Estimation of observation error statistics, using an optimality criterion

B.Chapnik*, G.Desroziers*, F.Rabier* and O.Talagrand**
* Météo France, CNRM **CNRS, LMD

Desroziers and Ivanov's method (2001)

Let B and R be the background and observation error covariance matrices specified in an operational data assimilation system and let B_t and R_t be the "true" matrices. Assuming that one can write

$$B_t = s_b B,\ R_t = s_o R,\ \text{or}\ R_{tk} = s_{ok} R_k,\ B_{tl} = s_{bl} B_l,$$

where the k and l subscript refer to statistically independent subparts of the observations and of the control vector , the aim of this work is to evaluate the tuning coefficients: s_o and s_b (s_{ok} and s_{bl}) Desroziers and Ivanov (2001) proposed to use an optimality criterion found by Talagrand (1999). The tuning coefficients are those for which this criterion if fulfilled.

If $\boldsymbol{x_a}$ is the minimizer of J_t then, following Talagrand (1999), the expectations of the subparts of the cost function at the minimum are:

$$\mathbf{E(2J_{ok}(\mathit{x_a})/s_{ok})=Tr\ [\pi_k(I_p - HK)\pi_k^T]}$$

$$\mathbf{E(2J_{bl}(\mathit{x_a})/s_{bl})=Tr\ (\pi_l KH\pi_l^T)}.$$

Where **E** is the expectation operator, **K** is the gain matrix, **H** is the observation operator, and $\boldsymbol{\pi_k}$ and $\boldsymbol{\pi_l}$ are the projections onto the k^{th} type of observations and to the l^{th} independent subpart of the control vector. The tuning coefficients are computed as the limit of a fixed point algorithm, going from step *i* to step *i+1* using the following relations:

$$s^{(i+1)}_{ok} = 2J_{ok}(\boldsymbol{x_a}(\mathbf{s}^{(i)}))/\ \mathrm{Tr}\ [\pi_k(\mathbf{I}_p - \mathbf{HK}^{(i)})\pi_k^T],\ \forall k$$

$$s^{(i+1)}_{bl} = 2J_{bl}(\boldsymbol{x_a}(\mathbf{s}^{(i)}))/\ \mathrm{Tr}\ (\pi_l \mathbf{K}^{(i)}\mathbf{H}\pi_l^T),\ \forall l.$$

Chapnik *et al*. (2003) have shown that the method is equivalent to a Maximum likelihood tuning of the variances. (Dee and da Silva 1998); therefore, the quality of the estimates depends on the number of observations and on the quality of the *a priori* modelization of covariances. The computed values are temporally stable (up to four years); on the contrary they react quickly and increase when the quality of observations is degraded: they behave like variances are supposed to.

Moreover, as already stated by Desroziers and Ivanov, the first fixed-point iteration yields a good approximation of the final result making the following implementation of the algorithm feasible: only one fixed point iteration is used ands several situations were "concatenated" to increase the accuracy of the estimate. The estimation of the kth observational tuning coefficient becomes:

$$s_{ok} = \left(\boldsymbol{\Sigma_i}\ \mathbf{J_{ok}^i}(\boldsymbol{x_a})\right) / \left(\boldsymbol{\Sigma_i}\ \mathrm{Tr}\ [\pi_k(\mathbf{I}_{pi} - \mathbf{H^iK^i})\pi_k^T]\ \right),$$

where i refers to the i^{th} situation. The different situations used in the computation are separated by at least 5 days in order to prevent time correlation.

Results with simulated and real satellite radiances

Figure 1 shows the ability of the method to retrieve optimal variances in a simulated case. In this case the true standard deviations are the operational values and the mis-specified standard deviations are equal to the square root of the operational values; six dates, separated by more than five days, between 03/15/2003 and 05/19/2003 were used. Another computation was carried out with more thinning of the data in order to check the impact of a smaller number of observations. The standard deviations were computed for each of the three satellites NOAA15, NOAA16 and NOAA17, for sea pixel observations. In all cases the computed deviations are fairly close to the expected ones.

The result of the same computations, carried out with actual data, is shown in Fig. 2. Roughly, all the standard deviations are over estimated by a factor of 2. Satellite NOAA16 instrument seems to have a larger standard deviation for channel 8 than the other satellites. The standard deviations computed with a twice larger thinning interval are almost always larger than those computed with the operational thinning, which does not appear in the simulated case, possibly due to spatial or inter-channel correlation.

Similar computations were also carried out with well-documented data (TEMP profiles). The tuned variances remained close to the prescribed ones (not shown).

Conclusions and future directions.

The first iteration of Desroziers and Ivanov's algorithm, cumulating the observations over several dates, has been shown to be able to produce reliable estimates in a simulated case, its application to ATOVS radiances show several possibly useful and unexpected features but the role of possible correlations has to be clarified. Its application to well-documented data is encouraging. Future work will extend to the tuning of all observation types and a level by level tuning of B in order to evaluate the impact of this tuning on the analysis and on the forecasts.

References

B. Chapnik, G. Desroziers, F.Rabier and O.Talagrand. 2003,
Properties and first applications of an error statistic tuning method in variationnal assimilation. Submitted to Q.J.R.M.S

G. Desroziers and S. Ivanov. 2001,
Diagnosis and adaptive tuning of information error parameters in a variational assimilation.
Quart. J. Roy. Meteor. Soc., 127, 1433--1452

D. Dee and A. da Silva. 1998,
Maximumlikelihood estimation of forecast and observation error covariance parameters. part I: Methodology.
Mon. Wea. Rev., 124:1822--1834.

O. Talagrand. 1999;
A posteriori verification of analysis and assimilation algorithms.
In Proceedings of the ECMWF Workshop on Diagnosis of Data Assimilation Systems, 24 November pages 17--28, Reading

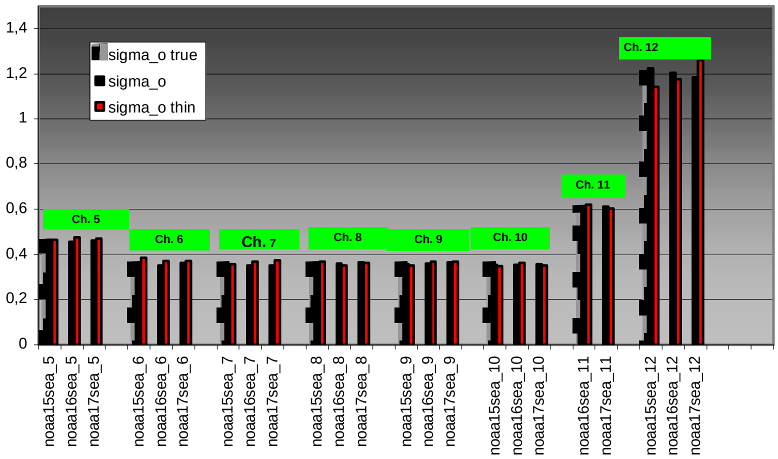

Figure. 1: Standard deviations of AMSU A channels obtained by the method in a simulated case. The black bars are computed with the operational thinning between obs. and the red bars with a twice larger thinning interval. A different deviation is computed for each satellite , a difference is also made between sea and land observations. The grey bars with dots show the simulated « true » standard deviations.

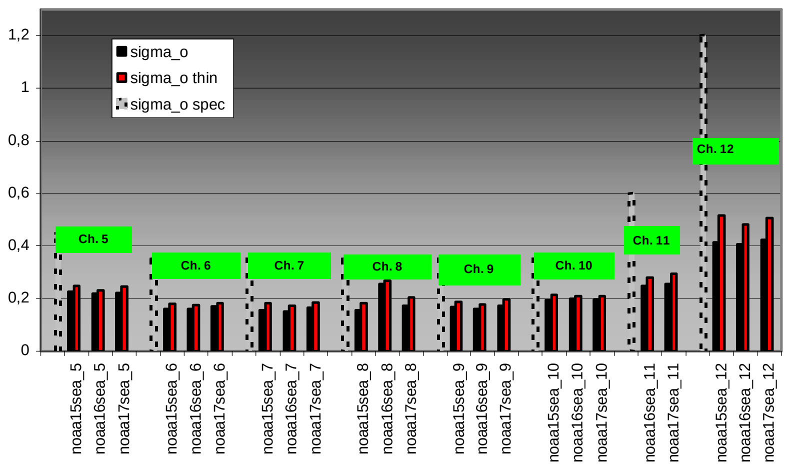

Figure 2: Standard deviations of AMSU A channels obtained by the method in a true case. Plotting conventions are the same as in Fig. 1 but this time the grey bars with black dots are the prescribed standard deviations