Theoretical foundation of error covariance estimation methods based on analysis residuals with an example of realistic surface observation network

Richard Ménard and Martin Deshaies-Jacques
Richard.Menard@canada.ca

Abstract

We examine the theoretical foundation of the method to estimate error covariances based on analysis residuals in observation space, also known as the Desroziers method. Our analysis also includes a combined estimation of the correlation length using the maximum likelihood method. We first review the analysis of convergence that was carried out with a simplified regular observation network in Ménard (2015). We then apply the method to surface observation network of air quality observations. The challenge here is to estimate spatially varying and non-stationary error statistics with a relatively small sample size. When using the statistics strictly at the station we observe that the scheme is usually non-convergent in the length-scale estimates. But by performing a local average of the analysis residuals statistics in the Desroziers' iterative scheme we resolve this issue and obtain a locally averaged error variance estimates that agrees with other estimation methods.

1 Theory

Theorem on error covariance estimates in obs space

Assuming that the observation and background errors are uncorrelated, the *necessary* and *sufficient* condition for error covariance estimates to be equal to the true observation and background error covariance in observation space is

$$1. \mathbf{H}\widetilde{\mathbf{K}} = \mathbf{H}\mathbf{K} \quad \text{the gain is equal to the Kalman gain in obs space} \tag{1}$$

$$2. \widetilde{\mathbf{D}} \overset{\triangle}{=} \mathbf{H}\widetilde{\mathbf{B}}\mathbf{H}^T + \widetilde{\mathbf{R}} = \mathbf{D} \quad \text{the } \textit{innovation covariance consistency} \tag{2}$$

where $\mathbf{K}$ is the *Kalman gain* matrix, $\mathbf{R}$ and $\mathbf{B}$ are the observation and background error covariances, and $\mathbf{D}$ is the *innovation matrix*. The proof is given in Ménard 2015. We stress the fact that having the gain equal to the Kalman gain in observation space or an optimal analysis is not a sufficient condition for the estimates to be equal to the true covariances, one need to have also $\widetilde{\mathbf{D}} = \mathbf{D}$.

Desroziers' scheme

Desroziers *et al.* (2005) showed how analysis residuals in observation space, specifically the *observation-minus-analysis* residuals $(O - A)$ and the *analysis-minus-background* residuals $(A - B)$, could be used to obtain estimates of $\widetilde{\mathbf{R}}$ and $\mathbf{H}\widetilde{\mathbf{B}}\mathbf{H}^T$ with an iterative scheme

$$\widetilde{\mathbf{R}}_{i+1} = \langle (O - A_i)(O - B)^T \rangle, \tag{3}$$

$$\mathbf{H}\widetilde{\mathbf{B}}_{i+1}\mathbf{H}^T = \langle (A_i - B)(O - B)^T \rangle, \tag{4}$$

Properties and consequences

Property 1 One step of the *full* error covariance estimation scheme (3-4) produces new estimates that meet the *innovation consistency* condition (2).

Property 2 The estimation of both full covariances $\widetilde{\mathbf{R}}$ and $\widetilde{\mathbf{B}}$ converges in a single iteration. Doing further iterations of the full covariances will not change the estimates.

Propery 3 The estimated observation and background error covariance in observation space are equal to the truth $\widetilde{\mathbf{R}} = \mathbf{R}$ and $\mathbf{H}\widetilde{\mathbf{B}}\mathbf{H}^T = \mathbf{H}\mathbf{B}\mathbf{H}^T$ *if and only if* the initial gain is the Kalman gain.

Consequence The estimation of the both full error covariance $\mathbf{R}$ and $\mathbf{H}\mathbf{B}\mathbf{H}^T$ is ill-posed with the Desroziers scheme. Feasible estimation is possible only if *not all* the elements (or simply parameters) of error covariances are estimated.

Property 4 The Desroziers' estimation of the error *variances* is well defined and gives an innovation *variance* consistency in a single step.

Remark 1 The iterative variance scheme will converge to the true error variances only if the other elements that makes up the Kalman gain, in particular, the error correlations are correctly specified.

Consequence A viable option to the estimation of error statistics is to simultaneously estimate the error variance and the correlation lengths.

2 Maximum likelihood estimation

Maximum likelihood method is used to estimate the correlation length. The error variances are provided by the Desrpziers method.

In the maximum likelihood method, we assume that the conditional p.d.f. of the innovations for a given parameter value (L_c) (the correlation length) is Gaussian-distributed,

It is usual to define the log-likelihood function $\mathcal{L}(L_c|\mathbf{d}_b^o)$ as $p(\mathbf{d}_b^o|L_c) = \exp[-\mathcal{L}(L_c|\mathbf{d}_b^o)]$, so that the maximum likelihood is equivalent to the minimum of the log-likelihood function,

$$\mathcal{L}(L_c|\mathbf{d}_b^o) \quad \propto \quad \log[\det(\widetilde{\mathbf{D}}(L_c))] + tr(\widetilde{\mathbf{D}}^{-1}(L_c)\mathbf{D}). \tag{5}$$

Remark When both variances and correlation length are iteratively estimated, the algorithm may not converge when the correlation model or other statistical assumptions are inadequate.

3 Application to surface air quality observations

We now apply the Desroziers method of estimation of error variances with the maximum likelihood estimation of correlation length to the AirNow air quality measurement network over North America. Innovations are obtained with respect to the ECCC operational air quality model, GEM-MACH. There is over 1,300 O_3 (ozone) and nearly 750 $PM_{2.5}$ (fine mode aerosols) measurements reported in real time each hour.

The constraints are:

- *Significant spatial inhomogeneity*
- *Strong diurnal cycles for photochemical species*

Thus only a small number of measurements per site per hour is available (typically 60).

While Desroziers method is generally used to estimate a single variance scaling factor, the challenge here is to estimate non-homogeneous time-varying variances with relatively small samples.

For each hour at each station, statistics are built from a 60 day (two months) sample. Using raw statistics at stations results in a lack of convergence of both Desroziers and maximum likelihood estimates. By using instead locally averaged statistics of (3) and (4), new estimates are obtained that now converge.

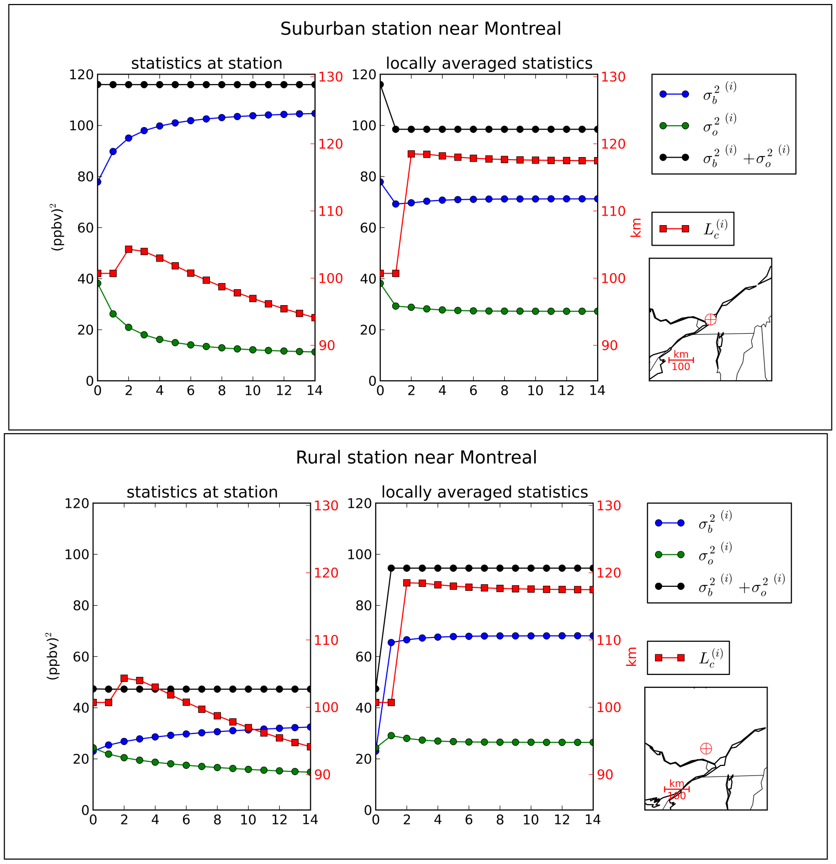
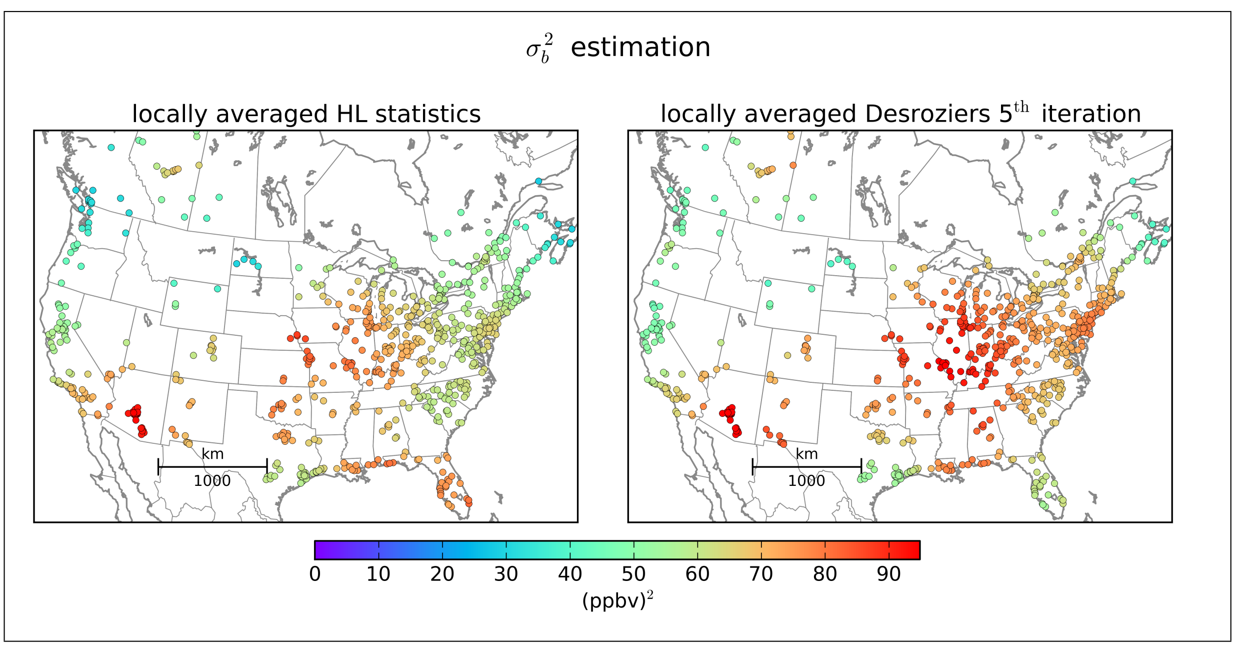


Figure 1: Desroziers iterates using statistics at stations (left). Locally averaged with an exponentially decaying fuction (right).

The resulting variance estimates are compared with a different approach based on a locally averaged Hollingsworth and Lonnberg method (Ménard *et al.* 2016).



4 Conclusion

We have applied the Desroziers method to an air quality surface measurements network to obtain non-homogeneous and time-varying estimates with a relatively small sample size. We found that there when we apply a locally average to the equations (3) and (4) we get convergent estimates with estimated values that agrees with other estimation methods.

5 Reference

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

Ménard R., 2015: Error covariance estimation methods based on analysis residuals: Theoretical foundation and convergence properties derived from simplified observation networks, *Q. J. R. Meteorol. Soc.*, **142**:257–273. DOI:10.1002/qj.2650

Ménard R., M. Deshaies-Jacques, and N. Gasset, 2016: A comparison of correlation-length estimation methods for the objective analysis of surface pollutants at Environment and Climate Change Canada, *Journal of the Air and Waste Management Association*, DOI:10.1080/10962247.2016.1177620
(available from http://dx.doi.org/10.1080/10962247.2016.1177620)