On the Complementary Role of Data Assimilation and Machine Learning: An Example Derived from Air Quality Analysis

Richard Ménard[1(✉)], Jean-François Cossette[2], and Martin Deshaies-Jacques[2]

[1] Air Quality Research Division, Environment and Climate Change Canada, Dorval, QC H9P 1J3, Canada
richard.menard@canada.ca

[2] Canadian Meteorological Center, Environment and Climate Change Canada, Dorval, QC H9P 1J3, Canada

Abstract. We present a new formulation of the error covariances that derives from ensembles of model simulations, which captures terrain-dependent error correlations, without the prohibitive cost of ensemble Kalman filtering. Error variances are obtained from innovation variances empirically related to concentrations using a large data set. We use a *k*-fold cross-validation approach to estimate the remaining parameters. We note that by minimizing the cross-validation cost function, we obtain the optimal parameters for an optimal Kalman gain. Combined with the innovation variance consistent with the sum of observation and background error variances in observation space, yield a scheme that estimates the true error statistics, thus minimizing the true analysis error. Overall, this yield a new error statistics formulation and estimation outperforms the older optimum interpolation scheme using isotropic covariances with optimized covariance parameters. Yet, the analysis scheme is computationally comparable to optimum interpolation and can be used in real-time operational applications. These new error statistics comes as data-driven models, were we use validation techniques that are common to machine learning. We argue that the error statistics could benefit from a machine learning approach, while the air quality model and analysis scheme derives from physics and statistics.

Keywords: Air quality analysis · Cross-validation · Data driven model of error covariance

1 Introduction

Data assimilation was originally developed from a need to specify the initial conditions of numerical weather prediction models [1] that otherwise would have little predictive skill due to unstable dynamics of the atmosphere. Imperfectly known and incorrect assumptions (e.g. no model error covariance) on the (input) error statistics used for data assimilation, is not as critical for numerical weather prediction as opposed to other

V. V. Krzhizhanovskaya et al. (Eds.): ICCS 2020, LNCS 12142, pp. 212–224, 2020.
https://doi.org/10.1007/978-3-030-50433-5_17

applications, due to the presence of the unstable subspace in the forecast error. Indeed, it has been argued that by confining the forecast error corrections to the unstable and neutral subspace, four-dimensional variational data assimilation can better perform than without this confinement [2, 3]. However, not all application areas of data assimilation have an unstable subspace. Atmospheric chemistry models, for example, are known to be "slaved" by the meteorology [4, 5] and it is known that the precise knowledge of the chemical observation and model error covariances has a rather strong impact on the performance of the (chemical) data assimilation results [6, 7]. The estimation of correct error statistics is important for a truly optimal chemical data assimilation system. This has been recognized from the very beginning of Kalman filtering data assimilation using neutral or dissipative models (e.g. [8, 9]).

With dissipative models, the verification of the forecast as a measure of analysis accuracy has limited value. However, the true analysis error can be evaluated using cross-validation [10, 11]. In cross-validation we produce analyses with a subset of observations and verify the analysis with the remaining observations. There is no need to conduct a forecast. Since the optimality of analyses (measured against independent observations) depends on having the correct (input) error statistics, we may view the problem of estimating the true error statistics as an inverse problem of analyses (measured by cross-validation). But as with most inverse problems, the minimization of the verification error alone is not sufficient to determine the correct input error statistics [12]. Additional criteria are needed, such as the matching of the covariance of innovation with the sum of the background error covariance in observation space with the observation error covariance, called the innovation covariance consistency [13]. This is where (even elementary concepts) of machine learning can become useful.

Atmospheric models are based on the laws of physics, and in our case on chemical laws as well; they induce our prior knowledge of the system. On the other hand, the innovations or observation-minus-model residuals, are quantities that contain information that is unexplained by the model, and this is where machine learning can have a complementary role to data assimilation. Complementary roles of machine learning and data assimilation were also developed in a form of Kalman filtering (known as the Parametric Kalman filter) that requires closure on the form and parameters or error covariances [14].

At Environment and Climate Change Canada (ECCC) we have developed an operational surface analysis of atmospheric pollutants since 2002 [15] that provide a complete and more accurate representation of the air quality atmospheric chemistry, which has been used in several health impact studies (e.g. [16, 17]). Although the optimization of (isotropic) covariance model parameters improves the analysis [18], little is known about the more realistic and accurate covariance models suitable for chemical data assimilation [19, 20].

Since air quality models are computationally demanding (compared to numerical weather prediction models) the use of (ensemble) Kalman filtering data assimilation to obtain non-homogeneous non-isotropic error covariances is more in the realm of research than operational purposes. Recently at ECCC we have been developing a simple approach to generate non-homogeneous non-isotropic error correlations near the surface that does not involve rerunning the air quality model, by simply using pre-existing model simulations over a period of a few months. Note that chemical

simulations are driven (slaved) by meteorological analyses. The error correlations that are obtained are not flow-dependent but are non-homogeneous non-isotropic and terrain-dependent, which account for a large fraction of the error correlation signal near the surface. We will presents some examples of these error correlations in Sect. 2.1. As with Kalman filtering localization of these ensembles is needed, and is obtained by minimizing the analysis error evaluated by cross-validation. This will be presented in Sect. 2.3. For error variances we use essentially an innovation-driven (data-driven) representation of the error variances as a function concentration, which is simple but somewhat appropriate for these fields. This will be discussed in Sect. 2.2. We show that this new analysis scheme is superior to the currently operational implementation using optimum interpolation with homogeneous isotropic correlation models [21]. We realize that the method we are using, is similar (yet much simpler) to some of the methods used in (simple) machine learning. We argue that a data-driven approach to model error covariances, and more sophisticated machine learning algorithms [22] could potentially improve those representations and be beneficial for truly optimizing an assimilation system.

2 Description of the Method

2.1 Model Representors

In our current operational version of the analysis of surface pollutants, we use homogeneous isotropic correlation models [21] with tuning of covariance parameters [15, 18] to optimize the system. In this new version, we obtain anisotropic error correlations from an ensemble of pre-existing air quality simulations (i.e. forecasts with no chemical data assimilation). The ensemble is in fact climatological, where the air quality simulations are collected over a period of two months. For each hour of the day, we thus have an ensemble of about 60 realizations over that time period that captures non-homogenous and non-isotropic correlations. As in the ensemble Kalman filter, there is a need for localization to avoid spurious correlations at large distances that infiltrate the analysis and thus significantly reduce its optimality. The idea behind using an ensemble of pre-existing model forecasts is that those error correlations will be able to capture stationary and terrain-dependent structures such as those induced by the proximity of water surfaces, mountain ranges, valleys, chemical sources, and so on, since these features are always present. This method does not capture the day-to-day variability (or flow dependence) of the error correlations like in an ensemble Kalman filter, but may capture the effect of prevailing winds for example.

In the example presented below, we computed the spatial correlations for each observation sites using a time series of the Canadian operational air quality model (GEM-MACH) output of PM2.5 at 21 UTC over a two-month period (July-August 2016) and applied a compact support correlation function (the 5^{th}-order piecewise rational function of Gaspari and Cohn [23]) as a method for localization.

Examples of spatial correlation of PM2.5 are presented in Fig. 1, with respect to an observation site in Toronto (Canada) and in Los Angeles (USA). We note that the correlation around Toronto (upper panel) is more or less oval, while strong anisotropic

structures extending along the coast and influenced by the presence of water and terrain are depicted with the correlation about Los Angeles (lower panel). Correlations at Winnipeg in the Canadian prairies over a flat terrain are nearly circular or isotropic (result not shown). Likewise, the spatial correlation with respect to Surrey (in the neighborhood of Vancouver, Canada) shows a minimum over the Coast Mountain range and over the Rockies with a maximum in between over the central interior plateau (results not shown). In general, the spatial correlation structures shows the presence of mountain ranges, valleys, and extended water surfaces. These correlation structures were obtained after localization using a Schur product with a compact support correlation function. The length-scale of the compact correlation model is estimated by cross-validation (see Sect. 2.4).

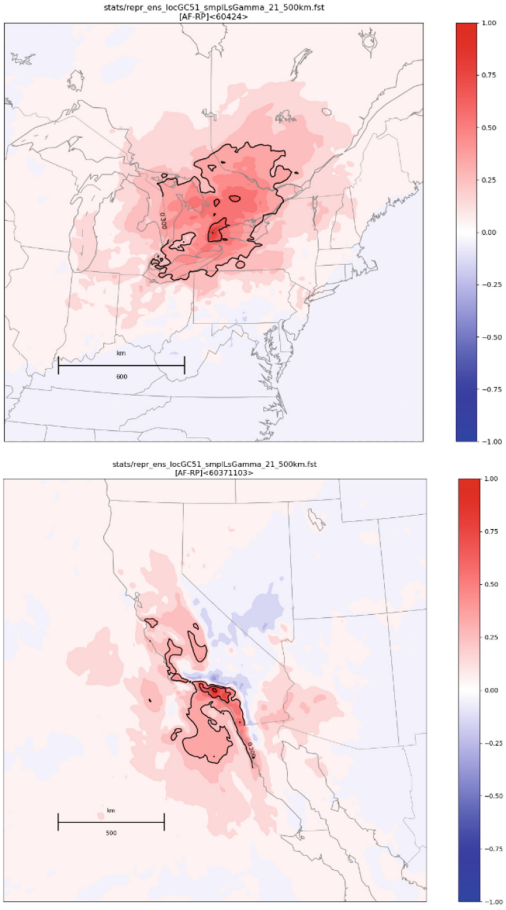

Fig. 1. Error correlation obtained for a time series of the air quality model GEM-MACH simulations valid at 21 UTC over a period of 2 months (July-August 2016), after localization using the compact support correlation function. Upper panel is the correlation with respect to Toronto (Canada) and lower panel with respect to Los Angeles (USA) stations. The solid black line depicts the correlation contour of 0.3.

2.2 Innovation Scatter Plot

The innovation (i.e. the observation-minus-model residual) variance has been plotted against the mean concentration for each station at a given local time (21 UTC) using a 2-months' time series to generate the statistics for each station. The variance of the innovation gives the sum of observation and model (i.e. background) error variances, but its partition into observation and model errors is yet unknown. Furthermore, the observation error is not simply the instrument error but should also include; errors due to the mismatch in scales being sampled in a single observation vs that of the model grid box, subgrid scale variability that may be captured by the observation but not by the model, and missing modeling processes, etc., that collectively we call representativeness error. The observation error is thus not well known and needs to be estimated. This will be done in the Sect. 2.4 using cross-validation and assuming that observation error variance is simply a fraction of the variance of the innovation.

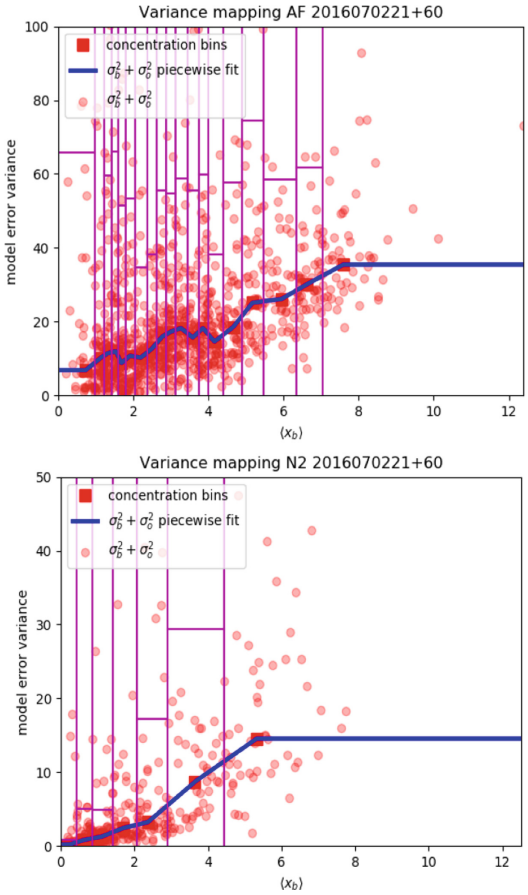

Fig. 2. Variance of the innovation as a function of the mean concentration, station by station. Upper panel is for $PM_{2.5}$ and lower panel for NO_2 (same time period as in Fig. 1). The red squares represent the median in each bin, and the horizontal line in each bin determines the third inter quartile of the distribution in a bin. (Color figure online)

The behavior as a function of concentration is different for $PM_{2.5}$ than for NO_2. We note that for NO_2 the innovation variance goes to zero when the concentration goes to zero, which is not the case for $PM_{2.5}$. The $PM_{2.5}$ is nearly linear, while NO_2 (especially at night - result not shown) has a shape of a quadratic for low concentrations and saturate at higher concentrations.

The fitting of the innovation variance as a function of concentration (as in Fig. 2) is important for two reasons. First, as it will become relevant in the following subsection, we need to have the sum of observation and model (background) error variances match the innovation variance - a property known as the innovation variance consistency. Secondly, after obtaining the portion of the innovation variance due to the model (background) error variance, we can then apply the relationship between background error variance with concentrations, to determine the background error variance at each model surface grid point (not only at the observation locations).

2.3 Estimation of the True Error Statistics in Observation Space

Under the assumptions of uncorrelated observation and background errors, it has been established that two necessary and sufficient conditions to estimate the true observation and background error covariances are [13]. One, is that the Kalman gain in observation space (i.e. **HK** where **H** is the observation operator and **K** the Kalman gain) is optimal in the sense that the true analysis error (in observation space) is minimum. The second condition is that the innovation covariance matches the sum of the background error covariance in observation space plus the observation error covariance, i.e. $\mathbf{HBH}^T + \mathbf{R}$ where **B** is the background error covariance and **R** the observation error covariance.

The fit of the innovation variance presented in Fig. 2 assures by construction that the sum of error variances $\sigma_o^2 + \sigma_b^2$ is consistent with innovation variance (not the innovation covariance), thus at least partly fulfilling the second condition above.

The first condition about the optimal Kalman gain is obtained by using cross-validation [10, 11]. As a way to illustrate this, we use a geometric interpretation where random variables y_1 , y_2 are represented in a Hilbert space whose inner product is defined as the covariance between the two random variables.

$$\langle y_1, y_2 \rangle := \mathrm{E}\left[\left(y_1 - \mathrm{E}(y_1) \right) \left(y_2 - \mathrm{E}(y_2) \right) \right], \tag{1}$$

where E is the mathematical expectation. In this framework, random variables form an Hilbert space. For example, uncorrelated random variables are represented as perpendicular "vectors", and the error variance as the norm squared of that "vector" (see [11] and references therein).

Figure 3 illustrates geometrically the cross-validation. Let the active observation y^o be illustrated as O in the Fig. 3, the prior or background y^b (illustrated as B), the analysis y^a (illustrated as A), and the independent (or passive) observation y^c (illustrated as O_c). The origin T corresponds to the truth. Note that although the illustration is made for a (scalar) random variable, the same principle holds for random vectors for each components. We assume that the background error is uncorrelated with observations errors, and that observations are spatially uncorrelated horizontally. Then, the background error, the active and the passive observation errors are uncorrelated with

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 axes ε^o and ε^b is the space where the analysis takes place, and is called the analysis plane. Since we define the analysis to be linear and unbiased, the analysis is a linear combination of the observation and background. Thus, in this illustration the analysis A lies on the line (B,O). The thick lines in Fig. 4 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, $\overline{(y^o - y^b)^2} := \langle (O-B), (O-B) \rangle = \sigma_o^2 + \sigma_b^2$. This is the usual statement that the innovation variance is the sum of background and observation error variances. The analysis is optimal $\hat{A}$ when the analysis error $||\varepsilon^a||^2 = \sigma_a^2$ is minimum, that is when the point A is closest to T. In this case the line $(T, \hat{A})$ is perpendicular to line (O,B). Since passive observations are never collocated with active observations, they have errors uncorrelated to active observations and background errors, and thus ε_c^o lies perpendicular to the analysis plane. So when the point A is closest to T it is also closest to O_c. Since the distance (O_c,A) squared corresponds to the variance of $(O_c - A)_{CV}$ evaluated in cross-validation, we conclude that when Var $(O_c - A)_{CV}$ is minimum, the analysis is optimal (i.e. minimal error), $A = \hat{A}$ [11].

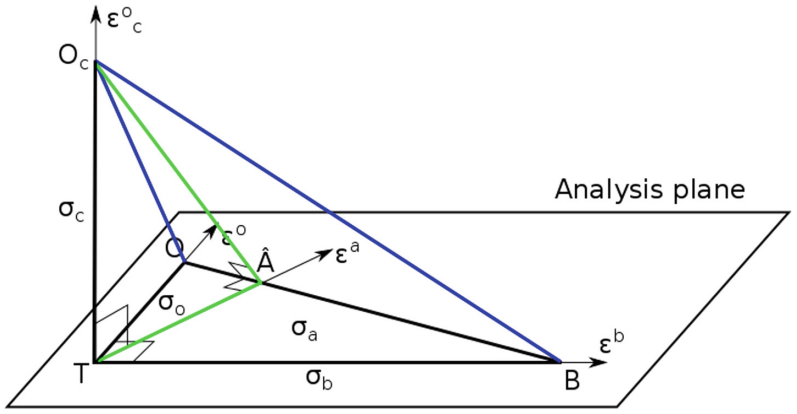

Fig. 3. Geometric representation (i.e. Hilbert space) of a scalar analysis with cross-validation (reproduced from [24] Figure 37.2 b)).

2.4 Optimization by Cross-Validation

An evaluation by cross-validation is performed to evaluate the quality of the analysis. The *k*-fold cross-validation methodology consists of separating the data set of observations into *k* equal size spatially random-distributed subsets and using each subset to evaluate an analysis while using the remaining observations in the *k*-1 subsets to

compute the analysis. Here we use $k = 3$ [16, 17]. The separation into spatially random-distributed subsets has been achieved by selecting one station over three in an ordered station ID number list. An illustration of the selection method for the PM2.5 surface monitoring stations is depicted in Fig. 4, below. Another method which makes this selection rigorous is the application Hilbert curves [25].

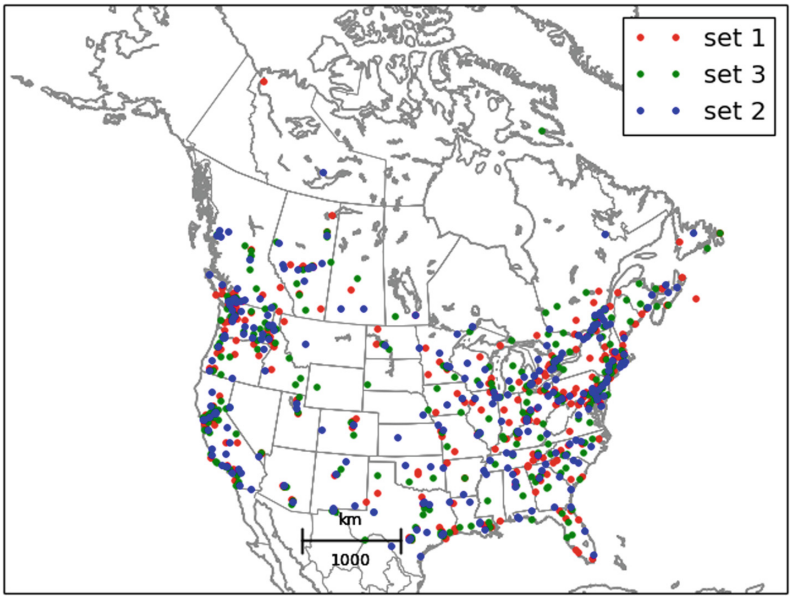

Fig. 4. Spatial distribution of three subsets of $PM_{2.5}$ observation sites (reproduced from [10] Figure 1)

Let $O_j(n;t)$ be the concentration measured at a station j at the local hour t, on day n. First, we calculate the verification statistics for each station at a given local time using an ensemble of days $\{n = 1, \ldots, N_{days}\}$. The average at a station j at the local time t, is simply

$$\bar{O}_j(t) \;=\; \frac{1}{N_{days}} \sum_{n=1}^{N_{days}} O_j(n;t). \tag{2}$$

The verification statistics are often defined over a region. For example, the mean concentration variance over an ensemble i, i.e. $\{O^i\}$, of stations, for a total number of N_s stations, is defined as

$$\mathrm{var}\big(O^i(t)\big) \;=\; \frac{1}{N_s} \sum_{j \in \{O^i\}} \left(\frac{1}{N_{days} - 1} \sum_{n=1}^{N_{days}} \big(O_j(n;t) - \bar{O}_j(t)\big)^2 \right). \tag{3}$$

In general the ensemble $\{O^i\}$ (or simply denoted by O^i) can either be an:

- ensemble over all stations in the whole domain
- ensemble over a region (or subdomain)
- ensemble over all passive stations (i.e. stations not used in the analysis)

or any variants or combination thereof. In this document, there is only a single domain consisting of the continental USA and Canada.

For cross-validation, the analyses are evaluated against passive observations, i.e. observations not used to construct the analyses. We recall that passive observation sites are never collocated with the active stations (stations used to construct the analyses). Specifically, the ensemble of observations (for each local time) is split into three disjoint subsets O^1, O^2, O^3, and we denote the cross-validation analysis by $A^{[1]} = A^{[1]}(O^2, O^3)$ as an analysis that uses O^2, O^3 and excludes the subset O^1. The interpolated analysis at the passive station $j \in O^1$ will be denoted by

$$A_j^{[1]} = A_j^{[1]}(O^2, O^3). \tag{4}$$

The cross-validation variance statistics are then given by the average over the 3 subsets $i = 1, 2, 3$, i.e.

$$\mathrm{var}(O-A)_{CV} = \frac{1}{3}\left\{\mathrm{var}\left(O^1 - A^{[1]}\right) + \mathrm{var}\left(O^2 - A^{[2]}\right) + \mathrm{var}\left(O^3 - A^{[3]}\right)\right\}, \tag{5}$$

where the statistics of each passive subset i are calculated by an average over all passive stations in the given subset. Specifically we have,

$$\begin{aligned}\mathrm{var}\left(O^i - A^{[i]}\right) &= \frac{1}{N_s}\sum_{j\in O^i}\mathrm{var}\left(O_j^i - A_j^{[i]}\right) \\ &= \frac{1}{N_s}\sum_{j\in O^i}\left\{\frac{1}{N_{days}-1}\sum_{n=1}^{N_{days}}\left[\left(O_j^i(n) - A_j^{[i]}(n)\right) - \overline{\left(O_j^i(n) - A_j^{[i]}(n)\right)}\right]^2\right\}.\end{aligned} \tag{6}$$

Note that a cross-validation statistic is evaluated for each local time and we have omitted the variable t, to keep the notation simple.

In our context where we enforce innovation variance consistency through the innovation variance fitting (see Sect. 2.2), we are left with only 2 parameters to estimate: 1) the ratio of observation error variance to background error variance σ_o^2/σ_b^2, and 2) the compact support correlation length-scale L_s. These parameters are estimated by minimizing $\mathrm{var}(O-A)_{CV}$ and thus result in the end in a nearly optimal analysis.

In Fig. 5, we plotted $\mathrm{var}(O-A)_{CV}$ for different values of σ_o^2/σ_b^2 and L_s. We find a single minimum of the fit of the analysis to independent (or passive observations) for an error variance ratio of 1.5 and for a compact support correlation length of 300 km, used to localized the raw model correlations.

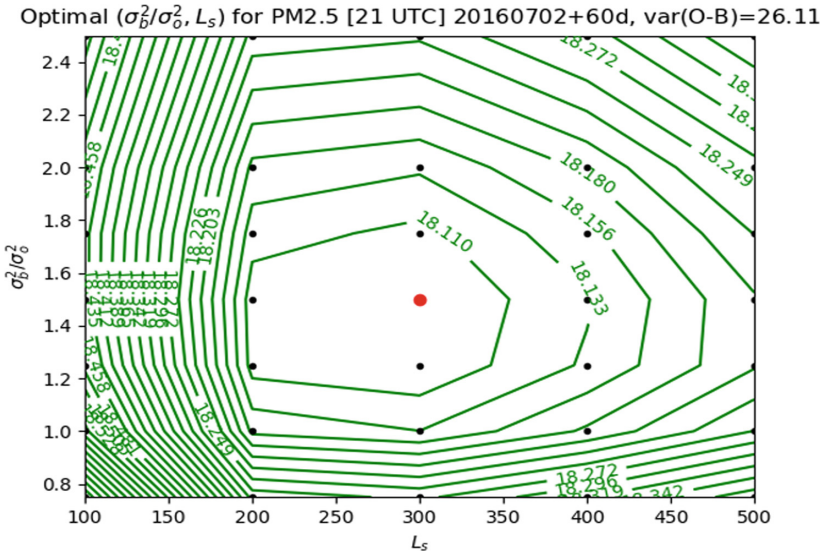

Fig. 5. Contours of the cross-validation of independent observation-minus-analysis (at the passive observation sites), as a function of compact support correlation length L_s and the ratio of model error variance to observation error variance.

3 Comparison with the Operational Analysis

Using these optimal parameter values and the modeling based in the innovation variance as function of concentration and with model output statistics to construct the spatial correlation structures, we evaluate the new analysis (i.e. Av2). We then compare it against the old analysis scheme (i.e. Av1) which is using homogenous isotropic correlation functions and χ^2-optimized error statistics [15]. The result evaluated by cross-validation is presented in the Fig. 6 below.

We observe a sensitivity to the error statistics used to generate the analysis, with superior analyses when the modeling is based on our methodology using the data rather than using some specified isotropic models.

It is by letting the data itself (model output and observations) provide the modeling elements of the observation and background error covariances that we arrive at an improved analysis. Thus, we thus argue that data-driven modeling of the observation and background error covariances plays a complementary role to data assimilation, resulting in a nearly optimal system.

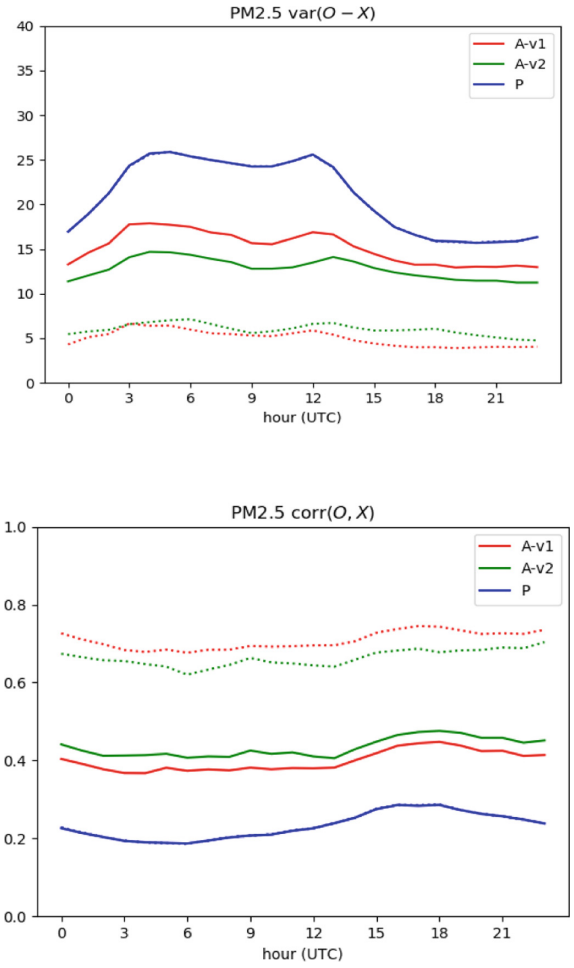

Fig. 6. Verification of the PM2.5 analysis against passive observations using cross-validation for Av1 (old scheme) and Av2 (new scheme). The solid line (green and red) uses independent observations, while the dotted lines are the statistics using the same observations as those used to construct the analysis. The solid blue line represent the verification of the model (i.e. no analysis). The upper panel displays the variance and lower panel the correlation. (Color figure online)

References

1. Daley, R.: Atmospheric Data Analysis. Cambridge University Press, New York (1991). 455 p.
2. Trevisan, A., D'Isidoro, M., Talagrand, O.: Four-dimensional variational assimilation in the unstable subspace and the optimal subspace dimension. Q. J. Roy. Meteorol. Soc. **136**, 487–496 (2010). https://doi.org/10.1002/qj.571

3. Grudzien, C., Carrassi, A., Bocquet, M.: Asymptotic forecast uncertainty and the unstable subspace in presence of additive model error. SIAM/ASA J. Uncertainty Quantification **6**(4), 1335–1363 (2018). https://doi.org/10.1137/17M114073X
4. Lahoz, W., Errera, Q.: Constituent Assimilation. In: Lahoz, W., Khattatov, B., Menard, R. (eds.) Data Assimilation, pp. 449–490. Springer, Heidelberg (2010). https://doi.org/10.1007/978-3-540-74703-1_18
5. Ménard, R., et al.: Coupled stratospheric chemistry-meteorology data assimilation. Part I: Physical background and coupled modeling aspects. Atmosphere **11**, 150 (2020). https://doi.org/10.3390/atmos11020150
6. Errera, Q., Ménard, R.: Technical Note: Spectral representation of spatial correlations in variational assimilation with grid point models and application to the Belgian Assimilation System for Chemical Observations (BASCOE). Atmos. Chem. Phys. **12**, 10015–10031 (2012). https://doi.org/10.5194/acp-12-10015-2012
7. Ménard, R., et al..: Coupled stratospheric chemistry-meteorology data assimilation. Part II: Weak and strong coupling. Atmosphere **10**(12), 798 (2019). https://doi.org/10.3390/atmos10120798
8. Daley, R.: The lagged innovation covariance: a performance diagnostic for atmospheric data assimilation. Mon. Wea. Rev. **120**, 178–196 (1992).
9. Daley, R.: The effect of serially correlated observation and model error on atmospheric data assimilation. Mon. Wea. Rev. **120**, 164–177 (1992).
10. Ménard, R., Deshaies-Jacques, M.: Evaluation of analysis by cross-validation. Part I: Using verification metrics. Atmosphere **9**(3) (2018). https://doi.org/10.3390/atmos9030086
11. Ménard, R., Deshaies-Jacques, M.: Evaluation of analysis by cross-validation. Part II: Diagnostic and optimization of analysis error covariance. Atmosphere **9**(2), 70 (2018). https://doi.org/10.3390/atmos9020070
12. Talagrand, O.: A posteriori verification of analysis and assimilation algorithms. In: Proceedings of the ECMWF Workshop on Diagnosis of Data Assimilation Systems, 2–4 November 1999, pp. 17–28. Reading, UK (1999)
13. Ménard, R.: Error covariance estimation methods based on analysis residuals: theoretical foundation and convergence properties derived from simplified observation networks. Q. J. Roy. Meteorol. Soc. **142,** 257–273 (2016). https://doi.org/10.1002/qj.2650. http://onlinelibrary.wiley.com/doi/10.1002/qj.2650/full
14. Pannekoucke, O., Fablet, R.: PDE-NetGen 1.0: from symbolic PDE representations of physical processes to trainable neural network representations. Geoscientific Model Development Discussion (2020). https://doi.org/10.5194/gmd-2020-35
15. Robichaud, A., Ménard, R.: 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. **14**, 1769–1800 (2014). https://doi.org/10.5194/acp-14-1769-201
16. Crouze, D.L., et al.: Ambient PM2.5, O3, and NO2 exposures and association with mortality over 16 years of follow-up in the Canadian Census Health and Environment Cohort (CanCHEC). Environ. Health Perspect. **123,** 1180–1186. https://doi.org/10.1289/ehp.1409276. Accessed 2 Nov 2015
17. To, T., et al.: Early life exposure to air pollution and incidence of childhood asthma, allergic rhinitis and eczema: Eur. Respir. J. pii, 1900913 (2019). https://doi.org/10.1183/13993003.00913-2019
18. Ménard, R., Deshaies-Jacques, M., Gasset, N.: A comparison of correlation-length estimation methods for the objective analysis of surface pollutants at Environment and Climate Change Canada. J. Air Waste Manag. Assoc. **66**(9), 874–895 (2016). https://doi.org/10.1080/10962247.2016.1177620

19. Constantinescu, E.M., Chai, T., Sandu, A., Carmichael, G.R.: Autoregressive models of background errors for chemical data assimilation. J. Geophys. Res. **112**, D12309 (2007). https://doi.org/10.1029/2006JD008103
20. Singh, K., Jardak, M., Sandu, A., Bowman, K., Lee, M., Jones, D.: Construction of non-diagonal background error covariance matrices for global chemical data assimilation. Geosci. Model Dev. **4**, 299–316 (2011). www.geosci-model-dev.net/4/299/2011, https://doi.org/10.5194/gmd-4-299-2011
21. Ménard, R., Robichaud, A.: The chemistry-forecast system at the Meteorological Service of Canada. In: The ECMWF Seminar Proceedings on Global Earth-System Monitoring, Reading, UK, 5–9 September 2005, pp. 297–308 (2005)
22. Rasmussen, C.E., Williams, C.K.: Gaussian Processes for Machine Learning. MIT Press, Cambridge (2006). 248 p.
23. Gaspari, G., Cohn, S.E.: Construction of correlation functions in two and three dimensions. Q. J. Roy. Meteorol. Soc. **125**, 723–757 (1999)
24. Ménard, R., Deshaies-Jacques, M.: Evaluation of air quality maps using cross-validation: Metrics, diagnostics and optimization. In: Mensink, C., Gong, W., Hakami, A. (eds.) Air Pollution Modelling and Its Application XXVI, pp. 237–242. Springer Proceedings in Complexity (2020). https://doi.org/10.1007/978-3-030-22055-6_37
25. De Pondeca, M.S.F.V., Park, S.-Y., Purser, J., DiMego, G.: Applications of Hilbert curves to the selection of subsets of spatially inhomogeneous observational data for cross-validation and to the construction of super-observations. Preprints, AGU Fall Meeting, San Francisco, CA, Amer. Geophys. Union, A31A-0868 (2006)