Role of observation and background error covariances in the retrieval

A retrieval is the 'optimal' estimate of atmospheric variables in a vertical column, at a particular time. The information used to form a retrieval is (i) a set of observations and (ii) an a-priori state. Indirect observations, for instance radiance measurements from satellites, form the source of observational information in a retrieval, but other observations can also be considered.

All data are inexact (random errors, systematic errors, etc.). This includes the observations and the a-priori. Thus we have to deal with an inexact inverse problem where we have to consider the data itself and its error characteristics. Accurate knowledge of the error charcteristics is important. These notes set out to show how knowledge of the error characteristics can significantly affect the retrieval.

The solution to the inverse problem leads to a formula for the retrieval. In the linear case (Eq. (6) of Eyre (1989)), the retrieval can be written as the a-priori, $\boldsymbol{x}^b$, plus a correction term due to the information provided by the observations.

$$\boldsymbol{x} = \boldsymbol{x}^b + \mathbf{C}\mathbf{K}^T(\mathbf{K}\mathbf{C}\mathbf{K}^T + \mathbf{E})^{-1}(\boldsymbol{y}^m - \boldsymbol{y}(\boldsymbol{x}^b)) \tag{1}$$

A formula like this appears in data assimilation (optimal interpolation, Kalman filter, etc.). It is useful to understand the role of the error covariance matrices.

Definitions of the terms

$\boldsymbol{x}$ is the retrieval (R^n).

$\boldsymbol{x}^b$ is the background profile (R^n). It encompasses our knowledge of the system before observations are considered (for this reason, it is also formally known as the a-priori). In practice it is a vertical column extracted from the (3d) numerical forecast fields, valid at the instant of the retrieval. This represents our best knowledge of the atmospheric column before observations are considered.

$\boldsymbol{y}^m$ are the observations ('m' is for measurements) and $\boldsymbol{y}(\boldsymbol{x})$ are the observations predicted from a given state $\boldsymbol{x}$ (both $\boldsymbol{y}^m$ and $\boldsymbol{y}$ are R^m). $\boldsymbol{y}(\boldsymbol{x})$ is called the forward model or observation operator. For satellite radiance measurements it encompasses the radiative transfer equations that predict the radiances measured by the satellite in each part of the electromagnetic spectrum measured. For simplicity, point observations are considered in the examples below, for which $\boldsymbol{y}(\boldsymbol{x})$ is trivial.

$\mathbf{K}$ is the matrix of first derivatives of the forward model, $\boldsymbol{y}(\boldsymbol{x})$, with respect to $\boldsymbol{x}$ ($\mathrm{R}^{m\times n}$),

$$\mathbf{K} = \frac{\partial \boldsymbol{y}(\boldsymbol{x})}{\partial \boldsymbol{x}}, \qquad \text{ie} \qquad \mathbf{K}_{ij} = \frac{\partial \boldsymbol{y}_i}{\partial \boldsymbol{x}_j}. \tag{2}$$

N.B. think of this as the linear term of a multi-variable Taylor expansion of the forward model $\boldsymbol{y}$,

$$\boldsymbol{y}(\boldsymbol{x}) = \boldsymbol{y}(\boldsymbol{x}^0) + \mathbf{K}(\boldsymbol{x} - \boldsymbol{x}^0) + \ldots \tag{3}$$

In order to see this, expand for the ith component of $\boldsymbol{y}(\boldsymbol{x})$,

$$\boldsymbol{y}_i(\boldsymbol{x}) = \boldsymbol{y}_i(\boldsymbol{x}^0) + \sum_{j=1}^{n}\mathbf{K}_{ij}(\boldsymbol{x}_j - \boldsymbol{x}_j^0) = \boldsymbol{y}_i(\boldsymbol{x}^0) + \sum_{j=1}^{n}\frac{\partial \boldsymbol{y}_i}{\partial \boldsymbol{x}_j}(\boldsymbol{x}_j - \boldsymbol{x}_j^0). \tag{4}$$

$\mathbf{K}$ is called the Jacobian, and its rows are called weighting functions.

C is the background error covariance matrix ($R^{n \times n}$). It describes the uncertainty of $\boldsymbol{x}^b$. **C** has diagonal and off-diagonal terms, it is symmetric, $\mathbf{C} = \mathbf{C}^T$, and is positive definite,

$$\mathbf{C} = \begin{pmatrix} \mathbf{C}_{11} & \mathbf{C}_{12} & \dots & \mathbf{C}_{1n} \\ \mathbf{C}_{21} & \mathbf{C}_{22} & \dots & \mathbf{C}_{2n} \\ \dots & \dots & \dots & \dots \\ \mathbf{C}_{n1} & \mathbf{C}_{n2} & \dots & \mathbf{C}_{nn} \end{pmatrix} = \begin{pmatrix} \sigma_1^{b2} & \sigma_1^b\sigma_2^b\mathrm{cor}(1,2) & \dots & \sigma_1^b\sigma_n^b\mathrm{cor}(1,n) \\ \sigma_2^b\sigma_1^b\mathrm{cor}(2,1) & \sigma_2^{b2} & \dots & \sigma_2^b\sigma_n^b\mathrm{cor}(2,n) \\ \dots & \dots & \dots & \dots \\ \sigma_n^b\sigma_1^b\mathrm{cor}(n,1) & \sigma_n^b\sigma_2^b\mathrm{cor}(n,2) & \dots & \sigma_n^{b2} \end{pmatrix} \quad (5)$$

$\mathrm{cor}(i, j)$ is the statistical correlation of errors in $\boldsymbol{x}^b$ between components i and j.

σ_i is the standard deviation of component *i of* $\boldsymbol{x}^b$.

The diagonal elements describe variances (square of the standard deviations) of each component of $\boldsymbol{x}^b$, and the off-diagonal terms describe the covariances between different components. Background errors usually have important diagonal and off-diagonal terms.

E is the observation error covariance matrix ($R^{m \times m}$). It describes the uncertainty of the observations (and also, strictly speaking, of the forward model). It is usually considered to be diagonal (ie errors between observations are uncorrelated; the diagonal elements are the variances of the observation uncertainties).

$$\mathbf{E} = \begin{pmatrix} \sigma_1^{ob2} & 0 & \dots & 0 \\ 0 & \sigma_2^{ob2} & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & \sigma_m^{ob2} \end{pmatrix}. \quad (6)$$

One observation made at a level

Forget satellite observations for now, make a single point observation ($m = 1$) of one element of $\boldsymbol{x}$ (say component k). $\boldsymbol{y}(\boldsymbol{x})$ just picks-out element k of $\boldsymbol{x}$. The **K**-operator has the following form ($R^{1 \times n}$) (c.f. Eq. (2)),

$$\mathbf{K} = (0 \;\; 0 \;\; \dots \;\; 0 \;\; 1 \;\; 0 \;\; \dots \;\; 0 \;\; 0) \quad (7)$$

(the '1' is at position 'k'). **E** is $R^{1 \times 1}$, ie just a scalar and consists of the variance of the observation. Note the following parts of Eq. (1),

$$\boldsymbol{y}_1^m = y^m, \qquad \mathbf{E}_{11} = \sigma^{ob^2}, \quad (8)$$

$$\mathbf{KCK}^T = \mathbf{C}_{kk} = \sigma_k^{b^2} \quad (9)$$

Putting this into Eq. (1),

$$\boldsymbol{x} = \boldsymbol{x}^b + \mathbf{C}\begin{pmatrix} 0 \\ 0 \\ \dots \\ 0 \\ 1 \\ 0 \\ \dots \\ 0 \\ 0 \end{pmatrix} \frac{y^m - \boldsymbol{x}_k^b}{\sigma_k^{b2} + \sigma^{ob2}} = \boldsymbol{x}^b + \begin{pmatrix} \sigma_1^b \sigma_k^b \mathrm{cor}(1, k) \\ \sigma_2^b \sigma_k^b \mathrm{cor}(2, k) \\ \dots \\ \sigma_{k-1}^b \sigma_k^b \mathrm{cor}(k-1, k) \\ \sigma_k^{b2} \\ \sigma_{k+1}^b \sigma_k^b \mathrm{cor}(k+1, k) \\ \dots \\ \sigma_{n-1}^b \sigma_k^b \mathrm{cor}(n-1, k) \\ \sigma_n^b \sigma_k^b \mathrm{cor}(n, k) \end{pmatrix} \frac{y^m - \boldsymbol{x}_k^b}{\sigma_k^{b2} + \sigma^{ob2}} \tag{10}$$

Component-by-component, Eq. (10) is,

$$\boldsymbol{x}_l = \boldsymbol{x}_l^b + \mathbf{C}_{lk} \frac{y^m - \boldsymbol{x}_k^b}{\sigma_k^{b2} + \sigma^{ob2}}, \tag{11}$$

The correction to $\boldsymbol{x}^b$ due to the observation is proportional to a column of matrix **C**. Note:

- The observation increment undergoes a 'spreading' from component k - where the observation is made - to other components by the background error covariance matrix **C**.
- If the observation error is large ($\sigma^{ob} >> \sigma_k^b$), then there is little trust in the information contained in the observation. In that case the correction due to the observation is small, and the retrieval essentially reverts to the background, $\boldsymbol{x} = \boldsymbol{x}^b$.
- If the background error is large ($\sigma_k^b >> \sigma^{ob}$) then the value of the retrieval at the observation position will be set to the observation value, and other components will be modified according to the kth column of **C** - see Eq. (12) below. Furthermore if there are no background error correlations, then only level k will be modified.

$$\boldsymbol{x} = \boldsymbol{x}^b + \begin{pmatrix} \sigma_1^b / \sigma_k^b \ \mathrm{cor}(1, k) \\ \sigma_2^b / \sigma_k^b \ \mathrm{cor}(2, k) \\ \dots \\ \sigma_{k-1}^b / \sigma_k^b \ \mathrm{cor}(k-1, k) \\ 1 \\ \sigma_{k+1}^b / \sigma_k^b \ \mathrm{cor}(k+1, k) \\ \dots \\ \sigma_{n-1}^b / \sigma_k^b \ \mathrm{cor}(n-1, k) \\ \sigma_n^b / \sigma_k^b \ \mathrm{cor}(n, k) \end{pmatrix} (y^m - \boldsymbol{x}_k^b) \tag{12}$$

Two observations made at two levels

Make two point observations ($m = 2$), each of a different element of $\boldsymbol{x}$ (say components k_1 and k_2). $\boldsymbol{y}(\boldsymbol{x})$ just picks-out elements k_1 and k_2 of $\boldsymbol{x}$. The **K**-operator has the following form ($\mathrm{R}^{2\times n}$) (c.f. Eq. (2)),

$$\mathbf{K} = \begin{pmatrix} 0 & 0 & \dots & 0 & 1 & 0 & \dots & 0 & 0 & 0 & \dots & 0 & 0 \\ 0 & 0 & \dots & 0 & 0 & 0 & \dots & 0 & 1 & 0 & \dots & 0 & 0 \end{pmatrix} \tag{13}$$

(the '1' on the first row is at position 'k_1' and on the second row at position 'k_2'). **E** is $\mathrm{R}^{2\times 2}$. Note the following part of Eq. (1),

$$\mathbf{KCK}^T = \begin{pmatrix} \mathbf{C}_{k_1k_1} & \mathbf{C}_{k_1k_2} \\ \mathbf{C}_{k_2k_1} & \mathbf{C}_{k_2k_2} \end{pmatrix}, \qquad \mathbf{KCK}^T + \mathbf{E} = \begin{pmatrix} \sigma_{k_1}^{b\ 2} + \sigma_1^{ob2} & \mathbf{C}_{k_1k_2} \\ \mathbf{C}_{k_1k_2} & \sigma_{k_2}^{b\ 2} + \sigma_2^{ob2} \end{pmatrix} \tag{14}$$

Inverting this matric and putting it into Eq. (1),

$$\boldsymbol{x} = \boldsymbol{x}^b + \mathbf{CK}^T \begin{pmatrix} \sigma_{k_1}^{b\ 2} + \sigma_1^{ob2} & \mathbf{C}_{k_1k_2} \\ \mathbf{C}_{k_1k_2} & \sigma_{k_2}^{b\ 2} + \sigma_2^{ob2} \end{pmatrix}^{-1} \begin{pmatrix} \boldsymbol{y}_1^m - \boldsymbol{x}_{k_1}^b \\ \boldsymbol{y}_2^m - \boldsymbol{x}_{k_2}^b \end{pmatrix} \tag{15}$$

$$= \boldsymbol{x}^b + \mathbf{CK}^T \frac{1}{(\sigma_{k_1}^{b\ 2} + \sigma_1^{ob2})(\sigma_{k_2}^{b\ 2} + \sigma_2^{ob2}) - \mathbf{C}_{k_1k_2}^2} \begin{pmatrix} \sigma_{k_2}^{b\ 2} + \sigma_2^{ob2} & -\mathbf{C}_{k_1k_2} \\ -\mathbf{C}_{k_1k_2} & \sigma_{k_1}^{b\ 2} + \sigma_1^{ob2} \end{pmatrix} \begin{pmatrix} \boldsymbol{y}_1^m - \boldsymbol{x}_{k_1}^b \\ \boldsymbol{y}_2^m - \boldsymbol{x}_{k_2}^b \end{pmatrix} \tag{16}$$

$$= \boldsymbol{x}^b + \mathbf{CK}^T \frac{1}{(\sigma_{k_1}^{b\ 2} + \sigma_1^{ob2})(\sigma_{k_2}^{b\ 2} + \sigma_2^{ob2}) - \mathbf{C}_{k_1k_2}^2} \begin{pmatrix} (\sigma_{k_2}^{b\ 2} + \sigma_2^{ob2})(\boldsymbol{y}_1^m - \boldsymbol{x}_{k_1}^b) - \mathbf{C}_{k_1k_2}(\boldsymbol{y}_2^m - \boldsymbol{x}_{k_2}^b) \\ (\sigma_{k_1}^{b\ 2} + \sigma_1^{ob2})(\boldsymbol{y}_2^m - \boldsymbol{x}_{k_2}^b) - \mathbf{C}_{k_1k_2}(\boldsymbol{y}_1^m - \boldsymbol{x}_{k_1}^b) \end{pmatrix} \tag{17}$$

$$= \boldsymbol{x}^b + \mathbf{C} \frac{1}{(\sigma_{k_1}^{b\ 2} + \sigma_1^{ob2})(\sigma_{k_2}^{b\ 2} + \sigma_2^{ob2}) - \mathbf{C}_{k_1k_2}^2} \begin{pmatrix} 0 \\ 0 \\ \cdots \\ 0 \\ (\sigma_{k_2}^{b\ 2} + \sigma_2^{ob2})(\boldsymbol{y}_1^m - \boldsymbol{x}_{k_1}^b) - \mathbf{C}_{k_1k_2}(\boldsymbol{y}_2^m - \boldsymbol{x}_{k_2}^b) \\ 0 \\ \cdots \\ 0 \\ (\sigma_{k_1}^{b\ 2} + \sigma_1^{ob2})(\boldsymbol{y}_2^m - \boldsymbol{y}_2(\boldsymbol{x}^b)) - \mathbf{C}_{k_1k_2}(\boldsymbol{y}_1^m - \boldsymbol{x}_{k_1}^b) \\ 0 \\ \cdots \\ 0 \\ 0 \end{pmatrix} \tag{18}$$

Notes:

- As in the single observation case, the observation increment undergoes a spreading by the background error covariance matrix **C**.
- The two observations 'see' each other due to cross correlations in **C**, even though the observation errors are themselves uncorrelated. If, additionally, there were no correlations in the background errors, the two observations would not 'see' eachother.

R.N.B., Oct 06.