Approximations in Practical OI

(Kalnay, §5.4.2)

We have seen that the analysis is obtained from

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

We define increments from the background as

$$\delta\mathbf{x} = \mathbf{x} - \mathbf{x}_b \,.$$

Then the analysis increment is

$$\delta\mathbf{x}_a = \mathbf{W}\delta\mathbf{y}_o$$

The optimal weight matrix **W** that minimizes the analysis error covariance is given by

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

If all the statistical assumptions are accurate, i.e., the covariance matrices are known exactly, these formulas provide the OI analysis.

In that case, the analysis error covariance is given by

$$\mathbf{P}_a = \mathbf{A} = (\mathbf{I} - \mathbf{W}\mathbf{H})\mathbf{B}$$

In reality, the statistics are only approximations of the true statistics. Thus, the formulae provide a statistical interpolation, not necessarily an optimal one.

Some scientists argue that the name statistical interpolation should be used instead of optimal interpolation. But the latter is generally used.

OI in Physical Space

OI is typically performed in physical space, either grid point by grid point or over limited volumes.

In the implementation of OI, the equations are solved point by point in grid-point space.

To make the formulation in physical space clearer, we expand the matrix equations:

$$\left.\begin{array}{ll}\mathbf{B} = \begin{bmatrix} b_{11} & \dots & b_{1n} \\ \vdots & & \vdots \\ b_{n1} & \dots & b_{nn} \end{bmatrix} & \mathbf{H} = \begin{bmatrix} h_{11} & \dots & h_{1n} \\ \vdots & & \vdots \\ h_{p1} & \dots & h_{pn} \end{bmatrix} \\ \mathbf{R} = \begin{bmatrix} r_{11} & \dots & r_{1p} \\ \vdots & & \vdots \\ r_{p1} & \dots & r_{pp} \end{bmatrix} & \mathbf{W} = \begin{bmatrix} w_{11} & \dots & w_{1p} \\ \vdots & & \vdots \\ w_{n1} & \dots & w_{np} \end{bmatrix} \end{array}\right\}$$

Note the orders of these matrices.

The Observation Operator

H is the linear perturbation (Jacobian) of the forward observational model H, and $\mathbf{H}^T$ is its ***transpose*** or ***adjoint***.

- Multiplying by **H** on the left transforms grid-point increments into observation increments (**H** is $p \times n$);
- Multiplying by $\mathbf{H}^T$ transforms from observation points back to grid points ($\mathbf{H}^T$ is $n \times p$).

For a single variable, there are n grid points.

If we are considering several variables, n is the product of the number of grid points and the variables.

Simple Low-order Example

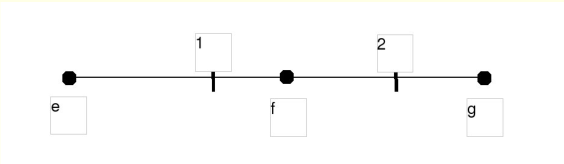
Consider again the OI equations in matrix form:

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

The optimal weight matrix is obtained by solving the system of equations

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

As an illustration, let us consider the simple case of three grid points e, f, g, and two observations, 1 and 2.



Simple example: three grid points and two observation points.

Then $\mathbf{x}^a = (x_e^a, x_f^a, x_g^a)^T$ and $\mathbf{x}^b = (x_e^b, x_f^b, x_g^b)^T$.

The observation vector is $\mathbf{y}^o = (y_1^o, y_2^o)^T$.

The background values at the observation points are

$$\mathbf{H}\mathbf{x}^b = \begin{pmatrix} h_{1e} & h_{1f} & h_{1g} \\ h_{2e} & h_{2f} & h_{2g} \end{pmatrix} \begin{pmatrix} x_e^b \\ x_f^b \\ x_g^b \end{pmatrix} = \mathbf{y}^b$$

The elements of the observation operator $\mathbf{H}$ represent interpolation from gridpoint to observation location.

For example, if we used linear interpolation, $\mathbf{H}$ would be

$$\mathbf{H} = \begin{pmatrix} \frac{x_f - x_1}{x_f - x_e} & \frac{x_1 - x_e}{x_f - x_e} & 0 \\ & & \\ 0 & \frac{x_g - x_2}{x_g - x_f} & \frac{x_2 - x_f}{x_g - x_f} \end{pmatrix}$$

Check: Verify that the correct answer is given when an observation is located at a grid point.

The background error covariance matrix elements are the covariances between grid points:

$$\mathbf{B} = \begin{pmatrix} b_{ee} & b_{ef} & b_{eg} \\ b_{fe} & b_{ff} & b_{fg} \\ b_{ge} & b_{gf} & b_{gg} \end{pmatrix}$$

The error covariance between grid points and observation points is

$$\mathbf{B}\mathbf{H}^T = \begin{pmatrix} b_{e1} & b_{e2} \\ b_{f1} & b_{f2} \\ b_{g1} & b_{g2} \end{pmatrix}$$

Then the background error covariance between observation points may be written

$$\mathbf{H}\mathbf{B}\mathbf{H}^T = \begin{pmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{pmatrix} .$$

The observation error covariance is

$$\mathbf{R} = \begin{pmatrix} r_{11} & r_{12} \\ r_{21} & r_{22} \end{pmatrix}$$

We often assume that measurement errors at different locations are uncorrelated.

Then $\mathbf{R}$ is a diagonal matrix:

$$\mathbf{R} = \begin{pmatrix} r_{11} & 0 \\ 0 & r_{22} \end{pmatrix}$$

Note that $\mathbf{R}$ includes not only the instrument error, but also the representativity error.

We can now write the OI equation for a particular (single) grid point g influenced by p observations as:

$$x_g^a = x_g^b + \sum_{j=1}^{p} w_{gj}\delta y_j$$

This is the grid-point version of the vector equation

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

The weights are the solution of the linear system

$$\sum_{j=1}^{p} w_{gj}(b_{jk}+r_{jk}) = b_{gk} \qquad k = 1, \ldots, p$$

This is the grid-point version of the matrix equation

$$\mathbf{W}(\mathbf{HBH}^T + \mathbf{R}) = \mathbf{BH}^T$$

There are equations like this for each grid point and, in multivariate analysis, for each variable at each point.

How big is the linear system?

A linear system of the form

$$\sum_{j=1}^{p} w_{gj}(b_{jk}+r_{jk}) = b_{gk} \qquad k = 1, \ldots, p$$

is solved for each gridpoint.

Each system comprises p equations, where p is the number of observations taken into account for the analysis at that gridpoint.

Clearly, a selection must be made. Nearby observations are chosen, whereas remote observations are ignored.

Depending on the computational power available, p may be as small as 10 or in the hundreds.

Recall that the computation required to solve a linear system of order N typically scales as the cube of N.

SCM *versus* OI

In SCM, the weights of the observational increments depend only on their distance to the grid point.

Therefore, all observations will be given similar weight, even if a number of them are "bunched up" in one region.

⋆ ⋆ ⋆

In OI and 3D-Var, the correlation between observational increments is taken into account.

Therefore, an isolated observational increment will be given more weight in the analysis than observations that are close together and therefore less independent.

The forecast error correlation $b_{jk}/\sqrt{b_{jj}b_{kk}}$ between the observation points j and k is large if the points are close.

Thus, isolated observations have more independent information than observations close together; OI allows for this.

Ill-conditioned matrices

When several observations are too close together, then the solution of

$$\sum_{j=1}^{p} w_{gj}(b_{jk}+r_{jk}) = b_{gk} \qquad k = 1, \ldots, p$$

becomes an ill-posed problem.

If two observations have essentially the same information, they are given similar weights. Thus, the matrix W has two rows almost equal, and is close to being singular.

In those cases, we can compute a super-observation, combining the close individual observations. This removes the ill-posedness.

The super-observation is given a weight that takes into account the relative errors of the original observations.

Conclusion of the foregoing

Estimation of the Matrices

To implement OI, we need to estimate the error covariances, **B** and **R** and the observation operator **H**.

The observation operator is an interpolator from the model to the observation location.

If the observed variable are different from the model variables, a conversion to observed variables is also included.

We will <u>not consider</u> the conversion process, *e.g.*, conversion of temperatures to radiance measurements.

⋆ ⋆ ⋆

The observational error covariance **R** is obtained from instrument error estimates.

If the measurements are independent, the matrix **R** is diagonal, which is a major advantage.

Background Error Covariance B

The background error covariance **B** is more difficult to estimate.

In OI it is common to normalize the background error covariance with **D**, the diagonal matrix of the variances:

$$\mathbf{B} = \mathbf{D}^{1/2}\mathbf{C}\mathbf{D}^{1/2} \quad \mathbf{D} = \begin{bmatrix} \sigma_1^2 & 0 & \dots & 0 \\ 0 & \sigma_2^2 & \dots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \dots & \sigma_n^2 \end{bmatrix} \quad \mathbf{C} = \begin{bmatrix} \mu_{11} & \mu_{12} & \dots & \mu_{1p} \\ \mu_{12} & \mu_{22} & \dots & \mu_{2p} \\ \vdots & \vdots & & \vdots \\ \mu_{1p} & \mu_{2p} & \dots & \mu_{pp} \end{bmatrix}$$

Here

$$\mu_{ij} = b_{ij}/(\sqrt{b_{ii}}\sqrt{b_{jj}}) = b_{ij}/(\sigma_i\sigma_j)$$

are the correlations of the background errors at two observational points i, j, and σ_i^2 are the error variances.

A number of additional simplifications are made in order to evaluate the background error covariance elements b_{ij}.

We normally assume that the background error correlations can be separated into the product of the horizontal correlation and the vertical correlation (this is illustrated below).

Moreover, the correlations are typically defined as functions of distance only.

⋆ ⋆ ⋆

We will assume that the background error correlation between two points in the same horizontal surface is homogeneous and isotropic.

Then the background error correlation of the geopotential height depends only on the distance between the two points.

We often use a Gaussian exponential function for the geopotential error correlation:

$$[\mu_{ij}(r_{ij})]_{\rm H} = e^{-r_{ij}^2/2L_\Phi^2}$$

Here $r_{ij}^2 = (x_i - x_j)^2 + (y_i - y_j)^2$ is the square of the distance between two points i and j, and L_Φ, typically of the order of 500 km, defines the background error correlation scale.

Gaussian functions can also be used for the vertical correlation functions:

$$[\mu_{ij}(z)]_{\rm V} = e^{-z^2/2L_z^2}$$

Then the total correlation is the product of horizontal and vertical:

$$\mu_{ij} = [\mu_{ij}(r_{ij})]_{\rm H} \times [\mu_{ij}(z)]_{\rm V}$$

Another important assumption is that the background wind error correlations are geostrophically related to the geopotential height error correlations.

This has two advantages:

- It avoids having to estimate independently the wind error correlation
- It imposes an approximate geostrophic balance of the wind and height analysis increments, and thus improves the balance of the analysis.

⋆ ⋆ ⋆

The multivariate correlation between heights and winds can be obtained from the height correlations.

For example, the background error correlation between δu and δv is:

$$E(\delta u_i \delta v_j) = -\frac{g}{f_i}\frac{g}{f_j} E\left(\frac{\partial(\delta z_i)}{\partial y_i}\frac{\partial(\delta z_j)}{\partial x_j}\right)$$

Since the geopotential error at the point x_j is independent of y_i, we can combine the derivatives and write

$$E(\delta u_i \delta v_j) = -\frac{g}{f_i}\frac{g}{f_j}\frac{\partial^2 E(\delta z_i \delta z_j)}{\partial y_i \partial x_j} = -\frac{g}{f_i}\frac{g}{f_j}\frac{\partial^2 b_{ij}}{\partial y_i \partial x_j} = -\frac{g^2}{f_i}\frac{\sigma_z^2}{f_j}\frac{\partial^2 \mu_{ij}}{\partial y_i \partial x_j}$$

The standard deviation of the wind increments can also be derived from the geostrophic relationship [*]:

$$\sigma_u = E(\delta u_i^2)^{1/2} = (g\sigma_z/f_i)\,, \qquad \sigma_v = E(\delta v_j^2)^{1/2} = (g\sigma_z/f_j)$$

So, we obtain the correlation of the increments of the two wind components by dividing by these standard deviations:

$$\rho_{u,v} = -\partial^2 \mu_{ij}/\partial y_i \partial x_j\,.$$

[*] Detail to be clarified later.

Similarly, we can obtain the correlations between the increments of any two of the variables at points i and j:

$$\rho_{h,h} = \mu_{ij}\,, \quad \rho_{h,u} = -\frac{\partial \mu_{ij}}{\partial y_i}\,, \quad \rho_{u,h} = -\frac{\partial \mu_{ij}}{\partial y_j}\,.$$

Exercise: Assume the height correlation function is Gaussian:

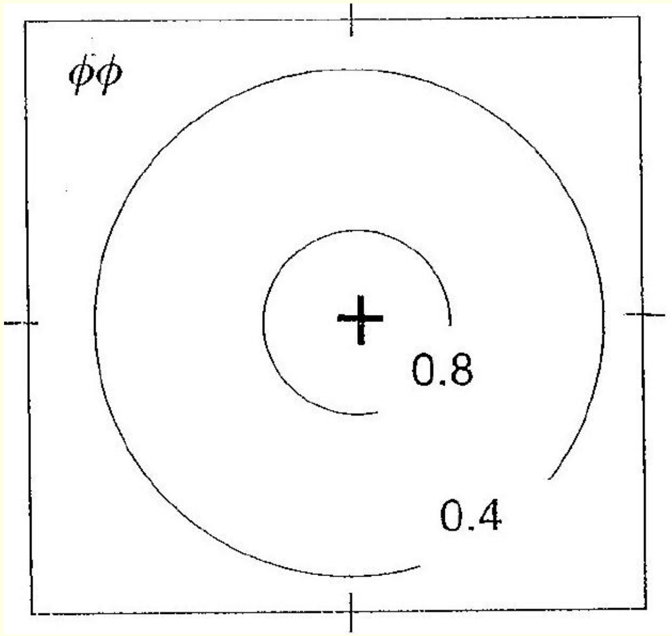
$$\mu_{ij} = e^{-r_{ij}^2/2L_\phi^2}$$

where $r_{ij}^2 = (x_i - x_j)^2 + (y_i - y_j)^2$.

Derive the expressions for the other correlations.

⋆ ⋆ ⋆

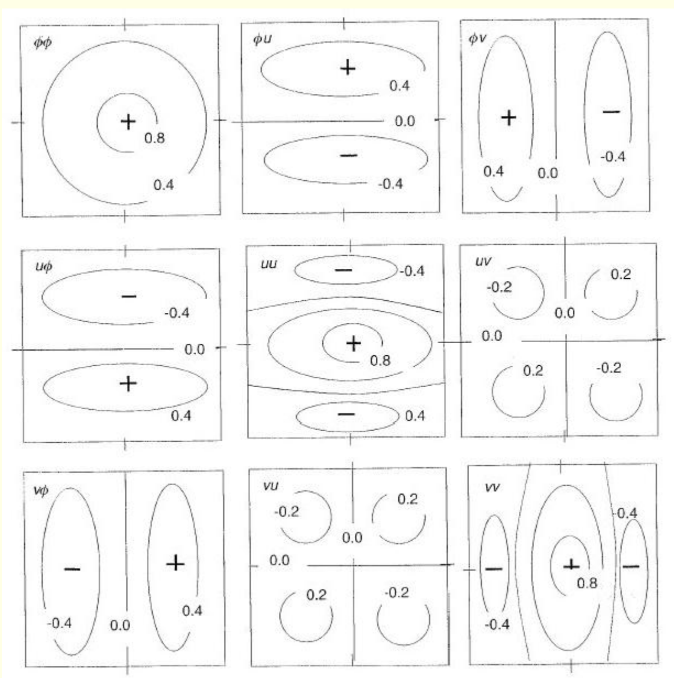
The following figure shows schematically the shape of of the correlation function for geopotential height used in OI.



Schematic illustration of the correlation of Φ-Φ.

The following figure shows schematically the shape of typical wind/height correlation functions used in OI.

Note that the u–h correlations have the opposite sign than the h–u correlations because the first and second variables correspond to the first and second points i and j respectively.



Correlation and cross-correlation functions.

Other Practical Limitations

Geostrophic balance does not hold near the equator, and additional approximations have to be made in the tropics to allow for a smooth decoupling of wind and height increments.

In addition, it is common to select the observations to be included in solving the linear system for the weight coefficients, depending on the computer resources available for the analysis, allowing for a maximum number of observations affecting each grid point.

Rules for the selection of the subset of observations to be used typically depend on the distance to the grid point (within a maximum radius of influence), the types of observations (giving priority to the most accurate) and their distribution.

End of §5.4