On Optimal Observation Quality Control Theory for Numerical Weather Prediction Systems

Frank Thomas Tveter
(P.O.BOX 43, N-0313 OSLO, NORWAY)

ABSTRACT

The purpose of this paper is to develop optimal observation quality control theory suitable for implementation in numerical weather prediction systems. The theory presented here is optimal (to 'first order') with respect to Mean Squared Error verification methods, and may be applied to observations that suffer from contamination and gross error. A practical criterion for accepting or rejecting observations in the data assimilation in numerical weather prediction systems, is derived using only common statistical assumptions and parameters that can be estimated from real data. The problem of selecting an optimal observation error covariance matrix is also briefly discussed here.

No. 7/2005 **Oslo, March 16th 2005**

Introduction

The focus in data assimilation is to identify the 'best' starting point for the numerical integration of the weather forecast model, using the available observations and a short term forecast as the 'first guess'. Any observation type, with known error statistics, can in principle be successfully assimilated in any data assimilation system if the observations are submitted to proper observation quality control. Observation quality control theory already exists for some special observation error cases. An optimal theory for observation quality control, which is suitable for more complicated error statistics, as we may have for instance for satellite radiance observations, is presented here. Bayesian decision theory provides a tool for making optimal decisions to achieve any specified objective. In this paper, the observation quality control decision is based on the objective that the resulting numerical weather prediction system should give the best Mean Squared Error verification results.

Some basic statistical theory on observation error is first presented to justify the observation error models. Next, the optimal method for assimilating an observation with non-Normal error statistics is identified. The optimal criterion for rejecting the observation, if the assimilation system assumes that it has Normal error characteristics, is derived. Finally the paper studies the problem of assimilating an observation using a poor estimate for the observation error.

The 'Law of Large Numbers' and the 'Central-Limit Theorem'

If we toss a large number (n) of coins independently, we anticipate that the number of heads (Y) that we get on average, will be equal to half the number of coins tossed altogether (i.e. $E[Y] = np, p = \frac{1}{2}$). This is a simple application of the 'Law of Large Numbers'. However, each time we repeat this experiment with the same large number of tosses (n), we will get a slightly different number of heads (Y). The exact distribution of Y is given by the binomial distribution, with a mean given by $E[Y] = np$, and a variance given by $\text{Var}[Y] = np(1-p)$. According to the 'Central Limit Theorem', when the number of tosses increases ($n \to \infty$), the probability distribution for the total number of heads, $P(Y)$, will converge toward a Normal (Gaussian) probability distribution with a mean given by $\mu = E[Y] = np$, and a standard deviation given by $\sigma^2 = \text{Var}[Y] = np(1-p)$, $P(Y) \to N[Y, np, np(1-p)]$. The 'Central Limit Theorem' is valid in this example since each toss is independent of the others and can only give a relatively small contribution to the total number of heads.

The Observation Error

Let us consider an observation made by a satellite, of radiation emitted from the atmosphere. The different parts of the satellite instrument system contribute a series of random errors to the measurement. Also, the radiation observation can not be exactly simulated by using an observation forward operator and the model representation of the atmosphere. The reason for this is that the actual radiation measurement gets contributions from the many air-segments along the ray-path in the atmosphere, and these contributions depend on small-scale properties of the air-segment which are not resolved in the model. If each of these effects (called error-sources) contribute to the deviation between the observed and simulated 'correct' radiation by some independent small amount, the probability distribution for the accumulated deviation will approach a Normal probability density distribution according to the 'Central Limit Theorem'. This is analogue to the results in the section above, where the probability distribution for the total number of heads converges toward the Normal distribution.

If the error-sources were identical for a sub-set of independent observations, and this sub-set was denoted i, one could, for any element in this sub-set, write

$$P[Y|\mathbf{X} \cap i] \quad = \quad N[Y, \mathbf{HX} + \mu_i, R_i], \tag{1}$$

where Y is our scalar observation, $\mathbf{X}$ is the model representation of the true atmospheric state, $\mathbf{HX}$ is the forward operator, R_i is the observation error covariance (matrix) and μ_i is the bias. Note here that both $\mathbf{X}$ and Y are random variables.

This paper is only considering the problem of adding a single observation with non-Normal error statistics to the assimilation system. The optimal quality control decision on a single, non-Normal observation is identified in the following theory, and this decision is only approximately optimal (or optimal to 'first order') when effects from other non-Normal observations also are present.

Error sources will vary from one observation to the other. Some observations may for instance be affected by precipitation while others are not. Let us assume that there is a limited number of different independent sets of error sources occurring in our data set. It is not known in advance exactly which sub-set of error sources a given observation is subjected to, but let us assume for now that the a-priori probability that an observation will be subjected to error sources of subset i, $\hat{q}_i = P[i]$, is known. We may write

$$P[Y|\mathbf{X}] \;=\; \sum_i P[i]P[Y|\mathbf{X}\cap i] = \sum_i \hat{q}_i N[Y, \mathbf{H}\mathbf{X}+\mu_i, R_i]. \tag{2}$$

It is convenient to divide the contributions to the observation error in Eq. (2) into three groups, each containing one or more elements. The first group contains only one element, $i = 1$, and this is our un-disturbed signal. This Normal distribution has high a-priori probability, zero mean and a relatively small error (R_1). The second group contains contaminated observations which are characterized by a higher error than the un-disturbed signal ($R > R_1$) and usually a significant bias. The third group contains observations that suffer from Gross error. These observations have large error, $R \gg R_1$ and $R \gg \mathbf{HBH}^T$ where $\mathbf{B}$ is the background error covariance matrix, and usually a small (un-detectable) bias compared to the large error, $|\mu| < \sqrt{R}$. We will later refer to $\mathbf{HBH}^T$ as 'the projection of the model error onto observation space'.

Data Assimilation

The *analysis* vector, $\mathbf{X}_a$, is in this context the state of the atmosphere that is used as a starting point for the integration of the future state of the atmosphere in a forecast model. An analysis is determined using observations, $\mathbf{Y}$, and a short forecast, also known as the first guess or background, $\mathbf{X}_b$. It can be shown (see the Appendix) that assimilating a set of non-Normal observations together with a set of Normal observations and a Normal first guess, is mathematically equivalent to simply assimilating the non-Normal observations together with a 'Gaussian' first guess ($\breve{\mathbf{X}}_b$) defined according to

$$\begin{aligned}
\breve{\mathbf{X}}_b &= \mathbf{X}_b + \mathbf{BH}^T(\mathbf{HBH}^T + R)^{-1}(\mathbf{Y} - \mathbf{HX}_b) \\
\breve{\mathbf{B}}^{-1} &= \mathbf{B}^{-1} + \mathbf{H}^T R^{-1}\mathbf{H},
\end{aligned} \tag{3}$$

where $\mathbf{R}$ is the Gaussian observation error covariance matrix and $\mathbf{H}$ is the observation operator for the Gaussian observations. It is assumed in the rest of this paper that the first guess has been redefined according to Eq. (3) so that it includes the effects from the observations with Normal error distribution. Our scalar, non-Normal observation is from now on referred to as Y. Note that the Gaussian first guess error covariance matrix is smaller than the first guess error covariance matrix in the scalar case.

Mean Squared Error (MSE) verification methods are commonly used to verify numerical weather prediction systems. For short forecasts, these verification methods mainly measure the quality of the analysis. In Bayesian decision theory, the MSE verification method for short forecasts translates to the quadratic *loss* function, $l[\mathbf{X}, \mathbf{X}_a] = (\mathbf{X} - \mathbf{X}_a)^T(\mathbf{X} - \mathbf{X}_a)$, where $\mathbf{X}$ represents the unknown true state of the atmosphere. The expected loss (also called *Bayes risk*) is then given by

$$\begin{aligned}
B(\mathbf{X}_a) &= E[l[\mathbf{X}, \mathbf{X}_a]] \\
&= E[E[(\mathbf{X} - \mathbf{X}_a)^T(\mathbf{X} - \mathbf{X}_a)|Y \cap \mathbf{X}_b]] \\
&= E[\mathrm{Var}[\mathbf{X}|Y \cap \mathbf{X}_b]] + \\
&\quad\; E[(E[\mathbf{X}|Y \cap \mathbf{X}_b] - \mathbf{X}_a)^T(E[\mathbf{X}|Y \cap \mathbf{X}_b] - \mathbf{X}_a)]
\end{aligned}$$

A large Bayes risk (expected loss) corresponds to a poor MSE verification (large MSE score). An important result from Bayesian decision theory is that the $\mathbf{X}_a^{**}$ function that minimizes this particular Bayes risk, is given by

$$\mathbf{X}_a^{**} \;=\; E[\mathbf{X}|Y \cap \mathbf{X}_b], \tag{4}$$

which is 'the expected state of the atmosphere'. It is seen from Eq. (4) that if two different analysis methods are compared, the one closest to 'the expected state of the atmosphere', will have lower risk, and therefore better verification score, than the other analysis method.

If the first guess, $\mathbf{X}_b$, and the observation, Y, have Normal error characteristics and no bias, and if the observation operator $\mathbf{H}$ is linear,

$$\mathbf{X}_a \;=\; \mathbf{X}_b + \mathbf{BH}^T(\mathbf{HBH}^T + R)^{-1}(Y - \mathbf{HX}_b). \tag{5}$$

This is a well known result from assimilation theory applied to numerical weather prediction.

Assimilating Observations with non-Normal Error Characteristics

The probability density function for the state of the atmosphere, given the first guess and the observation, can be written as

$$P[\mathbf{X}|Y \cap \mathbf{X}_b] = \frac{P[Y|\mathbf{X}]P[\mathbf{X}|\mathbf{X}_b]}{P[Y|\mathbf{X}_b]},$$

where it has been assumed that the observation error is independent of the first guess error. The expression for our probability function then becomes

$$\begin{aligned} P[\mathbf{X}|Y \cap \mathbf{X}_b] &= \sum_{k=1}^{n} q_k \frac{P[Y|\mathbf{X} \cap k]P[\mathbf{X}|\mathbf{X}_b]}{P[Y|\mathbf{X}_b \cap k]} \\ q_k &= \hat{q}_k \frac{P[Y|\mathbf{X}_b \cap k]}{P[Y|\mathbf{X}_b]} \\ &= \hat{q}_k \frac{P[Y|\mathbf{X}_b \cap k]}{\sum_{i=1}^{n} \hat{q}_i P[Y|\mathbf{X}_b \cap i]} \\ P[Y|\mathbf{X}_b \cap k] &= \int_{-\infty}^{\infty} P[Y|\mathbf{X} \cap k]P[\mathbf{X}|\mathbf{X}_b] d\mathbf{X} \\ &= N[Y, \mathbf{H}\mathbf{X}_b + \mu_k, v_k] \qquad (6) \\ v_k &= \mathbf{H}\mathbf{B}\mathbf{H}^T + R_k, \end{aligned}$$

where q_k is the a-posteriori probability that the observation is affected by the k^{th} error source. Note that $\sum_{k=1}^{n} q_k = 1$. The probability density distribution of the innovation, $Y - \mathbf{H}\mathbf{X}_b$, gives $P[Y|\mathbf{X}_b]$ under the assumption that the innovation is independent of the first guess. We have the following useful relationship

$$P[Y|\mathbf{X}_b] = \sum_{k=1}^{n} \hat{q}_k P[Y|\mathbf{X}_b \cap k], \qquad (7)$$

where $P[Y|\mathbf{X}_b \cap k]$ are normal distributions. One may in principle use the knowledge of the innovation probability density, $P[Y|\mathbf{X}_b]$, to determine the unknown parameters in Eq. (7), namely the a-priori probabilities, $\hat{q}_k$, and the bias, μ_k, and variance, v_k, of $P[Y|\mathbf{X}_b \cap k]$.

The best analysis with respect to a MSE verification method is, according to the Bayesian decision theory, expected to be given by Eq. (4),

$$\begin{aligned} \mathbf{X}_a^{**} &= E[\mathbf{X}|Y \cap \mathbf{X}_b] \\ &= \int_{-\infty}^{\infty} \mathbf{X} P[\mathbf{X}|Y \cap \mathbf{X}_b] d\mathbf{X} \\ &= \sum_{k=1}^{n} q_k \mathbf{X}_k \qquad (8) \\ \mathbf{X}_k &= \mathbf{X}_b + \mathbf{B}\mathbf{H}^T(\mathbf{H}\mathbf{B}\mathbf{H}^T + R_k)^{-1}(Y - \mathbf{H}\mathbf{X}_b - \mu_k). \end{aligned}$$

Observe that $\mathbf{X}_k \rightarrow \mathbf{X}_b$ for the Gross error term in Eq. (9) since we then have that $R_k \gg \mathbf{H}\mathbf{B}\mathbf{H}^T$. Having an observation that has a large a-posteriori probability of Gross error is in other words similar to ignoring the observation altogether, and sticking with the first guess, i.e. $\mathbf{X}_a^{**} \rightarrow \mathbf{X}_b$.

Optimal Quality Control

Contaminated observations and observations that suffer from Gross error are in principle not different from un-disturbed observations. However, contaminated observations and observations with Gross error have so large errors that they give little or no detectable positive impact on the verification score when they are assimilated in an optimal manner. The assimilation algorithm for optimal use of contaminated observations and observations with Gross error, is complex and requires an excessive amount of computer resources compared to algorithms for assimilating un-disturbed observations. The error characteristics of contaminated observations and observations that suffer from Gross error are usually not well known, so we may not know how to use the observations properly. These are the reasons for wanting to discard contaminated observations and observations with Gross error from data assimilation.

In this section we will study how to reject observations that are subjected to contamination and Gross error, in such a way that the resulting verification is optimal. The investigation begins by formulating the Bayes risk for the different analysis methods that we wish to choose between. We know from above that a low Bayes risk corresponds to good MSE verification. The Bayes risk for the optimal analysis (Eq. (8)) is given by

$$B_{\text{Optimal}}(\mathbf{X}_a) = C, \tag{9}$$

where C is a constant ($= E[\text{Var}[\mathbf{X}|Y \cap \mathbf{X}_b]]$). This is the best possible assimilation method for this problem, but it is too complicated to be used in practical problems.

The Normal analysis represents the assimilation method which is implemented when observations are assimilated in existing systems today, i.e. it is assumed that the observation has a Normal error distribution with no bias. The Bayes risk for the Normal analysis (Eq. (5)) is given by

$$\begin{aligned} B_{\text{Normal}}(\mathbf{X}_a) &= E[E[(\mathbf{X}_a^{**} - \mathbf{X}_a)^T(\mathbf{X}_a^{**} - \mathbf{X}_a)|Y \cap \mathbf{X}_b] + C \\ &= E[E[\sum_{k=2}^{n} q_k(\delta_k - \delta_1)(\mathbf{H}\mathbf{B}^T\mathbf{B}\mathbf{H}^T)\sum_{k=2}^{n} q_k(\delta_k - \delta_1)|Y \cap \mathbf{X}_b] + C \\ \delta_k &= v_k^{-1}(Y - \mathbf{H}\mathbf{X}_b - \mu_k). \end{aligned}$$

Note that q_k, v_k and μ_k can be estimated from the data as discussed above, and that δ_k can be calculated using the known innovation $(Y - \mathbf{H}\mathbf{X}_b)$.

If we choose to ignore the observations and not change the first guess, we get the following Bayes risk

$$B_{\text{First guess}}(\mathbf{X}_a) = E[E[\sum_{k=1}^{n} q_k\delta_k(\mathbf{H}\mathbf{B}^T\mathbf{B}\mathbf{H}^T)\sum_{k=1}^{n} q_k\delta_k|Y \cap \mathbf{X}_b] + C.$$

The assimilation system should choose between two strategies: either ignore the observation and use the first guess as the analysis, or use the observation assuming that it has Normal observation error characteristics. To choose between these two strategies, we compare the risk from ignoring the observation ($B_{\text{First guess}}$) with the risk from using the observation (B_{Normal}) and select the strategy that gives the lowest risk. Note that the unknown scalar $\mathbf{H}\mathbf{B}^T\mathbf{B}\mathbf{H}^T$ is positive. We may write

$$\begin{aligned} B_{\text{Normal}}(\mathbf{X}_a) - B_{\text{First guess}}(\mathbf{X}_a) &\propto E[E[\left(\sum_{k=2}^{n} q_k(\delta_k - \delta_1)\right)^2 - \left(\sum_{k=1}^{n} q_k\delta_k\right)^2 |Y \cap \mathbf{X}_b]] \\ &= E[E[\Delta b_{\text{Normal}-\text{First guess}}[Y - \mathbf{H}\mathbf{X}_b]|Y \cap \mathbf{X}_b]] \end{aligned}$$

We should in other words reject the observation if the risk increment is positive, $\Delta b_{\text{Normal}-\text{First guess}}[Y - \mathbf{H}\mathbf{X}_b] > 0$, and assimilate the observation otherwise. Figure 1 shows an example of innovation statistics for an observation with contamination and gross error. The figure also shows the risk increment as a function of the innovation, and we observe that in this case the risk increment is negative when $-3.8 < Y - \mathbf{H}\mathbf{X}_b < 2.9$. Observations with an innovation in this range will in other words have a positive effect on the analysis if they are assimilated. Observe that the risk increases dramatically if we assimilate observations outside this range. The negative impact from assimilating few 'poor' observations can in other words easily outweigh the positive impact from assimilating many 'good' observations.

Using a Poor Observation Error Covariance Matrix

In real life, we do not know the observation error covariance matrix nor the first guess error observation covariance matrix. It may be instructive to study a Normal observation with no contamination or Gross error (i.e. $n = 1$), that is assimilated using a poor estimate of the observation error covariance matrix. The Bayes risk is in this situation given by

$$\begin{aligned} B_{\text{Error}}(\mathbf{X}_a) &= E[E[(\mathbf{X}_a^{**} - \mathbf{X}_a)^T(\mathbf{X}_a^{**} - \mathbf{X}_a)|Y \cap \mathbf{X}_b] + C \\ \mathbf{X}_a^{**} - \mathbf{X}_a &= \mathbf{B}\mathbf{H}^T\left((\mathbf{H}\mathbf{B}\mathbf{H}^T + R_1)^{-1} - (\mathbf{H}\mathbf{B}\mathbf{H}^T + R_{\text{error}})^{-1}\right)(Y - \mathbf{H}\mathbf{X}_b) \end{aligned}$$

where R_1 is the true observation error covariance matrix and R_{error} is the matrix used in the assimilation. Let us compare this to not assimilating the observation at all. If we choose to ignore the observations and not change the first guess, we get the following risk

$$\begin{aligned} B_{\text{First guess}}(\mathbf{X}_a) &= E[E[(\mathbf{X}_a^{**} - \mathbf{X}_b)^T(\mathbf{X}_a^{**} - \mathbf{X}_b)|Y \cap \mathbf{X}_b] + C, \\ \mathbf{X}_a^{**} - \mathbf{X}_b &= \mathbf{B}\mathbf{H}^T(\mathbf{H}\mathbf{B}\mathbf{H}^T + R_1)^{-1}(Y - \mathbf{H}\mathbf{X}_b). \end{aligned}$$

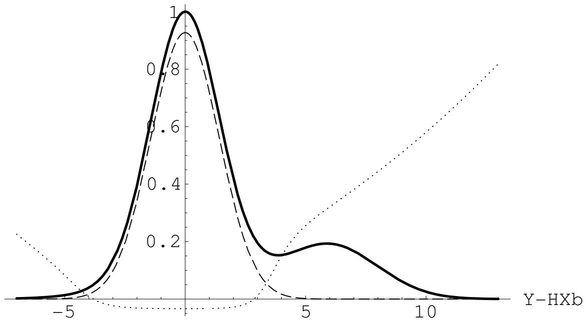

Figure 1: Example of Bayes risk increment ($B_{\text{Normal}} - B_{\text{First guess}}$, dotted), innovation probability (bold) and $q_1 N[Y, \mathbf{H}\mathbf{X}_b + \mu_1, R_1]$ (dashed) as a function of the innovation, $Y - \mathbf{H}\mathbf{X}_b$. The parameters in this example were $v_1 = 2$, $v_2 = 4$, $v_3 = 9$, $\mu_1 = 0$, $\mu_2 = 6$, $\mu_3 = 0$, $\hat{q}_1 = 0.7$, $\hat{q}_2 = 0.2$ and $\hat{q}_3 = 0.1$.

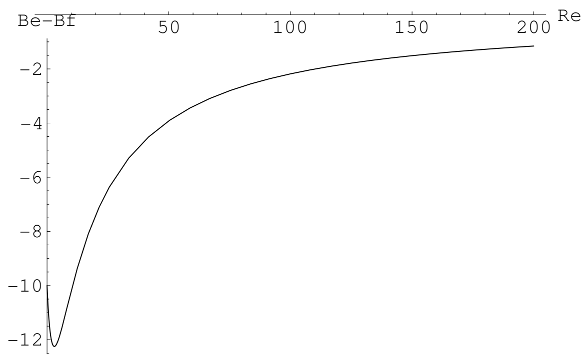

Figure 2: Bayes risk increment ($B_{\text{Error}} - B_{\text{First guess}}$) as a function of R_{Error} when $R_1 < \mathbf{HBH}^T$.

In the scalar case, it is obvious that $B_{\text{First guess}} > B_{\text{Error}}$ if

$$\begin{aligned} |(\mathbf{HBH}^T + R_1)^{-1}| &> |(\mathbf{HBH}^T + R_1)^{-1} - (\mathbf{HBH}^T + R_{\text{Error}})^{-1}| \\ 2 &> \frac{\mathbf{HBH}^T + R_1}{\mathbf{HBH}^T + R_{\text{Error}}} (> 0) \\ R_{\text{Error}} &> \frac{1}{2}(R_1 - \mathbf{HBH}^T). \end{aligned} \tag{10}$$

Figure 2 shows an example of how the risk increment, $B_{\text{Error}} - B_{\text{First guess}}$, varies as a function of R_{Error} in a case where $R_1 < \mathbf{HBH}^T$. We observe from Eq. (10) and Fig. 2 that if the observation error is small compared to the first guess error ($R_1 < \mathbf{HBH}^T$), any value for R_{Error} will give a positive effect of assimilating the observation. Figure 3 shows the corresponding plot if $R_1 > \mathbf{HBH}^T$. If the observation error is relatively large, caution must be made to not specify a too small observation error in the assimilation. When there is a doubt, it is better to use a too large observation error than risking a too small observation error, especially if it is known that the observation is not too accurate compared to the first guess. It is relatively easy to estimate the innovation covariance, $E[(Y - H(\mathbf{X}_b)^T (Y - H\mathbf{X}_b)] = \mathbf{HBH}^T + R_1$, compared to the observation error, R_1. Figure 4 shows the Bayes risk increment ($B_{\text{Error}} - B_{\text{First guess}}$) ratio between using the correct observation error ($R_{\text{Error}} = R_1$) and using the innovation covariance as observation error ($R_{\text{Error}} = \mathbf{HBH}^T + R_1$) plotted as a function of $R_1/(\mathbf{HBH}^T)$. We observe that using the innovation covariance as the observation error in the assimilation scheme results in a relative large positive impact compared to using the correct observation error in the assimilation ($\geq 75\%$), regardless of what the true observation error is. It can be shown that the ratio is exactly $3/4$ when $R_1 \to 0$. Note also that the relative impact resulting from this approach improves as the observation error increases relative to the first guess error, which is what we would expect, since $R_1 + \mathbf{HBH}^T \to R_1$ when $\mathbf{HBH}^T \ll R_1$.

Summary and Conclusions

The error probability density distribution of an observation type, may be modeled as a sum of Normal distributions, if the observations are subjected to different groups of error sources. Observations with non-Normal error characteristics may be assimilated optimally according to Bayesian decision theory, when the objective is to

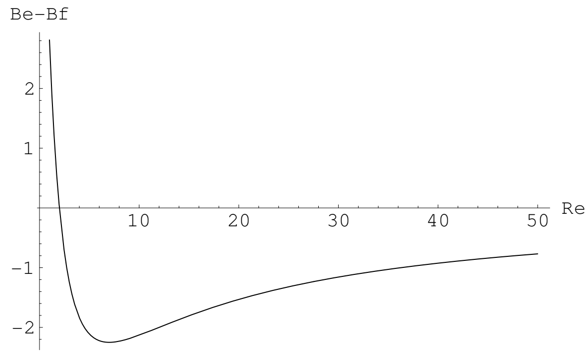

Figure 3: Bayes risk increment ($B_{\text{Error}} - B_{\text{First guess}}$) as a function of R_{Error} when $R_1 > \mathbf{HBH}^T$.

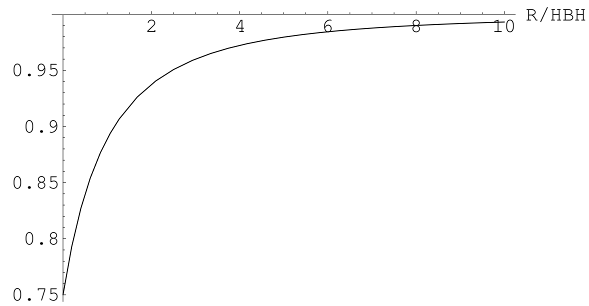

Figure 4: Bayes risk increment ($B_{\text{Error}} - B_{\text{First guess}}$) ratio between using the correct observation error ($R_{\text{Error}} = R_1$) and using the innovation covariance as observation error ($R_{\text{Error}} = \mathbf{HBH}^T + R_1$) plotted as a function of $R_1/(\mathbf{HBH}^T)$.

have the best Mean Squared Error verification results. There is in principle no mathematical difference between the optimal assimilation of un-disturbed observations and contaminated observations or observations with gross error. However, these other observations have large errors, and are not capable of giving noticeable positive effects on the verification, even if they are assimilated optimally. Optimal assimilation systems for observations with non-Normal error characteristics are complex and not practical to use, so it is better to assume that the observations have Normal error distributions, and reject those observations that have a negative effect on the verification. An observation quality control criterion for optimal rejection of observations was derived here, and it turns out that it did not require special assumptions nor parameters that can no in principle be estimated from real (innovation) data.

The value of the observation error is usually unknown, and it was shown that using a too large observation error in the assimilation in the scalar problem, always ensures that the observation gives a positive effect on the verification. Using a too small observation error may result in a negative effect on the verification if the true observation error is larger than the model error projected onto observation space. The MSE verification improvements which result from assimilating an observation using the innovation covariance as the observation error covariance, is more than 75% of the improvement which results from assimilating optimally observations using the true observation error as the observation error.

Appendix: The Normal analysis

Let us assume that we have a set of observations with Normal error characteristics, $\mathbf{Y}$, and another set of independent non-Normal observations, $\tilde{\mathbf{Y}}$. Using Bayes rule we may write the probability density function for the true atmospheric state as

$$P[\mathbf{X}|\tilde{\mathbf{Y}} \cap \mathbf{Y} \cap \mathbf{X}_b] = \frac{P[\tilde{\mathbf{Y}}|\mathbf{X} \cap \mathbf{Y} \cap \mathbf{X}_b]P[\mathbf{X}|\mathbf{Y} \cap \mathbf{X}_b]}{P[\tilde{\mathbf{Y}}|\mathbf{Y} \cap \mathbf{X}_b]}. \quad (11)$$

Note that the normalization term, $P[\tilde{\mathbf{Y}}|\mathbf{Y} \cap \mathbf{X}_b]$, does not depend on $\mathbf{X}$. The probability density function for observations with linear observation operators and Normal error characteristics may be modeled according to

$$P[\mathbf{Y}|\mathbf{X} \cap \mathbf{X}_b] = N[\mathbf{Y}, \mathbf{HX}, \mathbf{R}].$$

In HIRLAM 3D-Var it is assumed that the background error is Normal, giving

$$P[\mathbf{X}|\mathbf{X}_b] = N[\mathbf{X}, \mathbf{X}_b, \mathbf{B}].$$

Next, let us define a 'Normal analysis' by

$$\begin{aligned} \breve{\mathbf{X}}_b &= \mathbf{X}_b + \mathbf{BH}^T(\mathbf{HBH}^T + \mathbf{R})^{-1}(\mathbf{Y} - \mathbf{HX}_b) \\ \breve{\mathbf{B}}^{-1} &= \mathbf{B}^{-1} + \mathbf{H}^T\mathbf{R}^{-1}\mathbf{H}. \end{aligned}$$

It can be shown, by using Bayes rule and some algebra, that

$$\begin{aligned} P[\mathbf{X}|\mathbf{Y} \cap \mathbf{X}_b] &= \frac{P[\mathbf{Y}|\mathbf{X} \cap \mathbf{X}_b]P[\mathbf{X}|\mathbf{X}_b]}{P[\mathbf{Y}|\mathbf{X}_b]} \\ &= P[\mathbf{X}|\breve{\mathbf{X}}_b] \\ P[\mathbf{X}|\breve{\mathbf{X}}_b] &= N[\mathbf{X}, \breve{\mathbf{X}}_b, \breve{\mathbf{B}}]. \end{aligned}$$

Further, since our observations are independent of each other and the first guess, we have that

$$\begin{aligned} P[\tilde{\mathbf{Y}}|\mathbf{Y} \cap \mathbf{X} \cap \mathbf{X}_b] &= \frac{P[\tilde{\mathbf{Y}} \cap \mathbf{Y} \cap \mathbf{X}_b|\mathbf{X}]P[\mathbf{X}]}{P[\mathbf{Y} \cap \mathbf{X} \cap \mathbf{X}_b]} \\ &= \frac{P[\tilde{\mathbf{Y}}|\mathbf{X}]P[\mathbf{Y} \cap \mathbf{X}_b|\mathbf{X}]P[\mathbf{X}]}{P[\mathbf{Y} \cap \mathbf{X}_b|\mathbf{X}]P[\mathbf{X}]} \\ &= P[\tilde{\mathbf{Y}}|\mathbf{X}]. \end{aligned}$$

We may now rewrite Eq. (11), using the Normal analysis

$$\begin{aligned} P[\mathbf{X}|\tilde{\mathbf{Y}} \cap \mathbf{Y} \cap \breve{\mathbf{X}}] &= \frac{P[\tilde{\mathbf{Y}}|\mathbf{X}]P[\mathbf{X}|\breve{\mathbf{X}}_b]}{P[\tilde{\mathbf{Y}}|\breve{\mathbf{X}}_b]} \\ &= P[\mathbf{X}|\tilde{\mathbf{Y}} \cap \breve{\mathbf{X}}_b], \end{aligned} \quad (12)$$

where we see that the normalization term now has been written as $P[\tilde{\mathbf{Y}}|\breve{\mathbf{X}}_b]$. We observe that the probability density function $P[\mathbf{X}|\tilde{\mathbf{Y}} \cap \mathbf{Y} \cap \mathbf{X}_b]$ is equivalent to $P[\mathbf{X}|\tilde{\mathbf{Y}} \cap \breve{\mathbf{X}}_b]$, where the effects of the Normal observations, $\mathbf{Y}$, have been included in the Normal analysis, $\breve{\mathbf{X}}_b$.