

Type of the Paper (Article)

Evaluation of Analysis by Cross-Validation.

Part II: Diagnostic and Optimization of Analysis Error Covariance

Richard Ménard ^{1*} and Martin Deshaies-Jacques ¹

¹ Air Quality Research Division, Environment and Climate Change Canada; martin.deshaies-jacques@canada.ca

* Correspondence: richard.menard@canada.ca; Tel.: +1-514-421-4613,
2121 Transcanada Highway, Dorval, (QC), CANADA, H9P 1J3

Abstract: We present a general theory of estimation of analysis error covariances based on cross-validation as well as a geometric interpretation of the method. In particular we use the variance of *passive observation-minus-analysis residuals* and show that the true analysis error variance can be estimated, without relying on the optimality assumption. This approach is used to obtain near optimal analyses that are then used to evaluate the air quality analysis error using several different methods at active and passive observation sites. We compare the estimates according to the method of Hollingsworth-Lönnberg, Desroziers *et al.*, a new diagnostic we developed, and the perceived analysis error computed from the analysis scheme, to conclude that, as long as the analysis is near optimal, all estimates agree within a certain error margin.

Keywords: data assimilation; statistical diagnostics of analysis residuals; estimation of analysis error, air quality model diagnostics; Desroziers *et al.* method; cross-validation

1. Introduction

At Environment and Climate Change Canada (ECCC) we have been producing hourly surface pollutants analyses covering North America [1, 2, 3] using an optimum interpolation scheme which combines the operational air quality forecast model GEM-MACH output [4] with real-time hourly observations of O₃, PM_{2.5}, PM₁₀, NO₂, and SO₂ from the AirNow gateway with additional observations from Canada. These analyses are not used to initialize the air quality model and we wish to evaluate them by cross-validation, that is by leaving out a subset of observations from the analysis to use them for verification. Observations used to produce the analysis are called *active observations* while those used for verification are called *passive observations*.

In a first part paper of this study, i.e. Ménard and Deshaies-Jacques [5], we have examined different verification metrics using either active or passive observations. As we changed the ratio of observation error to background error variances $\gamma = \sigma_o^2 / \sigma_b^2$, while keeping the sum $\sigma_o^2 + \sigma_b^2$ equal to $\text{var}(O - B)$, we found a minimum in $\text{var}(O - A)$ in the passive observation space. In this second part paper, we formalize this result, develop the principles of estimation of the analysis error covariance by cross-validation, and apply it to estimate and optimize the analysis error covariance the our surface analyses of O₃ and PM_{2.5}.

The evaluation of the analysis error covariance using its own active observations is a misleading problem unless the analysis is already optimal. Hollingsworth and Lönnberg [6] addressed this issue for the first time where they noted that in the case of an optimal gain (i.e. optimal analysis), the statistics of observation-minus-analysis residuals $O - \hat{A}$ are related to the analysis error by $E[(O - \hat{A})(O - \hat{A})^T] = \mathbf{R} - \mathbf{H}\hat{\mathbf{A}}\mathbf{H}^T$, where $\hat{\mathbf{A}}$ is the optimal analysis error covariance and \mathbf{H} and \mathbf{R} are

the observation operator and observation error covariance respectively. The caret (^) over \mathbf{A} indicates that the analysis uses an optimal gain. In the context of spatially uncorrelated observation errors, the off-diagonal elements of $\mathbf{E}[(\mathbf{O} - \hat{\mathbf{A}})(\mathbf{O} - \hat{\mathbf{A}})^T]$ would then give the analysis error covariance in observation space. Hollingsworth and Lönnberg [6] argued that for most practical purposes, the negative intercept of $\mathbf{E}[(\mathbf{O} - \hat{\mathbf{A}})(\mathbf{O} - \hat{\mathbf{A}})^T]$ at zero distance and the prescribed observation weight should be nearly equal, and thus could be used as an assessment of optimality of an analysis. However, in case where such agreement does not exist, an estimate of the actual analysis error is not possible. Another method, proposed by Desroziers *et al.* [7], argued that the diagnostic $\mathbf{E}[(\mathbf{O} - \hat{\mathbf{A}})(\hat{\mathbf{A}} - \mathbf{B})^T]$ should be equal to the analysis error covariance in observation space but, again only if the gain is optimal and the innovation covariance consistency is respected [8].

Generally, a robust approach that does not require an optimal analysis is to use observations whose errors are uncorrelated with the analysis error. With observations that have errors that are temporarily (serially) uncorrelated, an estimation of the analysis error can be made with the help of a forecast model initialized by the analysis by verifying the forecast against these observations. This is the essential assumption used traditionally in meteorological data assimilation to assess indirectly the analysis error by comparing the resulting forecast with observations valid at the forecast time. As forecast error grows with time, the observation-minus-forecast can be used to assess whether an analysis is better than another. In a somewhat different method but making the same assumption, Daley [9] used the temporal (serial) correlation of the innovations to diagnose the optimality of the gain matrix. This property was first established in the context of Kalman filter estimation theory by Kailath [10]. However, both the traditional meteorological forecast approach and the Daley method [9] are subject to limitations: they assume that the model forecast has no bias and the analysis corrections are made correctly on all the variables needed to initialize the model. In practice, improper initialization of unobserved meteorological variables gives rise to spin-up problems or imbalances. Furthermore with the traditional meteorological approach, compensation due to model error can occur, so that an optimal analysis does not necessarily yield an optimal forecast [11].

An alternative approach introduced by Marseille *et al.* [12], which we will follow here, is to use passive observations to assess the analysis error. The essential assumption of this method is that the observations have spatially uncorrelated errors, so that the observations used for verification, i.e. the passive observations, have uncorrelated errors with the analysis. The advantage of this approach is that it doesn't involve any model to propagate the analysis information to a later time. Marseille *et al.* [12] then showed that by multiplying the Kalman gain with an appropriate scalar value, one can reduce the analysis error. In this paper, we go further by using principles of error covariance estimation to obtain a near optimal Kalman gain. In addition we impose the innovation covariance consistency [8] and show that all diagnostics of analysis error variance nearly agree with one another. These include the Hollingsworth and Lönnberg [6], the Desroziers *et al.* [7] and new diagnostics that we will introduce.

The paper is organized as follows. First we present the theory and diagnostics of analysis error covariance in both passive and active observation spaces, as well as a geometrical representation. In §3, we present the experimental setup on how we obtain near optimal analyses and presents the results of several diagnostics in active and passive observation spaces, and compare with the analysis error variance obtained from the optimum interpolation scheme itself. In §4, we discuss the statistical assumptions being used, how and if they can be extended and how this formalism can be used in other applications such as the estimation of correlated observation errors with satellite observations. Finally we draw some conclusions in §5.

2. Theoretical framework

2.1. Diagnostic of analysis error covariance in passive observation space

Let us decompose the observation space in two disjoint sets; the active observation set or training set $\{y\}$ used to create the analysis, and the independent or passive observation set $\{y_c\}$ used to evaluate the analysis. An analysis built from prescribed background and observation error covariances, $\tilde{\mathbf{B}}$ and $\tilde{\mathbf{R}}$ respectively, is given by

$$\mathbf{x}^a = \mathbf{x}^f + \tilde{\mathbf{B}}\mathbf{H}^T(\mathbf{H}\tilde{\mathbf{B}}\mathbf{H}^T + \tilde{\mathbf{R}})^{-1}\mathbf{d} = \mathbf{x}^f + \tilde{\mathbf{K}}\mathbf{d}, \quad (1)$$

where $\tilde{\mathbf{K}}$ is the gain matrix built from the prescribed error covariances, \mathbf{H} is the observation operator for the active observation set, \mathbf{d} is the active innovation vector, $\mathbf{d} = \mathbf{y} - \mathbf{H}\mathbf{x}^f = \boldsymbol{\varepsilon}^o - \mathbf{H}\boldsymbol{\varepsilon}^f$, \mathbf{x}^f is the background state or model forecast, $\boldsymbol{\varepsilon}^o$ is the active observation error and $\boldsymbol{\varepsilon}^f$ the background error. The observation-minus-analysis residual $(O-A)$ for the active set is given by [7,8,13]

$$\begin{aligned} (O-A) &= \mathbf{y} - \mathbf{H}\mathbf{x}^a = \boldsymbol{\varepsilon}^o - \mathbf{H}\boldsymbol{\varepsilon}^a \\ &= \mathbf{d} - \tilde{\mathbf{H}}\tilde{\mathbf{B}}\mathbf{H}^T(\mathbf{H}\tilde{\mathbf{B}}\mathbf{H}^T + \tilde{\mathbf{R}})^{-1}\mathbf{d} \\ &= \tilde{\mathbf{R}}(\mathbf{H}\tilde{\mathbf{B}}\mathbf{H}^T + \tilde{\mathbf{R}})^{-1}\mathbf{d}, \end{aligned} \quad (2)$$

where $\boldsymbol{\varepsilon}^a$ is the analysis error. The analysis interpolated at the passive observation sites can be denoted by $\mathbf{H}_c\mathbf{x}^a$, where \mathbf{H}_c is the observation operator at passive observation sites. The observation-minus-analysis residual at the passive observation sites $(O-A)_c$ is then given by

$$\begin{aligned} (O-A)_c &= \mathbf{y}_c - \mathbf{H}_c\mathbf{x}^a = \boldsymbol{\varepsilon}_c^o - \mathbf{H}_c\boldsymbol{\varepsilon}^a \\ &= \mathbf{d}_c - \mathbf{H}_c\tilde{\mathbf{B}}\mathbf{H}^T(\mathbf{H}\tilde{\mathbf{B}}\mathbf{H}^T + \tilde{\mathbf{R}})^{-1}\mathbf{d}, \end{aligned} \quad (3)$$

where $\mathbf{d}_c = \mathbf{y}_c - \mathbf{H}_c\mathbf{x}^f = \boldsymbol{\varepsilon}_c^o - \mathbf{H}_c\boldsymbol{\varepsilon}^f$ is the innovation at the passive observation sites. Note that the formalism introduced here is general, and can be used with any set of independent observations such as different instruments or observation networks as long as the \mathbf{H}_c operator is properly defined. Consequently, for generality, we distinguish the passive observation errors or independent observation error, $\boldsymbol{\varepsilon}_c^o$, from the active observation error $\boldsymbol{\varepsilon}^o$.

There are two important statistical assumptions from which we derive cross-validation diagnostics. Assuming that the observation errors are spatially uncorrelated, it follows that

$$\mathbf{E}[\boldsymbol{\varepsilon}^o(\boldsymbol{\varepsilon}_c^o)^T] = \mathbf{0}, \quad (4)$$

where $\mathbf{E}[\]$ is the mathematical expectation that represents the mean over an ensemble of realizations. It has been argued by Marseille *et al.* [12] that representativeness error can violate this assumption for a close pair of active-passive observations, but we will neglect this effect. Also, assuming that observation errors are uncorrelated with background error, we have

$$\mathbf{E}[\boldsymbol{\varepsilon}^o(\mathbf{H}\boldsymbol{\varepsilon}^f)^T] = \mathbf{0}, \quad \mathbf{E}[\boldsymbol{\varepsilon}_c^o(\mathbf{H}_c\boldsymbol{\varepsilon}^f)^T] = \mathbf{0}. \quad (5)$$

We come now to the most important property: since the analysis is a linear combination of the active observations and the background state, the analysis error is then uncorrelated with the passive observation errors,

$$\mathbf{E}[(\mathbf{H}_c\boldsymbol{\varepsilon}^a)(\boldsymbol{\varepsilon}_c^o)^T] = \mathbf{0} \quad (6)$$

and thus we get the following cross-validation diagnostic in passive observation space,

$$E[(O-A)_c(O-A)_c^T] = \mathbf{R}_c + \mathbf{H}_c \mathbf{A} \mathbf{H}_c^T, \quad (7)$$

similarly to Marseille *et al* [12]. A very important point to note is that \mathbf{A} is the *true* analysis error covariance - it does not assume that the gain is optimal. The matrices in eq.(7) are of the dimension of the passive observation space and \mathbf{R}_c is the observation error covariance matrix for the passive or independent observations.

2.2. A complete set of diagnostics of error covariances in passive observation space

It is also possible to define a set of diagnostics that would determine, in principle, the true error covariances, \mathbf{R} , \mathbf{B} and \mathbf{A} . From eq.(4,5) we get another cross-validation diagnostic,

$$E[(O-B)_c(O-B)_c^T] = \mathbf{H}_c \mathbf{B} \mathbf{H}_c^T. \quad (8)$$

This diagnostic is related to the Hollingsworth-Lönnberg [14] estimation of the spatially correlated part of the innovation, and in practice can be dominated by sampling error. Note that it is not a square matrix and an estimation of parameters of \mathbf{B} may not be trivial.

We can also obtain the innovation covariance matrix in passive observation space (identical in fact to the one in active observations space) as

$$E[(O-B)_c(O-B)_c^T] = \mathbf{R}_c + \mathbf{H}_c \mathbf{B} \mathbf{H}_c^T. \quad (9)$$

The system of equations (7-9) gives a complete set of equations to determine the *true* \mathbf{R} , \mathbf{B} and \mathbf{A} at the passive observation sites provided that by interpolation/extrapolation we can obtain $\mathbf{H}_c \mathbf{B} \mathbf{H}_c^T$ from $\mathbf{H}_c \mathbf{B} \mathbf{H}^T$.

Nevertheless, and for sake of completeness, we also investigated the meaning of the statistics $E[(O-A)_c(O-B)_c^T]$ and came to the conclusion that it can be interpreted as a misfit to the Desroziers *et al.* estimate of \mathbf{B} [7] with the true value \mathbf{B} . We recall that the first iterate of the Desroziers *et al.* estimate for \mathbf{B} in active observation space is given by $\mathbf{H} \mathbf{B}^D \mathbf{H}^T = E[(\mathbf{H} \mathbf{x}^a - \mathbf{H} \mathbf{x}^f)(\mathbf{y} - \mathbf{H} \mathbf{x}^f)^T]$ [8], where we used the superscript D indicate the Desroziers *et al.* first iterate estimate. We can actually generalize this diagnostic to be a cross-covariance between state space and active observation space as

$$\mathbf{B}^D \mathbf{H}^T = E[(\mathbf{x}^a - \mathbf{x}^f)(\mathbf{y} - \mathbf{H} \mathbf{x}^f)^T]. \quad (10)$$

By applying \mathbf{H}_c to eq.(10) we can then introduce a *generalized Desroziers et al. estimate of the background error covariance \mathbf{B} between passive and active observation spaces* as,

$$E[(A-B)_c(O-B)_c^T] = \mathbf{H}_c \mathbf{B}^D \mathbf{H}^T. \quad (11)$$

Since $(O-A)_c = (O-B)_c - (A-B)_c$, we get with eq.(8,10),

$$E[(O-A)_c(O-B)_c^T] = \mathbf{H}_c (\mathbf{B} - \mathbf{B}^D) \mathbf{H}^T. \quad (12)$$

That is the difference between the true \mathbf{B} and the Desroziers *et al.* first estimate \mathbf{B}^D in the cross passive-active observation spaces. Similarly to eq.(8) this diagnostics requires spatial interpolation of error covariances from active sites to passive sites of basically noisy statistics. The estimation of \mathbf{B} from this diagnostics is further complicated by the fact that what is being interpolated, that is $\mathbf{B} - \mathbf{B}^D$, may not even be positive definite, but the augmented matrix,

$$\text{cov} \begin{pmatrix} (O-B) \\ (O-A)_c \end{pmatrix} = \begin{bmatrix} \mathbf{R} + \mathbf{H} \mathbf{B} \mathbf{H}^T & \mathbf{H} (\mathbf{B} - \mathbf{B}^D) \mathbf{H}_c^T \\ \mathbf{H}_c (\mathbf{B} - \mathbf{B}^D) \mathbf{H}^T & \mathbf{R}_c + \mathbf{H}_c \mathbf{A} \mathbf{H}_c^T \end{bmatrix} \quad (13)$$

is positive definite. Except for the diagonal of eq.(13), we have not attempted in this study to conduct this complete estimation of \mathbf{R} , \mathbf{B} and \mathbf{A} , but rather focused on getting a reliable estimate of the analysis error covariance \mathbf{A} .

2.3. Geometrical interpretation

A geometrical illustration of some of the relationships obtained above can be made by using a Hilbert space representation of random variables of observation space. A 2D representation for the analysis of a scalar quantity was used in Desroziers *et al.* [7] to illustrate their a posteriori diagnostics. We will generalize this approach to include passive observations by considering a 3D representation.

As in Desroziers *et al.* [7] let's consider the analysis of a scalar quantity. Several variables are to be considered in this observation space: y^o the active observation (or measurement) of the scalar quantity, y^b the background (or prior) value equivalent in observation space (i.e. $y^b = H x^b$), y^a the analysis in observation space (i.e. $y^a = H x^a$), and for verification y^c an independent observation (or passive observation) that is not used to compute the analysis. Each of these quantities are random variables as they contain random errors, and any linear combination of random variables in observation space also belong to observation space. For example, $y^o - y^b$ is the innovation (commonly denoted by O-B) and that belongs to observation space, $y^a - y^b$ is the analysis increment in observation space (commonly denoted by A-B), and $y^o - y^a$ is the analysis residual in observation space (commonly denoted by O-A). We can also define an inner product of any random variables in observation space. y_1, y_2 , as

$$\langle y_1, y_2 \rangle := E[(y_1 - E(y_1))(y_2 - E(y_2))]. \quad (14)$$

The squared norm then represents the variance,

$$\|y\|^2 := \langle y, y \rangle = \sigma_y^2, \quad (15)$$

so the inner product has the following geometric interpretation

$$\langle y_1, y_2 \rangle = \|y_1\| \|y_2\| \cos \theta, \quad (16)$$

where $\cos \theta$ is the correlation coefficient. Uncorrelated random variables are thus *statistically orthogonal*. With this inner product, the observation space forms a Hilbert space of random variables.

Figure 1 illustrates the statistical relationship in observation space between: the active observation y^o (illustrated as O in the figure), the prior or background y^b (i.e. B), the analysis y^a (i.e. A), and the independent observation y^c (i.e. O_c). The origin T corresponds to the truth of the scalar quantity, and also corresponds to the zero of the central moment of each random variables, e.g. $y - E[y]$, since each variables are assumed to be unbiased. We also assume that the background, active and passive observations errors are uncorrelated to one another, so the three axes; ε^o for the active observation error, ε^b for the background error, and ε_c^o for the passive observation error are orthogonal. The plane defined by ε^o and ε^b axes is the space where the analysis takes place, and is called the analysis plane. But since we define the analysis to be linear and unbiased, only linear combinations of the form $y^a = k y^o + (1 - k) y^b$ where k is a constant are allowed. The analysis A then lies on the line (B,O). The thick lines in Figure 1 represent the norm of the associated error. For example, the thick line along the ε^o axis depict the (active) observation standard deviation σ_o , and similarly for the other axes and other random variables. Since the active observation error is uncorrelated with the background error, the triangle ΔOTB is a right triangle, and by Pythagoras theorem we have, $(y^o - y^b)^2 := \langle (O - B), (O - B) \rangle = \sigma_o^2 + \sigma_b^2$. This is the usual statement that the innovation variance is the

sum of background and observation error variances. The analysis is optimum when the analysis error $\|\varepsilon^a\|^2 = \sigma_a^2$ is minimum, in which case the line (T,A) is perpendicular to line (O,B).

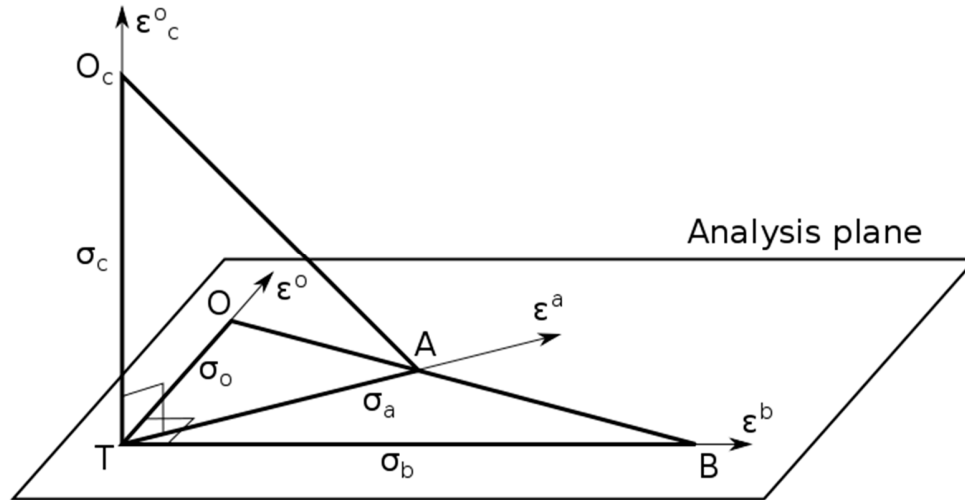


Figure 1. Hilbert space representation of a scalar analysis and cross-validation problem. The arrows indicate the directions of variability of the random variables, and the plane defined by the background and observation errors ε^b , ε^o defines the analysis plane. The thick lines represent the norm associated with the different random variables. T indicate the truth, O the active observation, B the background, A the analysis and O_c the passive observation.

Now let's consider the passive observation O_c . The passive observation error is perpendicular to the analysis plane, thus the triangle ΔO_cTA is a right triangle,

$$\overline{(y^c - y^a)^2} := \langle (O - A)_c, (O - A)_c \rangle = \sigma_c^2 + \sigma_a^2, \quad (17)$$

where σ_c^2 is the passive observation error variance. The most important fact to stress here is that the orthogonality expressed in eq.(17) is true whether or not the analysis is optimal. Furthermore, as the distance $(y^c - y^a)^2$ varies with the position of A along the line (O,B), the distance $(y^c - y^a)^2$ reaches a minimum value when σ_a^2 is minimum that is when the analysis is optimal. We thus also argue from this representation that there is always a minimum, and the minimum is unique. Finally we note that ΔBTO_c is also a right triangle so that $\overline{(y^c - y^b)^2} := \langle (O - B)_c, (O - B)_c \rangle = \sigma_c^2 + \sigma_b^2$, which is the scalar version of eq.(9).

To extend this formalism to random vectors requires to define a proper *matricial inner product* as briefly described in §1.2 of Caines [15]. It is important here to distinguish the stochastic metric space from the observation vector space. In the stochastic metric space a "scalar" needs not to be a number, but only a non-random quantity invariant with respect to $E[\cdot]$. Hence, we can define an Hilbert space with the following (stochastic scalar) matricial inner product $\langle \mathbf{y}, \mathbf{w} \rangle = E[(\mathbf{y} - E(\mathbf{y}))(\mathbf{w} - E(\mathbf{w}))^T] = \text{cov}(\mathbf{y}, \mathbf{w})$. The matrix nature of $\text{cov}(\mathbf{y}, \mathbf{w})$ pertains to the observation vector space, but it remains a scalar with respect to the stochastic Hilbert space herein defined. In order to obtain a true scalar ($\in \mathbb{R}$), one would need to define a metric on the observation space matrices as well, such as the trace. Also to be able to compare active and passive observations, projectors need to be introduced on the observation space, implying yet another metric structure on the observation space. We do not carry out this formalism here as it would represent a rather lengthy development which would distract us from the main purpose of this paper; this will be considered in a

future manuscript. Finally, we remark that Hilbert space representation of random variables in infinite dimensional space (i.e. continuous space) can also be defined, see Appendix 1 of Cohn [16].

2.4. Error covariance diagnostics in active observation space for optimal analysis

An analysis is optimal if the analysis error $E[(\boldsymbol{\varepsilon}^a)^T \boldsymbol{\varepsilon}^a]$ is minimum. This implies that the gain matrix using the prescribed error covariances, $\tilde{\mathbf{K}}$ in eq.(1), must be identical to the gain using the true error covariances, i.e. $\tilde{\mathbf{K}} = \mathbf{B}\mathbf{H}^T(\mathbf{H}\mathbf{B}\mathbf{H}^T + \mathbf{R})^{-1}$ [7,8]. It is important to mention that necessary and sufficient conditions to obtain the true error covariances in observation space that is $\mathbf{H}\mathbf{B}\mathbf{H}^T$ and \mathbf{R} , are: 1 - the Kalman gain condition, $\mathbf{H}\tilde{\mathbf{K}} = \mathbf{H}\mathbf{K}^{true}$ and 2- the innovation covariance consistency, $E[(O - B)(O - B)^T] = \mathbf{H}\tilde{\mathbf{B}}\mathbf{H}^T + \tilde{\mathbf{R}}$. For a proof see the Theorem on error covariance estimates in Ménard [8].

From the optimality of the analysis (or Kalman gain) alone, we derive that $E[(\hat{A} - T)(O - B)^T] = \mathbf{0}$ or $E[(\hat{A} - T)(O - \hat{A})^T] = \mathbf{0}$. Indeed, from eq.(2), we get $(O - \hat{A}) = \mathbf{R}(\mathbf{H}\mathbf{B}\mathbf{H}^T + \mathbf{R})^{-1}\mathbf{d}$, and for the analysis error in observation space we get, $(\hat{A} - T) = \mathbf{R}(\mathbf{H}\mathbf{B}\mathbf{H}^T + \mathbf{R})^{-1}\mathbf{H}\boldsymbol{\varepsilon}^f + \mathbf{H}\mathbf{B}\mathbf{H}^T(\mathbf{H}\mathbf{B}\mathbf{H}^T + \mathbf{R})^{-1}\boldsymbol{\varepsilon}^o$, from which we derive the expectations above. Using the geometrical representation in §2.3 the distance between A and T is minimum, when ΔTAO and ΔTAB (Figure 1) are right triangles. We should also note that for the scalar problem, the Kalman gain depends only on the ratio of observation to background error variances and thus the scalar Kalman gain is optimal if the ratio of the prescribed error variances is equal to the ratio of the true error variances.

If in addition to the optimality of the analysis or Kalman gain we add the innovation covariance consistency then we get three different statistical diagnostics of the (optimal) analysis error covariance. Hollingsworth and Lönnberg [6] was the first to introduce a statistical diagnostic of analysis error in the active observation space, as

$$E[(O - \hat{A})(O - \hat{A})^T] = \mathbf{R} - \mathbf{H}\hat{\mathbf{A}}_{HL}\mathbf{H}^T. \quad (18)$$

Here we use a subscript, *HL* to indicate that this is the Hollingsworth-Lönnberg estimate. Eq.(18) can be obtained from the covariance of $(O - \hat{A})$ and that, for an optimal gain matrix, $\mathbf{H}\hat{\mathbf{A}}\mathbf{H}^T = \mathbf{R} - \mathbf{R}(\mathbf{H}\mathbf{B}\mathbf{H}^T + \mathbf{R})^{-1}\mathbf{R}$ which derives from the usual formula, $\hat{\mathbf{A}} = \mathbf{B} - \mathbf{B}\mathbf{H}^T(\mathbf{H}\mathbf{B}\mathbf{H}^T + \mathbf{R})^{-1}\mathbf{H}\mathbf{B}$. Geometrically it derives from the fact that the triangle $\Delta T\hat{A}O$ is a right triangle, and from the innovation covariance consistency that implies that the triangle ΔOTB is a right triangle. Using data to construct $E[(O - \hat{A})(O - \hat{A})^T]$, an estimated analysis error covariance obtained from eq.(19) is symmetric but could be non-positive definite as it is obtained by subtracting two positive definite matrices. The effect of misspecification in the prescribed error covariances resulting from a lack of innovation covariance consistency will be discussed in the result section §3 and in Appendix B.

Inspired from the geometrical interpretation that ΔTAB is also be a right triangle we derived the following diagnostic,

$$E[(\hat{A} - B)(\hat{A} - B)^T] = \mathbf{H}\mathbf{B}\mathbf{H}^T - \mathbf{H}\hat{\mathbf{A}}_{MDJ}\mathbf{H}^T = \mathbf{H}(\mathbf{B} - \hat{\mathbf{A}}_{MDJ})\mathbf{H}^T, \quad (19)$$

where *MDJ* stands for Ménard-Deshaies-Jacques. This relationship is obtained by using the expression $(\hat{A} - B) = \mathbf{H}\mathbf{B}\mathbf{H}^T(\mathbf{H}\mathbf{B}\mathbf{H}^T + \mathbf{R})^{-1}\mathbf{d}$, the innovation covariance consistency and the formula for the optimal analysis error covariance $\hat{\mathbf{A}} = \mathbf{B} - \mathbf{B}\mathbf{H}^T(\mathbf{H}\mathbf{B}\mathbf{H}^T + \mathbf{R})^{-1}\mathbf{H}\mathbf{B}$. As for the *HL* diagnostics, the estimated error covariance obtained from this diagnostic is symmetric by construction but may not be positive definite. Another way of looking at eq.(19) is that it expresses, in observation space, the error reduction due to the use of observations.

Another diagnostic of analysis error covariance was proposed by Desroziers *et al.* [7]. By combining $(O - \hat{A})$ and $(\hat{A} - B)$ we get

$$E[(O - \hat{A})(\hat{A} - B)^T] = \mathbf{H}_D \hat{\mathbf{A}} \mathbf{H}_D^T, \quad (20)$$

where the subscript D denotes the Desroziers *et al.* estimate. By construction, the estimated analysis error covariance is not necessarily symmetric. A geometrical derivation is provided in Appendix A. We also provide in Appendix B a sensitivity analysis on the departure from innovation covariance consistency for each diagnostics, as well as for those in passive observation space which we will for our discussion.

2.5. Error covariance diagnostics in passive observation space for optimal analysis

We can also derive optimal analysis diagnostics in the passive observation space. Considering the 3D geometric interpretation, and in particular the tetrahedron (O_c, T, A, B) , we notice that since ΔTAB is a right triangle, so is $\Delta O_c AB$, which is a projection of the triangle ΔTAB on the plane passing through O_c , O and B . We thus have $E[(\hat{A} - B)_c(\hat{A} - B)_c^T] + E[(O - \hat{A})_c(O - \hat{A})_c^T] = E[(O - B)_c(O - B)_c^T]$. Combining this result with eq.(7) and using, $E[(O - B)_c(O - B)_c^T] = \mathbf{H}_c \mathbf{B} \mathbf{H}_c^T + \mathbf{R}_c$, we then get

$$E[(\hat{A} - B)_c(\hat{A} - B)_c^T] = \mathbf{H}_c \mathbf{B} \mathbf{H}_c^T - \mathbf{H}_c \hat{\mathbf{A}}_{MDJ} \mathbf{H}_c^T. \quad (21)$$

Note that our analysis diagnostic is the only diagnostic that is valid in both active and passive observation spaces, i.e. eq.(21) is similar to eq.(19). The other, less direct, diagnostic for optimal analysis is simply based on eq.(7) that is

$$\arg \min_{\gamma, L_c} \{ E[(O - A)_c(O - A)_c^T] \} = \mathbf{R}_c + \mathbf{H}_c \hat{\mathbf{A}}(\gamma, L_c) \mathbf{H}_c^T. \quad (22)$$

These five diagnostics will be used later in the results section.

3. Results with near optimal analyses

3.1. Experimental setup

We will just give here a short summary of the experimental setup we are using in this study. More details can be found in the Part I paper (Ménard and Deshaies-Jacques [5]). A series of hourly analyses of O_3 and $PM_{2.5}$ at 21 UTC for a period of 60 days (June 14 to August 12, 2014) were performed using an optimum interpolation scheme combining the operational air quality model GEM-MACH forecast and the real-time AirNow observations (see §2 of [5] for further details). The analyses are made off-line so they are not used to initialize the model. As input error covariances, we use uniform observation and background error variances, with $\tilde{\mathbf{R}} = \sigma_o^2 \mathbf{I}$ and $\tilde{\mathbf{B}} = \sigma_b^2 \mathbf{C}$, where \mathbf{C} is an homogeneous isotropic error correlation based on a second-order autoregressive model. The correlation length is estimated by using a maximum likelihood method using at first error variances obtained from a local Hollingworth-Lönnberg fit [7] (and only for the purpose of obtaining a first estimate of the correlation length). We then conduct a series of analyses by changing error variance ratio $\gamma = \sigma_o^2 / \sigma_b^2$ while at the same time respecting the innovation variance consistency condition, $\sigma_o^2 + \sigma_b^2 = \text{var}(O - B)$. This corresponds basically in minimizing the tr (trace) of eq.(7) while the trace of the innovation covariance consistency, $tr[\mathbf{R} + \mathbf{H} \mathbf{B} \mathbf{H}^T] = tr\{E[(O - B)(O - B)^T]\}$, is respected.

By leaving out 1/3rd of the observations for verification and constructing analyses with the remaining 2/3rd, we construct a cross-validation setup from which we can evaluate the diagnostic $\text{var}(O - A)_c = tr\{E[(O - A)_c(O - A)_c^T]\}$ in passive observation space. This constitutes our first guess experiment that we will refer to as **iter 0**. No observation or model bias was applied, and the innovation mean at the station were removed prior to perform the analysis. The statistic are first computed at the station using the 60-day members, and then was averaged over the domain so to give

the equivalent of a *tr* (trace). The mean value statistic for the three subsets are then averaged. More details can be found in §2 and beginning of §3 of Part I [5].

The red curve on Figure 2 illustrates how this diagnostic varies with γ and exhibits a minimum. This minimum can easily be understood by referring to Figure 1: as the analysis point *A* changes position along the line (*O*,*B*) the distance $\|O_c - A\|$ reaches a minimum, and this is what we observe in Figure 2.

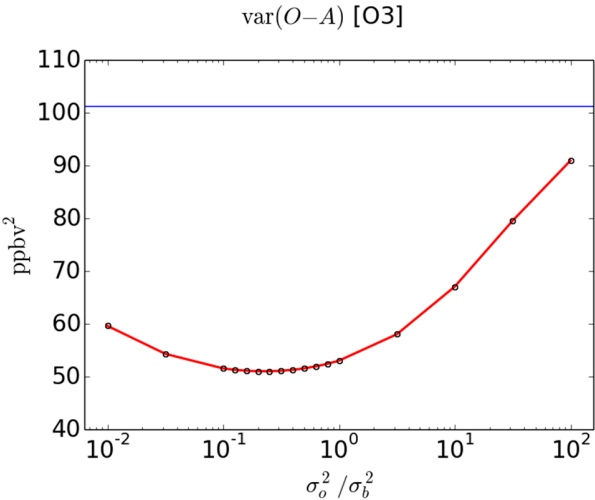


Figure 2. Variance of observation-minus-analysis of O₃ in passive observation space. Red line, and the variance of observation-minus-model, blue line.

In our next step, **iter 1**, we address the issue related to the spatial correlation that has not being optimized. Our approach consist in using a variance estimator sequentially with a maximum likelihood estimation of correlation length as was done in Ménard [8]. So, to refine our estimate of model covariance in **iter 1**, we use the variances verifying the optimal ratio $\hat{\gamma}$ of **iter 0** and the innovation variance consistency to re-estimate by, maximum likelihood, the correlation length of \tilde{C} . Then, in a procedure identical to **iter 0** we estimate the optimal $\hat{\gamma}$ (**iter 1**), which turn out to be very close to the value obtained in **iter 0**. By innovation variance consistency we thus get improved estimates for the observation and background error variances, from which we will compute a number of statistical diagnostics of the analysis error variance in the section that follows. However, as before, we do not attempt to provide a spatial distribution of the error statistics that are still assumed uniform. A summary of the main features of the two experiments are presented in Table 1. The χ^2/N_s value shown in Table 1 is the 60-day mean value of $\chi_k^2/N_s(k)$ where $N_s(k)$ is the total number of observations available at time t_k . We observe that there was a significant improvement from **iter 0** to **iter 1** in terms of χ^2/N_s but is still not equal to one. We thus refer the analysis of **iter 1** as *near* optimal.

Table1. Input error statistics for the first experiment and optimized variance ratio experiment

Experiment	L_c (km)	$\langle (O - B)^2 \rangle$	$\hat{\gamma} = \hat{\sigma}_o^2 / \hat{\sigma}_b^2$	$\hat{\sigma}_o^2$	$\hat{\sigma}_b^2$	χ^2 / N_s
O ₃ iter 0	124	101.25	0.22	18.3	83	2.23
O ₃ iter 1	45	101.25	0.25	20.2	81	1.36
PM _{2.5} iter 0	196	93.93	0.17	13.6	80.3	2.04
PM _{2.5} iter 1	86	93.93	0.22	16.9	77	1.25

This procedure converges really fast and in practice there is no need to go beyond **iter 1**. Figure 3 displays iterates 0 to 4 with our estimation procedure for O₃. With one iteration update we nearly converge. A similar procedure was used in Ménard [8], where the variance and correlation length (estimated by maximum likelihood) were estimated in sequence, which taught us that a slow and fictitious drift in estimated variances and correlation length can occur when the correlation model is not the true

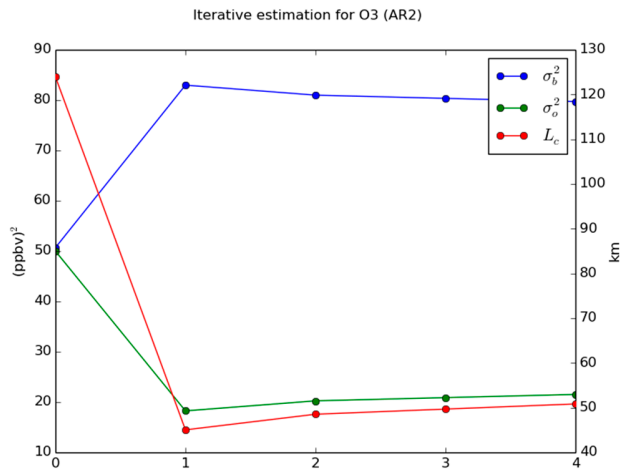


Figure 3. Optimal estimates of σ_o^2 , σ_b^2 and maximum likelihood estimate of correlation length L_c for the first four iterates. Blue, is the optimal background error variance, green, the optimal observation error variance and in red the correlation length (in km, with labels on the right side of the figure).

correlation. So in regard of similar considerations that may occur here, we do not extend our iteration procedure beyond the first iterate.

3.2. Statistical diagnostics of analysis error variance

For each of these experiments statistics related diagnostics for analysis error variance, discussed in §2, are computed and the results are presented in Table 2 for the verification made against active observations, and in Table 3 to the verification made against the passive observations.

Table2. Analysis statistics against active observations

Experiment	Active var($\hat{A}-B$)	Active $tr(\mathbf{H}\hat{\mathbf{A}}_{MDJ}\mathbf{H}^T)/N_s$	Active $tr(\mathbf{H}\hat{\mathbf{A}}_D\mathbf{H}^T)/N_s$	Active var($O-\hat{A}$)	Active $tr(\mathbf{H}\hat{\mathbf{A}}_{HL}\mathbf{H}^T)/N_s$
O ₃ iter 0	60.29	22.69	9.61	24.33	-6.03
O ₃ iter 1	67.66	13.32	13.68	11.26	8.94
PM _{2.5} iter 0	62.29	17.98	7.71	16.78	-3.18
PM _{2.5} iter 1	66.3	10.68	9.51	9.57	7.33

The second, third and last column of Table 2 are tabulated estimates of the analysis error variance at the active location sites, i.e. $tr(\mathbf{H}\mathbf{A}\mathbf{H}^T)/N_s$, obtained by three different methods. The second column is an estimate given with our method $\sigma_b^2 - \text{var}(\hat{A}-B) = tr(\mathbf{H}\hat{\mathbf{A}}_{MDJ}\mathbf{H}^T)/N_s$. The third column is the Desroziers *et al.* estimate of analysis error [7], eq.(20), and the last column is the estimate using the method proposed by Hollingsworth and Lönnberg [6], eq.(18). We note that the analysis error variance estimate provided by the first two methods is fairly consistent for an updated correlation length

estimate, i.e. **iter1** (but not **iter0**). We also note that χ^2/p is closer to one for **iter1**. These two facts indicate that the updated correlation length (**iter1**) with uniform error variances is closer to the innovation covariance consistency. The Hollingsworth and Lönnberg [6] method however, is very sensitive and negatively biased in the lack of innovation covariance consistency.

Estimate of the analysis error variance at the passive observation locations, i.e. $tr(\mathbf{H}_c \mathbf{A} \mathbf{H}_c^T)/N_s$, provided by two different methods are given by eq.(21) in column 3 and by eq.(22) in column 5 of table 3. As for the estimate at the active locations (Table 2), there is a general agreement on the analysis error estimates with the updated correlation length (**iter1**), although this distinction is not that clear for PM2.5.

Table3. Analysis statistics against passive observations

Experiment	Passive $var[(\hat{A}-B)_c]$	Passive $tr(\mathbf{H}_c \hat{\mathbf{A}}_{MDJ} \mathbf{H}_c^T)/N_s$	Passive $var[(O-\hat{A})_c]$	Passive $var[(O-\hat{A})_c]-\sigma_{oc}^2$
O ₃ iter 0	56.95	26.03	51.02	32.72
O ₃ iter 1	52.04	28.95	48.95	28.75
PM _{2.5} iter 0	62.29	22.65	38.09	24.49
PM _{2.5} iter 1	66.3	24.62	38.28	21.38

We note also that the analysis error variance at the active sites is smaller than the analysis error variance at the passive observation sites. This involves in particular the fact that since the passive observation are away from the active observation sites, the reduction of variance at the passive observation sites is smaller than at the active observation sites.

3.3. Comparison with the perceived analysis error variance

We computed the analysis error covariance \mathbf{A} resulting from the analysis scheme, the so-called perceived analysis covariance [9], using the expression,

$$\mathbf{A} = \tilde{\mathbf{B}} - \tilde{\mathbf{B}} \mathbf{H}^T (\mathbf{H} \tilde{\mathbf{B}} \mathbf{H} + \tilde{\mathbf{R}})^{-1} \mathbf{H} \tilde{\mathbf{B}} = \tilde{\mathbf{B}} - \mathbf{G} \mathbf{G}^T . \tag{23}$$

We then compared the perceived analysis error variance with the estimated active analysis error variance from the previous subsection.

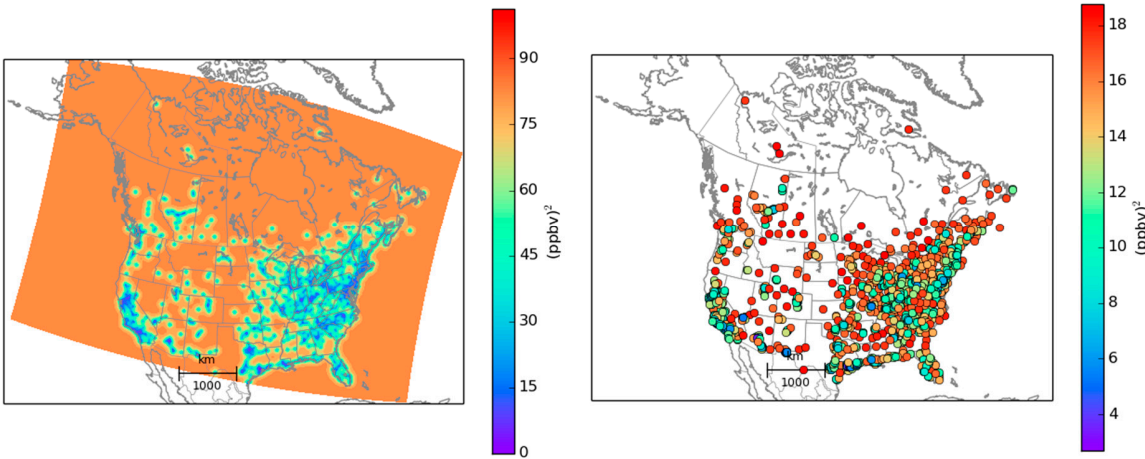


Figure 4. Analysis error variance for ozone optimal analysis case O₃ **iter1**. Left panel is the analysis error on the model grid and on the right panel at the active observation sites. Note that the color bar of the left and right panels are different. The maximum of the color bar for the left panel correspond to $\sigma_o^2 + \sigma_b^2$.

In order to calculate the perceived analysis error covariance eq.(23) we first perform a Choleski decomposition of $\mathbf{H}\tilde{\mathbf{B}}\mathbf{H}^T + \tilde{\mathbf{R}} = \mathbf{L}\mathbf{L}^T$, where \mathbf{L} is a lower triangular matrix. Then with a forward substitution we obtain \mathbf{L}^{-1} , from which we compute $\mathbf{G} = \tilde{\mathbf{B}}\mathbf{H}^T\mathbf{L}^{-T}$. The perceived analysis error variance for the ozone optimal analysis (i.e. O₃ **iter1**) is displayed in Figure 4 (A similar figure but for PM_{2.5} is given in supplementary material). We note that although the input statistics used for the analysis are uniform (i.e. uniform background and observation error variances, and homogeneous correlation model), the computed analysis error variance at the active observation location displays large variations, which is attributed to the non-uniform spatial distribution of the active observations. In Figure 5 we display a histogram of those variances for the ozone optimal analysis O₃ **iter1** (right panel) and for the first experiment O₃ **iter0** (left panel) without optimization (A similar figure is but for PM_{2.5} is given in supplementary material). Note that median or mean values of variances are significantly different between the optimal and non-optimal analysis cases. We also observe two maxima, one of which is actually due to isolated observation sites. At those sites the analysis error variance is simply obtained by the scalar equation $1/\sigma_a^2 = 1/\sigma_o^2 + 1/\sigma_b^2$. For O₃ **iter1** the scalar analysis error variance gives 16.2, and for O₃ **iter0** we get 15.0, thus explaining the secondary maxima.

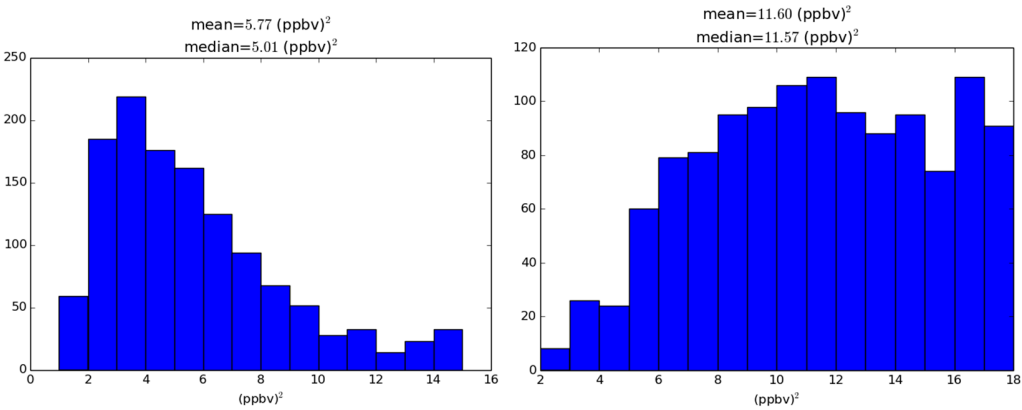


Figure 5. Distribution (histogram) of the ozone analysis error variance at the active observation locations. First analysis experiment O₃ **iter0** (no optimization) on the left panel, and optimal analysis case O₃ **iter1** on the right panel.

The mean perceived analysis error variance for all experiments is presented in Table 4. Comparing these values with the estimated values of analysis error variance based on diagnostics in Table 2 we note that for both optimal experiments, O₃ **iter1** and PM_{2.5} **iter1**, the perceived analysis error variance roughly agrees with all analysis error variances estimated with diagnostics (Table 2). But for the non-optimal analyses, O₃ **iter0** and PM_{2.5} **iter0**, there is a general disagreement between all estimated values.

Table 4. Perceived analysis error variance. Mean over active observation sites.

Experiment	Perceived $tr(\mathbf{H}\mathbf{A}_p\mathbf{H}^T)/N_s$
O ₃ iter 0	5.77
O ₃ iter 1	11.60
PM _{2.5} iter 0	4.37
PM _{2.5} iter 1	8.21

Looking more closely, however, we note that the agreement in the optimal case is not perfect. The perceived analysis error variance is about 20% lower than the best estimates $tr(\mathbf{H}\hat{\mathbf{A}}_{MDJ}\mathbf{H}^T)/N_s$ and $tr(\mathbf{H}\hat{\mathbf{A}}_D\mathbf{H}^T)/N_s$. The optimal χ^2/N_s values in the “optimal” cases are slightly above one, thus indicating that some more tuning of the error statistics could be done to between reach an innovation consistency. More on that matter will be presented in §4.5.

4. Discussion on the statistical assumptions and practical applications

4.1 Representativeness error with in situ observations

The statistical diagnostics presented in §2 derive from the assumption that the observation errors are horizontally uncorrelated and uncorrelated with the background error. Although this assumption is never entirely observed in reality, there are ways to work around it. In the case of in situ observations, and assuming that any systematic error have been removed, random errors are still present, due to the difference between the observation and the model’s equivalent of the observation – called representativeness error (see Janjic *et al.* [18] for a review). Representativeness error is due to unresolved scales and processes in the model and interpolation or forward observation model errors. These errors are typically roughly at the scale of the model grid [19,20], so typically a few tens of kilometers for air quality models. This should not be confused with the representativeness of an observation, where, for example, remote stations are representative of large area (e.g. several hundreds of kilometers), whereas urban and suburban stations are at the scale of human activity in the cities, traffic and industries, etc. and are, depending on the chemical specie, of a few kilometers and less.

Representativeness error of *in situ* measurements can be discarded altogether by simply filtering any pair of observations that are in the range of a few model grid size, both in assimilation and estimation of error statistics [1] or in pairs of passive-active observations for cross-validation [12]. Once this filtering is done, the assumption on observation errors being spatially uncorrelated and uncorrelated with the background error then applies.

4.2 Correlated observation-background errors

In any case, it is interesting to show how the different diagnostics, introduced in §2, depends on the statistical assumptions of the observation error. One way to get an understanding of the effect of these assumptions is to look at it from a geometrical point of view, using the representation introduced in §2.2. Note that the same results can be obtained analytically, but the geometrical interpretation gives a simple and appealing way of looking at the problem.

Let us consider the effect on the analysis of observation error correlated with background error. The case where the observation error is uncorrelated with background error is represented in Figure 6 on the left panel and when we they are correlated on the right panel. The observation and background error variances are kept unchanged, with the same (O,T) length and (B,T) length in both panels. In the case of correlated errors the angle $\angle BTO$ is no longer a right angle. Yet, it is still possible to obtain an optimal analysis, $\hat{\mathbf{A}}$, as a linear

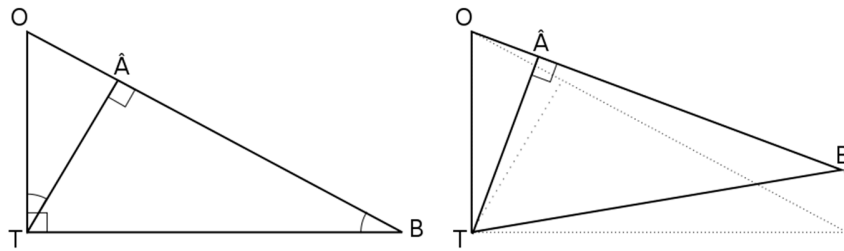


Figure 6. Geometrical representation of the analysis. Left panel, for observation errors uncorrelated with the background error. Right panel, with correlated errors. T indicate the truth, O the observation, B the background and \hat{A} the optimal analysis.

combination of the observation and the background, on the line (O,B) , for which the distance \hat{A} to T (i.e. the analysis error variance) is minimum. In this case, $(\hat{A},T) \perp (O,B)$. Note that for strongly correlated errors and when $\sigma_b^2 > \sigma_o^2$, although \hat{A} is still on the line (O,B) , it may actually lie outside the segment $[O,B]$. Yet, the principles and theory still hold in that case.

When the observation error is uncorrelated with the background error, $(O,T) \perp (B,T)$, the triangles $\triangle OT\hat{A}$ and $\triangle TB\hat{A}$ are similar and it follows that $\langle (O-\hat{A})(\hat{A}-B) \rangle = \|\hat{A}-T\|^2$, which is the Desroziers *et al.* [7] diagnostic for analysis error variance. But, when the observation error is correlated with the background error (right panel of Figure 6), the triangles $\triangle OT\hat{A}$ and $\triangle TB\hat{A}$ are no longer similar triangles and the Desroziers *et al.* [7] diagnostics for analysis error does not hold (see derivation in Appendix A). However, the HL, eq.(18), and MDJ diagnostic, eq.(19), depend only on having right triangles $\triangle OT\hat{A}$ and $\triangle TB\hat{A}$, and not on the orthogonality of (B,T) with (O,T) . Therefore, the HL and MDJ diagnostics are valid with or without correlated observation-background errors.

4.3 Estimation of satellite observation errors with *in situ* observation cross-validation

One of the important problems in satellite assimilation is the estimation of the satellite observation error, which could be addressed with a simple modification of our cross-validation procedure. Let us assume that we have *in situ* observations that we assume to have uncorrelated errors between themselves (or use a filter with a minimum distance as discussed in §4.1), with the background errors and the satellite observation errors. Yet, the satellite observation errors could be correlated with the background error. Satellite observations could come from a multi-channel instrument with channel-correlated observation errors, as found with many instruments, and yet our validation procedure can still be used. Let us consider that the analyses comprise of satellite and *in situ* observations but, for the purpose of cross-validation, we use only 2/3rd of the *in situ* observations in the analysis, and keep the remaining 1/3rd as passive to carry out the cross-validation procedure.

The first thing to note is that the passive *in situ* observations have uncorrelated errors with the analysis error (the analysis is composed of satellite observations and 2/3rd of the *in situ* observations). We then use eq.(7) where the interpolation of the analysis is made only at the *in situ* active observations. Minimizing $E[(O-A)_c^T(O-A)]$ (i.e. the trace of the l.h.s. of eq.(7)) results in finding the optimal *in situ* observation weight. Then, computing the analysis error covariance in the satellite observation space from the analysis scheme (either from a Hessian of a variational cost function, or with an explicit gain as in eq.(23)), i.e. $\mathbf{H}_{sat}\hat{\mathbf{A}}\mathbf{H}_{sat}^T$, we use the HL formulation eq.(18) to obtain the satellite observation error covariance,

$$\mathbf{H}_{sat}\hat{\mathbf{A}}\mathbf{H}_{sat}^T + e[(O-A)_{sat}(O-A)_{sat}^T] = \mathbf{R}_{sat}. \quad (24)$$

The equation (24) has the important properties that the estimated observation error covariance is symmetric and positive definite by construction. Then, a new analysis could be carried out to obtain a more realistic $\mathbf{H}_{sat} \hat{\mathbf{A}} \mathbf{H}_{sat}^T$, with a resulting updated \mathbf{R}_{sat} , and so forth until convergence.

4.4 Remark on cross-validation of satellite retrievals

As a last remark, it appears that cross-validation of satellite retrieval observations using a k -fold approach where the observations are used as passive observations to validate the analysis can be a difficult problem. Retrievals from passive remote sensing at nadir generally involve a prior or climatology or a model assumption over different regions, and is thus likely to have spatially correlated errors and errors correlated with the background error. It doesn't mean, however, that nothing can be done in that case. For example, for certain sensors, such as infrared sensors, it is possible to disentangle the prior from the retrieval, so that by an appropriate transformation of the measurements, observations can be practically decorrelated from the background [21,22]. However to the authors' knowledge, such an approach have never been undertaken for visible measurements such as for NO₂ or AOD's.

4.5 Lack of innovation covariance consistency and its relevance to the statistical diagnostics

The error covariance diagnostics for optimal analysis, presented in §2.4 and §2.5, depends on the innovation covariance consistency, $E[(O - B)(O - B)^T] = \mathbf{H}\tilde{\mathbf{B}}\mathbf{H}^T + \tilde{\mathbf{R}}$, and our results presented in §4 have shown that the different estimates for the optimal analysis error variance are close, but do not strictly agreeing to each other. This disagreement is related to the lack of innovation consistency as follows.

Let us introduce a departure matrix Δ from innovation covariance consistency as,

$$E[\mathbf{d}\mathbf{d}^T] (\mathbf{H}\tilde{\mathbf{B}}\mathbf{H}^T + \tilde{\mathbf{R}})^{-1} = \mathbf{I} + \Delta. \quad (25)$$

The trace of eq.(25), which is related to χ^2 , is given by

$$E[\chi^2] = \text{tr}\{E[\mathbf{d}\mathbf{d}^T] (\mathbf{H}\tilde{\mathbf{B}}\mathbf{H}^T + \tilde{\mathbf{R}})^{-1}\} = N_s + \text{tr}(\Delta). \quad (26)$$

We recall that in the experiment **iter 1** we got χ^2 / N_s values of 1.36 for O₃ and 1.25 for PM_{2.5} (see Table 1), indicating that the innovation covariance consistency is deficient, although less serious than with the experiment **iter 0** where values of 2 and higher have been obtained.

If we take into account the fact that there can be a difference between $E[\mathbf{d}\mathbf{d}^T]$ and $(\mathbf{H}\tilde{\mathbf{B}}\mathbf{H}^T + \tilde{\mathbf{R}})$ and we rederive the (active) analysis error covariance for *HL*, *MDJ* and *D* schemes, we get (see Appendix B)

$$\text{tr}\{\mathbf{H}\hat{\mathbf{A}}_{HL}\mathbf{H}^T\} = \text{tr}\{\mathbf{H}\hat{\mathbf{A}}^{true}\mathbf{H}^T\} + \text{tr}\{\tilde{\mathbf{R}}\Delta\} - \text{tr}\{\mathbf{H}\tilde{\mathbf{B}}\mathbf{H}^T\Delta\} + \text{tr}\{\mathbf{error}_{MDJ}\} \quad (27)$$

$$\text{tr}\{\mathbf{H}\hat{\mathbf{A}}_{MDJ}\mathbf{H}^T\} = \text{tr}\{\mathbf{H}\hat{\mathbf{A}}^{true}\mathbf{H}^T\} - \text{tr}\{\mathbf{error}_{MDJ}\} \quad (28)$$

$$\text{tr}\{\mathbf{H}\hat{\mathbf{A}}_D\mathbf{H}^T\} = \text{tr}\{\mathbf{H}\hat{\mathbf{A}}^{true}\mathbf{H}^T\} - \text{tr}\{\mathbf{error}_D\}, \quad (29)$$

where $\text{tr}\{\mathbf{error}_{MDJ}\} = \text{tr}\{(\mathbf{H}\tilde{\mathbf{B}}\mathbf{H}^T)(\mathbf{H}\tilde{\mathbf{B}}\mathbf{H}^T + \tilde{\mathbf{R}})^{-1}\Delta(\mathbf{H}\tilde{\mathbf{B}}\mathbf{H}^T)\}$, $\text{tr}\{\mathbf{error}_D\} = \text{tr}\{\tilde{\mathbf{R}}(\mathbf{H}\tilde{\mathbf{B}}\mathbf{H}^T + \tilde{\mathbf{R}})^{-1}\Delta(\mathbf{H}\tilde{\mathbf{B}}\mathbf{H}^T)\}$. We note that although the error terms are complex expressions, they all depend linearly on Δ . Thus, the disagreement between the *HL*, *MDJ* and *D* analysis error variance estimates is due to lack of innovation covariance consistency.

5. Conclusions

We show that analysis error variance can be estimated and optimized, without using a model forecast, by partitioning the original observation data set into a training set, to create the analysis, and an independent (or passive) set, used to evaluate the analysis. This kind of evaluation by partitioning is called cross-validation. The method derives from assuming that the observations have spatially uncorrelated errors or, minimally, that the independent (or passive) observations have uncorrelated errors with the active observation, and are uncorrelated the background error. This leads to the important property that passive observations are uncorrelated with the analysis error and can then be used to evaluate the analysis [12].

We have developed a theoretical framework and a geometric interpretation that has allowed us to derive a number of statistical estimation formulas of analysis error covariance that can be used in both passive and active observation spaces. It is shown that by minimizing the variance of observation-minus-analysis residuals in passive observation space we actually identify the optimal analysis. This has been done with respect to a single parameter, namely the ratio of observation to background error variances, to obtain a near optimal Kalman gain. The optimization is also done under the constraint of the innovation covariance consistency [7,8]. This optimization could have been done with more than one error covariance parameter but this has not been attempted here. The theory does suggest, however, that the minimum is unique.

Once an optimal analysis is identified we conduct an evaluation of the analysis error covariance using several different formulas; Desroziers *et al.* [7], Hollingsworth Lönnberg [6], and one that we develop in this paper which works in either active or passive observation spaces. As a way to validate the analysis error variance computed by the analysis scheme itself, the so-called perceived analysis error variance [9], we compare it with the values obtained from the different statistical diagnostics of analysis error variance.

We applied this methodology to surface air quality analyses using our operational air quality model GEM-MACH and real-time surface observations of O₃ and PM_{2.5}. Our method applied the theory in a simplified way. First by considering the averaged observation and background error variances and finding an optimal ratio $\gamma = \sigma_o^2 / \sigma_b^2$ using as a constraint the trace of the innovation covariance consistency [8]. Second, using a single parameter correlation model, its correlation length, we used the maximum likelihood estimation [17] to obtain near optimal analyses. Also we did not attempt to account for representativeness error in the observations by, for example, filtering observations that are close. Despite all these limitations, our results show that with near optimal analyses, all estimates of analysis error variance roughly agree with each other, while disagreeing strongly when the input error statistics are not optimal. This first check on estimating the analysis error variance gives us confidence that the method we propose is reliable and that by being more specific in the spatial distribution of error variances, background error correlation and possible thinning to remove effects of representativeness errors, we could get an even better agreement between all the diagnostic measure of analysis error variances.

The methodology introduced here for estimating analysis error variances is general and not restricted to the case of surface pollutant analysis. It would be desirable to investigate other areas of applications, such as surface analysis in meteorology and oceanography. The method could, in principle, provide guidance for any assimilation system. By considering the observation space subdomain [23], proper scaling, local averaging [24], or other methods discussed in Janjic *et al.* [18] it may also be possible to extend this methodology to spatially varying error statistics. Based on our verification results in Part I [5], we found that there is a dependence between model values and error variances, which we will investigate further in view of our next operational implementation of the Canadian surface air quality analysis and assimilation.

One strong limitation of the optimum interpolation scheme we are using (i.e. homogeneous isotropic error correlation and uniform error variances), which is also the case for most 3D-Var implementations, is the lack of innovation covariance consistency. Ensemble Kalman filters seems, however, much better in that regard although they have their own issues with localization and inflation. Experiments with chemical data assimilation using an ensemble Kalman filter does gives χ^2/N_s values very close to unity after simple adjustments for observation and model error variances [25]. We thus argue that ensemble methods, such as the ensemble Kalman filter, would produce analysis error variance estimates that are much more consistent between the different diagnostics.

Estimates of analysis uncertainties can also be obtained by resampling techniques, such as the jackknife method and bootstrapping [26]. In bootstrapping with replacement, the distribution of the analysis error is obtained by creating new analyses by replacing and duplicating observations from an existing set of observations [26]. This technique relies on the assumption that each member of the dataset is independent and identically distributed. For surface ozone analyses where there is persistence to next day and the statistics is spatially inhomogeneous, the assumption of statistical independence may not be adequate. The comparison of these resampling estimates of analysis uncertainties could be compared with our analysis error variance estimates to help us identify limitations and areas of improvement.

Supplementary Materials: The following are available online at www.mdpi.com/link, Figure S6: Analysis error variance for ozone optimal analysis case PM_{2.5} **iter1**. Left panel is the analysis error on the model grid and on the right panel at the active observation sites. Note that the color bar of the left and right panels are different. The maximum of the color bar for the left panel correspond to $\sigma_o^2 + \sigma_b^2$, Figure S7: Distribution (histogram) of the ozone analysis error variance at the active observation locations. First analysis experiment PM_{2.5} **iter0** (no optimization) on the left panel, and optimal analysis case PM_{2.5} **iter1** on the right panel.

Acknowledgments: We are grateful to the US/EPA for the use of the AIRNow database for surface pollutants and to all provincial governments and territories of Canada for kindly transmitting their data to the Canadian Meteorological Centre to produce the surface analysis of atmospheric pollutants. We are also thankful for the proof read by Kerill Semeniuk, and for two anonymous reviewers for their comments and help in improving the manuscript.

Author Contributions: This research was conducted as a joint effort by both authors. R.M. contributed to the theoretical development and wrote the paper, and M.D.-J. conducted all experiments design and execution, proof reading and introduced a new diagnostic of optimal analysis error that was further extended to passive observations space.

Conflicts of Interest: The authors declare no conflict of interest. The founding sponsors, which is the government of Canada, had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, and in the decision to publish the results.

Appendix A – A geometrical derivation of the Desroziers *et al.* diagnostic

Let u and v be two random variables of a real Hilbert space as defined in §2.3. Two properties of Hilbert spaces are : the polarization identity

$$\langle u, v \rangle = \frac{1}{4} \{ \|u+v\|^2 - \|u-v\|^2 \} \quad (\text{A.1})$$

and the parallelogram identity [15]

$$\|u+v\|^2 + \|u-v\|^2 = 2(\|u\|^2 + \|v\|^2). \quad (\text{A.2})$$

Combining these two equations, we get:

$$\langle u, v \rangle = \frac{1}{2} \left\{ \|u + v\|^2 - \|u\|^2 - \|v\|^2 \right\}. \quad (\text{A.3})$$

640

641

642 Let $u = O - A$ and $v = A - B$, then $u + v = O - B$, and so we have

$$\langle (O - A), (A - B) \rangle = \frac{1}{2} \left\{ \|O - B\|^2 - \|O - A\|^2 - \|A - B\|^2 \right\}. \quad (\text{A.4})$$

644 If we assume that the analysis is optimal, so that $\angle OAT$ is a right triangle (see Figure 1), then

$$\|\hat{A} - T\|^2 + \|O - \hat{A}\|^2 = \|O - T\|^2, \quad (\text{A.5})$$

646 and similarly that $\angle TAB$ is a right triangle,

$$\|\hat{A} - T\|^2 + \|\hat{A} - B\|^2 = \|B - T\|^2, \quad (\text{A.6})$$

648 and substitute these expression into (A.4) we get

$$\langle (O - \hat{A}), (\hat{A} - B) \rangle = \frac{1}{2} \left\{ \|O - B\|^2 - \|O - T\|^2 - \|B - T\|^2 - 2\|\hat{A} - T\|^2 \right\}. \quad (\text{A.7})$$

650 If in addition we assume that we have uncorrelated observation-background errors, that is

$$\|O - B\|^2 = \|O - T\|^2 + \|B - T\|^2, \quad (\text{A.8})$$

652 we then get

$$\langle (O - \hat{A}), (\hat{A} - B) \rangle = \|\hat{A} - T\|^2. \quad (\text{A.9})$$

654 Note that the difference with the result obtained in Appendix A (property 6) of Ménard [8] where it is
 655 shown that the necessary and sufficient condition for the *perceived* analysis error covariance in active
 656 observation space meet the Desroziers *et al.* diagnostics for analysis error, is to have innovation
 657 covatiance consistency and uncorrelated observation-background errors. The result obtained here is
 658 different, it concerns the *optimal* analysis error covariance.

659 Appendix B – Diagnostics of analysis error covariance and the innovation covariance consistency

660 Let us introduce a departure matrix Δ from innovation covariance consistency as,

$$E[\mathbf{d}\mathbf{d}^T] (\mathbf{H}\tilde{\mathbf{B}}\mathbf{H}^T + \tilde{\mathbf{R}})^{-1} = \mathbf{I} + \Delta. \quad (\text{B.1})$$

662 The HL diagnostic, eq.(18), can be expanded as

$$\begin{aligned} E[(O - \hat{A})(O - \hat{A})^T] &= E[\mathbf{d}\mathbf{d}^T] + \mathbf{H}\tilde{\mathbf{B}}\mathbf{H}^T (\mathbf{H}\tilde{\mathbf{B}}\mathbf{H}^T + \tilde{\mathbf{R}})^{-1} E[\mathbf{d}\mathbf{d}^T] (\mathbf{H}\tilde{\mathbf{B}}\mathbf{H}^T + \tilde{\mathbf{R}})^{-1} \\ &\quad - E[\mathbf{d}\mathbf{d}^T] (\mathbf{H}\tilde{\mathbf{B}}\mathbf{H}^T + \tilde{\mathbf{R}})^{-1} \mathbf{H}\tilde{\mathbf{B}}\mathbf{H}^T - \mathbf{H}\tilde{\mathbf{B}}\mathbf{H}^T (\mathbf{H}\tilde{\mathbf{B}}\mathbf{H}^T + \tilde{\mathbf{R}})^{-1} E[\mathbf{d}\mathbf{d}^T] \\ &= \tilde{\mathbf{R}} - \mathbf{H}(\tilde{\mathbf{B}} - \tilde{\mathbf{B}}\mathbf{H}(\mathbf{H}\tilde{\mathbf{B}}\mathbf{H}^T + \tilde{\mathbf{R}})^{-1} \mathbf{H}\tilde{\mathbf{B}})\mathbf{H}^T + \mathbf{error}_{HL} \\ &= \tilde{\mathbf{R}} - \mathbf{H}\hat{\mathbf{A}}\mathbf{H}^T + \mathbf{error}_{HL} \end{aligned} \quad (\text{B.2})$$

664 where the \mathbf{error}_{HL} is given as

$$\mathbf{error}_{HL} = \Delta \tilde{\mathbf{R}} - (\mathbf{H}\tilde{\mathbf{B}}\mathbf{H}^T) \Delta^T + \mathbf{H}\tilde{\mathbf{B}}\mathbf{H}^T (\mathbf{H}\tilde{\mathbf{B}}\mathbf{H}^T + \tilde{\mathbf{R}})^{-1} \Delta (\mathbf{H}\tilde{\mathbf{B}}\mathbf{H}^T), \quad (\text{B.3})$$

666 which includes three error terms. A scalar version of eq.(B.2) is given as

$$error_{HL} = \frac{\sigma_b^2 \Delta}{1 + \gamma} - \sigma_b^2 \Delta + \sigma_o^2 \Delta. \quad (B.4)$$

The error analysis for the MDJ diagnostic, eq. (19) is

$$E[(\hat{A} - B)(\hat{A} - B)] = \mathbf{H}\tilde{\mathbf{K}} E[\mathbf{d}\mathbf{d}^T] \tilde{\mathbf{K}}\mathbf{H}^T \\ = \mathbf{H}\tilde{\mathbf{B}}\mathbf{H}^T (\mathbf{H}\tilde{\mathbf{B}}\mathbf{H}^T + \tilde{\mathbf{R}})^{-1} \mathbf{H}\tilde{\mathbf{B}}\mathbf{H}^T + \mathbf{error}_{MDJ}, \quad (B.5)$$

where the \mathbf{error}_{MD} is given as

$$\mathbf{error}_{MDJ} = \mathbf{H}\tilde{\mathbf{B}}\mathbf{H}^T (\mathbf{H}\tilde{\mathbf{B}}\mathbf{H}^T + \tilde{\mathbf{R}})^{-1} \Delta (\mathbf{H}\tilde{\mathbf{B}}\mathbf{H}^T). \quad (B.6)$$

There is only one term, and its scalar version is given as,

$$error_{MDJ} = \frac{\sigma_b^2 \Delta}{1 + \gamma}. \quad (B.7)$$

The error analysis for the Desroziers *et al.* diagnostic, eq.(20) is

$$E[(O - \hat{A})(\hat{A} - B)] = E[\mathbf{d}\mathbf{d}^T] (\mathbf{H}\tilde{\mathbf{B}}\mathbf{H}^T + \tilde{\mathbf{R}})^{-1} \mathbf{H}\tilde{\mathbf{B}}\mathbf{H}^T - \mathbf{H}\tilde{\mathbf{B}}\mathbf{H}^T (\mathbf{H}\tilde{\mathbf{B}}\mathbf{H}^T + \tilde{\mathbf{R}})^{-1} \mathbf{H}\tilde{\mathbf{B}}\mathbf{H}^T \\ = \mathbf{H}\tilde{\mathbf{B}}\mathbf{H}^T - \mathbf{H}\tilde{\mathbf{B}}\mathbf{H}^T (\mathbf{H}\tilde{\mathbf{B}}\mathbf{H}^T + \tilde{\mathbf{R}})^{-1} \mathbf{H}\tilde{\mathbf{B}}\mathbf{H}^T + \mathbf{error}_D, \quad (B.8)$$

where the \mathbf{error}_D is given as

$$\mathbf{error}_D = \{\mathbf{I} - \mathbf{H}\tilde{\mathbf{B}}\mathbf{H}^T (\mathbf{H}\tilde{\mathbf{B}}\mathbf{H}^T + \tilde{\mathbf{R}})^{-1}\} \Delta (\mathbf{H}\tilde{\mathbf{B}}\mathbf{H}^T) = \tilde{\mathbf{R}} (\mathbf{H}\tilde{\mathbf{B}}\mathbf{H}^T + \tilde{\mathbf{R}})^{-1} \Delta (\mathbf{H}\tilde{\mathbf{B}}\mathbf{H}^T). \quad (B.9)$$

Again only one error term that is similar to the MDJ diagnostic, and its scalar version is given as,

$$error_D = \frac{\sigma_o^2 \Delta}{1 + \gamma}. \quad (B.10)$$

Error analysis for the diagnostics using passive observations can also be derived. For the passive MDJ diagnostics, eq.(21), we have similarly to (B.5),

$$E[(\hat{A} - B)_c (\hat{A} - B)_c] = \mathbf{H}_c \tilde{\mathbf{K}} E[\mathbf{d}\mathbf{d}^T] \tilde{\mathbf{K}}\mathbf{H}_c^T \\ = \mathbf{H}_c \tilde{\mathbf{B}}\mathbf{H}^T (\mathbf{H}\tilde{\mathbf{B}}\mathbf{H}^T + \tilde{\mathbf{R}})^{-1} \mathbf{H}\tilde{\mathbf{B}}\mathbf{H}_c^T + \mathbf{error}_{MDJ_passive} \quad (B.11)$$

with a single error term given as,

$$\mathbf{error}_{MDJ_passive} = \mathbf{H}_c \tilde{\mathbf{B}}\mathbf{H}^T (\mathbf{H}\tilde{\mathbf{B}}\mathbf{H}^T + \tilde{\mathbf{R}})^{-1} \Delta (\mathbf{H}\tilde{\mathbf{B}}\mathbf{H}_c^T). \quad (B.12)$$

To express a scalar version of this equation we need to account for the background error correlation ρ between the active observation location and the passive observation location, and thus expressed as,

$$error_{MDJ_passive} = \frac{\rho^2 \sigma_b^2 \Delta}{1 + \gamma}. \quad (B.13)$$

Finally, the fundamental diagnostic of cross-validation eq.(7) does not depend explicitly on the innovation covariance consistency. However, attaining its true minimum by tuning only γ and L_c as would suggest eq.(22), does introduce some innovation in-consistency, which all other optimal diagnostics eq.(18-21) has to account for.

References

1. Ménard, R., and A. Robichaud. The chemistry-forecast system at the Meteorological Service of Canada. In *ECMWF Seminar Proceedings on Global Earth-System Monitoring*, September 5–9, 2005, Reading, UK, 297–308.

2. Robichaud, A. and R. Ménard. Multi-year objective analysis of warm season ground-level ozone and PM2.5 over North-America using real-time observations and Canadian operational air quality models. *Atmos. Chem. Phys.* **2014**, 14:1769-1800. DOI: 10.5194/acp-14-1769-2014.

3. Robichaud, A., R. Ménard, Y. Zaitseva, and D. Anselmo. Multi-pollutant surface objective analyses and mapping of air quality health index over North America. *Air Qual. Atmos. Health.* **2016**, DOI 10.1007/s11869-015-0385-9.

4. Moran, M.D., S. Ménard, R. Pavlovic, D. Anselmo, S. Antonopoulos, A. Robichaud, S. Gravel, P.A. Makar, W. Gong, C. Stroud, J. Zhang, Q. Zheng, H. Landry, P.A. Beaulieu, S. Gilbert, J. Chen, and A. Kallaur. Recent advances in Canada’s national operational air quality forecasting system. 32nd NATO-SPS ITM, Utrecht, NL, 7-11 May **2012**.

5. Ménard, R. and M. Deshaies-Jacques. Evaluation of analysis by cross-validation. Part I: Using verification metrics. *Atmosphere*, 2017, 8, x

6. Hollingsworth, A. and P. Lönnberg. The verification of objective analyses: Diagnostics of analysis system performance. *Meteorol. Atmos. Phys.* 1989, 40, 3-27.

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

8. Ménard, R. Error covariance estimation methods based on analysis residuals: theoretical foundation and convergence properties derived from simplified observation networks. *Q. J. R. Meteorol. Soc.* 2016, 142, 257-273, DOI:10.1002/qj.2650.

9. Daley, R.. The lagged-innovation covariance: A performance diagnostic for atmospheric data assimilation. *Mon. Wea. Rev.* 1992, 120, 178-196.

10. Kailath, T.. An innovation approach to least-squares estimation. Part I: Linear filtering in additive white noise. *IEEE Transactions on Automatic Control*, 1968, AC-13:6, 646-655.

11. Daley, R. *Atmospheric Data Analysis*. Cambridge University Press: New York, 1991, 457 pp.

12. Marseille, G.-J., J. Barkmeijer, S. de Haan, and W. Verkley. Assessment and tuning of data assimilation systems using passive observations. *Q. J. R. Meteorol. Soc.*, 2016, 142, 3001-3014, DOI:10.1002/qj.2882.

13. Waller, J.A., S.L. Dance, and N.K. Nichols. Theoretical insight into diagnosing observation error correlations using observation-minus-background and observation-minus-analysis statistics. *Q. J. R. Meteorol. Soc.*, 2016, 142, 418-431, DOI:10.1002/qj.2661.

14. 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 A* 38:111–36. doi:10.1111/j.1600-0870.1986.tb00460.x

15. Caines, P.E.. *Linear Stochastic Systems*. John Wiley and Sons, New York, 1988, 874 pp.

16. Cohn, S.E.. The principle of energetic consistency in data assimilation. W. Lahoz et al. (eds), *Data Assimilation*, DOI 10.1007/978-3-74703-1_7, Springer-Verlag Berlin Heidelberg, 2010.

17. Ménard, R., M. Deshaies-Jacques, and N. Gasset. A comparison of correlation-length estimation methods for the objective analysis of surface pollutants at Environment and Climate Change Canada. *Journal of the Air & Waste Management Association*. 2016, 66:9, 874-895, DOI: 10.1080/10962247.2016.1177620.

18. Janjic, T., N. Bormann, M. Bocquet, J.A. Carton, S.E. Cohn, S.L. Dance, S.N. Losa, N.K. Nichols, R. Potthast, J.A. Waller, and P. Weston. On the representation error in data assimilation. *Q. J. R. Meteorol. Soc.*, 2017, DOI:10.1002/qj.3130.

19. Mitchell, H. L., and R. Daley. Discretization error and signal/error correlation in atmospheric data assimilation. (I). All scales resolved. *Tellus*, 1997, 49A, 32–53.

20. Mitchell, H. L., and R. Daley. Discretization error and signal/error correlation in atmospheric data assimilation. (II). The effect of unresolved scales. *Tellus*, 1997, 49A, 54–73.

21. Joiner, J., and A. da Silva. Efficient methods to assimilate remotely sensed data based on information content. *Q. J. R. Meteorol. Soc.* 1998, 124, 1669–1694.

22. Migliorini, S. On the quivalence between radiance and retrieval assimilation. *Mon. Wea. Rev.* 2012, 140, 258-265, DOI: 10.1175/MWR-D-10-05047.1
23. Chapnik, B., G. Desroziers, F. Rabier, and O. Talagrand. Properties and first application fo an error-statistics tuning method in variational assimilation. *Q. J. Meteorol. Soc.* 2005, 130, 2253-2275. DOI:10.1256/qj.03.26
24. Ménard, R. and M. Deshiaux-Jacques. Error covariance estimation methods based on analysis residuals and its application to air quality surface observation networks. *Air Pollution and Its Application XXV*, Eds. C. Mensink and G. Kallos; Springer International AG, 2017, DOI 10.1007
25. Skachko, S., Q. Errera, R. Ménard, Y. Christophe, and S. Chabrillat. Comparison of the ensemble Kalman filter and 4D-Var assimilation methods using a stratospheric tracer transport model. *Geosci. Model Dev.* **2014**, 7, 1451–1465, www.geosci-model-dev.net/7/1451/2014/ doi:10.5194/gmd-7-1451-2014
26. Efron, B. An introduction to bootstrap. Chapman & Hall: New York, 1993, 436 pp.