METR 5303 – Lecture #11

Iteration of the SCM to the Optimal Solution

Reading: Daley, p. 93-96

We have seen that the successive correction method (SCM) using the Barnes weight function has the following features:

a. No background field is needed (although could use one if available)
b. We can pre-determine the amount of detail present in the analysis as a function of the Gaussian filter scale parameter κ and station spacing Δn.
c. No need for an influence radius except to save computer time
d. Only 2 passes or scans needed (although perhaps not the best procedure)
e. Time-weighting of observations is possible

However, the following problems remain:

a. The method works poorly in both data cluster areas and data sparse regions
b. Treats all data the same.
c. (Here) makes no use of background information

The Daley textbook and Barnes references have shown that SCM algorithms converge to the observations at the obs. stations. This is desired only if the data are perfect and there is no background field - a situation that never occurs. Thus, with imperfect obs. and a background field (also imperfect), an "optimal analysis" should be a linear combination of both. In the following, we will use the equations and techniques learned so far to derive an equation for this optimal analysis.

Recall the iteration SCM equation (17) from Lecture 7:

$$f_A^{j+1}(\mathbf{r}_i) = f_A^{j}(\mathbf{r}_i) + \mathbf{W}_i^T [\, \mathbf{f}_o - \mathbf{f}_A^j \,] \qquad (1)$$

(this is the scheme that converges to the observations)

where

$$\mathbf{W}_i^T = \frac{\sum_{k=1}^{K_i} w(r_{ik})}{\sum_{k=1}^{K_i} w(r_{ik}) + \varepsilon_o^2} = \frac{\sum_{k=1}^{K_i} E_b^2\, w(r_{ik})}{\sum_{k=1}^{K_i} E_b^2\, w(r_{ik}) + E_o^2} \qquad (2)$$

since $\varepsilon_o^2 = E_o^2 / E_b^2$, and E_b is not a function of k.

Now multiply (2) by E_o^{-2} / E_o^{-2} and substitute result in Eq. (1):

$$f_A^{j+1}(\mathbf{r}_i) - f_A^{j}(\mathbf{r}_i) = \frac{\sum_{k=1}^{K_i} E_b^2 w\,(r_{ik})\, E_o^{-2}\, [\mathbf{f}_o(\mathbf{r}_k) - \mathbf{f}_A^j(\mathbf{r}_k)\,]}{1 + \sum_{k=1}^{K_i} E_b^2 w\,(r_{ik})\, E_o^{-2}}$$

or

$$f_A^{j+1}(\mathbf{r}_i) - f_A^j(\mathbf{r}_i) = (1 + q_i)^{-1} \sum_{k=1}^{K_i} E_b^2\, w(r_{ik})\, E_o^{-2}\, [\mathbf{f}_o(\mathbf{r}_k) - \mathbf{f}_A^j(\mathbf{r}_k)] \qquad (3)$$

where $q_i = \sum_{k=1}^{K_i} E_b^2\, w(r_{ik})\, E_o^{-2}$

Note that $f_A^{j=0}(\mathbf{r}_i) = f_B(\mathbf{r}_i)$: the initial background field

$f_A^{j=0}(\mathbf{r}_k) = f_B(\mathbf{r}_k)$: forward interpolated background value at station

Eq. (3) is still equivalent to eq. (1). Now we are going to modify (3) in **3** significant ways to produce the "optimal" analysis equation. **First**, we add a "fit to background" term:

$$f_A^{j+1}(\mathbf{r}_i) - f_A^j(\mathbf{r}_i) = (1 + q_i)^{-1} \sum_{k=1}^{K_i} E_b^2\, w(r_{ik})\, E_o^{-2}\, [\mathbf{f}_o(\mathbf{r}_k) - \mathbf{f}_A^j(\mathbf{r}_k)]$$

$$+ (1 + q_i)^{-1} [f_B(\mathbf{r}_i) - f_A^{\mathbf{j}}(\mathbf{r}_i)] \qquad (4)$$

Thus the analysis is now a weighted combination of observed and background values. Unlike Eq. (3), we see that the sum of all weights is now 1. We also note that when $j = 0$, the second term vanishes, therefore if only one pass is used, this term does not appear.

See Fig. 3.10 in Daley for an example of this (when $\varepsilon_o^2 = 1$).

The **second** step is to specify the weights $w(r_{ik})$ based on the background error covariance, rather than a distance-based function, i.e.,

$$w\,(r_{ik}) \;=\; E_b^{-2} < \varepsilon_B(\mathbf{r}_k)\, \varepsilon_B(\mathbf{r}_i) >$$

where $\varepsilon_B(\mathbf{r}_k)$, $\varepsilon_B(\mathbf{r}_i)$ are background errors at $\mathbf{r}_k$, $\mathbf{r}_i$ respectively.

That is, we are assuming that the weight between station k and grid point i is given by the <u>background error covariance normalized by the expected (uniform) background error variance.</u> $w\,(r_{ik})$ is effectively the background error correlation coefficient between background values at locations $\mathbf{r}_k$ and $\mathbf{r}_i$.

We will now put this back into full matrix form.

Define $\mathbf{O}$ as a diagonal obs. error covariance matrix with elements $< \varepsilon_o(\mathbf{r}_k)\, \varepsilon_o(\mathbf{r}_i) >$

Define $\mathbf{B_i}$ as a background error column vector of length K (no. of stations) with elements $< \varepsilon_B(\mathbf{r}_k)\, \varepsilon_B(\mathbf{r}_i) >$.

Now rewrite eq. (4) as

$$f_A^{j+1}(\mathbf{r}_i) \;-\; f_A^{j}(\mathbf{r}_i) \;=\; (1 + q_i)^{-1} \{ \mathbf{B_i^T}\, \mathbf{O}^{-1} [\, \mathbf{f}_o - \mathbf{f}_A^{j} \,] \;+\; [f_B - f_A^{\mathbf{j}}] \} \qquad (5)$$

Now rewrite eq. (5) for the case where the analysis is performed at the observation stations:

$$\mathbf{f}_A^{j+1} - \mathbf{f}_A^{j} = (\mathbf{I} + \mathbf{Q})^{-1} \{ \mathbf{B}\,\mathbf{O}^{-1} [\mathbf{f}_o - \mathbf{f}_A^{j}] + [\mathbf{f}_B - \mathbf{f}_A^{\mathbf{j}}] \} \qquad (6)$$

where **B** is a background error covariance matrix with elements

$b_{kl} = < \varepsilon_B(\mathbf{r}_k)\, \varepsilon_B(\mathbf{r}_l) >$ where $\mathbf{r}_k$ and $\mathbf{r}_l$ are observation locations.

and **Q** is a diagonal matrix with elements $q_k = \sum_{l=1}^{K_i} b_{kl}\, E_o^{-2}$

In the limit, if eq. (6) converges, $\mathbf{f}_A^{j} = \mathbf{f}_A^{j+1} = \mathbf{f}_A^{\infty}$, then the left hand side of (6) is 0.

$$(\mathbf{I} + \mathbf{Q})^{-1} \{ \mathbf{B}\,\mathbf{O}^{-1} [\mathbf{f}_o - \mathbf{f}_A^{j}] + [\mathbf{f}_B - \mathbf{f}_A^{\mathbf{j}}] \} = 0,$$

$$\mathbf{B}\,\mathbf{O}^{-1} [\mathbf{f}_o - \mathbf{f}_A^{\infty}] + \mathbf{f}_B - \mathbf{f}_A^{\mathbf{j}} = 0\,,$$

$$\mathbf{f}_A^{\infty} + \mathbf{B}\,\mathbf{O}^{-1}\mathbf{f}_A^{\infty} = \mathbf{f}_B + \mathbf{B}\,\mathbf{O}^{-1}\,\mathbf{f}_o$$

$$\mathbf{B}^{\mathbf{-1}}\,\mathbf{f}_A^{\infty} + \mathbf{O}^{-1}\mathbf{f}_A^{\infty} = \mathbf{B}^{\mathbf{-1}}\,\mathbf{f}_B + \mathbf{O}^{-1}\,\mathbf{f}_o$$

$$\mathbf{f}_A^{\infty} = \frac{\mathbf{B}^{-1}}{\mathbf{B}^{-1} + \mathbf{O}^{-1}}\mathbf{f}_B + \frac{O^{-1}}{\mathbf{B}^{-1} + \mathbf{O}^{-1}}\mathbf{f}_o$$

As pointed out in lecture 4 going from eq. (38) to eq. (39), the above equation the mathematically equivalent to

$$\mathbf{f}_A^{\infty} = \mathbf{f}_B + \mathbf{B}\,[\,\mathbf{B} + \mathbf{O}\,]^{-1}\,[\,\mathbf{f}_o - \mathbf{f}_B\,] \qquad (7)$$

Eq.(7) is the same as eq. (39) in Lecture 4, i.e. it is an optimal solution (at the observation points) where the analysis is a weighted average of the background and observation, and the weights depend on their respective error covariance.

Recall that eq. (7) is an analysis equation valid at observation locations.

Now go back to eq. (5) (an analysis valid at grid points), and we also have in the limit if the analysis converges

$$f_A^j(\mathbf{r}_i) = f_A^{j+1}(\mathbf{r}_i) = \mathbf{f}_A^{\infty}(\mathbf{r}_i) \quad \text{and eq. (5) becomes}$$

$$(1 + q_i)^{-1} \{ \mathbf{B_i^T} \mathbf{O}^{-1} [\mathbf{f}_o - \mathbf{f}_A^j] + [f_B - f_A^{\mathbf{j}}] \} = 0$$

therefore

$$f_A(\mathbf{r}_i) = f_B(\mathbf{r}_i) + \mathbf{B_i^T} \mathbf{O}^{-1} [\mathbf{f}_o - \mathbf{f}_A]. \qquad (8)$$

Note that in (8), the left hand side $f_A(\mathbf{r}_i)$ is a scalar defined at a grid point, while $\mathbf{f}_A$ on the right hand side is a vector of analysis at the observation locations. We can use eq.(7) to substitute for $\mathbf{f}_A$ in (8),

$$f_A(\mathbf{r}_i) = f_B(\mathbf{r}_i) + \mathbf{B_i^T} \mathbf{O}^{-1} \{\mathbf{f}_o - \mathbf{f}_B - \mathbf{B} [\mathbf{B} + \mathbf{O}]^{-1} [\mathbf{f}_o - \mathbf{f}_B]\},$$

$$f_A(\mathbf{r}_i) = f_B(\mathbf{r}_i) + \mathbf{B_i^T} \mathbf{O}^{-1} \{ [\mathbf{B} + \mathbf{O}] [\mathbf{B} + \mathbf{O}]^{-1} - \mathbf{B} [\mathbf{B} + \mathbf{O}]^{-1} \} [\mathbf{f}_o - \mathbf{f}_B],$$

$$f_A(\mathbf{r}_i) = f_B(\mathbf{r}_i) + \mathbf{B_i^T} \mathbf{O}^{-1} \mathbf{O} [\mathbf{B} + \mathbf{O}]^{-1} [\mathbf{f}_o - \mathbf{f}_B],$$

Therefore $$f_A(\mathbf{r}_i) = f_B(\mathbf{r}_i) + \mathbf{B_i}^\mathbf{T} [\ \mathbf{B} + \mathbf{O}\ ]^{-1} [\mathbf{f}_o - \mathbf{f}_B]. \quad (9)$$

Eq. (9) is our desired "optimal analysis" defined at grid point $\mathbf{r}_i$, obtained by iterating a specific form of the SCM algorithm.

Bratseth (1986) developed another version of a SCM algorithm that converge to the same optimal analysis. See Appendix F in Daley.

Later in this course, we will derive the above "optimal analysis" as a minimum variance solution, where it will be call "optimal interpolation (OI)" solution.