Estimating observational, background and model error of AROME using a posteriori diagnostics

RC-LACE supported stay at Météo-France
15th March - 5th June 2010

Benedikt Strajnar
EARS, Ljubljana

Supervisor: Loïk Berre

June 4, 2010

Contents

1 Introduction

The quality of analysis provided by data assimilation relies on providing good error characteristics of all information sources, i.e. the background and observations. The former can be estimated by several error simulation techniques, e.g. with an ensemble of forecasts. It accounts for the errors, originating from a small perturbations in the initial and/or boundary conditions, but not for the model imperfections. The background covariances deduced from

ensemble assimilations are thus somewhat too small because they do not take into account the error introduced by imperfect model equations, physical parametrizations, and resolution. In operational data assimilation applications, they are usualy inflated by multiplicative factors.

This work relies on the idea, that one can model model error as a fraction of background error. So that model error $\mathbf{Q} = \alpha\mathbf{B}$, where $\mathbf{B}$ is the background error covariance matrix. The fraction α could be estimated by the comparison of expected background covariances in a perfect model framework with diagnosed values. Most of the work, presented in this report, is devoted to finding the true backgorund and observational error standard deviations, by using a posteriori diagnostics. Two suitable methods, introduced by Desroziers and Ivanov (2001) and Desroziers et al. (2005), are briefly presented in the next section, and then used in both simplified framework and then AROME high resolution model data assimilation system. The presented a posteriori diagnostic methods are known to be efficient only in case that the correlation structures in background and observational error covariance matrices are different enough. For covariances of residuals, this can be shown with the spectral interpretation (Desroziers et al., 2005). While the separation in the correlation structure is not questionable for a global model like ARPEGE (with the background error length scales of few hundreds of kilometers), it becomes an issue for high resolution models.

Using these findings, we can finally obtain an indirect estimate of model error. Once α has been computed, one could consider their usage in an assimilation ensemble, to aproprietly inflate the short-range forecasts, used as a background for analysis.

2 Methods of a posteriori diagnostics

This section presents two methods that enable a posteriori diagnostics. They are relatively related, but they rely on two different optimality criteria. The first method, used in this study, is based on ideas of Talagrand (1999) and was introduced by Desroziers and Ivanov (2001). It compares the value of the variational cost function (J) after the analysis (at the end of iterative minimization process) with its theoretical expectation, which equals to (Desroziers and Ivanov, 2001):

$$E[J(\mathbf{x})] = E[J_b(\mathbf{x})] + E[J_o(\mathbf{x})] = \frac{1}{2}Tr(\mathbf{HK}) + \frac{1}{2}(p - Tr(\mathbf{HK})) = \frac{p}{2}. \tag{1}$$

Here p is the number of assimilated observations, $Tr(.)$ denotes trace operator, $E[.]$ is expectation operator, and $\mathbf{K}$ and $\mathbf{H}$ represent analysis gain matrix and linearized observation operator, respectively. Basic diagnostics of this kind involves two scalar parameters, used to present the balance between diagnosed and predefined error standard deviations. Following Desroziers and Ivanov (2001), they are defined as

$$S_o = \frac{2J_o(\mathbf{x}_a)]}{Tr(I_{p*p} - \mathbf{HK})} \tag{2}$$

$$S_b = \frac{2J_b(\mathbf{x}_a)]}{Tr(\mathbf{KH})}. \tag{3}$$

The values of $J_o(\mathbf{x}^a)$ and $J_b(\mathbf{x}^a)$, observational and background cost functions, are obtained directly as an output of the analysis. On the other hand, $Tr(\mathbf{HK})$ $(= Tr(\mathbf{KH}))$ is not explicitly present in the analysis procedure. An estimate of it is therefore needed. Several methods are possible; here we use the more recent approach introduced by Desroziers et al. (2009). The trace of HK is computed in the ensemble data assimilation framework. It is

possible to estimate traces on a subsets of observations (e.g. by type or location). The expression to estimate trace on a subset of observations with assimilation ensemble is

$$Tr(\Pi_i \mathbf{HK} \Pi_i^T) = \frac{1}{2L} \sum_{l,l'} (\mathbf{y}^o_{l',i} - \mathbf{y}^o_{l,i})^T (\frac{1}{\sigma^o_i})^2 (H_i(\mathbf{x}^a_{l'}) - H_i(\mathbf{x}^a_l)), \tag{4}$$

where Π_i is so called projection operator, which extracts a set of p_i observations from a total observational vector $\mathbf{y}$ with p observations, $\mathbf{x}_a$ is the analysis, and σ^o_i is the observational error standard deviation of the subset. The sum is over different ensemble members l and l'. Because this method uses the value of cost function at the end of minimization process as the optimality criterion, we will further refer to it as Jmin method.

The second approach, proposed by Desroziers et al. (2005), is based on the information provided by first guess, background, and analysis departures. Noting that

$$E[(\mathbf{y} - \mathbf{x}_a)(\mathbf{y} - H(\mathbf{x}_b))^T] = \mathbf{R} \tag{5}$$

$$E[(\mathbf{x}_a - H(\mathbf{x}_b))(\mathbf{x}_a - H(\mathbf{x}_b))^T] = \mathbf{HBH}^T, \tag{6}$$

it is possible to diagnose diagonal parts of covariance matrices, in observation space, for any subset of p_i observations, as

$$(\sigma^d_o)^2 = (\mathbf{y}_i - \mathbf{x}_{a,i})(\mathbf{y}_i - H_i(\mathbf{x}_{b,i}))^T \tag{7}$$

$$(\sigma^d_b)^2 = (\mathbf{x}_{a,i} - H_i(\mathbf{x}_{b,i}))(\mathbf{x}_{a,i} - H_i(\mathbf{x}_{b,i}))^T. \tag{8}$$

To compare diagnosed error standard deviations (σ^d_o, σ^d_b) with the predefined ones, like before, we introduce the scalar parameters

$$S_o = \frac{\sigma^d_o}{\sigma_o} \qquad\qquad S_b = \frac{\sigma^d_b}{\sigma_b}, \tag{9}$$

where σ_o and σ_b are predefined observational and background error standard deviations (i.e. those used in the analysis process). We will further refer to this method as covariances of residuals.

Both described methods are iterative. One should redo the analysis, with factors So and Sb applied to error covariances, to obtain the new estimates of So and Sb. However, it has been shown by the cited authors that the iteration converges quickly and so the first estimate is not to far from the final converged value. Moreover, the speed of convergence increases with the number of observations (Chapnik et al., 2004). Both methods will be applied and compared throughout this report. They are applied first to a toy model within a simplified framework, and then to AROME ensemble assimilation system.

3 Simplified framework runs

A very crude validation of performance of the tuning methods, in the conditions similar to those of the target AROME high resolution assimilation system, can be provided by toy model experiments. Such 1D simple model, used to investigate the behavior of the methods, is the same to that used in the studies of Chapnik et al. (2004), Chapnik et al. (2006) and, to illustrate their spectral response, also in Desroziers et al. (2005).

The predefined background error length scales, diagnosed from AROME $\mathbf{B}$ matrix specification, range from less then ten to few tens of kilometers, at 2.5 km horizontal resolution. To mimic the situation like this, a periodic domain with 601 grid points (300 wave numbers) and

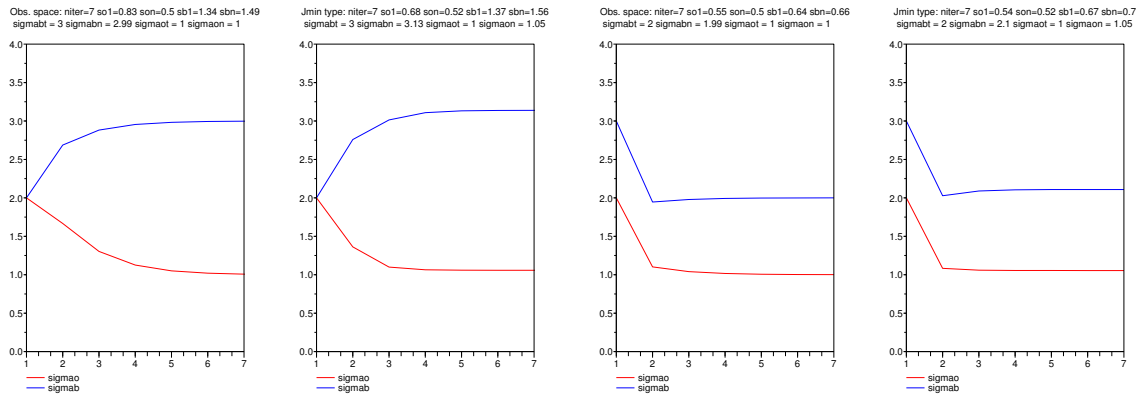

Figure 1: Diagnosed σ_o (red) and σ_b(blue) as a function of number of iteration (starting point is denoted by 1,first iteration by 2), computed using covariances of residuals (first and third from left) and Jmin method (second and forth from left). Presented are situations with overestimated σ_o and underestimated σ_b(leftmost two panels) or overestimated σ_b(two rightmost panels). Prescribed and true correlation length scales are L_b=25 km, L_b^t=25 km, L_o=0 km and L_o^t=0 km.

dimension of 1500 km has been set-up. There are 150 equally spaced observations. Random noise, generated from the prescribed $\mathbf{B}$ matrix, is added to the first guess. Similarly, the observational noise, drawn from specified observational error matrix $\mathbf{R}$, is added to observations. We tried to choose prescribed error variances and correlation length scales (L) such that possible real cases would be contained in the experimentations. Instead of using variational algorithms, the size of the problem enables the exact computation of analysis, using best linear unbiased estimate (BLUE). The main advantage of such toy model is, however, that the true values are known and can be controlled.

As indicated by the previous diagnostics with ARPEGE and ALADIN, the prescribed observational error seems to be overestimated in these two models (e.g., Desroziers et al. (2009) and Sadiki and Fischer (2005)). So the specified observational error variance was set to be larger than the true one. For background terms, both larger and smaller error variances, compared to their true values, were tried. The behavior of the two methods in this simple case is illustrated on figures 1 - 3. The values of true ($\sigma_o^t = 1$) and specified observational error standard deviations are fixed: $\sigma_o^t = 1$ and $\sigma_o = 2$. The prescribed and true values for background errors (σ_b^t) are set to 2 and 3; both combinations are tried. Various situations are the simulated by choosing different length scales of (Gaussian) correlation functions, used to form the covariance matrices.

Figure 1 presents the situation with well specified background error correlation, and no correlation in observational errors. Both diagnostics are able to find the true error standard deviations after a few iterations. Even the first estimate, especially in the case with both σ_o and σ_b overestimated, is already fairly close to the true final value. The speed of convergence is similar for both methods. Covariances of residuals seem to provide slightly more accurate converged value.

However, observational errors are also likely to be (at least weakly) correlated. Figure 2 shows the situation with true observational error not equally zero, but with about twice smaller length scales compared to background error, which is still perfectly modeled. Both methods still converge, but the convergence is slower in less accurate. In particular, this holds for Jmin type method. In the case with underestimated background error standard

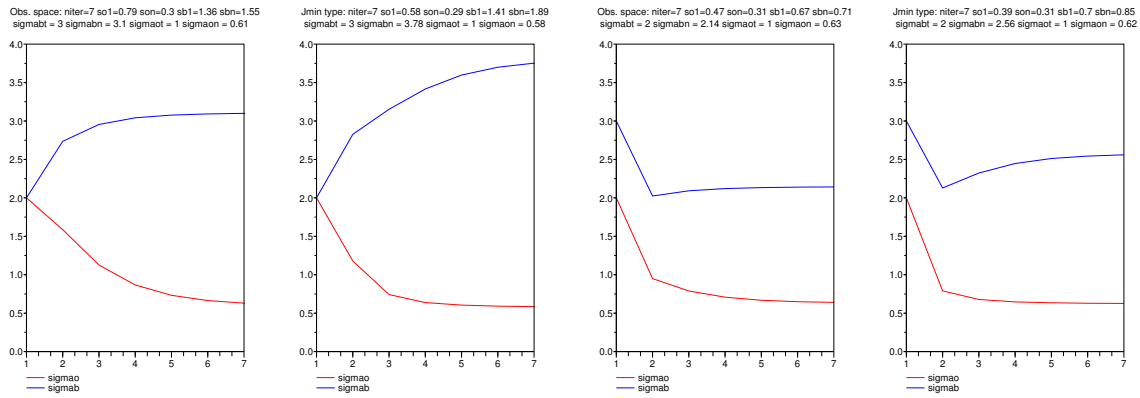

Figure 2: Same as figure 1, but with L_b=25 km, L_b^t=25 km, L_o=0 km and L_o^t=10 km.

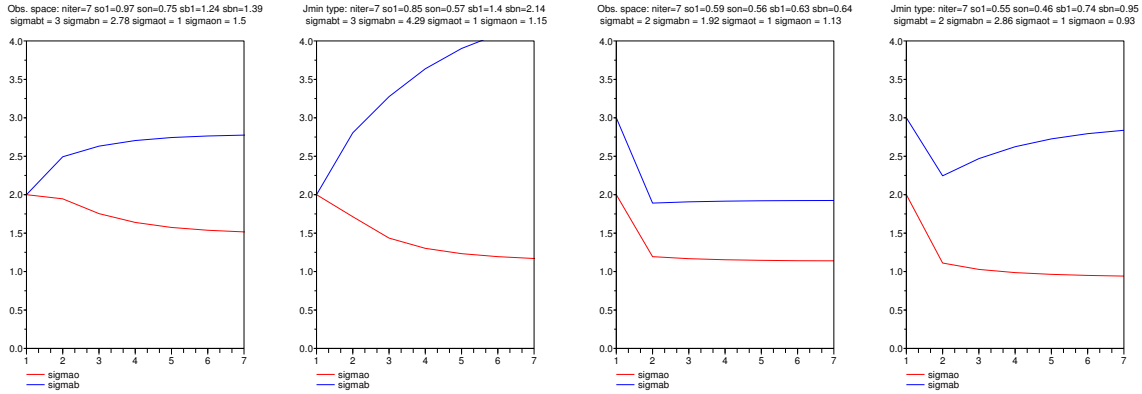

Figure 3: Same as figure 1, but withL_b=25 km, L_b^t=15 km, L_o=0 km and L_o^t=10 km.

deviations, it is moving away from the true value from the second iteration on. But still, the first iteration is reasonable for both methods.

In the last case (figure 3), both background and observational error correlations are set to be miss specified, true observational error correlations still being relatively small. The performance of the methods is now even slightly worse, in particular the final convergence point is wrong. Again, the first iteration is still in the right direction. If one increases observational error correlations further, the methods are not able to provide reasonable estimates (not shown).

The simple experiments show that the methods perform comparably and that they can be used, even if the observational or background error correlations are not exactly true. The covariances of residuals appears to be a little more robust. The observed robustness is expected to further increase with the number of observations, which is at least few thousands in a real assimilation system. In this toy case, we are limited to around 150 observations; in order to mimic the observation density and to keep a reasonable number of grid point on the other hand. The experiments with the toy model can be therefore affected by particular realization of a random process, used to generate background and observational errors.

4 Data set

The further diagnostics are carried out with 2.5 km AROME ensemble data assimilation system. Its 6 assimilation cycles are generated by explicitly perturbing observations; backgrounds are only perturbed explicitly. The system is coupled to ALADIN assimilation ensemble with similar characteristics, and horizontal resolution of about 10 km. The testing period covers the period of 23th February - 18th March 2008. The variety of synoptic situations is covered by this period.

5 Diagnostic experiments with AROME

5.1 Application of Jmin method

The quantity $Tr(\mathbf{HK})$ has been computed separately for different observation types, using the methodology presented in section 2 (see equation 4). It may be interesting to note that this quantity also represents the degree of freedom for signal (DFS), as also explained by Desroziers and Ivanov (2001) and Desroziers et al. (2009). DFS measures the impact of a given observation subset on the final analysis. It is an important tool to assess the impact of different observation types. A temporal distribution of this quantity for different observation types is shown on figure 4, top right panel. It can be seen that the impact on the AROME analysis is largest for surface (SYNOP) observations, followed by aircraft, radar, satellite and radiosonde observations. The impact of radiosondes, as presented here, is reduced by their sparse temporal resolution within 3-hour assimilation cycle. The impact of single observation, however, is largest for satellite derived winds (SATOB), satellite radiances, scatterometer and radar observations (top right panel of figure 4). It can be also seen that the impact of observation is largest for 12 UTC analysis, and smallest for 03 UTC analysis.

Using the computation of DFS ($Tr(\mathbf{HK})$) product and final values of J_o and J_b after the minimization, the time series of the coefficients S_o and S_b have been computed (figure 5, bottom right panel). The first surprising feature to observe is the similarity of evolutions of the two diagnostic coefficients. S_o exhibits some more variation, and is for some cases (for the network times with more observations in particular) somewhat larger than S_b. The average value of both is around 0.57. This behavior of diagnostics, unseen in experiments with global models, may indicate possible weaknesses of the method. This is the reason to use also diagnostics in observation space, namely covariance of residuals.

5.2 Application of covariances of residuals

The covariances of residuals are computed in the observation space, observation type by observation type. In order to be able to compare the diagnosed error standard deviations with those found with Jmin method, one needs to compute a global average. We use the following approach (in case of S_o, but similarly for S_b):

$$S_o = \sqrt{\sum_i \frac{p_i}{p} (S_o^i)^2} \tag{10}$$

This means that we average variance rather then standard deviations. Weights are proportional to the number of observations for a given observation type. It has been found that the ratio S_o is very close to the one provided by Jmin method. This provides a confirmation that the two methods perform successfully, as also suggested by toy model example. Diagnostic

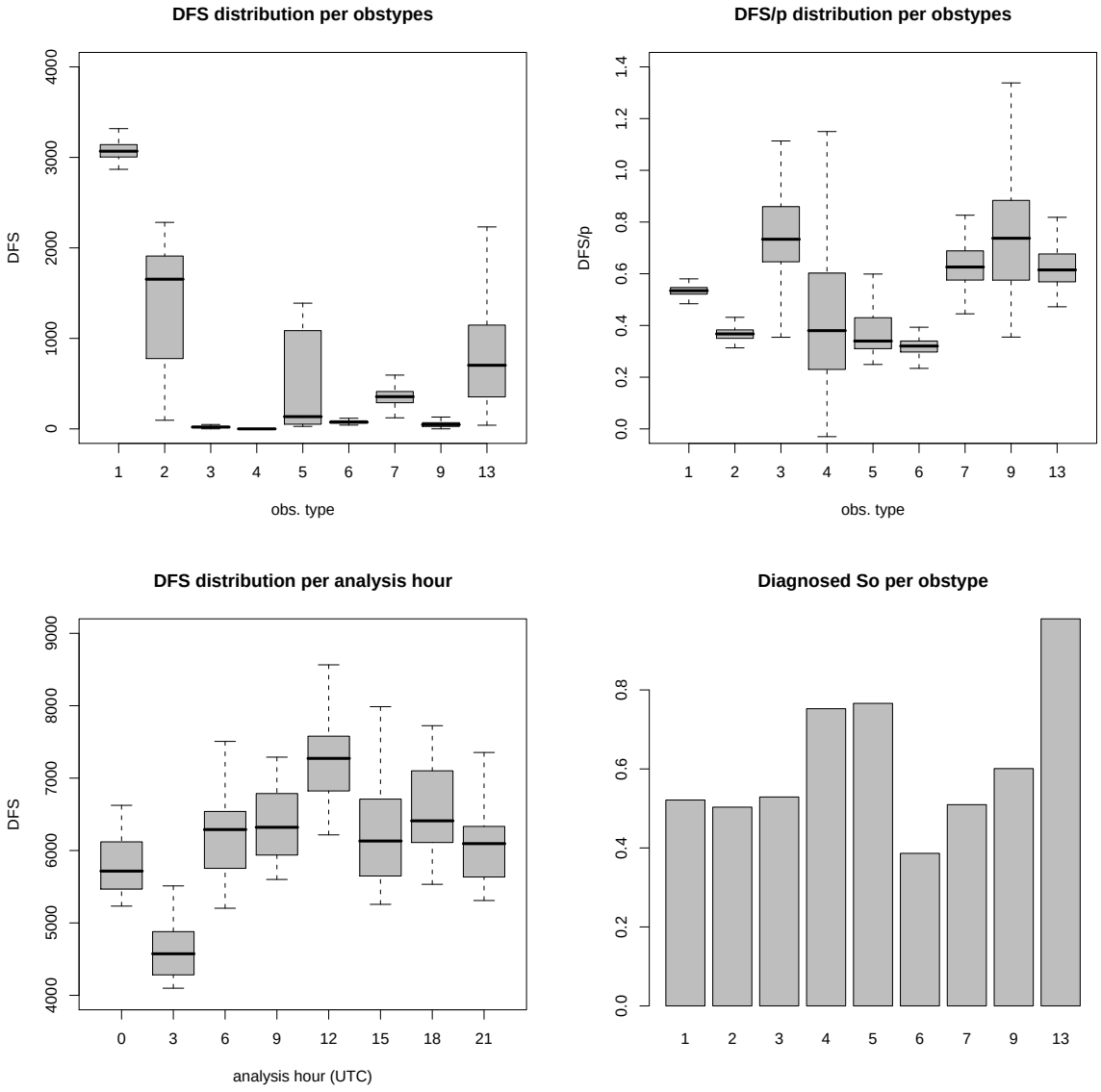

Figure 4: Distributions of DFS over the period (top left), DFS per single observation (top right), total DFS per analysis time (bottom left), and average diagnostic factor So diagnosed with the Jmin method (bottom right) for different observation types. Observational type numbers are: 1 - SYNOP, 2 - AIREP, 3 - SATOB, 4 - DRIBU, 5 - TEMP, 6 - PROFILER, 7 - satellite radiances (AMSU, HIRS, SEVIRI), 9 - SCATT, 13 - RADAR (radial winds only).

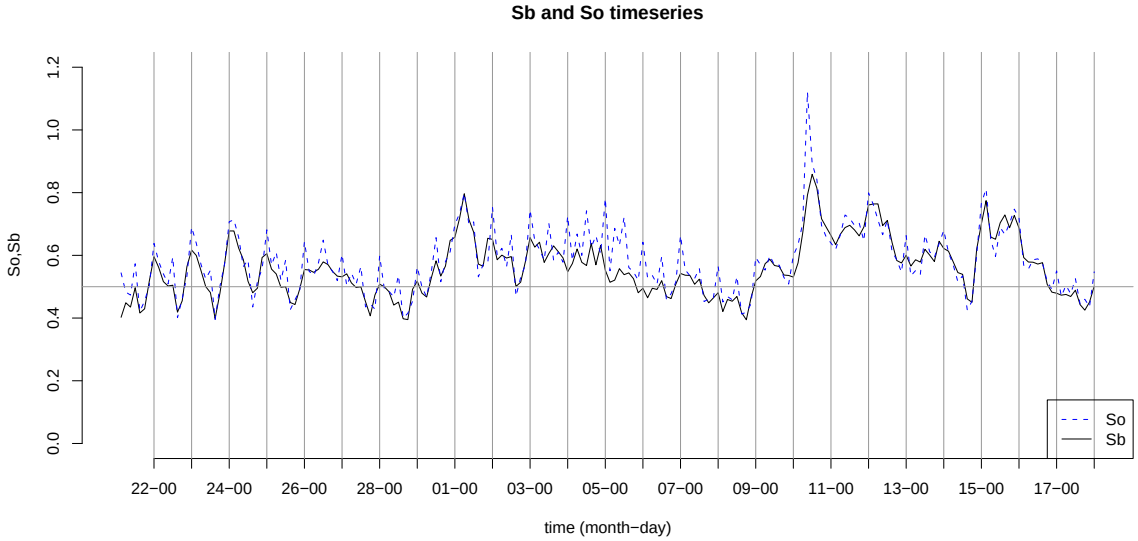

Figure 5: Temporal evolution of diagnosed So and Sb, using Jmin method. The mean values are So=0.58 and Sb=0.56.

factors S_b, however, were quite far from the Jmin counterparts in the first trials. However, an error was introduced in the computation by obtaining (prescribed) σ_b from observational database (ODB). Those value, written into ODB during observation screening, correspond to ARPEGE background errors, but are used also in AROME screening. As the background errors, really used in the minimization, are provided in specific **B** matrix files (and not written in the ODB), these files had to be examined to provide the correct a priori background error specifications.

This raises a question about computation of prescribed wind background error variances. They are provided by **B** in terms of divergence and vorticity, as spectral coefficients. Those can be transformed to expected full wind background errors, using the spectral relation

$$E[(uv)^2] = -\frac{1}{\triangle(k^*)} \sum_{k^*} \frac{1}{2}(\zeta^2(k^*) + \eta^2(k^*)), \tag{11}$$

where k^* is the total 2D wave number ($k^* = \sqrt{m^2 + n^2}$). This approach has been used before for diagnostics with ALADIN model (Bölöni, 2006).The quantity uv is a squared sum of vorticity and divergence variances over all wave numbers and can be seen as an average wind standard deviation. In observational space, it corresponds to

$$(uv)^2 = \frac{1}{2}\sqrt{u^2 + v^2}, \tag{12}$$

and this quantity can be compared to error standard deviations, diagnosed by the method. The presented technique has been implemented inside *fediacov*, a program to diagnose covariances in the **B** matrix files.

Additionally, the covariances of residuals are computed in observation space, while background errors are provided on model levels. An interpolation to a common vertical grid is then needed. A sparse vertical grid was used, each segment spanning more model vertical levels. Finally, a temporal average of the vertical profiles of diagnosed and prescribed variances were calculated, in order to be able to carry out the comparison and to estimate the S_b values for

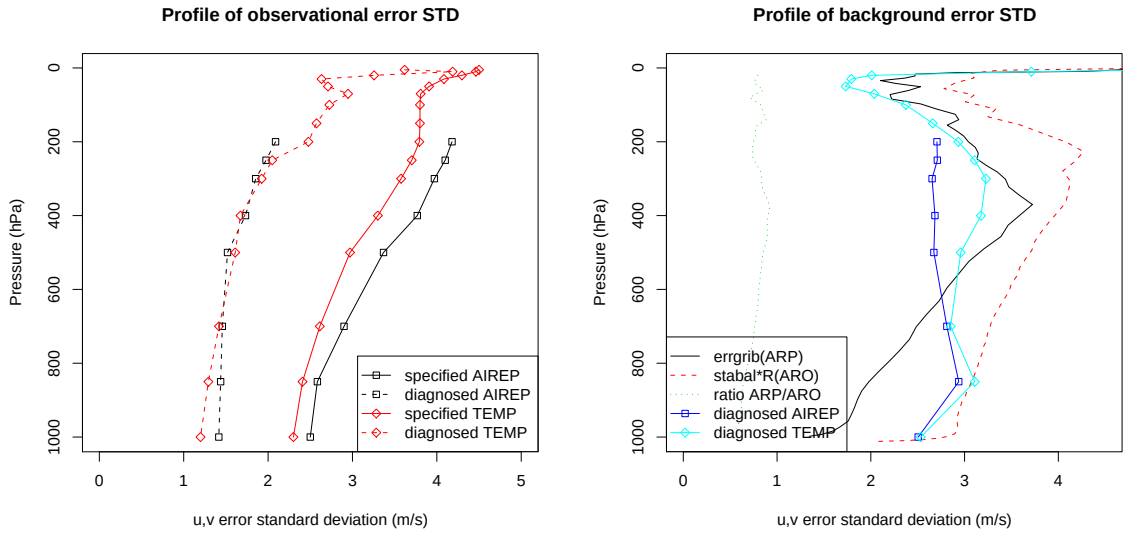

Figure 6: Estimated and predefined profiles of σ_o (left) and σ_b (right) for wind (average of both components). For background errors, both predefined AROME (*stabal*) and ARPEGE (*errgrib*) profiles are plotted. R stands for multiplication by *REDNMC*.

this method. As a reference, ARPEGE background errors were also included into analysis. Since they are specified in a non-homogeneous way on a global grid, a horizontal averages (in terms of variances) over the AROME domain were calculated, for each model level.

The wind standard error deviations, diagnosed by aircraft observations (AIREP) and radio sounding (TEMP), are presented on figure 6. The diagnostics show consistent over estimations of observational error (diagnosed σ_o being approximately one half of predefined) and also background error, at the levels above 800 hPa. Above 300 hPa, the diagnosed errors are closer to ARPEGE background errors.

The temperature diagnostics (figure 7) show that background errors are either almost perfectly specified (TEMP) or just slightly overestimated (AIREP) in the AROME $\mathbf{B}$ matrix. Observational error seem to be fairly overestimated again, for both observation types. Moreover, the difference between TEMP and AIREP in terms ob observational error seems to be smaller than predefined.

Specific humidity errors, again, seem to be overestimated (both background and observation parts, figure 8). Particularly in this case, it seems that ARPEGE background error profile (as provided in *errgrib* for screening in AROME), match the diagnosed profile much better than that of AROME.

5.3 Comparing ARPEGE and AROME error profiles

To further validate the presented diagnostic, the diagnosed σ_o and σ_b can be compared to the corresponding values of ARPEGE model. It is expected to observe larger σ_o values in ARPEGE than AROME because the former contains less represented scales, which help model to be able to represent the observed variability. On the other hand, background errors are expected to be larger in AROME because there is more uncertainty in the smaller scales, represented by this high resolution model. The comparisons for wind, temperature and specific humidity observation and background error are plotted on figures 9 - 11. It can be noticed

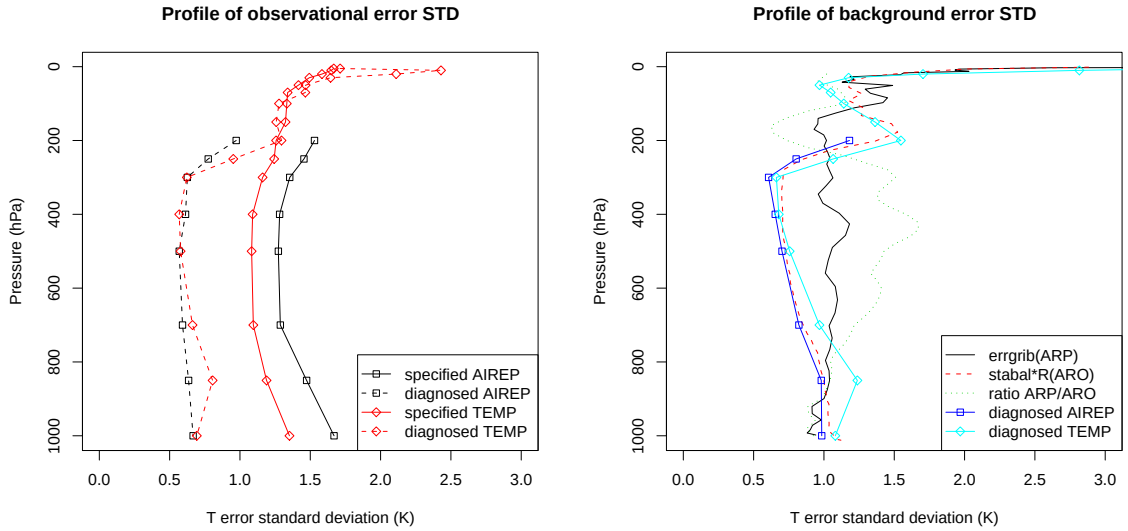

Figure 7: Estimated and predefined profiles of σ_o(left) and σ_b(right) for temperature. For background errors, both predefined AROME (*stabal*) and ARPEGE (*errgrib*) profiles are plotted. R stands for multiplication by *REDNMC*.

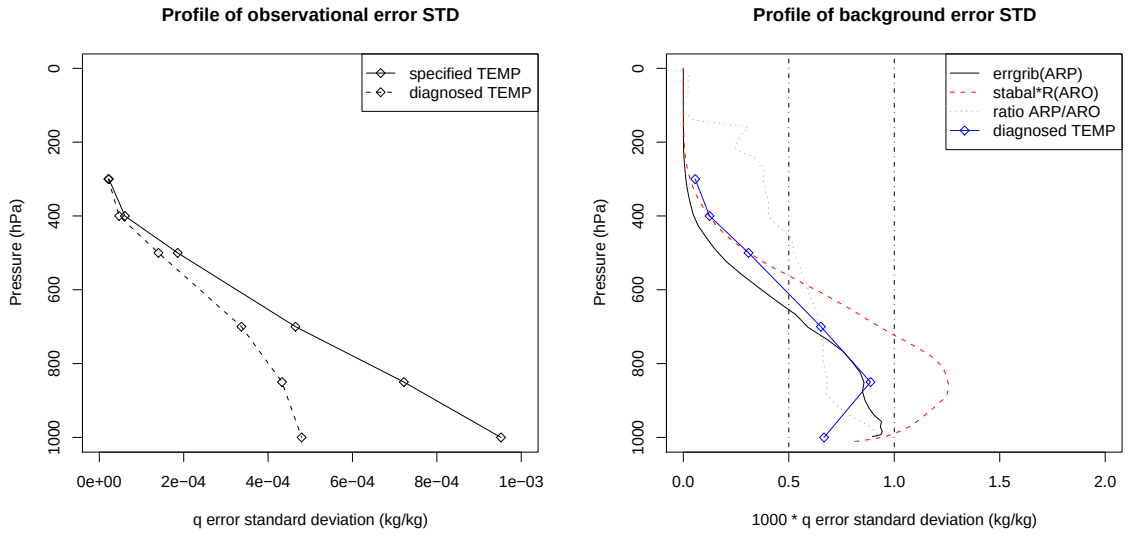

Figure 8: Estimated and predefined profiles of σ_o(left) and σ_b(right) for specific humidity. For background errors, both predefined AROME (*stabal*) and ARPEGE (*errgrib*) profiles are plotted. R stands for multiplication by *REDNMC*.

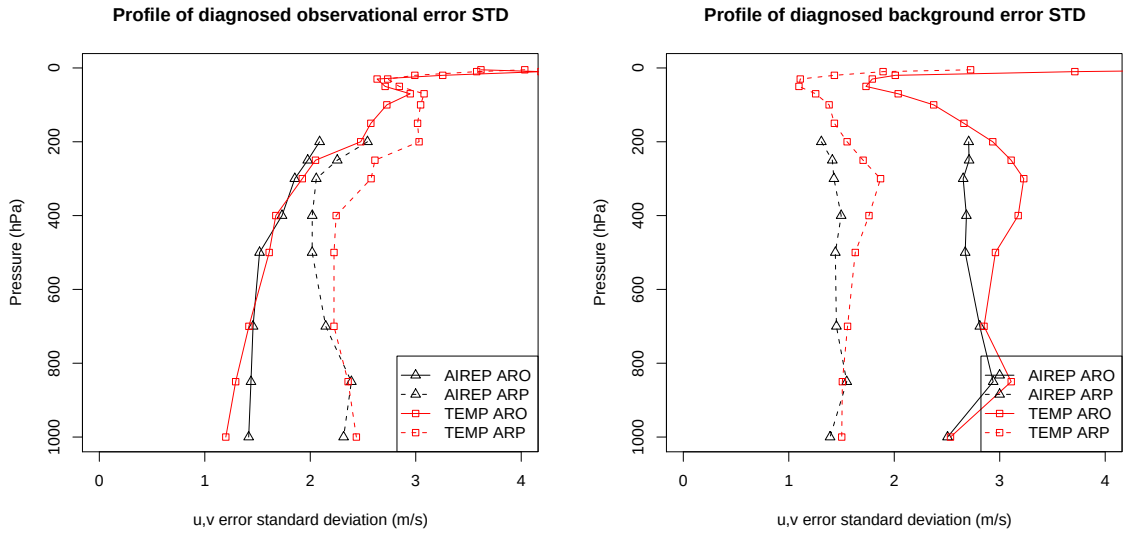

Figure 9: Comparison of observation and background errors, diagnosed for AROME and ARPEGE.

that the results fully support this hypothesis.

Additionally, it may be of interest to compare the standard deviation of background departures, because they represent the sum of background and observational error, in an optimal analysis system. The covariances of residuals basically try to determine the parts corresponding to both error contributions. Results, plotted in figure 12, indicate that standard deviation of background departures are higher in AROME than ARPEGE, especially those of wind and humidity.

6 Tuning experiments

6.1 Impact runs

Previous chapter presents some findings about the ratios of diagnosed and prescribed error standard deviations in AROME assimilation system. In this chapter, we present the tuning approach used in this work. In contract to type or height dependent tuning, we choose a constant tuning for background and observation errors. This decision is based on the experiences with tuning ARPEGE model. Finally, two choices of S_o and S_b are suggested by the diagnostics:

- S_o=0.58 and S_b=0.56, diagnosed with Jmin method
- S_o=58, and S_b=0.86, based on covariances of residuals. The observation error tuning is the same as for Jmin, and for background error we consider a rather subjective combination of the factors for wind (0.75-0.85), temperature (0.9-1) and humidity (0.7).

From the technical point of view, one also has to consider the following in the process of tuning:

- there is a parameter in ARPEGE/ALADIN/AROME code which multiplies (in a constant way) the background error covariances, called *REDNMC*, which is equal to 1.5

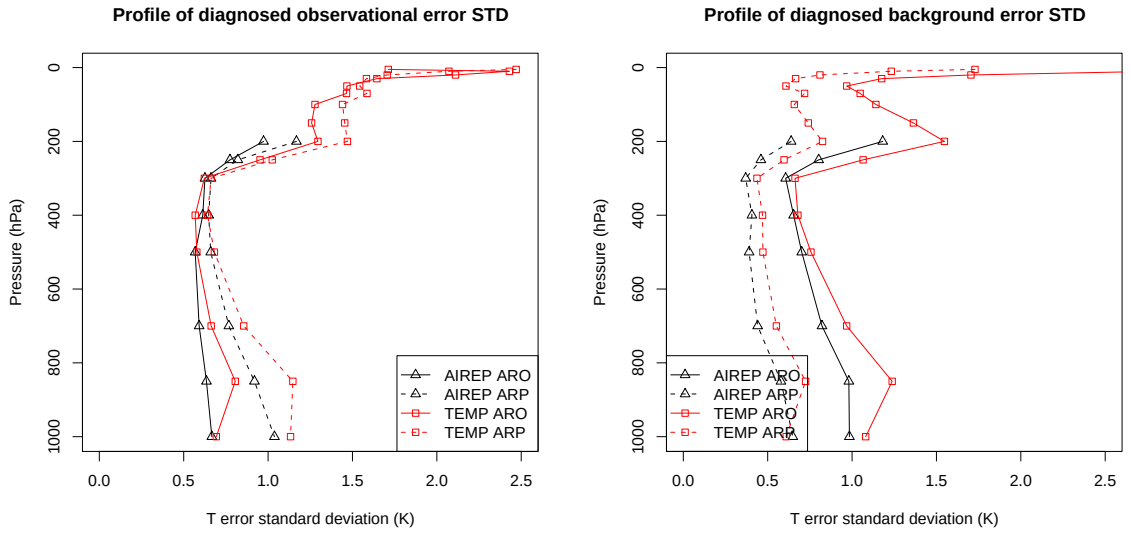

Figure 10: Comparison of observation and background errors, diagnosed for AROME and ARPEGE.

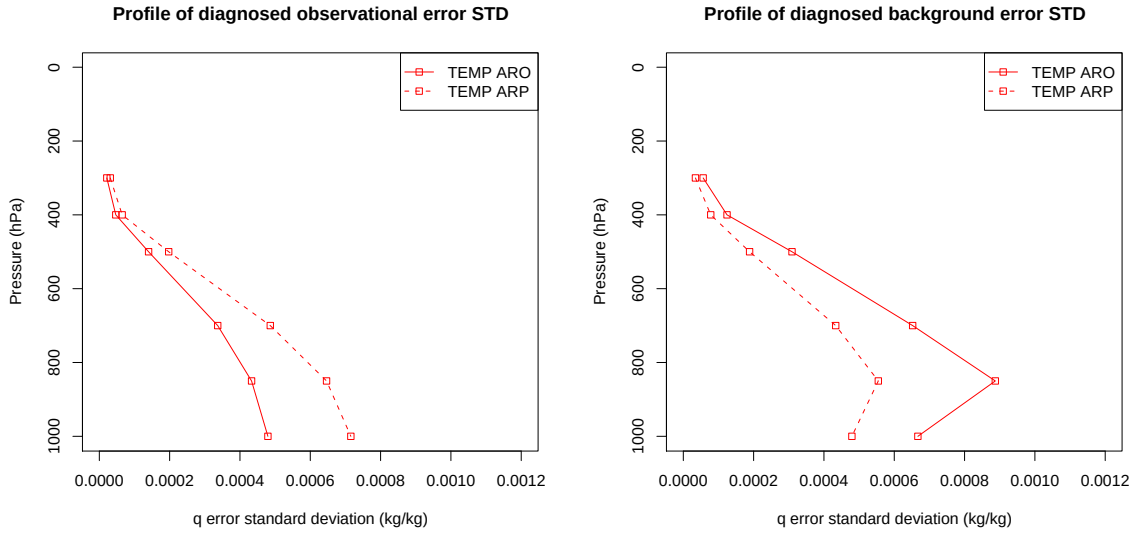

Figure 11: Comparison of observation and background errors, diagnosed for AROME and ARPEGE.

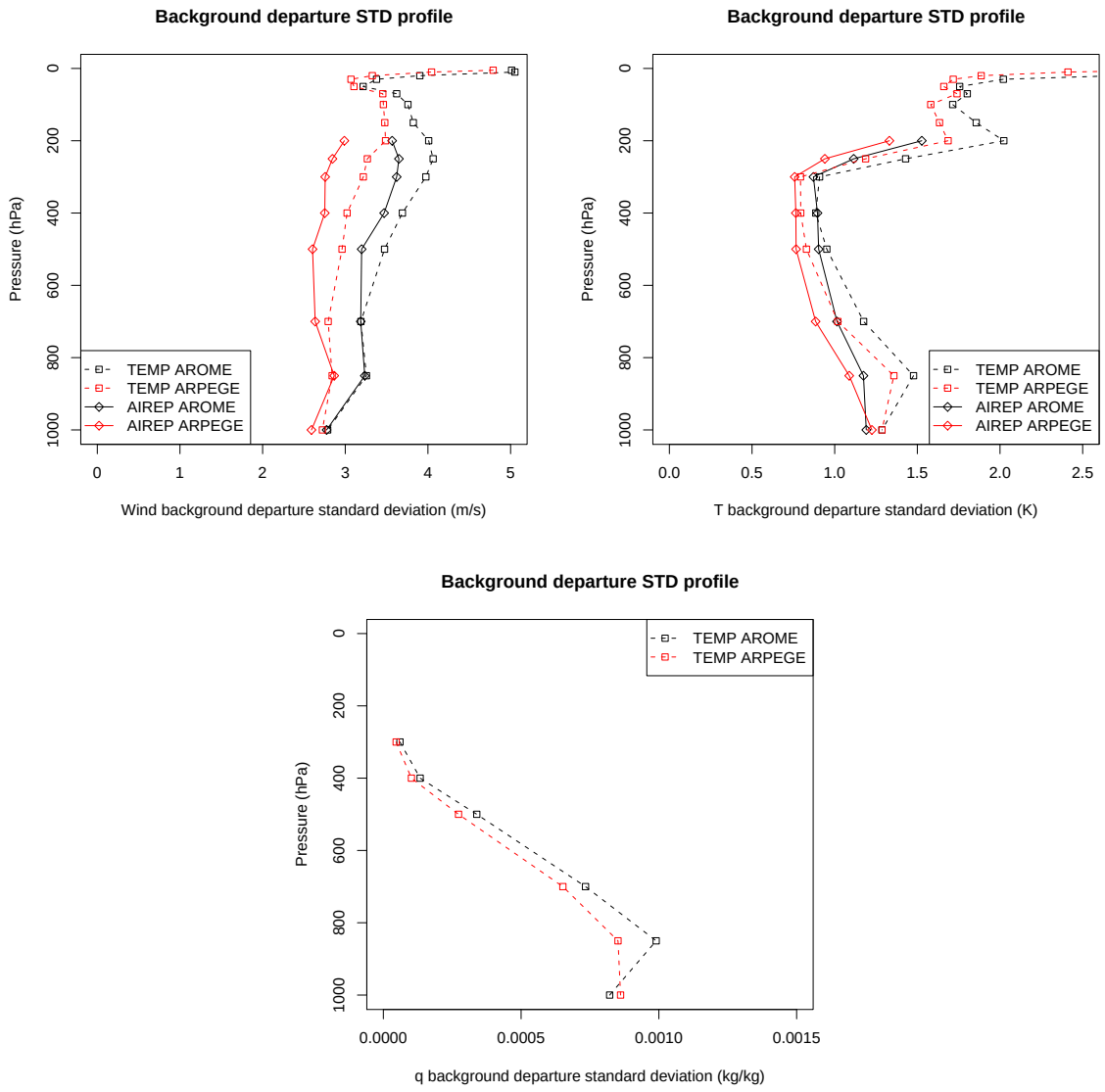

Figure 12: Comparison standard deviation of background departures, for AROME and ARPEGE.

in our reference experiments. So the new *REDNMC* becomes 0.84 and 1.3, for Jmin method and covariances of residuals, respectively.

- Observation error tuning is achieved by a factor called *SIGMAO_COEF*, which has to be applied at the times of ODB creation (*bator* program) and during observational screening. It was equal to 1 in the reference experiment and it is equal to S_o for both tuning options.

Two unperturbed assimilation cycles and 6-member ensembles of assimilation cycles (using perturbed observations) with applied tuning has then been considered. The impact of tunings has been estimated by running 24-hour forecasts twice a day (00 and 12 UTC). This production runs started from the unperturbed analyses. As a reference, a set of forecasts from the non-tuned and unperturbed assimilation cycle has been also run.

For the Jmin type of tuning, the scores are almost perfectly neutral, for all forecast ranges. In fact, by choosing similar values for S_o and S_b tuning coefficients, the relative fit of the analysis to observations cannot be modified. However, some decisions about accepting or rejecting observations could be changed since they are based on comparison of background departures with the weighted sum of observational and background error.

For background of residuals, the forecast scores against radiosounding and surface station data shows improved fit to observations in the analysis (for 0 h forecast range). The improvements, however decays to quickly in the forecast to be observed for other forecast ranges, except slight improvement of scores of wind forecasts, computed against surface stations (figure 13).

Another complementary approach to short range verification is to examine the analysis and background departures. The later can be seen also as a 3-hour forecasts. Figure 14 presents this kind of profiles, diagnosed by TEMP observations. For temperature and Jmin method, it can be seen that there is no visible difference in both analysis and background departures. This supports the idea, that the balance between background and observations is not modified. A small difference seems to appear in the very high levels (above 30 hPa), where more observations are used in the assimilation (top left panel of figure 14). The same same holds for covariances of residuals, too. The bottom panel of figure 14 presents a typical situation of tuning with covariances of residuals. The analysis fit has improved, but the background departures in the tuned experiment are again very similar to the reference. There is a slight change in the number of used observations - a consequence of modified screening criteria. The same holds also for other quantities (temperature, specific humidity).

6.2 Diagnostics of tuned analyses

After the tuning has been done, and evaluated in terms of forecast scores, it is worth to redo the diagnostics; to check that the tuning improves the agreement between specified and diagnosed errors. The degree of convergence can also be estimated with this approach. Figure 15 presents error profiles, diagnosed by this second iteration of covariances of residuals. It can be seen, that we generally obtain more fit to the prescribed statistics. However, some discrepancies still exist. For wind, the observational errors seem to be still overestimated (for around 20-30%) under 200 hPa. Background errors are now closed to prescribed ones. Their shapes are different, but further adjustment could be only achieved by vertically dependent tuning. The observational error is still overestimated for temperature, based on AIREP measurements, but is seems to be more or less well defined to TEMP observations. The correspondence of diagnosed background errors and specified ones is now a bit degraded, again due to uniform tuning. The agreement appears to be improved for specific humidity:

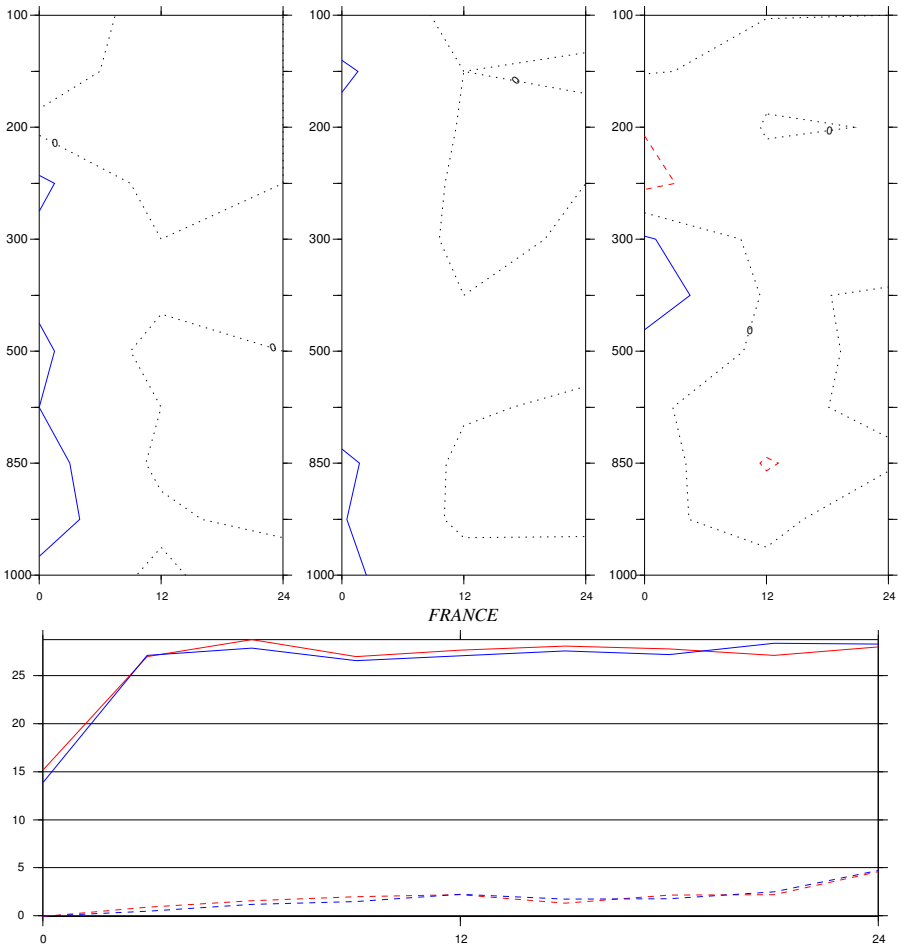

Figure 13: Verification scores of tuned assimilation cycle (using covariances of residuals). Upper panels: time-vertical cross sections of RMSE for 12 UTC temperature, 00 UTC wind and 12 UTC humidity forecasts with respect to radiosounde observations (TEMP), from left to right. Bottom panel: verification of surface winds against surface station data over France.

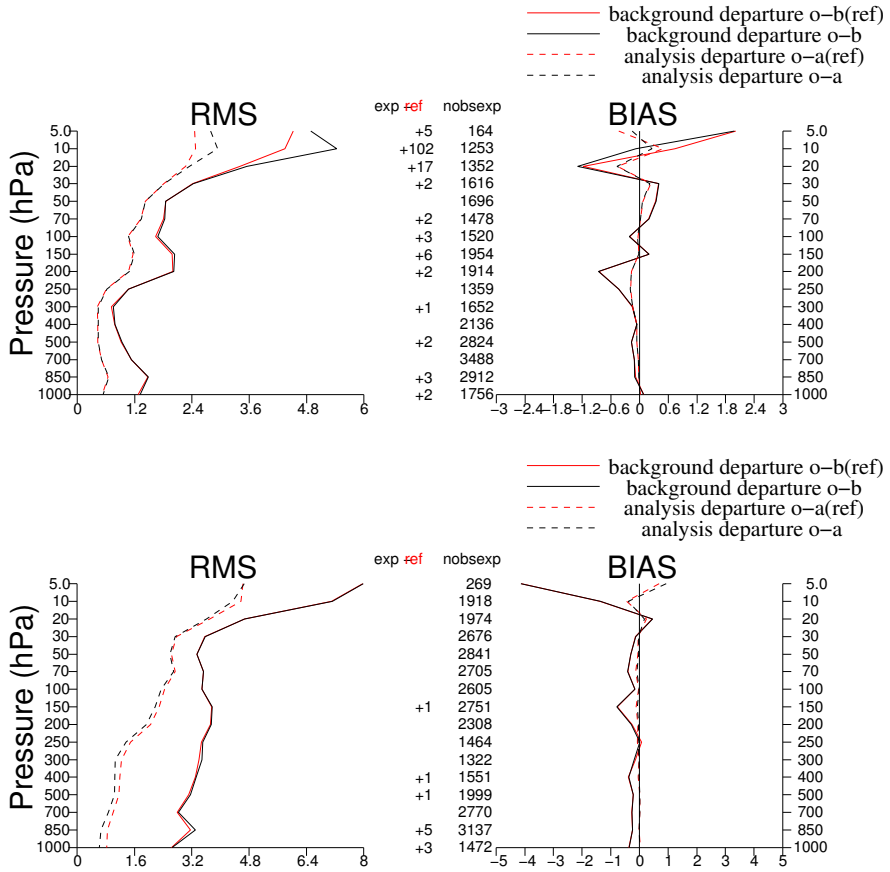

Figure 14: The diagnosed RMS and BIAS profiles of background and analysis departures, and number of observations, in the space of TEMP observations. Jmin tuning (top) for temperature and tuning of wind with respect to covariances of residuals (bottom), compared to untuned reference.

only slight overestimation of observational error remains below 800 hPa, and the diagnosed background error are now a bit larger and correspond better to prescribed errors.

The repeated diagnosis with Jmin method (figure 16) shows much improved S_o and S_b, with respect to their theoretical expectation of 1. Analysis to analysis variability is around 20%. The S_o estimates are, again, slightly larger than those of S_b. A daily cycle can be observed, before 25th February, which correspond to winter anticyclonic weather with low clouds and fog. This feature could be observed also in the original diagnostics.

7 Discussion and conclusions

We have presented a posteriori diagnosis and tuning of background and observational errors in the AROME data assimilation system. Two recognized and widely used a posteriori methods were used to provide two, slightly different estimates of error statistics.

First, it has to be taken into account that our **B** matrix covariances are estimated from a set of differences between perturbed ensemble members rather then the difference with respect to unperturbed reference. It means that their variance is twice the expected background error variances, and therefore, the expected σ_b equals

$$\sigma_b^{exp} = \frac{1}{\sqrt{2}}\sigma_b^{ens\_diff}, \tag{13}$$

$\sigma_b^{ens\_diff}$ being the background error standard deviation, computed from ensemble member differences, in the case where the model is perfect. Then, accumulated contribution of model error can be estimated as the remaining fraction of diagnosed background errors. The diagnosed background error variance, here denoted with σ_b^d, is then

$$(\sigma_b^d)^2 = (\sigma_b^{exp})^2 + (\sigma_m)^2 = 2REDNMC^2(\sigma_b^{exp})^2, \tag{14}$$

so that the accumulated model error fraction α becomes

$$\alpha = \sqrt{\frac{(\sigma_m)^2}{(\sigma_b^{exp})^2}} = \sqrt{2REDNMC^2 - 1}. \tag{15}$$

Parameter $REDNMC$ is the updated one (i.e. one equal to that used in the tuning experiments, with S_b already applied). The accumulated contribution of model error can be thus estimated in two possible ways (using two kinds of diagnostics):

- Jmin method: α=0.64
- covariances of residuals: α=1.54.

The diagnosed accumulated model error contribution is positive, and of similar magnitude than expected background error, provided by the perfect framework approach. There is a noticeable difference between two used methods; the diagnosed α for Jmin method approximately two and a half times smaller than that of covariances of residuals. It is not easy to conclude which result is more reliable. However, based on the experience in the simplified framework, and because of suspicious correlation of S_o and S_b for Jmin method, we tend to trust more in the covariances of residuals.

Of course, this kind of diagnostics cannot not tell anything about other model error characteristics like model error correlations. To conclude, the applied diagnostics and tuning seems to provide reasonable estimated values of the true statistics, which are more close to the optimality criteria (e.g. more consistent with observed background departures). It has been shown that the tuning does not significantly modify the quality of the subsequent forecasts.

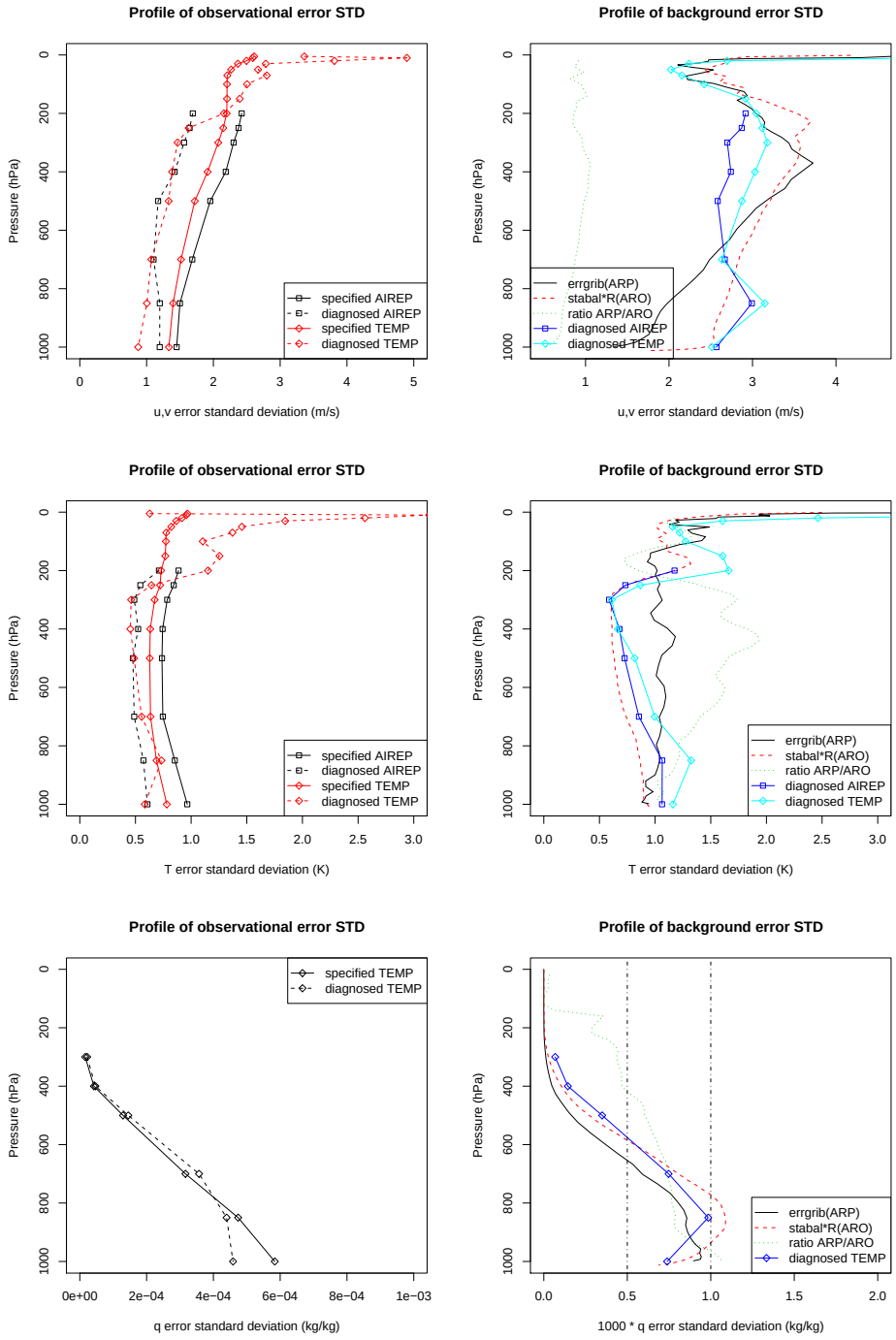

Figure 15: The diagnosed profiles of σ_o (left column) and σ_b (right column) for wind, temperature, and specific humidity in the tuned assimilation experiment. The tuning is with respect to covariances of residuals. R stands for multiplication by *REDNMC*.

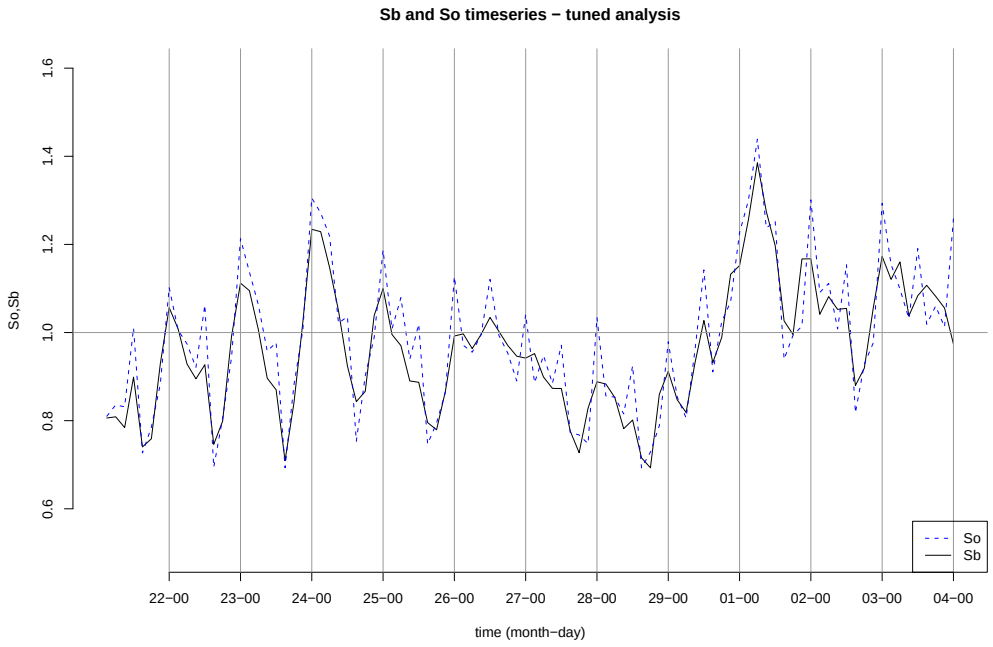

Figure 16: Temporal evolution of So and Sb (Jmin method), recomputed on the tuned analyses. The new mean values are So=0.95 and Sb=0.92.

References

G. Bölöni. Diagnosis of covariance residuals and tuning of background and observation errors in the ALADIN/HU 3DVAR. "`http://www.rclace.eu/File/Data_Assimilation/2006/`"„ 2006.

B. Chapnik, G. Desroziers, F. Rabier, and O. Talagrand. Properties and first application of an error statistics tuning method in variational assimilation. *Quart. J. Roy. Meteor. Soc.*, 130:2253–2275, 2004.

B. Chapnik, G. Desroziers, F. Rabier, and O. Talagrand. Diagnosis and tuning of observational error statistics in a quasi operational data assimilation setting. *Quart. J. Roy. Meteor. Soc.*, 132:543–565, 2006.

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

G. Desroziers, L. Berre, B. Chapnik, and P. Poli. Diagnosis of observation, background and analysis-error statistics in observation space. *Quart. J. Roy. Meteor. Soc.*, 131:3385–3396, 2005.

G. Desroziers, L. Berre, V. Chabot, and B. Chapnik. A posteriori diagnostics in an ensemble of perturbed analyses. *Monthly Weather Review*, 137(10):3420–3436, 2009.

W. Sadiki and C. Fischer. A posteriori validation applied to the 3d-var arpège and aladin data assimilation systems. *Tellus A*, 57:21–34, 2005.

O. Talagrand. A posteriori verification of analysis and assimilation algorithms. In *Proc. ECMWF Workshop on Diagnosis of Data Assimilation Systems*, page 17–28. ECMWF, 1999.

A Experiment list

Exp. ID	owner	description
63SV	mrpm613	6-member assimilation ensemble, untuned
64MY	mrpm613	unperturbed assimilation cycle, untuned
64P6	mrpa683	6-member assimilation ensemble, Jmin tuning
76A7	mrpa683	unperturbed assimilation cycle, Jmin tuning
76I9	mrpa683	6-member assimilation ensemble, tuning with covariances of residuals
76IE	mrpa683	unperturbed assimilation cycle, tuning with covariances of residuals
7DID	mrpa683	24-h production run, untuned reference
76GX	mrpa683	24-h production run, Jmin tuning
76J5	mrpa683	24-h production run, tuning with covariances of residuals

Table 1: List of used experiments in the OLIVE system.

B Code modifications

Some ALADIN and related code modifications have been applied to perform a posteriori diagnostics:

- application of *SIGMAO_COEFF* in *bator* and screening configuration has been backphased to cycle 33t1. Additionally, such a coefficient has been applied to radar Doppler winds (routines bator_ecritures_mod.F90 and bator_init_mod.F90). The pack can be found on yuki: /cnrm/gp/mrpa/mrpa683/pack/tune_sigmao/.

- a program for computation Tr(**HK**), written by Gérald Desroziers, has been modified to treat correctly the radar observations: a different vertical variable (*press_rl*) was used. Some corrections were applied that prevent the program to crash if all observation types are not available. Computation of normalized departures and averaging S_o and S_b over observation types were applied.

- The conversion of vorticity and divergence background errors to full wind errors has been coded in *fediacov*. Sources are available from yuki pack: /cnrm/gp/mrpa/mrpa683/pack/ALDODB_FEDIACOV_35T2_NAMEL_KE.