A NON-GAUSSIAN APPROACH TO 1D-VAR

H. Schyberg
Norwegian Meteorological Institute
Oslo, Norway

F.T.Tveter
Norwegian Meteorological Institute
Oslo, Norway

Abstract

The purpose of this paper is to show how 1D-VAR may be formulated so that it better handles the non-Gaussian error characteristics of the ATOVS (AMSU-A) satellite sounding observations. The paper shows how a Gaussian transformation may be found for the observation error vector based on the error statistics from a reference data set. The paper then demonstrates how this transformation may be applied to the 1D-VAR minimization problem.

1. INTRODUCTION

The Norwegian Meteorological Institute (DNMI) receives ATOVS data over the North-Atlantic from NOAA-15 through its local antenna. These data are processed with the EUMETSAT AAPP package, and the AMSU-A data are used in a 1D-Var system based on that of Eyre et al (1993). This work (see Breivik et al, 1999) is part of a project to assimilate ATOVS data in the HIRLAM (High Resolution Limited Area Model) 3D-Var scheme.

The 1D-Var least squares estimator expressed in terms of quadratic cost terms is only optimal if the observations (and also the parameters that are adjusted) all have Gaussian error characteristics. However, this is usually not the case for the error characteristics of AMSU-A (ATOVS) satellite sounding data. In particular, the effect of clouds and precipitation not accounted for in the forward radiative transfer model (see Eyre, 1991) contributes to inflating the tail in the error probability distribution.

In this paper we introduce a new approach to implicitly take care of the data control necessary to deal with the cloud and precipitation contamination. The paper demonstrates how a control variable transformation can be applied to the observation cost term and the background cost term in order to bring the variables closer to following Gaussian error characteristics. The paper then shows how the least squares estimator and 1D-Var may be applied to these transformed variables. This approach ensures a better agreement with the Gaussian assumptions of the variational method. For the observation term the method can also be applied in 3D- or 4D-Var systems.

2. THE ORIGINAL APPROACH

Variational schemes are based on minimization of cost functions with terms that can be derived from probabilities (see e.g. Lorenc, 1986). The background cost term can be expressed as

$$J_b(\mathbf{x}) = -\ln P(\mathbf{x} = \mathbf{x}_{true}),$$

where $P(\mathbf{x} = \mathbf{x}_{true})$ is the a priori probability that the control variable $\mathbf{x}$ is near the true value, $\mathbf{x}_{true}$ and the observation term is expressed as

$$J_o(\mathbf{x}) = -\ln P(\mathbf{y}|\mathbf{x} = \mathbf{x}_{true}), \tag{1}$$

which is the probability of obtaining the observation $\mathbf{y}$ given that $\mathbf{x}$ is the true value of the control variable.

If the above probabilities follow Gaussian distribution laws (multi-normal distributions), we obtain the 1D-Var least squares estimator, and the scheme is then based on the minimization of quadratic cost terms. The cost function can be written as

$$J(\delta\mathbf{x}) = \frac{1}{2}\delta\mathbf{x}^T \cdot \mathbf{B}^{-1} \cdot \delta\mathbf{x} + \frac{1}{2}\delta\mathbf{y}(\delta\mathbf{x})^T \cdot \mathbf{O}^{-1} \cdot \delta\mathbf{y}(\delta\mathbf{x}) \tag{2}$$

where

$$\begin{aligned}
\delta\mathbf{x} &= \mathbf{x} - \mathbf{x}_b \\
\delta\mathbf{y}(\delta\mathbf{x}) &= \mathbf{y} - \mathbf{H}(\mathbf{x}_b + \delta\mathbf{x}) \\
\mathbf{B} &= \left\langle (\delta\mathbf{x}_{true}) \cdot (\delta\mathbf{x}_{true})^T \right\rangle \\
\mathbf{O} &= \left\langle (\delta\mathbf{y}(\delta\mathbf{x}_{true})) \cdot (\delta\mathbf{y}(\delta\mathbf{x}_{true}))^T \right\rangle \\
\delta\mathbf{x}_{true} &= \mathbf{x}_{true} - \mathbf{x}_b
\end{aligned} \tag{3}$$

where the bold symbols refer to vectors, the pointed brackets refer to the mean value over a large reference data set. $\mathbf{y}$ refers to an observation (consisting of many channels), $\mathbf{x}$ refers to an estimate of the atmospheric state, $\mathbf{x}_b$ is a prognosis of the state (background). The forward model $\mathbf{H}$ takes a given atmospheric state and estimates the corresponding observation value vector.

The least squares estimator gives the optimal solution of the problem if the statistical characteristics of the two error functions $\delta\mathbf{x}_{true}$ and $\delta\mathbf{y}(\delta\mathbf{x}_{true})$ are Gaussian. The best estimate is found from the iterative equation (see Eyre et al 1993)

$$\begin{aligned} \delta \mathbf{x}_{j+1} &= \delta \mathbf{x}_j + \mathbf{B} \cdot \mathbf{H}'^T_j \left[\mathbf{H}'_j \cdot \mathbf{B} \cdot \mathbf{H}'^T_j + \mathbf{O} \right]^{-1} \cdot \delta \mathbf{y} \left(\delta \mathbf{x}_j \right) \\ \mathbf{H}'_j &= \left(\frac{\partial \mathbf{H}}{\partial \left(\delta \mathbf{x} \right)} \right)_j , \end{aligned} \tag{4}$$

where the index j here refers to the iteration number. The idea is to apply the iterative Eq. (4) until $\delta\mathbf{x}$ has converged. This solution is used together with the background field ($\mathbf{x}_b$) in Eq. (3) to get the final estimate of the state of the atmosphere. This final estimate is usually referred to as *the analysis.*

3. THE GENERAL VARIABLE TRANSFORMATION

We want to make a transformation

$$\begin{aligned} \delta \hat{\mathbf{y}} &= \mathbf{g} \left(\delta \mathbf{y} \right) \\ \delta \mathbf{y} &= \mathbf{g}^{-1} \left(\delta \hat{\mathbf{y}} \right) . \end{aligned}$$

so that the transformed quantity $\delta\hat{\mathbf{y}}$ has a Gaussian probability distribution. We could also define a similar transform for the background term,

$$\begin{aligned} \delta \hat{\mathbf{x}} &= \mathbf{f} \left(\delta \mathbf{x} \right) \\ \delta \mathbf{x} &= \mathbf{f}^{-1} \left(\delta \hat{\mathbf{x}} \right) . \end{aligned}$$

In fact, a stochastic variable with any probability distribution can be transformed into a variable with Gaussian probability distribution by applying a change of variable based on the variable's own probability distribution. This is best expressed in terms of cumulative probabilities. If the vector $(\delta X_1, \delta X_2, \ldots, \delta X_N)$ is a certain realization of the $\delta\mathbf{x}$ vector, the cumulative distribution F is defined as the following simultaneous probability P,

$$F(\delta \mathbf{x}) = P\left(\delta X_1 \leq \delta x_1, \delta X_2 \leq \delta x_2, \ldots, \delta X_N \leq \delta x_N \right) .$$

A standard N-dimensional simultaneous Gaussian cumulative distribution for a vector with zero means, unity standard deviations and zero covariances is defined by the equation

$$\Phi^{(N)}(\delta \hat{\mathbf{x}}) = \frac{1}{(2\pi)^{\frac{N}{2}}} \int_{-\infty}^{\delta \hat{x}_1} \cdots \int_{-\infty}^{\delta \hat{x}_N} e^{-\frac{1}{2} \sum_{k=1}^{N} x_k^2} \, dx_1 \cdots dx_N .$$

We can define a transform $\delta\mathbf{x} \rightsquigarrow \delta\hat{\mathbf{x}}$ by requiring that

$$\Phi^{(N)}(\delta\hat{\mathbf{x}}) = F(\delta\mathbf{x}) \tag{5}$$

is fulfilled. The procedure for $\delta\hat{\mathbf{y}}$ is completely analogous.

A transform $\delta\mathbf{x} \rightsquigarrow \delta\hat{\mathbf{x}}$ defined by Eq. (5) is not unique when $\delta\hat{\mathbf{x}}$ has more than one element, $N > 1$, since only a single requirement on a scalar function is to be fulfilled. Note that the equation can even be satisfied using a $\delta\hat{\mathbf{x}}$ containing only a single element, which then will have a one-dimensional Gaussian probability distribution ($N = 1$). This reflects the fact that the observation cost term is uniquely defined by a scalar probability as defined in Eq. 1.

However, such a transformation from a vector to a scalar is clearly not invertible, and we shall see later on that it is only applicable to transforming the observation term in the cost function. Another problem here is the determination of the cumulative distribution F. The observation vector used at DNMI has dimension 10 (only 10 AMSU-A channels are used), which means that the available dataset must be grouped into intervals along 10 different axes. If K intervals is used for subdivision for counting of occupancies along each axis, this means that the data must be grouped into K^{10} boxes. For values of K large enough to give fairly detailed descriptions of the distribution, this means that even for a very large dataset, there will be very few points in each box, and it is not realistic to estimate the distribution this way. We shall below suggest an alternative and approximate approach to making the transformation.

4. THE INDEPENDENT VARIABLE TRANSFORMATION

The transformation becomes particularly simple if the components of the variable $\delta\mathbf{x}$ are statistically independent. The cumulative probability distribution may in this case be formulated as $F_k(\delta x_k) = P(\delta X_k \leq \delta x_k)$ where x_k refers to the particular variable. Each component can then be transformed separately, and the transformation function f_k depends only on one element.

The transformations are then written

$$\begin{aligned} \delta\hat{y}_k &= g_k(\delta y_k) \\ \delta y_k &= g_k^{-1}(\delta\hat{y}_k)\,. \end{aligned} \tag{6}$$

In this equation k refers to a given element in the observation vector, and for the state vector transformation we get,

$$
\begin{aligned}
\delta\hat{x}_k &= f_k\left(\delta x_k\right) \\
\delta x_k &= f_k^{-1}\left(\delta\hat{x}_k\right),
\end{aligned} \tag{7}
$$

where the index k refers to a given element in the atmospheric state vector.

We can transform each variable δx_k into a variable $\delta\hat{x}_k$ with a standard one-dimensional (unity standard deviation) normal distribution defined by the cumulative distribution

$$
\Phi(\delta\hat{x_k}) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\delta\hat{x}_k} e^{-\frac{x^2}{2}}\, dx.
$$

We then simply require that

$$
\Phi(\delta\hat{x_k}) = F_k(\delta x_k).
$$

For finding the transformation function f_k and its inverse in Eq. (7), we only need the inverses of the scalar functions F_k and Φ (which exist, since they are strictly growing functions), F_k^{-1} and Φ^{-1}. The Φ function and its inverse is independent of the problem, and can be tabulated based on the numerical integration of the Gauss function. F_k is found empirically by making statistics for a large reference data set. The transformation and its inverse is finally given by

$$
\begin{aligned}
\delta\hat{x_k} &= f_k\left(\delta x_k\right) = \Phi^{-1}\left(F_k\left(\delta x_k\right)\right) \\
\delta x_k &= f_k^{-1}\left(\delta\hat{x}_k\right) = F_k^{-1}\left(\Phi\left(\delta\hat{x}_k\right)\right).
\end{aligned} \tag{8}
$$

The result can be implemented as a table with corresponding values of $\delta\hat{x}_k$ and δx_k or, in the case of the observation vector, corresponding values of $\delta\hat{y}_k$ and δy_k ranging over all possible values that these parameters may have. Note that you may have to smoothen this table *slightly* to ensure that the gradient f_k' does not vary too much from one table element to the next. The purpose of smoothening is to ensure that the least squares method does not become unstable. The reader should also note that $\delta^k\hat{\mathbf{x}}$ does not get any bias in this approach, regardless of the bias that may be present in $\delta^k\mathbf{x}$.

The simple procedure for independent variables outlined above can also be performed variable by variable on correlated variables. However, separating the problem in such a way, does no longer guarantee that Eq. (5) is fulfilled nor that the transformed variables have a multi-normal distribution (which is assumed by the assimilation method). However, it is now realistic to estimate the required distributions with a reference dataset, since distributions are only needed along one δy_k axis at a time, and the full simultaneous distribution is not estimated.

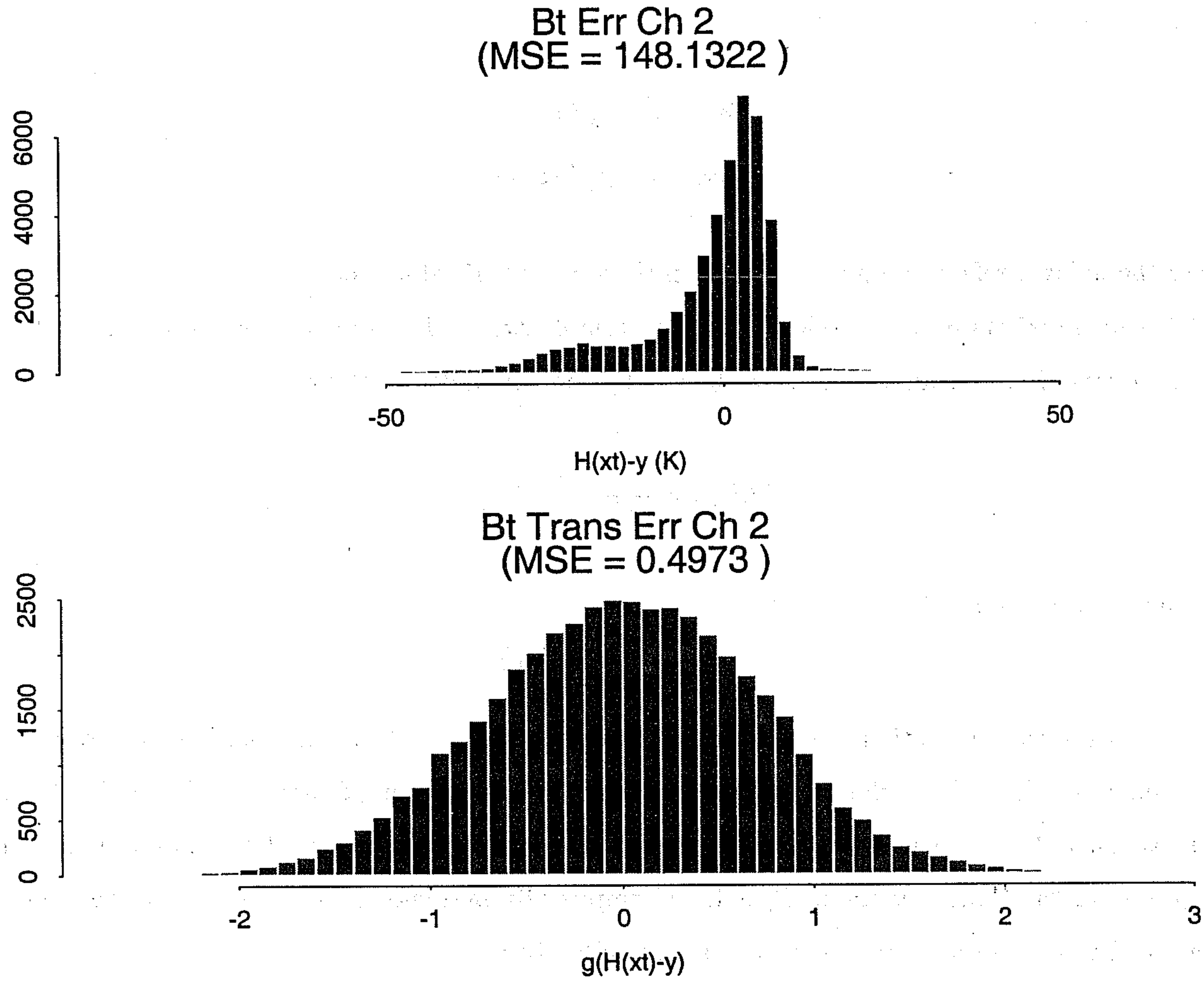

Figure 1: Distribution of the error statistics in AMSU-A channel 2, before (δy_2) and after ($\delta \hat{y}_2$) the transformation has been applied. The HIRLAM analysis has been used as reference ($\mathbf{x}_{true}$).

The hope is that such a simplified transform would be advantageous and bring the Gaussian error assumption on the control variables closer to being fulfilled. In particular, one would hope that the result of the least square minimization would be less sensitive to the contaminated observations, which are usually found in the tails of the error distribution.

Fig. 1 shows the effect of the transformation to the distribution of δy_k for AMSU-A channel 2 for the North-Atlantic for a period in October 1999. Note the secondary maximum at $-20K$ in the error distribution of the original variable, δy_k. This maximum is probably caused by precipitation contamination. We observe that the error distribution of the transformed variable, $\delta \hat{y}_k$, is completely Gaussian.

5. APPLICATION TO THE LEAST SQUARES ESTIMATOR

We now formulate the least squares estimator using the procedure based on independent variables. The cost function is re-defined to use the transformed variables,

$$\begin{aligned} \hat{J}(\delta\hat{\mathbf{x}}) &= \frac{1}{2}\delta\hat{\mathbf{x}}^T \cdot \hat{\mathbf{B}}^{-1} \cdot \delta\hat{\mathbf{x}} \\ &+ \frac{1}{2}\mathbf{g}\left(\delta\mathbf{y}\left(\mathbf{f}^{-1}(\delta\hat{\mathbf{x}})\right)\right)^T \cdot \hat{\mathbf{O}}^{-1} \cdot \mathbf{g}\left(\delta\mathbf{y}\left(\mathbf{f}^{-1}(\delta\hat{\mathbf{x}})\right)\right) \end{aligned}$$

where we now have defined

$$\begin{aligned} \hat{\mathbf{B}} &= \left\langle (\mathbf{f}(\delta\mathbf{x}_{true})) \cdot (\mathbf{f}(\delta\mathbf{x}_{true}))^T \right\rangle \\ \hat{\mathbf{O}} &= \left\langle (\mathbf{g}(\delta\mathbf{y}(\delta\mathbf{x}_{true}))) \cdot (\mathbf{g}(\delta\mathbf{y}(\delta\mathbf{x}_{true})))^T \right\rangle . \end{aligned} \tag{9}$$

The iterative equations for the least squares estimator for this cost function, Eqs. (4) are now replaced by,

$$\begin{aligned} \delta\hat{\mathbf{x}}_{j+1} &= \delta\hat{\mathbf{x}}_j + \hat{\mathbf{B}} \cdot \hat{\mathbf{H}}'^T_j \left[\hat{\mathbf{H}}'_j \cdot \hat{\mathbf{B}} \cdot \hat{\mathbf{H}}'^T_j + \hat{\mathbf{O}}\right]^{-1} \cdot \mathbf{g}\left(\delta\mathbf{y}\left(\delta\mathbf{x}_j\right)\right) \\ \hat{\mathbf{H}}'_j &= \left(\frac{\partial(\delta\hat{\mathbf{y}})}{\partial(\delta\mathbf{y})} \cdot \mathbf{H}' \cdot \frac{\partial(\delta\mathbf{x})}{\partial(\delta\hat{\mathbf{x}})}\right)_j . \end{aligned}$$

The interested reader should note that we do not actually need to know the function g^{-1}, unlike f, g and f^{-1}. As suggested earlier, the problem may therefore in theory also be solved by using $\Phi(\delta\hat{y}) = G(\delta\mathbf{y})$, where $G(\delta\mathbf{y})$ is the (non-invertible) cumulative distribution of $\delta\mathbf{y}$ and $\delta\hat{y}$ is a *scalar*. The observation transformation is in this case given by $\delta\hat{y} = g(\delta\mathbf{y}) = \Phi^{-1}(G(\delta\mathbf{y}))$. However, as explained earlier, it is not possible to use this approach since the determination of $G(\delta\mathbf{y})$ requires an unrealistically large reference data set.

6. CONCLUSION

It is possible to formulate a Gaussian variable transformation of the observation error based on the statistics of a large reference data set. The 1D-VAR problem may be reformulated using the transformed observation error. The transformation ensures a better agreement with the Gaussian assumptions of the 1D-VAR method. Preliminary statistical results using 1D-Var on 3 to 9 hours HIRLAM forecasts show that the proposed method makes the retrieved profiles verify better against independent radiosonde data than without doing the transform.

References

Breivik, L.A. et al, 1999: A 1D-Var system for ATOVS data at the Norwegian Meteorological Institute. *DNMI Technical report no. 88*, ISSN 0332-9879. Available from DNMI.

Eyre, J., 1991: A fast radiative transfer model for satellite sounding systems. *ECMWF Technical Memorandum No. 176.* March 1991.

Eyre, J. et al, 1993: Assimilation of TOVS radiance information through one-dimensional variational analysis *Q. J. R. Meteorol. Soc, 119*, pp.1427–1463.

Lorenc, A.C., 1996: Analysis methods for numerical weather prediction. *Q. J. R. Meteorol. Soc, 112*, pp.1177–1194.