beta_2 \\ \alpha_3 & \beta_3 \end{bmatrix} \begin{bmatrix} c_{11} & c_{12} & c_{13} \\ c_{21} & c_{22} & c_{23} \end{bmatrix}.$$

$$= \begin{bmatrix} \alpha_1 c_{11} + \beta_1 c_{21} & \alpha_1 c_{12} + \beta_1 c_{22} & \alpha_1 c_{13} + \beta_1 c_{23} \\ \alpha_2 c_{11} + \beta_2 c_{21} & \alpha_2 c_{12} + \beta_2 c_{22} & \alpha_2 c_{13} + \beta_2 c_{23} \\ \alpha_3 c_{11} + \beta_3 c_{21} & \alpha_3 c_{12} + \beta_3 c_{22} & \alpha_3 c_{13} + \beta_3 c_{23} \end{bmatrix} \equiv \begin{bmatrix} d_{11} & d_{12} & d_{13} \\ d_{21} & d_{22} & d_{23} \\ d_{31} & d_{32} & d_{33} \end{bmatrix}.$$

Here, covariances c_{ij} are interpolated again, from grid point (indicated by the first index) to the observational point.
For example, d_{12} represents background error covariance between y_1 and y_2 point, and is equal to the interpolated value (using weight α_1 and β_1) of the covariance between x_1 point and y_1 point and the covariance between x_2 point and y_1 point.

The error variances and covariances for x are defined as

$$b_{11} = \frac{1}{N-1}\sum_{i=1}^{N}\left[(x_{1i} - x_{1t})(x_{1i} - x_{1t})\right] = \frac{1}{N-1}\sum_{i=1}^{N}\left[\varepsilon_{1i}\varepsilon_{1i}\right]$$

$$b_{12} = \frac{1}{N-1}\sum_{i=1}^{N}\left[(x_{1i} - x_{1t})(x_{2i} - x_{2t})\right] = \frac{1}{N-1}\sum_{i=1}^{N}\left[\varepsilon_{1i}\varepsilon_{2i}\right]$$

$$b_{21} = \frac{1}{N-1}\sum_{i=1}^{N}\left[(x_{2i} - x_{2t})(x_{1i} - x_{1t})\right] = \frac{1}{N-1}\sum_{i=1}^{N}\left[\varepsilon_{2i}\varepsilon_{1i}\right]$$

$$b_{22} = \frac{1}{N-1}\sum_{i=1}^{N}\left[(x_{2i} - x_{2t})(x_{2i} - x_{2t})\right] = \frac{1}{N-1}\sum_{i=1}^{N}\left[\varepsilon_{2i}\varepsilon_{2i}\right]$$

where the summations are over the number of samples, and $\varepsilon_{1i} = x_{1i} - x_{1t}, \varepsilon_{2i} = x_{2i} - x_{2t}$ are the errors of *x* at points 1 and 2. x_t denotes the truth.

Covarance d_{12} is then

$$d_{12} = \alpha_1 c_{12} + \beta_1 c_{22} = \frac{1}{N-1}\left[\alpha_1(\alpha_2 \sum_{i=1}^{N}[\varepsilon_{1i}\varepsilon_{1i}] + \beta_2 \sum_{i=1}^{N}[\varepsilon_{1i}\varepsilon_{2i}]) + \beta_1(\alpha_2 \sum_{i=1}^{N}[\varepsilon_{2i}\varepsilon_{1i}] + \beta_2 \sum_{i=1}^{N}[\varepsilon_{2i}\varepsilon_{2i}]) \right]$$

$$= \frac{1}{N-1}\sum_{i=1}^{N}[\alpha_1\alpha_2\varepsilon_{1i}\varepsilon_{1i} + \alpha_1\beta_2\varepsilon_{1i}\varepsilon_{2i} + \beta_1\alpha_2\varepsilon_{2i}\varepsilon_{1i} + \beta_1\beta_1\varepsilon_{2i}\varepsilon_{2i}]$$

$$= \frac{1}{N-1}\sum_{i=1}^{N}[(\alpha_1\varepsilon_{1i} + \beta_1\varepsilon_{2i})(\alpha_2\varepsilon_{1i} + \beta_2\varepsilon_{2i})]$$

$$= \frac{1}{N-1}\sum_{i=1}^{N}[(y_{1i} - y_{1t})(y_{2i} - y_{2t})]$$

where $y_{1i} = \alpha_1 x_{1i} + \beta_1 x_{2i}$ and $y_{2i} = \alpha_2 x_{1i} + \beta_2 x_{2i}$.
which is clearly the covariance between the *x* errors interpolated to observation points 1 and 2. For this reason, in the ensemble Kalman filter procedure where we need $\mathbf{HBH}^T$ from ensemble samples, we apply the observation operators H to *x* first, then directly calculate the background error covariance between observation points, which is much cheaper than calculating **B** first then multiple **H** on the left and $\mathbf{H}^T$ on the right.

Here, we have two grid points and three observations, therefore *n*=2 and *p*=3, therefore B is a 2x2 array while $\mathbf{HBH}^T$ is a 3x3 array.

Suppose the background error covariance is zero, i.e., there is no correlation between errors at different grid points, then

$$\mathbf{B} = \begin{bmatrix} b_{11} & 0 \\ 0 & b_{22} \end{bmatrix},$$

$$\mathbf{BH}^T = \begin{bmatrix} \alpha_1 b_{11} & \alpha_2 b_{11} & \alpha_3 b_{11} \\ \beta_1 b_{22} & \beta_2 b_{22} & \beta_3 b_{22} \end{bmatrix},$$

and

$$\mathbf{HBH}^T = \begin{bmatrix} \alpha_1 & \beta_1 \\ \alpha_2 & \beta_2 \\ \alpha_3 & \beta_3 \end{bmatrix} \begin{bmatrix} \alpha_1 b_{11} & \alpha_2 b_{11} & \alpha_3 b_{11} \\ \beta_1 b_{22} & \beta_2 b_{22} & \beta_3 b_{22} \end{bmatrix}$$

$$= \begin{bmatrix} \alpha_1\alpha_1 b_{11} + \beta_1\beta_1 b_{22} & \alpha_1\alpha_2 b_{11} + \beta_1\beta_2 b_{22} & \alpha_1\alpha_3 b_{11} + \beta_1\beta_3 b_{22} \\ \alpha_2\alpha_1 b_{11} + \beta_2\beta_1 b_{22} & \alpha_2\alpha_2 b_{11} + \beta_2\beta_2 b_{22} & \alpha_2\alpha_3 b_{11} + \beta_2\beta_3 b_{22} \\ \alpha_3\alpha_1 b_{11} + \beta_3\beta_1 b_{22} & \alpha_3\alpha_2 b_{11} + \beta_3\beta_2 b_{22} & \alpha_3\alpha_3 b_{11} + \beta_3\beta_3 b_{22} \end{bmatrix}.$$

Consider the special case where y_1 is located at x_1 point, and y_3 is located at x_2 point, and y_2 is located between x_1 and x_2, then $\alpha_1 = 1, \beta_1 = 0, \alpha_3 = 0, \beta_3 = 1.$

$$\mathbf{BH}^T = \begin{bmatrix} \alpha_1 b_{11} & \alpha_2 b_{11} & 0 \\ 0 & \beta_2 b_{22} & \beta_3 b_{22} \end{bmatrix},$$

$$\mathbf{HBH}^T = \begin{bmatrix} b_{11} & \alpha_2 b_{11} & 0 \\ \alpha_2 b_{11} & \alpha_2\alpha_2 b_{11} + \beta_2\beta_2 b_{22} & \beta_2 b_{22} \\ 0 & \beta_2 b_{22} & b_{22} \end{bmatrix}.$$

Assuming

$$\mathbf{R} = \begin{bmatrix} \sigma_o^2 & 0 & 0 \\ 0 & \sigma_o^2 & 0 \\ 0 & 0 & \sigma_o^2 \end{bmatrix}$$

Derive the formula for x_1^a, x_2^a, and x_3^a (**you homework – no need to turn it in but make sure you do it!)**

Consider three situations: (1) y_2^o is located at equal distance from x_1 and x_2, (2) y_2^o is located at the same point as y_1^o and x_1, (3) y_2^o does not exist. **Discuss the your results.**