Impact of observation density in data assimilation: A study with simulated observations

By Zhiquan Liu [1] and Florence Rabier [2]

[1] *National Satellite Meteorological Center, Beijing, China (zqliu@nsmc.cma.gov.cn)*
[2] *Centre National de Recherches Météorologiques, Toulouse, France*

Abstract

The potential of high-density observations is studied in a practical context of the 4DVAR assimilation. A series of observing system simulation experiments (OSSEs) are carried out. Observations with both uncorrelated and correlated observation errors are simulated in sensitive areas. The results show that: for the observations with uncorrelated error, increasing the observation density generally improves the analysis and the forecast; for the observations whose error is correlated and by using a sub-optimal scheme (i.e., no modelling of the error correlation), the assimilation system can still extract useful information and one can determine an observation density leading to a minimum error of analysis and forecast. A risk of using horizontal high-density observations is that it could produce unrealistic increments and degrade the analysis on the levels without observations in the case of inappropriate background error correlations.

Introduction

In a 1D study, Liu and Rabier (2002) showed that the optimal density of observations in data assimilation depends mainly on the correlation of observation error. Figure 1 (Liu 2002, p13) shows the analysis error as a function of the observation interval. One can see the results for the 3 cases: (1) for the observations with uncorrelated error, increasing the observation density generally improves the analysis (black-mixed line); (2) for the observations with correlated error and by using an optimal scheme which models the error correlation of observation in the cost function, increasing the observation density beyond a certain threshold will bring little or no improvement of the analysis quality (red-solid line); (3) for the observations with correlated error but by using a sub-optimal scheme which assimilates the correlated data as if they were not, one can find an optimal interval of observations, which reaches a compromise between the risks to have a too low data density and to be affected by the correlated error of observations (blue-dashed line). In this study, one wants to check the results for the cases (1) and (3) in a practical context of the 4DVAR assimilation (the French model "ARPEGE"). As the modelling of the observation error correlation is not coded in ARPEGE, the case (2) is not examined. Two different weather situations poorly predicted by ARPEGE, are studied to examine the impact of various densities of observations on the analysis and the forecast. A series of OSSEs are carried out.
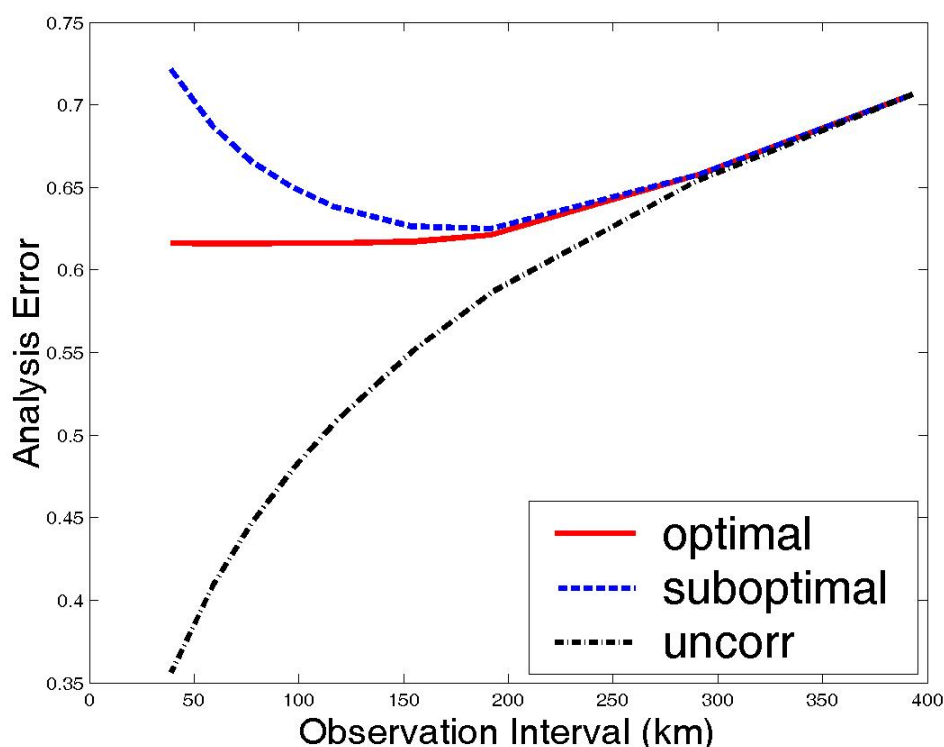
Fig. 1: Variation of the analysis error as a function of the observation interval. The analysis mesh △x is equal to 100km, the length of the background error correlation r_b=200km. The errors of the background and the observations have the same variance 1. The observation error correlation is of Gaussian type with a correlation length of 100km. The observations and their error covariances for each configuration are extracted from a network of observation with an interval of 20km. From Liu (2002, p13).

Assimilation and forecast system used for OSSEs

The assimilation and forecast system with which the observing-system simulation experiments (OSSEs) are performed is the French ARPEGE (Action de Recherche Petite Echelle et Grande Echelle) model. It is a global spectral model. A specific feature of the model is the use of a stretched grid in the horizontal direction to obtain an increased resolution over a geographical area of interest. The resolution used in the study is taken as the operational configuration in 2001. In the horizontal, it uses a triangular truncation T199. A terrain-following pressure-based hybrid vertical coordinate η with 31 levels is used and the top of the model is about 5hPa. A stretching coefficient 3.5 gives a spectral resolution varying from T696 over France to T57 over New Zealand. The assimilation component of the system includes a multi-incremental 4DVAR assimilation (Veersé and Thépaut, 1998) with a 6h assimilation window for the upper-air and surface pressure fields and an optimal interpolation (OI) analysis for other surface fields. The incremental analysis is performed on a regular unstretched grid in the current implementation. The minimization of the incremental cost function is performed with a successively increased resolution T42, T63 and T95. The configuration of resolution can be denoted in

short by T199C3.5L31/T42-63-95C1.0L31, where "T" represents the spectral truncation, "C" the stretching coefficient and "L" the vertical levels. That corresponds to a distance between the Gaussian mesh points varied from about 20km to 200km for the forecast and 200km for the analysis.

The first case: the Christmas storm in 1999

The first case examined is the storm having hit Europe and in particular France at 18UTC on 27th December 1999. This case is known because of the serious damage and the poor forecasts produced by most of the NWP centers. The trajectory of the 54h forecast started from the pre-operational 4DVAR analysis at 12UTC on 25th December is considered as the true in the OSSEs. This forecast has a low with a central pressure of 968hPa with perfect positioning. The operational forecast at that time (3DVAR) is considered as the background (the central pressure is equal to 975hPa and is poorly positioned). The simulated observations (temperature at all the 31 levels and surface pressure) are in the sensitive area (3066 Gaussian points) and at 48h before the storm. The assimilations of the simulated data with various samplings (0.3, 0.6 and 1.0 spherical degree) are carried out at 18UTC on 25th and followed by a 48h forecast. Two kinds of observation noise models are considered: uncorrelated and horizontally correlated. No vertical correlation is considered. Three models of correlation are considered: the correlation between two points separated by 0.6^o reaches 0.6, 0.3 and 0.15 respectively for strong, moderate and weak correlation models.
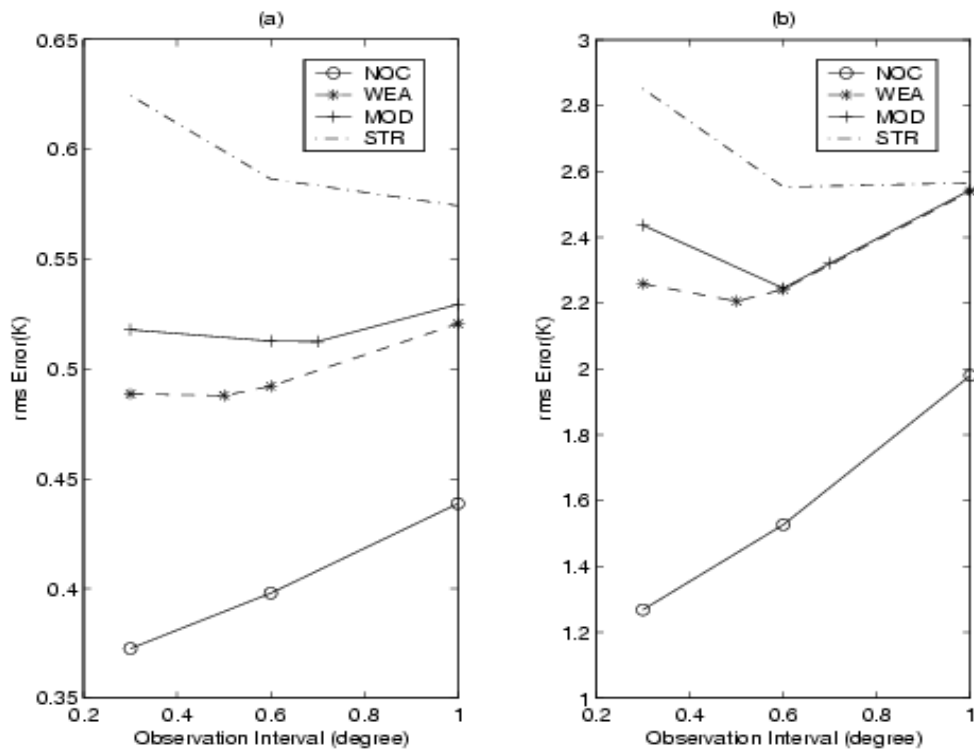

Figure 2: Level-averaged temperature r.m.s. error of (a) the analysis and (b) the 48h forecast for experiments with the weakly correlated (WEA), moderately correlated (MOD), strongly correlated (STR) and uncorrelated (NOC) observation error.

Figure 2 shows the RMS (root-mean-square) of error for the temperature analysis (in the sensitive area) and forecast at 48h (in the targeted area), averaged on all the model levels. The results are the average of 10 OSSEs with different realizations of random error. The assimilation scheme does not take into account the correlation of observation error. One can see that the errors of analysis and forecast increase regularly with the strenthening of the correlation of observation error. Moreover, the variation of the error with the observation interval shows a progressive transition from the uncorrelated case (denoted by "NOC") to the strongly correlated case (denoted by "STR"). For example, the error decreases in a monotonous way with the decrease of the observation interval for the uncorrelated case and increases for the strongly correlated case. A minimum error is located at an intermediate interval in the case of weak correlation (denoted by "WEA") and moderate correlation (denoted by "MOD"). This intermediate interval producing the minimum error of analysis is 0.5° for the case "WEA" and is 0.7° for the case "MOD". For these two optimal intervals, the adjacent observations have the same correlation equal to approximately 0.2 which is close to the value 0.15 found in the 1D study (Liu and Rabier, 2002).

It is noted that the errors of the 48h forecast can be approximately 4-5 times larger than the initial errors of analysis. The difference of the forecast error between various intervals of observation can be of an order of magnitude more significant than that of the analysis error. For example, the reduction of 0.04K of the temperature analysis error for experiment NOC0.6 compared to experiment NOC1.0 leads to a reduction of 0.45K of the forecast error.

Figure 3: Mean sea level pressure (hPa) of the 48h forecast for 6 experiments: (a) NOC0.3; (b) NOC0.6;   (c) NOC1.0; (d) MOD0.3; (e) MOD0.6; (f) MOD1.0.
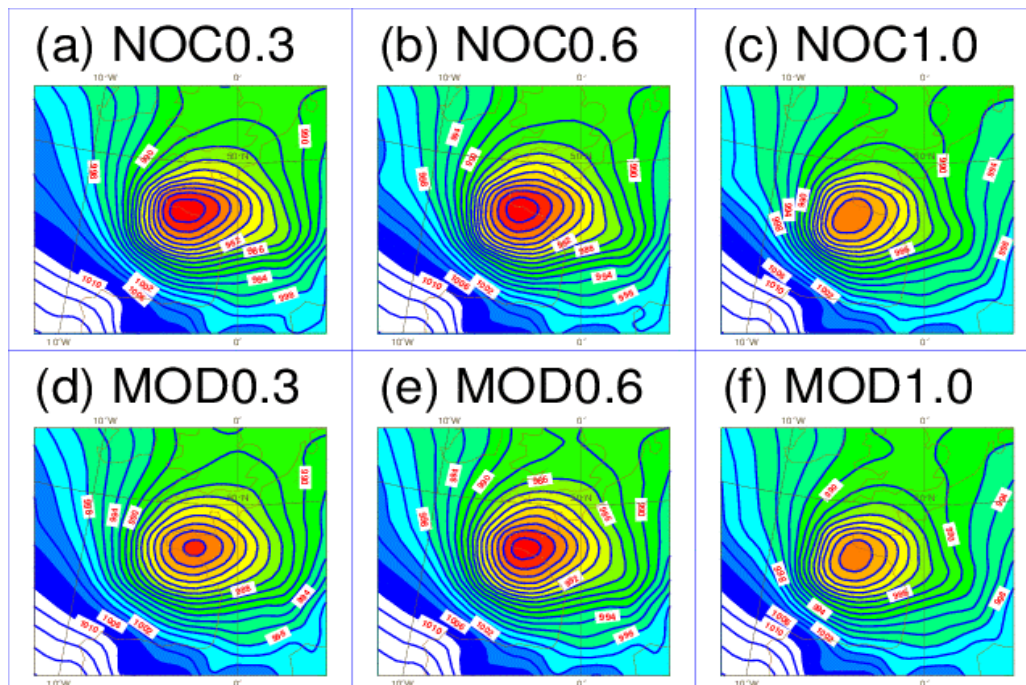
Figure 3 shows the pressure reduced to the sea level predicted for the 6 experiments. They are consistent with the results in Figure 2. One can see that the three worse forecasts are those of experiment MOD1.0 (3f), NOC1.0 (3c) and MOD0.3 (3d). Their central pressures are respectively 974.6hPa, 974.4hPa and 973.4hPa.

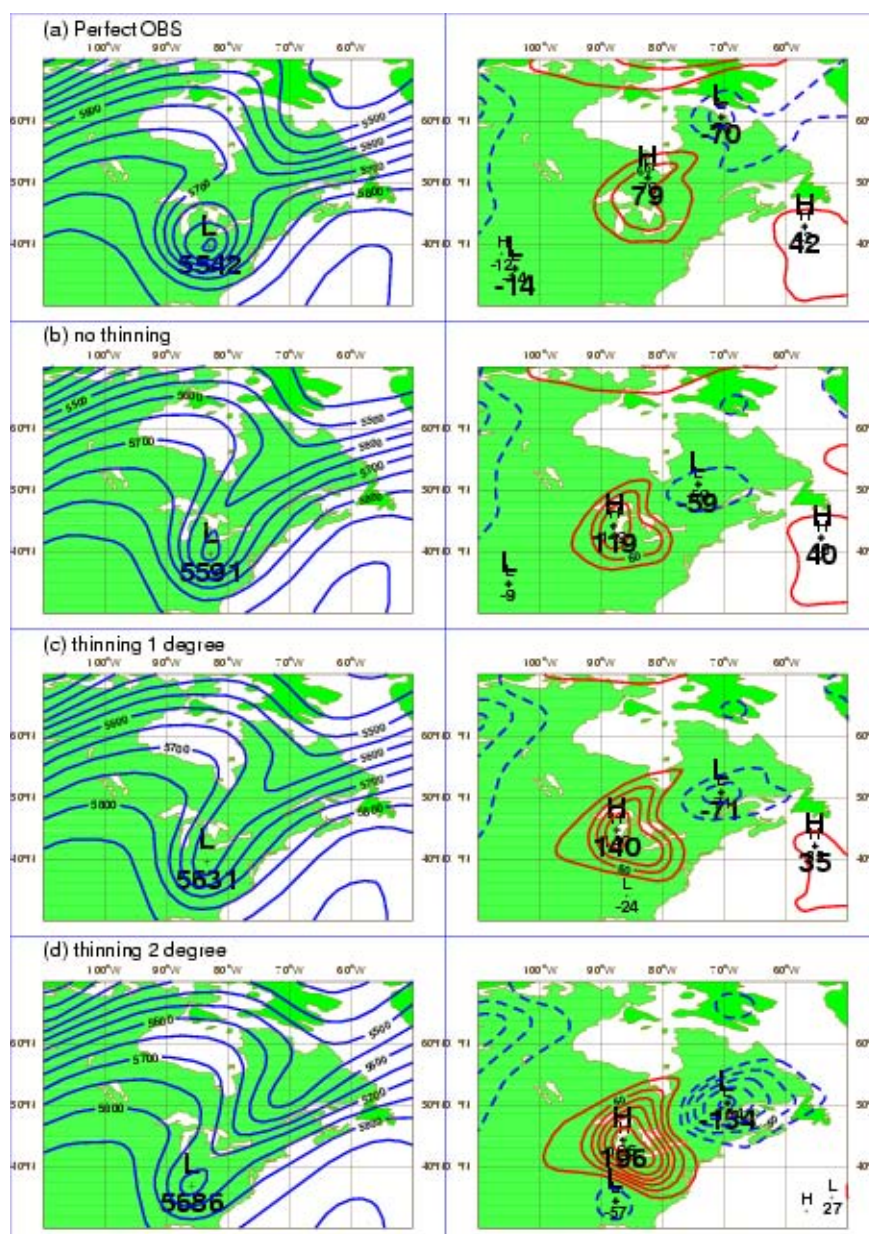## The synoptic case at   00UTC on 25th September 2001

Figure 4: 96h forecast (left panel) and forecast error (right panel) of 500hPa geopotential height (in m) valid at 00UTC on 25th September 2001 for experiments (a). Perfect observations without thinning; (b). Noisy observations without thinning; (c). Noisy observations with a thinning of 1°; (d). Noisy observations with a thinning of 2°.

The 96h forecast of the ARPEGE from the analysis at 00UTC on 21st September 2001 has an

enormous error (the order of 300m for 500hPa Z) compared to the verified analysis. On the other hand, ECMWF made an almost perfect forecast (thus is used as the true in OSSEs like the first case). The simulated observations with uncorrelated error are radiances of AMSU-A channels 5~12 in the north of North-America at 00UTC on 21st September. Figure 4 shows the 96h forecast of 500hPa Z for 4 experiments. For the 2 cases without thinning, the maximum error is reduced from 300m to 79m and to 119m respectively for the perfect and noisy observations. The forecast is degraded when the number of observations decreases.
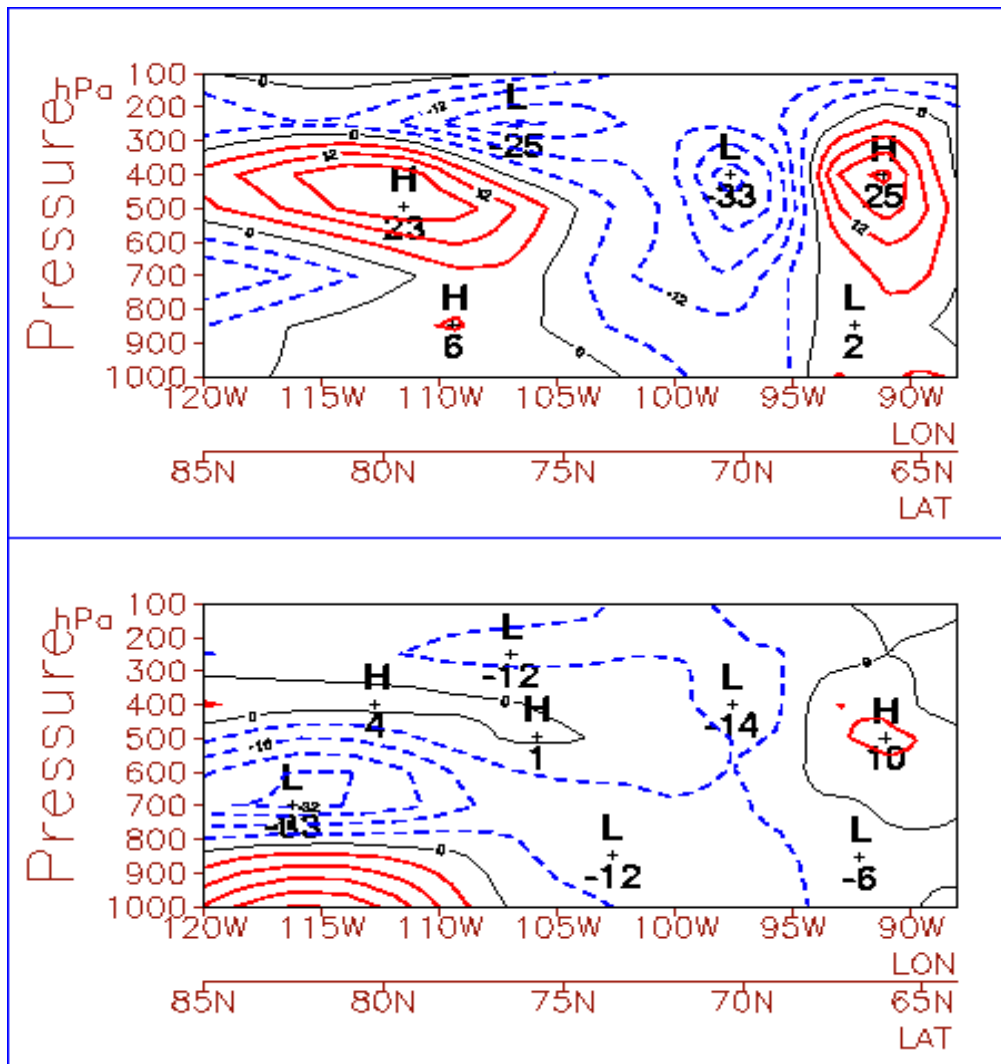

Figure 5: Cross-section of geopotential height (in m). Top panel: for background error; bottom panel: for analysis error with perfect observations.

However, the error on other levels is not always consistent with 500hPa. That can be illustrated by the cross-section in the Figure 5. One can see that the error between 300hPa to 500hPa is considerably reduced but the error at the lower levels is reinforced. We think that this problem comes mainly from the weakness of modelling the background error covariance matrix, which in operational practice is often constant and can propagate the increment in an incorrect way from a place to another

place or from a variable to another variable in certain situations. That represents a risk: employing observations with horizontal high density could accentuate unrealistic increments and degrade the analysis on the levels without observations, although this degradation does not necessarily influence the forecast in the interesting area as shown the Figure 4.

Conclusions and Discussions

This study confirms the results in the 1D study. A correlation of the order 0.2 is suggested for the determination of optimal thinning in the data assimilation. The data with high-density in the sensitive areas are important, particularly for some "tricky" conditions. Inconsistency between horizontal and vertical resolutions introduces certain risks because of imperfections in the background error covariance matrix. The specification of the statistics of observation errors and in particular their correlation, is necessary in order to model them and also to find the optimal thinning of observations. This subject of study constitutes an important work. Some additional results with respect to the statistical diagnostic and tuning of the observation error correlation are not presented here. The interested readers are referred to the recent paper of Liu and Rabier (2003).
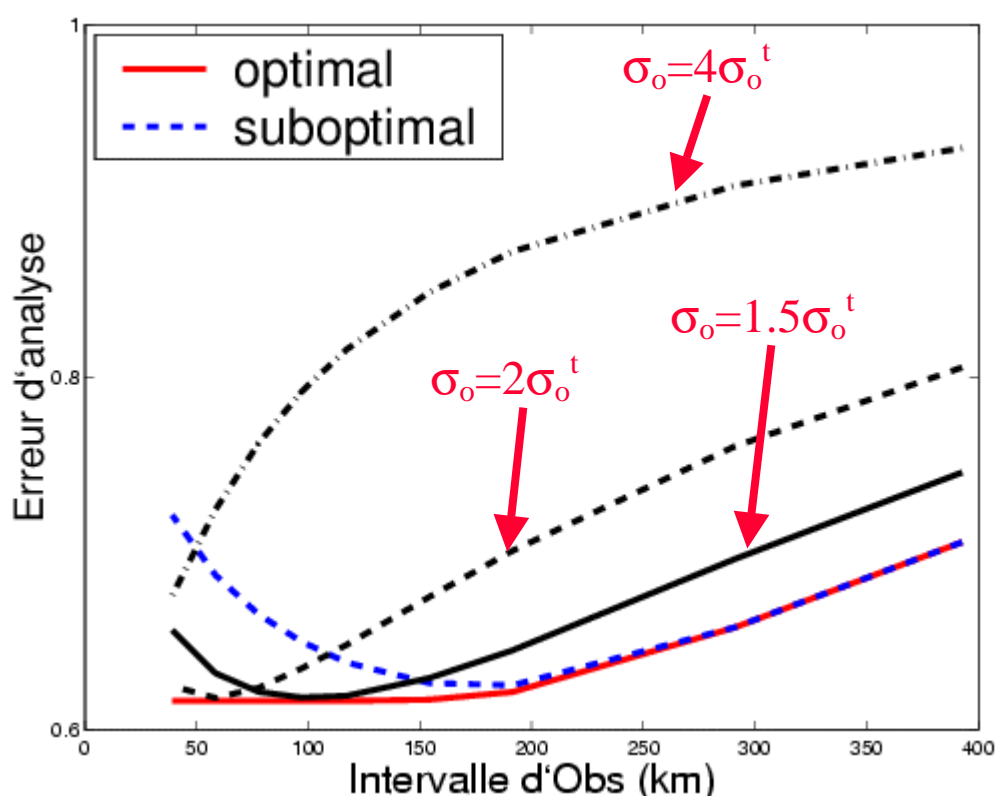

Figure 6: Impact of inflating the observation error. Two color curves are the same as those in the Figure 1, three black curves are the results with different inflating factors (1.5, 2 and 4) for the sub-optimal scheme.

In practice, we have other options to compensate the impact of the observation error correlation in a sub-optimal scheme. For example, inflating the observation error standard deviation σ_o specified

in the assimilation system (i.e., one gives less confidence to these correlated data). This has been applied to the practice but in an empirical way. One can expect that the optimal weight of observation is related to the correlation of observation error. The bigger the error correlation is, the more one should inflate σ_o. The figure 6 illustrates this in the 1D context as shown the figure 1. Two curves in color are the same as those in the figure 1. Three black curves are the results with different inflating factor (1.5, 2 and 4 times) for the sub-optimal scheme. One can sees that the optimal interval decreases when σ_o is inflated (respectively 100km for 1.5 times and 60km for 2 times). We find a correlation of 0.5 and 0.85 for the interval 100km and 60km. Moreover, the analysis error is exactly the same as that for the optimal scheme (red curve). That shows that inflating σ_o is a simple and efficient way to reduce deficiency resulted from the sub-optimality of the assimilation scheme. As expected, the analysis would be worse than that of the optimal scheme if one inflates too much σ_o (that is the case with 4 times inflating). The result in figure 6 implies that the optimal thinning interval depends not only the error correlation value itself but also the observation error standard deviation specified in the assimilation system.

References

Liu, Z.-Q., 2002. Influence of the observation resolution in data assimilation. Thesis of L'Université de Paul Sabatier, Toulouse, France, pp123, available from the author.

Liu, Z.-Q. and Rabier, F., 2002. The interaction between model resolution, observation resolution and observation density in data assimilation: A one-dimensional study. Q. J. R. Meteorol. Soc., 128, 1367-1386.

Liu, Z.-Q. and Rabier, F., 2003. The potential of high-density observations on Numerical Weather Prediction: A study with simulated observations. Q. J. R. Meteorol. Soc., 129, 3013-3035.

Veersé, F. and Thépaut, J.-N., 1998. Multiple-truncation incremental approach for four-dimensional variational data assimilation. Q. J. R. Meteorol. Soc., 124, 1889-1908

Proceedings
of the
Thirteenth International
TOVS Study Conference

INTERNATIONAL
ATOVS
WORKING GROUP
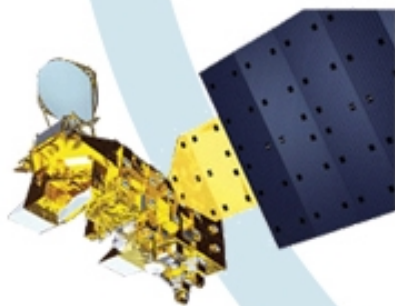
Sainte-Adèle, Québec, Canada
29 October – 4 November 2003