15B.8 ADAPTIVE ESTIMATION OF BACKGROUND AND OBSERVATION ERROR STATISTICS IN VARIATIONAL DATA ASSIMILATION

Zhuo Liu*[1], Mark Buehner[2] and Pierre Gauthier[1,2]

1. Department of Earth and Atmospheric Sciences, Université du Québec à Montréal, Montreal, Québec, Canada
2. Data Assimilation and Satellite Meteorology Division, Meteorological Service of Canada, Dorval, Québec, Canada

1. INTRODUCTION

Variational data assimilation system aims at providing an accurate estimation of the current state of the atmosphere by means of minimizing a cost function measuring the distance to the background and the observations.

The background error covariance matrix **B** and observational error covariance matrix **R** are specified and play a very important role. However, the matrix **B** and **R** are not very well known. Generally, background errors can be constructed by using the so-called NMC method or Ensemble Kalman Filter (Parrish and Derber, 1992; Houtekamer *et al.*, 1998). Both of them suffer from some deficiencies. One needs to tune these error statistics to make them consistent with the comparison to observations. Hollingsworth and Lönnberg (1986) used information from innovations (departures between observations and 6-h forecast). Expressed in observation space, innovations covariances are the sum of background and observation covariance matrix. Assuming that observation error is not spatially correlated, while background error is, it is then possible to estimate the background error correlations and variances, and observation error variances. To achieve this, some assumptions must be made about the background error correlations. Dee and da Silva (1998) used the framework of *maximum likelihood* estimation to determine various parameters, such as bias, correlation length or variance in data assimilation. Wahba *et.al* (1995) used randomized General Cross Validation (GCV) method to tune the error statistics.

In the context of the variational formulation, Talagrand (1999) investigated the statistical properties of the cost function and its components to establish an an optimality criterion that should be met if the error statistics were perfectly specified. If they are not, one can use these diagnostics to tune the specified covariance matrices by applying scaling factors to the background and the observational error statistics respectively, so that the total cost function meets the optimality criterion with respect to the new error statistics (Desroziers and Ivanov, 2001; Chapnik *et al.*, 2004, 2005). The advantages of this method are (1) scaling factors for the background and observation error statistics are estimated simultaneously, (2) for observations, the tuning coefficients can be divided into subsets allowing for more freedom of the choice of parameters to be estimated, (3) the tuning process is done on-line.

In this study, the method proposed by Desroziers and Ivanov (2001) is used to tune background and observational error statistics of the 3D-Var assimilation system of the Canadian Meteorological Centre (Gauthier *et al.*, 1999a-b). Several experiments were carried out to investigate the method and highlight some of its limitations.

2. THE METHOD OF DESROZIERS AND IVANOV

The variational data assimilation minimizes the following cost function:

$$
\begin{aligned}
J(\delta\mathbf{x}) &= \tfrac{1}{2}\delta\mathbf{x}^T\mathbf{B}^{-1}\delta\mathbf{x} + \tfrac{1}{2}(\mathbf{H}\delta\mathbf{x}-\mathbf{d})^T\mathbf{R}^{-1}(\mathbf{H}\delta\mathbf{x}-\mathbf{d}) \\
&= \tfrac{1}{2}\mathrm{Tr}(\mathbf{B}^{-1}\delta\mathbf{x}\delta\mathbf{x}^T) + \tfrac{1}{2}\mathrm{Tr}(\mathbf{R}^{-1}(\mathbf{H}\delta\mathbf{x}-\mathbf{d})(\mathbf{H}\delta\mathbf{x}-\mathbf{d})^T) \\
&= J_b(\delta\mathbf{x}) + J_o(\delta\mathbf{x})
\end{aligned}
$$

where $\delta\mathbf{x} = \mathbf{x} - \mathbf{x}^b$, $\mathbf{x}$ **is** the model state, $\mathbf{x}^b$ is the background state, **H** is the linearized observation operator $\mathbf{d} = \mathbf{y}^o - \mathbf{H}\mathbf{x}^b$ is the innovation vector, and $\mathbf{y}^o$ is the observation vector.

The solution $\mathbf{x}^a$ that minimizes ***J*** gives

$$\delta\mathbf{x}^a = \mathbf{x}^a - \mathbf{x}^b = \mathbf{K}\mathbf{d}$$

where $\mathbf{K} = \mathbf{B}\mathbf{H}^T(\mathbf{H}\mathbf{B}\mathbf{H}^T + \mathbf{R})^{-1}$ is the *gain matrix*.

One can derive the expectation of the innovation covariances to be:

$$\mathbf{D} = \langle \mathbf{d}\mathbf{d}^T \rangle = \mathbf{H}\mathbf{B}\mathbf{H}^T + \mathbf{R}$$

where brackets stand for statistical expectancy. So for a "well tuned" system, this equality should hold. Furthermore, we have

Corresponding author address: Dr. Zhuo Liu, Department of Earth and Atmospheric Sciences, Université du Québec à Montréal (UQAM), P.O. Box 8888, Montréal, Québec, CANADA H3C 3P8

$$
\begin{aligned}
\mathbf{B}^{-1}\left\langle \delta\mathbf{x}^a \delta\mathbf{x}^{aT} \right\rangle &= \mathbf{B}^{-1}\mathbf{K}\left\langle \mathbf{d}\mathbf{d}^T \right\rangle \mathbf{K}^T \\
&= \mathbf{B}^{-1}\mathbf{K}(\mathbf{HBH}^T + \mathbf{R})\mathbf{K}^T \\
&= \mathbf{B}^{-1}\mathbf{BH}^T(\mathbf{HBH}^T + \mathbf{R})^{-1}(\mathbf{HBH}^T + \mathbf{R})\mathbf{K}^T \\
&= \mathbf{H}^T\mathbf{K}^T = (\mathbf{KH})^T
\end{aligned}
$$

Thus,

$$E(J_b(\delta\mathbf{x}^a)) = \tfrac{1}{2}\mathrm{Tr}(\mathbf{HK})^T = \tfrac{1}{2}\mathrm{Tr}(\mathbf{HK})$$

Similarly, it can be shown that

$$
\begin{aligned}
E(J_o(\delta\mathbf{x}^a)) &= \tfrac{1}{2}\mathrm{Tr}(\mathbf{I}_p\text{-}(\mathbf{HK})^T) \\
&= \tfrac{1}{2}(p\text{-}\mathrm{Tr}(\mathbf{HK}))
\end{aligned}
\qquad (1)
$$

where p is the number of observations.

The expectation of the total cost function J satisfies:

$$E(J(\delta\mathbf{x}^a)) = E(J_b(\delta\mathbf{x}^a)) + E(J_o(\delta\mathbf{x}^a)) = \tfrac{1}{2}p$$

If this equality does not hold, this may mean the specified error variances are over (<) or under (>) estimated.

So one can define

$$S^b = 2J^b(\delta\mathbf{x}^a)/\mathrm{Tr}(\mathbf{KH})$$

and

$$S^o = 2J^o(\delta\mathbf{x}^a)/\mathrm{Tr}(\mathbf{I}_p - \mathbf{KH})$$

After multiplication with the covariance matrices **B** and **R** respectively, new covariances are redefined as

$$\widetilde{\mathbf{B}} = S^b\mathbf{B} \text{ and } \widetilde{\mathbf{R}} = S^o\mathbf{R}$$

so that

$$E(J^b(\delta\mathbf{x}^a)) = E(\tfrac{1}{2}\delta\mathbf{x}^{aT}\widetilde{\mathbf{B}}^{-1}\delta\mathbf{x}^a) = \mathrm{Tr}(\mathbf{KH})$$

and

$$E(J^o(\delta\mathbf{x}^a)) = E(\tfrac{1}{2}\delta\mathbf{x}^{aT}\widetilde{\mathbf{R}}^{-1}\delta\mathbf{x}^a) = \mathrm{Tr}(\mathbf{I}_p - \mathbf{KH})$$

Thus, one can define a new cost function:

$$J(\delta\mathbf{x}) = \frac{1}{s^{b2}}J_b(\delta\mathbf{x}) + \frac{1}{s^{o2}}J_o(\delta\mathbf{x})$$

where $s^{b^2} = S^b(s^b, s^o)$ and $s^{o^2} = S^o(s^b, s^o)$ are the scaling coefficients that can be obtained iteratively as follows:

1. Set $\mathbf{B}^{(1)} = \mathbf{B}$ and $\mathbf{R}^{(1)} = \mathbf{R}$.
2. Perform an analysis on the following cost function:

$$
\begin{aligned}
J(\delta\mathbf{x}) &= \frac{1}{2}\delta\mathbf{x}^T(\mathbf{B}^{(n)})^{-1}\delta\mathbf{x} \\
&\quad + \frac{1}{2}(\mathbf{d} - \mathbf{H}\delta\mathbf{x})^T(\mathbf{R}^{(n)})^{-1}(\mathbf{d} - \mathbf{H}\delta\mathbf{x}) \\
&= J_b^{(n)}(\delta\mathbf{x}) + J_o^{(n)}(\delta\mathbf{x})
\end{aligned}
$$

3. Compute

$$s_n^{o^2} = 2J_o^{(n)}(\delta\mathbf{x}^a)/(p - \mathrm{Tr}(\mathbf{HK}^{(n)}))$$

$$s_n^{b^2} = 2J_b^{(n)}(\delta\mathbf{x}^a)/\mathrm{Tr}(\mathbf{HK}^{(n)})$$

where

$$\mathbf{K}^{(n)} = \mathbf{B}^{(n)}\mathbf{H}^T(\mathbf{HB}^{(n)}\mathbf{H}^T + \mathbf{R}^{(n)})^{-1}$$

4. If $s_n^{b^2}$ and $s_n^{o^2}$ are not close to 1, we rescale the previous covariance matrices as

$$\mathbf{B}^{(n+1)} = s_n^{b^2}\mathbf{B}^{(n)},\ \mathbf{R}^{(n+1)} = s_n^{o^2}\mathbf{R}^{(n)}$$

5. Repeat step 2

Numerical experiments show that the iterative process is convergent, *i.e*.

$$s_n^{b^2} \to 1 \text{ and } s_n^{o^2} \to 1 \text{ as } n \to \infty$$

Finally, we obtain:

$$
\begin{aligned}
\mathbf{B}^{(n+1)} &= s_n^{b^2}\mathbf{B}^{(n)} = s_n^{b^2}s_{n-1}^{b^2}\cdots s_1^{b^2}\mathbf{B}^{(1)} = s^{b^2}\mathbf{B} \\
\mathbf{R}^{(n+1)} &= s_n^{o^2}\mathbf{R}^{(n)} = s_n^{o^2}s_{n-1}^{o^2}\cdots s_1^{o^2}\mathbf{R}^{(1)} = s^{o^2}\mathbf{R}
\end{aligned}
$$

and these are the scaling coefficients.

In the procedure, one needs to calculate $\mathrm{Tr}(\mathbf{HK}^{(n)})$. This can be done by a randomized estimation:

$$
\begin{aligned}
rand\left(Tr(\mathbf{HK})\right) &= (\mathbf{R}^{-1/2}\xi)^T\mathbf{HKR}^{1/2}\xi \\
&= (\mathbf{R}^{-1/2}\xi)^T(\mathbf{H}\delta\mathbf{x}^a_{(y^o+R^{1/2}\xi)} - \mathbf{H}\delta\mathbf{x}^a_{(y^o)})
\end{aligned}
.
$$

A detailed description of the method can be found in Desroziers and Ivanov (2001).

In this paper, we will focus mostly on the estimation of the observation error statistics for which different observation types have uncorrelated observation error in which case, eq.(1) still applies to individual sub-components of $\mathbf{J_o}$. Since rescaling will be applied to each sub-component, the subdivision should be made to group together elements that can be reasonably assumed to have the same error statistics. However, background error being highly correlated; tuning subcomponents of the $\mathbf{J_b}$ term cannot be done in a straightforward manner. We will then use a single tuning coefficient for $\mathbf{J_b}$.

Thus, instead of using a single tuning coefficient, one can use multiple tuning coefficients and the total cost function become:

$$J(\delta\mathbf{x}) = \frac{1}{s^{b^2}}J_b(\delta\mathbf{x}) + \sum_k \frac{1}{s_k^{o^2}}J_o^k(\delta\mathbf{x})$$

with

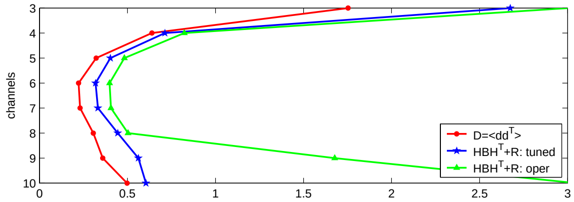

Figure 1 Stddev of innovation **D** and **HBH**' +**R** of Brightness Temperature at 12UTC(TOVS)

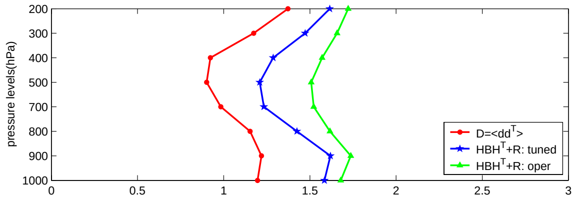

Figure 2 same as Figure 1, but for temperature and aircraft report

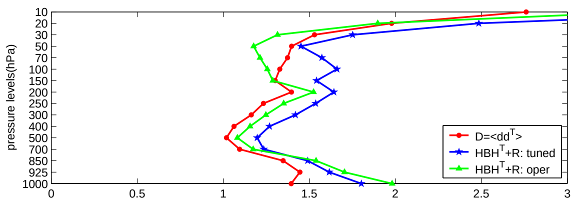

Figure 3 same as Figure 2, but for radiosonde

$$s_k^{o^2} = \frac{2J_o^k(\mathbf{x}^a)}{Tr(\mathbf{I}_{p^k} - \mathbf{H}_k\mathbf{K}_k)}, \forall k \in \{1, 2, \cdots, \nu_o\}$$

$$s^{b^2} = \frac{2J_b(x^a)}{Tr(HK)}$$

To increase the size of the sample to have more robust statistics, Sadiki and Fischer (2004) assumed these statistics to be stationary, in which case the size of the sample can be increased by considering several cases (at different dates and times). Instead of using one situation, they use *m* situations to calculate the tuning coefficients:

$$s_k^{o^2} = \frac{\sum_{i=1}^{m} 2J_{o_i}^k(\mathbf{x}_i^a)}{\sum_{i=1}^{m} Tr(\mathbf{I}_{p_i^k} - \mathbf{H}_k^i\mathbf{K}_k^i)}$$
$$\forall k \in \{1, 2, \cdots, \nu_o\}$$

$$s^{b^2} = \frac{\sum_{i=1}^{m} 2J_{b_i}(\mathbf{x}_i^a)}{\sum_{i=1}^{m} Tr(\mathbf{H}^i\mathbf{K}^i)},$$

where *m* refers to those distinct analysis times. For the same reasons, we also assumed the error statistics to be stationary, which makes it possible to perform the statistical averaging over larger data sets.

3. Estimation of observation error statistics

Several experiments were carried out based on the 3D-Var data assimilation system of the Canadian Meteorological Centre (Gauthier *et al.*, 1999,a-b) over a period of December 1, 2003, 12UTC, until December 31, 2003, 12UTC. Over that period, we retained one day every 5 days to insure statistical independency so that the data set comprises the data from seven different cases.

The observation data of upper air radiosondes (UA), aircraft report (AI), SATWind (SW) and radiance (TOVS) with the channel from AMSU-A (CH3) to AMSU- A (CH10) were used to tune the error statistics. Surface observation data (SF) were also used, but its error statistics were not tuned. For each observation family, a different scaling coefficient was introduced and computed. For the radiosondes, aircraft and SATWINDS, the error variance for winds and temperature, where applicable, at each level was scaled separately. For ATOVS, the error variance of each channel was also tuned independently. However, the background error being highly correlated, only one coefficient was used to scale globally the background error covariances.

Figure 1-3 show both the variances associated with the tuned and operational total error statistics **R** + **HBH**^{T}, and the innovation **D**. The results show that, except for UA, the tuned error statistics are more consistent with that of the innovations **D** compared with those used in the operational system.

Figure 4 shows the tuning coefficients for each variable. For the upper air observations (UA), the tuning coefficients for

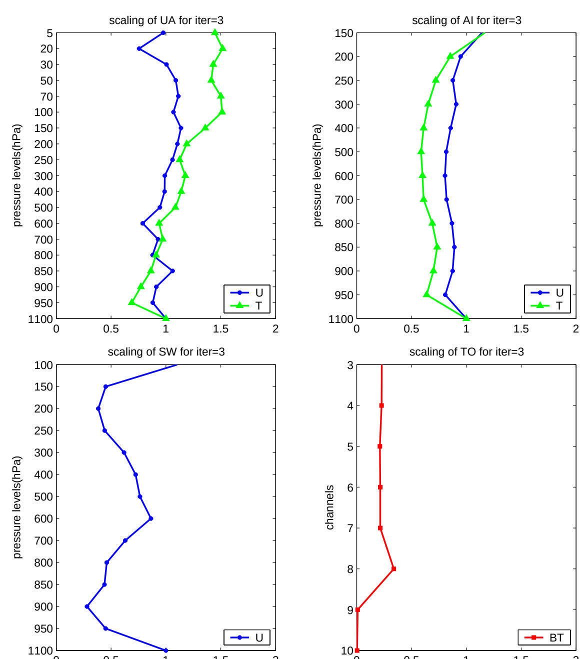

Figure 4 Tuning coefficients of variables after 3 iterations

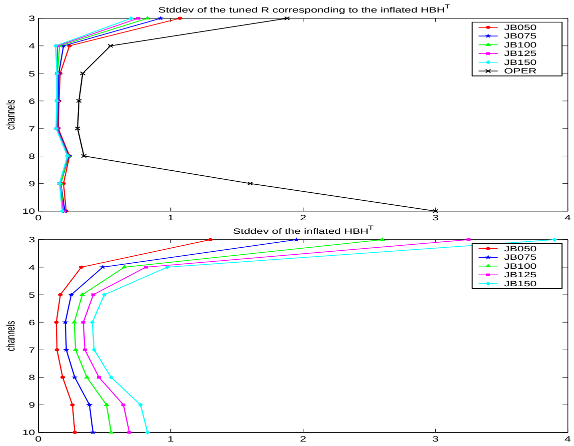

Figure 5: Tuned observation variances (top panel) obtained for TOVS for different background variances (bottom panel) artificially multiplied by factors ranging from 0.5 to 1.5. JB050 means multiplying the background error statistics by 0.5 and so on; OPER is operational observation error variance

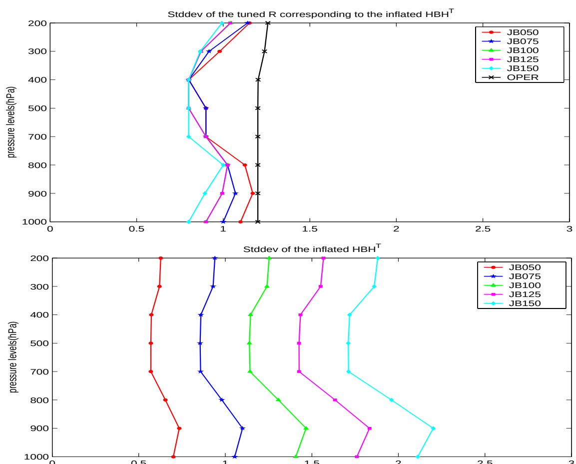

Figure 6 same as Figure 5, but for aircraft data

temperature are less than 1 near the surface and grow with height to become larger than 1. This means that the error statistics of the operational system were overestimated in the lower levels and underestimated in the middle and upper levels. While the tuning coefficients of wind components are more or less equal to 1, this indicates that they were already well-tuned. In the context of geostrophic balance, the error statistics of the background term are mostly controlled by the specified wind error statistics, which had to be consistent with the innovation statistics used.

For aircraft report and SATWIND observations, the tuning coefficients of wind components are smaller than 1 for all levels below 200 hPa, which implies that the error statistics were overestimated for those observations.

For the TOVS observations, the errors of operational for all channels were severely overestimated.

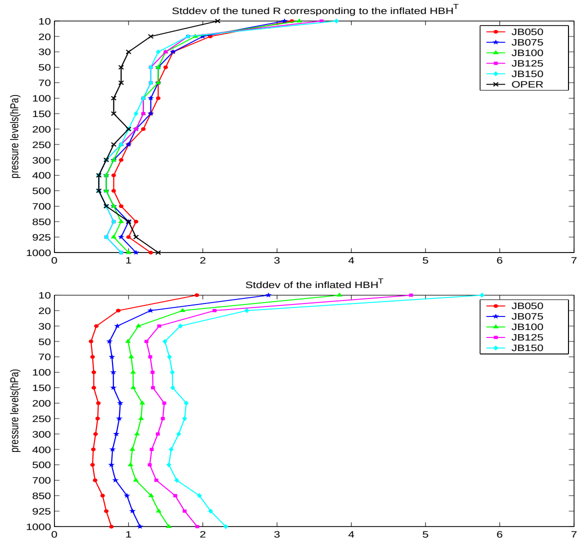

Figure 7 same as Figure 5, but for radiosonde data

4. Tuning observation error independently from the background-error

In the experiment presented above, it is far from obvious that the resulting estimates correspond to reality since the background error statistics are not optimal and were only rescaled globally with a single scaling coefficient. As discussed in Chapnik *et al.* (2004), under some circumstances it is possible to estimate the observation error statistics independently of the background-error statistics.

In order to see the impact of the background error on the tuning process, we carried out a series of experiments in which the background error was artificially inflated and hold fixed while the observations error statistics were tuned. If the argument of Chapnik *et al.* (2004) is true, the resulting estimates should reveal the conditions under which it is indeed possible to independently estimate the observation error statistics.

Figures 5-7 show the results of tuning the observation error statistics without tuning the background error statistics. The upper figure shows the tuned observation error statistics when the background error statistics are artificially changed (those are shown in the bottom panel for reference). The results show that for different scaling of the background error statistics, the changes in the tuned observation error statistics are small compared to the change induced in the background error. For radiance, the results are rather robust, which indicates that the scaling is not significantly affected by the background error statistics used in the assimilation.

5. Validation experiments

To test the limitations of applicability of the method, we adopted the method of Desroziers and Ivanov (2001) to generate an idealized experiment. It consists of generating

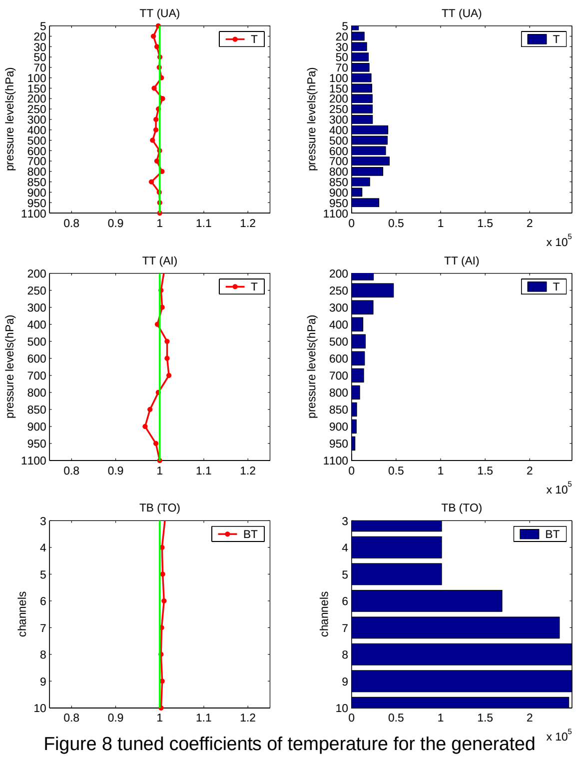

Figure 8 tuned coefficients of temperature for the generated observations and number of observations

synthetic observations by using the background state, i.e., first taking

$$x^t = x^b - B^{1/2}\zeta^b$$

to be the true value, where ζ^b is an n-dimensional random vector with Gaussian distribution and unit variance 1 and adding a known error coherent with the observation error statistics used in the assimilation. Those synthetic observations are then

$$\mathbf{y}^o = H(\mathbf{x}^t) + \mathbf{R}^{1/2}\zeta^o$$

where ζ^o is a p-dimensional random vector with Gaussian distribution and unit variance 1. The generated observation vector $\mathbf{y}^o$ has covariance matrix **R**, and is consistent with the background state, the background and observation errors statistics, *i.e.*, the statistical properties of the cost function are satisfied by construction. Using these observations, we went through the scaling process from which one should expect the tuning coefficients to be approximately 1.

Figure 8 shows the results for temperature for the generated observations (left) and the number of observations (right) used to obtain this estimate. It is seen that for TOVS data, the tuned coefficients are very close to 1, which is expected; while for other types of observations, there are some variations in the vertical. For TOVS, the number of observations is much higher than for the other types of observations. This is in agreement with our results that the results are

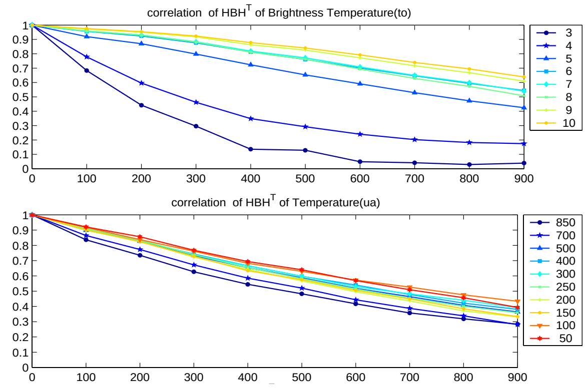

Figure 9 Correlation length of $\mathbf{HBH}^T$ Upper: Brightness temperature (TOVS); Lower: temperature (UA)

more reliable when there is a large number of data. This shows that the number of observations is a key point in the tuning process and the reliability of the estimate obtained.

To further investigate why the tuning results for TOVS are much better than other observations, we analyzed the correlation length of the background covariance matrix **B** for radiosonde (UA) and radiance (TOVS), using the innovations at every 12-h for UA and every 36-h for TOVS. This was to show if the observation coverage was sufficient to resolve the background error correlations. The region is from 20N to 90N, the increment of bin is 100km to represent the distance between the observations used.

Figure 9 shows the calculated background error correlations for brightness temperature in each channel for the radiance data (upper panel) and temperature at each level for radiosonde data (lower panel). The corresponding number of pairs of observations used is shown in the corresponding panels in figure 10.
It shows that for AMSU-A channel 5-10, the correlations are broader than others and the counts of the pairs of observations in each bin are also much higher than for others.

This confirm the conclusion of Chapnik *et al.* (2004) that when the correlations in the background covariance matrix

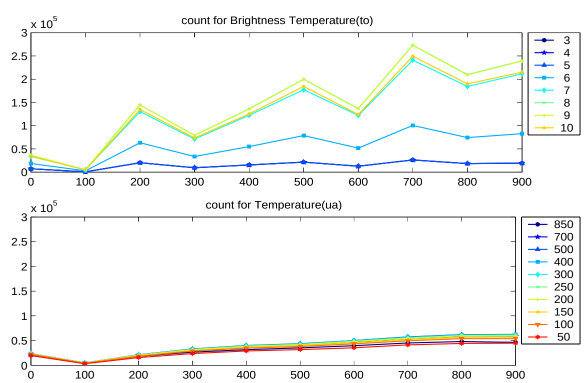

Figure 10 same as Figure 9, but for number of pair of observations

B are broad and well resolved by the observations, tuning the error statistics of observation without tuning that of the background error does not affect the evaluation of the optimal observation error tuning coefficients, as long as the number of observations is large enough. This has given us confidence that the estimate of observation errors obtained for the TOVS data are reliable and can be trusted and introduced in the assimilation.

6. Summary and conclusion

We conclude that using the statistical properties of the cost function to tune **R** can improve the estimate of error statistics of observations. The tuning process for TOVS is more robust than for other types of observations primarily because of the broader error correlations lengths that can be resolved by the huge number of observations. However, the estimation of the background-error statistics with this method remains problematic. In Buehner *et al.* (2005), an ensemble approach is used to estimate the background-error statistics. This method was shown to benefit from having better observation error estimates that were obtained as described in this paper.

Acknowledgements. This work was funded in part by the *Canadian Foundation for Climate and Atmospheric Sciences* through the research network on *Quantitative Precipitation Forecasting*.

7. References

Buehner, M., P. Gauthier and Z. Liu, 2005: Evaluation of new estimates of background and observation error covariances for variational assimilation. (submitted to *Quart. J.R. Meteor. Soc.*)

Chapnik, B and G. Desroziers, F. Rabier and O. Talagrand, 2004: Properties and first application of an error statistics and tuning method in variational assimilation, Quart. J. R. Meteorol. Soc., 130, pp.2253-2275

————, G. Desroziers, F. Rabier and O. Talagrand, 2005: Diagnosis and tuning of observational error statistics and in a quasi operational data assimilation setting. *Quart. J. R. Meteorol. Soc*., (to appear)

Dee, D.P. and A. da Silva, 1998, Maximum- likelihood estimation of forecast and observation error covariance parameters, Part I: Methodology, *Mon. Wea. Rev.*, **124**, 1822--1834

Desroziers, G. and S. Ivanov, 2001: Diagnosis and adaptive tuning of observation-error parameters in a variational assimilation, *Quart. J. R. Meteorol. Soc*., **127**, pp. 1433-1452

Gauthier, P., C. Charette, L. Fillion, P. Koclas and S. Laroche, 1999a: Implementation of a 3D variational data assimilation system at the Canadian Meteorological Centre. Part I: The global analysis. *Atmosphere-Ocean*, **37**, 103-156.

———— , M. Buehner and L. Fillion, 1999b: Background-error statistics modelling in a 3D variational data assimilation scheme: estimation and impact on the analyses. Proceedings of the *ECMWF Workshop on diagnosis of data assimilation systems*, Reading UK, p. 131-145.

Hollingsworth, A. and P. Lönnberg, 1986, The statistical structure of short-range forecast errors as determined from radiosonde data. Part I: The wind field, *Tellus*, **38A**, 111-136

Houtekamer, P.L. and H.L. Mitchell, 1998: Data assimilation using an ensemble Kalman filter technique. *Mon. Wea. Rev.*, **126**, 796-811.

Parrish, D.F. and J.C. Derber, 1992: The National Meteorological Center's spectral statistical interpolation analysis system. *Mon. Wea. Rev.,* **120**, 1747-1763.

Sadiki, W. and C. Fischer, 2005: A posteriori validation applied to the 3D-VAR Arpège and Aladin data assimilation systems. *Tellus*, **57A**, 21-34

Talagrand, O., 1999, A posteriori verification of analysis and assimilation algorithms, pp. 17-28 in Proceedings of the *ECMWF Workshop on Diagnosis of data assimilation systems*, 2-4 November 1999, Reading, UK

Wahba, G., D.R. Johnson, F. Gao and J. Gong, 1995, Adaptive Tuning of Numerical Weather Prediction Models: Randomized GCV in Three- and Four-Dimensional Data Assimilation, *Mon. Wea. Rev.*, **123**, 3358-3369