<u>Statistical Objective Analysis</u> (cont.)

Now that we have a statistical objective analysis algorithm, eq. (14) from Lecture #13, in which $f_A(\mathbf{r}_i)$, provides a linear unbiased estimate of $f_T(\mathbf{r}_i)$, we should also formulate an expression for the expected analysis error variance, E_A^2. This requires manipulation of our previous equation for E_A^2, eq. (5) in Lecture 13. The result, without derivation is

$$E_A^2 = E_B^2 - \underline{W}_i^T B_i = E_B^2 - \underline{\mathbf{B}}_i^T [\underline{\underline{\mathbf{B}}} + \underline{\underline{\mathbf{O}}}]^{-1} \underline{\mathbf{B}}_i \qquad (15)$$

All terms on the RHS of (14) are known from quantities used in the analysis procedure. Thus the expected analysis error variance is the background error variance minus a positive term proportional to the weights used on the observations - which means that the analysis error is smaller than the background error in regions where there are observations.

In summary, eq. (12) is the minimum variance estimate of $f_T(\mathbf{r}_i)$, and if the optimal weights given by eq. (13) are used for W_i, then the system (12) and (13) is called <u>optimal interpolation</u> *provided* $\underline{\underline{\mathbf{B}}}$, $\underline{\underline{\mathbf{O}}}$, $\underline{\mathbf{B}}_i$ *are correct.* Since these are not perfectly known, Daley prefers the term statistical interpolation or SOA. The estimation of $\underline{\underline{\mathbf{B}}}$, $\underline{\underline{\mathbf{O}}}$, $\underline{\mathbf{B}}_i$ is a major current area of research, especially for mesoscale and storm-scale data assimilation.

<u>Normalized form of the SOA Equations</u>

We now obtain a non-dimensional (normalized) version of the SOA analysis equations. We will need to introduce some new notation.

Define $\underline{\underline{\boldsymbol{\sigma}}}_B = \langle (B_k - T_k)^2 \rangle^{1/2}$ to be a diagonal matrix of standard deviations at <u>observation</u> stations. We also define a new set of normalized weights $\underline{\boldsymbol{\omega}}_i$ via

$$\underline{W}_i^T = E_B \, \underline{\boldsymbol{\omega}}_i^T \underline{\underline{\boldsymbol{\sigma}}}_B^{-1}, \quad \text{where } E_B = \langle \varepsilon_B^2(\mathbf{r}_k) \rangle^{1/2} = \langle (B_i - T_i)^2 \rangle^{1/2}$$

is the background error standard deviation at the grid points.

If we use this expression in eq. (12), we have

$$f_A(\mathbf{r}_i) = f_B(\mathbf{r}_i) + E_B \, \underline{\boldsymbol{\omega}}_i^T \underline{\underline{\boldsymbol{\sigma}}}_B^{-1} [\, \underline{f}_O - \underline{f}_B \,] \qquad (16)$$

Dividing (16) by $E_B = \langle (B_i - T_i)^2 \rangle^{1/2}$ and reverting back to summation notion, we can rewrite (16) as

$$\frac{A_i - B_i}{\langle (B_i - T_i)^2 \rangle^{1/2}} = \sum_{k=1}^{K} \frac{\omega_{ik} [O_k - B_k]}{\langle (B_k - T_k)^2 \rangle^{1/2}} \tag{17}$$

This is a normalized (nondimensional) version of eq. (12). We now wish to rewrite the weight equation (13) using our normalized weights $\underline{\omega}_i$:

$$[\underline{\underline{B}} + \underline{\underline{O}}] E_B \underline{\omega}_i^T \underline{\underline{\sigma}}_B^{-1} = \underline{B}_i$$

Now left multiply by $\underline{\underline{\sigma}}_B^{-1}$ and divide by E_B to get

$$\underline{\underline{\sigma}}_B^{-1} [\underline{\underline{B}} + \underline{\underline{O}}] \underline{\omega}_i^T \underline{\underline{\sigma}}_B^{-1} = E_B^{-1} \underline{\underline{\sigma}}_B^{-1} \underline{B}_i$$

Interchanging $\underline{\omega}_i^T$ and $\underline{\underline{\sigma}}_B^{-1}$ and bringing $\underline{\underline{\sigma}}_B^{-1}$ inside bracket, we have

$$[\underline{\underline{\sigma}}_B^{-1} \underline{\underline{B}} \underline{\underline{\sigma}}_B^{-1} + \underline{\underline{\sigma}}_B^{-1} \underline{\underline{O}} \underline{\underline{\sigma}}_B^{-1}] \underline{\omega}_i = E_B^{-1} \underline{\underline{\sigma}}_B^{-1} \underline{B}_i \tag{18}$$

Note that $\underline{\underline{\sigma}}_B^{-1} \underline{\underline{B}} \underline{\underline{\sigma}}_B^{-1}$ is the definition of the background error correlation matrix at stations, denoted $\underline{\underline{\rho}}_B$ $(-1 \le \underline{\underline{\rho}}_B \le 1)$, and written in summation form as

$$\underline{\underline{\sigma}}_B^{-1} \underline{\underline{B}} \underline{\underline{\sigma}}_B^{-1} = \frac{\langle (B_k - T_k)(B_l - T_l) \rangle}{[\langle (B_k - T_k)^2 \rangle \langle (B_l - T_l)^2 \rangle]^{1/2}} = \underline{\underline{\rho}}_B$$

Similarly,

$$E_B^{-1} \underline{\underline{\sigma}}_B^{-1} \underline{B}_i = \frac{\langle (B_k - T_k)(B_i - T_i) \rangle}{[\langle (B_k - T_k)^2 \rangle \langle (B_l - T_l)^2 \rangle]^{1/2}} = \underline{\rho}_B^i$$

is the background error correlation vector (between stations and grid point).

Therefore, we can now write eq. (18) as

$$[\underline{\underline{\rho}}_B + \underline{\underline{\sigma}}_B^{-1} \underline{\underline{\Omega}} \, \underline{\underline{\sigma}}_B^{-1}] \, \underline{\omega}_i \quad = \quad \underline{\rho}^i_B \qquad (19)$$

We will use a version of eq. (19) later for simple analysis examples.

The normalized equation for the expected analysis error variance is

$$\varepsilon^2_A \;=\; 1 \;-\; \underline{\omega}_i^T \, \underline{\underline{\rho}}^i_B \qquad \text{where} \qquad \varepsilon^2_A \;=\; E_A^2 / E_B^2$$

So now our normalized SOA algorithm is eq. (16) or (17) written as

$$\frac{f_A(\mathbf{r}_i) \;-\; f_B(\mathbf{r}_i)}{E_B} \;=\; \frac{\underline{\omega}_i^T [\, \underline{f}_o \;-\; \underline{f}_B \,]}{\underline{\underline{\sigma}}_B} \qquad \text{with weights given by eq. (19).}$$

<u>Conversion between Daley and Kalnay Notation</u>

In Sec. 5.6 of Daley's book, he summarizes some of the limitations of the SOA algorithm as presented in his Chap. 4 and 5. Many of these limitations are valid today, but some can be addressed by provided the proper formulation of $\underline{f}_B$, the values of the background field at the stations, which, up to now, have been assumed to be obtained simply by interpolation from surrounding grid points. However, suppose the observations are not the same variable as the ones on the grid (e.g., radiances instead of temperature). Then we need to introduce a linear operator that acts on the model variables to both convert those variables to the observed variables and interpolates to the observation locations. In Daley, the symbol Ω is used to denote this operator, whereas Kalnay uses the conventionally accepted symbol H. Thus the analysis equation (12) in Lecture 13 becomes

$$f_A(\mathbf{r}_i) \;=\; f_B(\mathbf{r}_i) \;+\; \underline{W}_i^T [\, \underline{f}_o \;-\; \underline{\underline{\Omega}} \underline{f}_B \,] \qquad (1)$$

which is the same as eq. 5.4.1 in Kalnay:

$$\mathbf{x}_a \;=\; \mathbf{x}_b \;+\; \mathbf{W} [\, \mathbf{y}_o \;-\; H(\mathbf{x}_b) \,] \qquad (2)$$

where $\mathbf{x}_a$ is the maximum likelihood analysis on the grid, $\mathbf{x}_b$ is the background field on the grid, and $\mathbf{y}_o$ are the observations at the stations. The $H(\mathbf{x}_b)$ term operates on the background fields (H could simply be interpolation formulas or the equations of radiative transfer) to produce background estimates at the stations.

Daley then repeats the SOA derivation for a new expression for $\underline{W}_i^T$, which expresses the weights associated with the analysis at the i^{th} grid point. The analysis weights for _all_ analysis gridpoints produces a (K x I) rectangular matrix $\underline{\underline{W}}$ that replaces eq. (13) and is written

$$\underline{\underline{W}} = [\underline{\underline{O}} + \underline{\underline{\Omega}}\,\underline{\underline{B}}\,\underline{\underline{\Omega}}^T]^{-1}\,\underline{\underline{\Omega}}\,\underline{\underline{B}} \qquad (3)$$

which is equivalent to eq. 5.4.19a in Kalnay:

$$\mathbf{W} = \mathbf{B}\,\mathbf{H}^T\,[\mathbf{R} + \mathbf{H}\,\mathbf{B}\,\mathbf{H}^T]^{-1} \qquad (4)$$

where **R** is the observation error covariance matrix.