# Selection of Data Assimilation Consistency Diagnostics

R.N.B. February 2011, March 2012

## 1. "Bennett-Talagrand theorem"

(Based on notes from T. Payne, MetO.) There is a very simple result that gives the expected value of the cost function at the minimum which shall now be developed. Assume data assimilation system is optimal (e.g. all error covariance matrices are correctly specified). Then

$$\mathbf{x}_\mathrm{a} - \mathbf{x}_\mathrm{b} = \mathbf{K}(\mathbf{y} - \mathbf{H}\mathbf{x}_\mathrm{b}), \tag{1.1}$$

$$\text{where} \qquad \mathbf{K} = \mathbf{B}\mathbf{H}^\mathrm{T}(\mathbf{H}\mathbf{B}\mathbf{H}^\mathrm{T} + \mathbf{R})^{-1}. \tag{1.2}$$

We wish to evaluate the expected value of the cost function at $\mathbf{x} = \mathbf{x}_\mathrm{a}$. This expected value is written $E[J(\mathbf{x}_\mathrm{a})]$ and the cost function at the analysis is (given a specific background state and set of observations)

$$J(\mathbf{x}_\mathrm{a}) = J_\mathrm{b}(\mathbf{x}_\mathrm{a}) + J_\mathrm{o}(\mathbf{x}_\mathrm{a}), \tag{1.3}$$

$$\text{where} \qquad J_\mathrm{b}(\mathbf{x}_\mathrm{a}) = \frac{1}{2}(\mathbf{x}_\mathrm{a} - \mathbf{x}_\mathrm{b})^\mathrm{T}\mathbf{B}^{-1}(\mathbf{x}_\mathrm{a} - \mathbf{x}_\mathrm{b}), \tag{1.4}$$

$$\text{and} \qquad J_\mathrm{o}(\mathbf{x}_\mathrm{a}) = \frac{1}{2}(\mathbf{y} - \mathbf{H}\mathbf{x}_\mathrm{a})^\mathrm{T}\mathbf{R}^{-1}(\mathbf{y} - \mathbf{H}\mathbf{x}_\mathrm{a}). \tag{1.5}$$

The analysis, background and observation errors are (again given a specific background state and set of observations)

$$\varepsilon_\mathrm{a} = \mathbf{x}_\mathrm{a} - \mathbf{x}_\mathrm{t}, \tag{1.6}$$

$$\varepsilon_\mathrm{b} = \mathbf{x}_\mathrm{b} - \mathbf{x}_\mathrm{t}, \tag{1.7}$$

$$\varepsilon_\mathrm{o} = \mathbf{y} - \mathbf{H}\mathbf{x}_\mathrm{t}. \tag{1.8}$$

The analysis error can be developed as follows (using (1.1), (1.7) and (1.8))

$$\begin{aligned}
\varepsilon_\mathrm{a} &= \mathbf{x}_\mathrm{a} - \mathbf{x}_\mathrm{b} + \varepsilon_\mathrm{b}, \\
&= \mathbf{K}(\mathbf{y} - \mathbf{H}\mathbf{x}_\mathrm{b}) + \varepsilon_\mathrm{b}, \\
&= \mathbf{K}(\mathbf{y} - \mathbf{H}(\mathbf{x}_\mathrm{b} - \mathbf{x}_\mathrm{t}) - \mathbf{H}\mathbf{x}_\mathrm{t}) + \varepsilon_\mathrm{b}, \\
&= \mathbf{K}(\varepsilon_\mathrm{o} - \mathbf{H}\varepsilon_\mathrm{b}) + \varepsilon_\mathrm{b}, \\
&= (\mathbf{I} - \mathbf{K}\mathbf{H})\varepsilon_\mathrm{b} + \mathbf{K}\varepsilon_\mathrm{o}.
\end{aligned} \tag{1.9}$$

Equations (1.4) and (1.5) are inner products. To evaluate them, the following identity is useful

$$\mathbf{u}^\mathrm{T}\mathbf{C}\mathbf{v} = \sum_{i,j}\mathbf{u}_i\mathbf{C}_{ij}\mathbf{v}_j = \mathrm{tr}(\mathbf{C}\mathbf{v}\mathbf{u}^\mathrm{T}). \tag{1.10}$$

The expectation of the <u>background term</u> (1.4) is

$$\begin{aligned}
E[J_\mathrm{b}(\mathbf{x}_\mathrm{a})] &= \frac{1}{2}E[\mathrm{tr}(\mathbf{B}^{-1}(\mathbf{x}_\mathrm{a} - \mathbf{x}_\mathrm{b})(\mathbf{x}_\mathrm{a} - \mathbf{x}_\mathrm{b})^\mathrm{T})], \\
&= \frac{1}{2}\,\mathrm{tr}(\mathbf{B}^{-1}E[(\mathbf{x}_\mathrm{a} - \mathbf{x}_\mathrm{b})(\mathbf{x}_\mathrm{a} - \mathbf{x}_\mathrm{b})^\mathrm{T}]), \\
&= \frac{1}{2}\,\mathrm{tr}(\mathbf{B}^{-1}E[(\varepsilon_\mathrm{a} - \varepsilon_\mathrm{b})(\varepsilon_\mathrm{a} - \varepsilon_\mathrm{b})^\mathrm{T}]),
\end{aligned} \tag{1.11}$$

where (1.6) and (1.7) have been used for the last line. Part of the last line is the expression

$E[(\varepsilon_a - \varepsilon_b)(\varepsilon_a - \varepsilon_b)^T]$ which may be developed using (1.9)

$$
\begin{aligned}
E[(\varepsilon_a - \varepsilon_b)(\varepsilon_a - \varepsilon_b)^T] &= E[\varepsilon_a\varepsilon_a^T + \varepsilon_b\varepsilon_b^T - \varepsilon_a\varepsilon_b^T - \varepsilon_b\varepsilon_a^T], \\
&= (\mathbf{I} - \mathbf{KH})E[\varepsilon_b\varepsilon_b^T](\mathbf{I} - \mathbf{KH})^T + \mathbf{K}E[\varepsilon_o\varepsilon_o^T]\mathbf{K}^T + E[\varepsilon_b\varepsilon_b^T] - \\
&\quad (\mathbf{I} - \mathbf{KH})E[\varepsilon_b\varepsilon_b^T] - E[\varepsilon_b\varepsilon_b^T](\mathbf{I} - \mathbf{KH})^T, \\
&= (\mathbf{I} - \mathbf{KH})\mathbf{B}(\mathbf{I} - \mathbf{KH})^T + \mathbf{KRK}^T + \mathbf{B} - (\mathbf{I} - \mathbf{KH})\mathbf{B} \\
&\quad -\mathbf{B}(\mathbf{I} - \mathbf{KH})^T, \\
&= \mathbf{B} + \mathbf{KHB}(\mathbf{KH})^T - \mathbf{B}(\mathbf{KH})^T - \mathbf{KHB} + \mathbf{KRK}^T + \mathbf{B} - \mathbf{B} \\
&\quad +\mathbf{KHB} - \mathbf{B} + \mathbf{B}(\mathbf{KH})^T, \\
&= \mathbf{KHB}(\mathbf{KH})^T + \mathbf{KRK}^T. \qquad (1.12)
\end{aligned}
$$

These steps assume that background and observation errors are mutually uncorrelated. Using the definition of **K** (1.2)

$$
\begin{aligned}
E[(\varepsilon_a - \varepsilon_b)(\varepsilon_a - \varepsilon_b)^T] &= \mathbf{BH}^T(\mathbf{HBH}^T + \mathbf{R})^{-1}\mathbf{HB}(\mathbf{BH}^T(\mathbf{HBH}^T + \mathbf{R})^{-1}\mathbf{H})^T + \\
&\quad \mathbf{BH}^T(\mathbf{HBH}^T + \mathbf{R})^{-1}\mathbf{R}(\mathbf{BH}^T(\mathbf{HBH}^T + \mathbf{R})^{-1})^T, \\
&= \mathbf{BH}^T(\mathbf{HBH}^T + \mathbf{R})^{-1}\mathbf{HBH}^T(\mathbf{HBH}^T + \mathbf{R})^{-1}\mathbf{HB} + \\
&\quad \mathbf{BH}^T(\mathbf{HBH}^T + \mathbf{R})^{-1}\mathbf{R}(\mathbf{HBH}^T + \mathbf{R})^{-1}\mathbf{HB} \\
&= \mathbf{BH}^T(\mathbf{HBH}^T + \mathbf{R})^{-1}\mathbf{HB} = \mathbf{KHB}. \qquad (1.13)
\end{aligned}
$$

Inserting (1.13) into (1.11) gives

$$E[J_b(\mathbf{x}_a)] = \frac{1}{2}\,\mathrm{tr}(\mathbf{B}^{-1}\mathbf{KHB}). \qquad (1.14)$$

Note the following identity, which holds for matrices **E** and **F**, where **E** is $r \times s$ and **F** is $s \times r$

$$\mathrm{tr}(\mathbf{EF}) = \sum_{j=1}^{r}\sum_{i=1}^{s}\mathbf{E}_{ji}\mathbf{F}_{ij} = \sum_{i=1}^{s}\sum_{j=1}^{r}\mathbf{F}_{ij}\mathbf{E}_{ji} = \mathrm{tr}(\mathbf{FE}), \qquad (1.15)$$

ie, the order of the operators inside the trace can be reversed. Applying (1.15) to (1.14) gives

$$E[J_b(\mathbf{x}_a)] = \frac{1}{2}\,\mathrm{tr}(\mathbf{KH}). \qquad (1.16)$$

Using (1.10) the expectation of the <u>observation term</u> (1.5) is

$$
\begin{aligned}
E[J_o(\mathbf{x}_a)] &= \frac{1}{2}E[\mathrm{tr}(\mathbf{R}^{-1}(\mathbf{y} - \mathbf{Hx}_a)(\mathbf{y} - \mathbf{Hx}_a)^T)], \\
&= \frac{1}{2}\,\mathrm{tr}(\mathbf{R}^{-1}E[(\mathbf{y} - \mathbf{Hx}_a)(\mathbf{y} - \mathbf{Hx}_a)^T]), \\
&= \frac{1}{2}\,\mathrm{tr}(\mathbf{R}^{-1}E[(\varepsilon_o - \mathbf{H}\varepsilon_a)(\varepsilon_o - \mathbf{H}\varepsilon_a)^T]). \qquad (1.17)
\end{aligned}
$$

where (1.6) and (1.8) have been used for the last line. Part of the last line is the expression $E[(\varepsilon_o - \mathbf{H}\varepsilon_a)(\varepsilon_o - \mathbf{H}\varepsilon_a)^T]$ which may be developed using (1.9)

$$
\begin{aligned}
E[(\varepsilon_o - \mathbf{H}\varepsilon_a)(\varepsilon_o - \mathbf{H}\varepsilon_a)^T] &= \mathbf{H}E[\varepsilon_a\varepsilon_a^T]\mathbf{H}^T + E[\varepsilon_o\varepsilon_o^T] - \mathbf{H}E[\varepsilon_a\varepsilon_o^T] - E[\varepsilon_o\varepsilon_a^T]\mathbf{H}^T, \\
&= \mathbf{H}\{(\mathbf{I} - \mathbf{KH})E[\varepsilon_b\varepsilon_b^T](\mathbf{I} - \mathbf{KH})^T + \mathbf{K}E[\varepsilon_o\varepsilon_o^T]\mathbf{K}^T\}\mathbf{H}^T \\
&\quad +E[\varepsilon_o\varepsilon_o^T] - \mathbf{HK}E[\varepsilon_o\varepsilon_o^T] - E[\varepsilon_o\varepsilon_o^T]\mathbf{K}^T\mathbf{H}^T,
\end{aligned}
$$

$$
\begin{aligned}
&= \mathbf{H}\{(\mathbf{I} - \mathbf{KH})\mathbf{B}(\mathbf{I} - \mathbf{KH})^{\mathrm{T}} + \mathbf{KRK}^{\mathrm{T}}\}\mathbf{H}^{\mathrm{T}} + \mathbf{R} \\
&\quad -\mathbf{HKR} - \mathbf{RK}^{\mathrm{T}}\mathbf{H}^{\mathrm{T}}, \\
&= \mathbf{HBH}^{\mathrm{T}} + \mathbf{HKHB}(\mathbf{KH})^{\mathrm{T}}\mathbf{H}^{\mathrm{T}} - \mathbf{HB}(\mathbf{KH})^{\mathrm{T}}\mathbf{H}^{\mathrm{T}} - \mathbf{HKHBH}^{\mathrm{T}} \\
&\quad +\mathbf{HKRK}^{\mathrm{T}}\mathbf{H}^{\mathrm{T}} + \mathbf{R} - \mathbf{HKR} - \mathbf{RK}^{\mathrm{T}}\mathbf{H}^{\mathrm{T}}.
\end{aligned} \tag{1.18}
$$

These steps assume that background and observation errors are mutually uncorrelated. Using the definition of **K** (1.2)

$$
\begin{aligned}
E[(\varepsilon_{\mathrm{o}} - \mathbf{H}\varepsilon_{\mathrm{a}})(\varepsilon_{\mathrm{o}} - \mathbf{H}\varepsilon_{\mathrm{a}})^{\mathrm{T}}] &= \mathbf{HBH}^{\mathrm{T}} + \mathbf{HBH}^{\mathrm{T}}(\mathbf{HBH}^{\mathrm{T}} + \mathbf{R})^{-1}\mathbf{HBH}^{\mathrm{T}}(\mathbf{HBH}^{\mathrm{T}} + \mathbf{R})^{-1}\mathbf{HBH}^{\mathrm{T}} \\
&\quad -\mathbf{HBH}^{\mathrm{T}}(\mathbf{HBH}^{\mathrm{T}} + \mathbf{R})^{-1}\mathbf{HBH}^{\mathrm{T}} - \mathbf{HBH}^{\mathrm{T}}(\mathbf{HBH}^{\mathrm{T}} + \mathbf{R})^{-1}\mathbf{HBH}^{\mathrm{T}} \\
&\quad +\mathbf{HBH}^{\mathrm{T}}(\mathbf{HBH}^{\mathrm{T}} + \mathbf{R})^{-1}\mathbf{R}(\mathbf{HBH}^{\mathrm{T}} + \mathbf{R})^{-1}\mathbf{HBH}^{\mathrm{T}} + \mathbf{R} \\
&\quad -\mathbf{HBH}^{\mathrm{T}}(\mathbf{HBH}^{\mathrm{T}} + \mathbf{R})^{-1}\mathbf{R} - \mathbf{R}(\mathbf{HBH}^{\mathrm{T}} + \mathbf{R})^{-1}\mathbf{HBH}^{\mathrm{T}}.
\end{aligned}
$$

Merging the 2nd and 5th terms leads to

$$
\begin{aligned}
E[(\varepsilon_{\mathrm{o}} - \mathbf{H}\varepsilon_{\mathrm{a}})(\varepsilon_{\mathrm{o}} - \mathbf{H}\varepsilon_{\mathrm{a}})^{\mathrm{T}}] &= \mathbf{HBH}^{\mathrm{T}} + \mathbf{HBH}^{\mathrm{T}}(\mathbf{HBH}^{\mathrm{T}} + \mathbf{R})^{-1}\mathbf{HBH}^{\mathrm{T}} \\
&\quad -\mathbf{HBH}^{\mathrm{T}}(\mathbf{HBH}^{\mathrm{T}} + \mathbf{R})^{-1}\mathbf{HBH}^{\mathrm{T}} - \mathbf{HBH}^{\mathrm{T}}(\mathbf{HBH}^{\mathrm{T}} + \mathbf{R})^{-1}\mathbf{HBH}^{\mathrm{T}} \\
&\quad +\mathbf{R} - \mathbf{HBH}^{\mathrm{T}}(\mathbf{HBH}^{\mathrm{T}} + \mathbf{R})^{-1}\mathbf{R} - \mathbf{R}(\mathbf{HBH}^{\mathrm{T}} + \mathbf{R})^{-1}\mathbf{HBH}^{\mathrm{T}}.
\end{aligned}
$$

Further simplifications can be made by merging the 3rd and 6th terms and the 4th and 7th terms

$$
\begin{aligned}
E[(\varepsilon_{\mathrm{o}} - \mathbf{H}\varepsilon_{\mathrm{a}})(\varepsilon_{\mathrm{o}} - \mathbf{H}\varepsilon_{\mathrm{a}})^{\mathrm{T}}] &= \mathbf{HBH}^{\mathrm{T}} + \mathbf{HBH}^{\mathrm{T}}(\mathbf{HBH}^{\mathrm{T}} + \mathbf{R})^{-1}\mathbf{HBH}^{\mathrm{T}} - \mathbf{HBH}^{\mathrm{T}} - \mathbf{HBH}^{\mathrm{T}} + \mathbf{R}, \\
&= \mathbf{HBH}^{\mathrm{T}}\{(\mathbf{HBH}^{\mathrm{T}} + \mathbf{R})^{-1}\mathbf{HBH}^{\mathrm{T}} - \mathbf{I}\} + \mathbf{R}.
\end{aligned} \tag{1.19}
$$

Consider the term inside the curly brackets in (1.19)

$$
\begin{aligned}
(\mathbf{HBH}^{\mathrm{T}} + \mathbf{R})^{-1}\mathbf{HBH}^{\mathrm{T}} - \mathbf{I} &= (\mathbf{HBH}^{\mathrm{T}} + \mathbf{R})^{-1}\mathbf{HBH}^{\mathrm{T}} - (\mathbf{HBH}^{\mathrm{T}} + \mathbf{R})^{-1}(\mathbf{HBH}^{\mathrm{T}} + \mathbf{R}), \\
&= (\mathbf{HBH}^{\mathrm{T}} + \mathbf{R})^{-1}[\mathbf{HBH}^{\mathrm{T}} - \mathbf{HBH}^{\mathrm{T}} - \mathbf{R}], \\
&= -(\mathbf{HBH}^{\mathrm{T}} + \mathbf{R})^{-1}\mathbf{R}.
\end{aligned} \tag{1.20}
$$

Using (1.20) to rewrite (1.19)

$$
E[(\varepsilon_{\mathrm{o}} - \mathbf{H}\varepsilon_{\mathrm{a}})(\varepsilon_{\mathrm{o}} - \mathbf{H}\varepsilon_{\mathrm{a}})^{\mathrm{T}}] = -\mathbf{HBH}^{\mathrm{T}}(\mathbf{HBH}^{\mathrm{T}} + \mathbf{R})^{-1}\mathbf{R} + \mathbf{R}. \tag{1.21}
$$

Substituting (1.21) into (1.17) and then using (1.15) gives

$$
\begin{aligned}
E[J_{\mathrm{o}}(\mathbf{x}_{\mathrm{a}})] &= \frac{1}{2}\,\mathrm{tr}\,(\mathbf{R}^{-1}[-\mathbf{HBH}^{\mathrm{T}}(\mathbf{HBH}^{\mathrm{T}} + \mathbf{R})^{-1}\mathbf{R} + \mathbf{R}]), \\
&= \frac{1}{2}\,\mathrm{tr}\,(-\mathbf{HBH}^{\mathrm{T}}(\mathbf{HBH}^{\mathrm{T}} + \mathbf{R})^{-1} + \mathbf{I}), \\
&= \frac{1}{2}\,\mathrm{tr}\,(-\mathbf{HK} + \mathbf{I}) = \frac{1}{2}\big(-\,\mathrm{tr}\,(\mathbf{HK}) + p\big) = \frac{1}{2}\big(-\,\mathrm{tr}\,(\mathbf{KH}) + p\big),
\end{aligned} \tag{1.22}
$$

where $p$ is the number of observations. The <u>sum of the background and observation terms</u> is (using (1.3), (1.16) and (1.22))

$$
\begin{aligned}
E(J(\mathbf{x}_{\mathrm{a}})) &= E(J_{\mathrm{b}}(\mathbf{x}_{\mathrm{a}})) + E(J_{\mathrm{o}}(\mathbf{x}_{\mathrm{a}})), \\
&= \frac{1}{2}\big(\mathrm{tr}\,(\mathbf{KH}) - \mathrm{tr}\,(\mathbf{KH}) + p\big) = \frac{p}{2}.
\end{aligned} \tag{1.23}
$$

The involved derivation leads to the simple result that the expectation of the minimum of the cost function has value equal to half the number of observations. Some people have called this the Bennett-Talagrand theorem.

## 2. "Desrozier diagnostics"

How can we test the consistency of VAR?

Analysis increments

$$\delta \mathbf{x}_a = \mathbf{x}_a - \mathbf{x}_b = \hat{\mathbf{K}} \mathbf{d}_b^o. \tag{2.1}$$

Kalman gain

$$\hat{\mathbf{K}} = \hat{\mathbf{B}} \mathbf{H}^T (\mathbf{H} \hat{\mathbf{B}} \mathbf{H}^T + \hat{\mathbf{R}})^{-1}. \tag{2.2}$$

$\hat{\mathbf{B}}$, $\hat{\mathbf{R}}$ are the background and observation error covariance matrices specified in the data assimilation (and **K** is the Kalman gain that follows). These may be mis-specfied. **B**, **R** are the correct background and observation error covariance matrices which may be diagnosed by looking at actual statistics and **K** is the correct Kalman gain that follows.

O-B, A-B, O-A EXPRESSIONS

Innovation (observation minus background difference in observation space)

$$\mathbf{d}_b^o = \mathbf{y} - \mathbf{h}(\mathbf{x}_b) \approx \varepsilon_o - \mathbf{H} \varepsilon_b, \tag{2.3}$$

$\varepsilon_o$ observation error, $\varepsilon_b$ background error.

Now express other important differences in terms of the innovations.

Analysis minus background difference in observation space

$$\mathbf{d}_b^a = \mathbf{H} \delta \mathbf{x}_a = \mathbf{H} \hat{\mathbf{K}} \mathbf{d}_b^o. \tag{2.4}$$

Residual (observation minus analysis in observation space)

$$\begin{aligned} \mathbf{d}_a^o &= \mathbf{y} - \mathbf{h}(\mathbf{x}_a) = \mathbf{y} - \mathbf{h}(\mathbf{x}_b + \delta \mathbf{x}_a) \\ &\approx \mathbf{y} - \mathbf{h}(\mathbf{x}_b) - \mathbf{H} \delta \mathbf{x}_a = \mathbf{d}_b^o - \mathbf{H} \hat{\mathbf{K}} \mathbf{d}_b^o = (\mathbf{I} - \mathbf{H} \hat{\mathbf{K}}) \mathbf{d}_b^o. \end{aligned} \tag{2.5}$$

MEASURED STATISTICS

<u>O-B -- O-B statistics</u> (assume background and observation errors are uncorrelated)

$$\begin{aligned} E\{\mathbf{d}_b^o \mathbf{d}_b^{o^T}\} &= E\{(\varepsilon_o - \mathbf{H} \varepsilon_b)(\varepsilon_o - \mathbf{H} \varepsilon_b)^T\}, \\ &= E\{\varepsilon_o \varepsilon_o^T\} - E\{\varepsilon_o \varepsilon_b^T\} \mathbf{H}^T - \mathbf{H} E\{\varepsilon_b \varepsilon_o^T\} + \mathbf{H} E\{\varepsilon_b \varepsilon_b^T\} \mathbf{H}^T, \\ &= \mathbf{R} + \mathbf{H} \mathbf{B} \mathbf{H}^T. \end{aligned} \tag*{$*$ (2.6)}$$

<u>A-B -- O-B statistics</u> (use (2.4), (2.6), (2.2))

$$\begin{aligned} E\{\mathbf{d}_b^a \mathbf{d}_b^{o^T}\} &= \mathbf{H} \hat{\mathbf{K}} E\{\mathbf{d}_b^o \mathbf{d}_b^{o^T}\} = \mathbf{H} \hat{\mathbf{K}} (\mathbf{R} + \mathbf{H} \mathbf{B} \mathbf{H}^T), \\ &= \mathbf{H} \hat{\mathbf{B}} \mathbf{H}^T (\mathbf{H} \hat{\mathbf{B}} \mathbf{H}^T + \hat{\mathbf{R}})^{-1} (\mathbf{R} + \mathbf{H} \mathbf{B} \mathbf{H}^T). \end{aligned} \tag{2.7}$$

If $\hat{\mathbf{B}} = \mathbf{B}$ and $\hat{\mathbf{R}} = \mathbf{R}$ then (2.7) becomes

$$E\{\mathbf{d}_b^a \mathbf{d}_b^{o\,T}\} = \mathbf{HBH}^T(\mathbf{HBH}^T + \mathbf{R})^{-1}(\mathbf{R} + \mathbf{HBH}^T) = \mathbf{HBH}^T. \qquad * (2.7a)$$

<u>O-A -- O-B statistics</u> (use (2.5), (2.6), (2.2))

$$E\{\mathbf{d}_a^o \mathbf{d}_b^{o\,T}\} = (\mathbf{I} - \mathbf{H}\hat{\mathbf{K}})E\{\mathbf{d}_b^o \mathbf{d}_b^{o\,T}\} = (\mathbf{I} - \mathbf{H}\hat{\mathbf{K}})(\mathbf{R} + \mathbf{HBH}^T),$$
$$= (\mathbf{I} - \mathbf{H}\hat{\mathbf{B}}\mathbf{H}^T(\mathbf{H}\hat{\mathbf{B}}\mathbf{H}^T + \hat{\mathbf{R}})^{-1})(\mathbf{R} + \mathbf{HBH}^T). \qquad (2.8)$$

If $\hat{\mathbf{B}} = \mathbf{B}$ and $\hat{\mathbf{R}} = \mathbf{R}$ then (2.8) becomes

$$E\{\mathbf{d}_a^o \mathbf{d}_b^{o\,T}\} = (\mathbf{I} - \mathbf{HBH}^T(\mathbf{HBH}^T + \mathbf{R})^{-1})(\mathbf{R} + \mathbf{HBH}^T) = \mathbf{R}. \qquad * (2.8a)$$

<u>A-B -- O-A statistics</u> (use (2.4), (2.5), (2.6), (2.2))

$$E\{\mathbf{d}_b^a \mathbf{d}_a^{o\,T}\} = \mathbf{H}\hat{\mathbf{K}}E\{\mathbf{d}_b^o \mathbf{d}_b^{o\,T}\}(\mathbf{I} - \mathbf{H}\hat{\mathbf{K}})^T = \mathbf{H}\hat{\mathbf{K}}(\mathbf{R} + \mathbf{HBH}^T)(\mathbf{I} - \mathbf{H}\hat{\mathbf{K}})^T,$$
$$= \mathbf{H}\hat{\mathbf{B}}\mathbf{H}^T(\mathbf{H}\hat{\mathbf{B}}\mathbf{H}^T + \hat{\mathbf{R}})^{-1}(\mathbf{R} + \mathbf{HBH}^T)(\mathbf{I} - \mathbf{H}\hat{\mathbf{B}}\mathbf{H}^T(\mathbf{H}\hat{\mathbf{B}}\mathbf{H}^T + \hat{\mathbf{R}})^{-1})^T. \qquad (2.9)$$

If $\hat{\mathbf{B}} = \mathbf{B}$ and $\hat{\mathbf{R}} = \mathbf{R}$ then (2.9) becomes

$$E\{\mathbf{d}_b^a \mathbf{d}_a^{o\,T}\} = \mathbf{HBH}^T(\mathbf{I} - \mathbf{HBH}^T(\mathbf{HBH}^T + \mathbf{R})^{-1})^T.$$

Write $\mathbf{I} = (\mathbf{HBH}^T + \mathbf{R})(\mathbf{HBH}^T + \mathbf{R})^{-1}$

$$E\{\mathbf{d}_b^a \mathbf{d}_a^{o\,T}\} = \mathbf{HBH}^T((\mathbf{HBH}^T + \mathbf{R})(\mathbf{HBH}^T + \mathbf{R})^{-1} - \mathbf{HBH}^T(\mathbf{HBH}^T + \mathbf{R})^{-1})^T,$$
$$= \mathbf{HBH}^T([\mathbf{HBH}^T + \mathbf{R} - \mathbf{HBH}^T](\mathbf{HBH}^T + \mathbf{R})^{-1})^T,$$
$$= \mathbf{HBH}^T(\mathbf{R}(\mathbf{HBH}^T + \mathbf{R})^{-1})^T,$$
$$= \mathbf{HBH}^T(\mathbf{HBH}^T + \mathbf{R})^{-1}\mathbf{R}. \qquad * (2.9a)$$

Note that the inverse Hessian has the form $\mathbf{A}^{-1} = \mathbf{B}^{-1} + \mathbf{H}^T\mathbf{R}^{-1}\mathbf{H}$ and the Sherman-Morrison-Woodbury formula in terms of $\mathbf{A}^{-1}$ is $\mathbf{A}^{-1}\mathbf{BH}^T = \mathbf{H}^T\mathbf{R}^{-1}(\mathbf{R} + \mathbf{HBH}^T)$. This makes (2.9a)

$$E\{\mathbf{d}_b^a \mathbf{d}_a^{o\,T}\} = \mathbf{HAH}^T. \qquad (2.9b)$$