

Ozone ensemble forecasts:

2. A Kalman filter predictor bias correction

Luca Delle Monache,^{1,2} Thomas Nipen,¹ Xingxiu Deng,^{1,3} Yongmei Zhou,^{1,4} and Roland Stull¹

Received 1 June 2005; revised 11 August 2005; accepted 29 August 2005; published 7 March 2006.

[1] The Kalman filter (KF) is a recursive algorithm to estimate a signal from noisy measurements. In this study it is tested in predictor mode, to postprocess ozone forecasts to remove systematic errors. The recent past forecasts and observations are used by the KF to estimate the future bias. This bias correction is calculated separately for, and applied to, 12 different air quality (AQ) forecasts for the period 11–15 August 2004, over five monitoring stations in the Lower Fraser Valley, British Columbia, Canada, a population center in a complex coastal mountain setting. The 12 AQ forecasts are obtained by driving an AQ Model (CMAQ) with two mesoscale meteorological models (each run at two resolutions) and for three emission scenarios (Delle Monache et al., 2006). From the 12 KF AQ forecasts an ensemble mean is calculated (EK). This ensemble mean is also KF bias corrected, resulting in a high-quality estimate (KEK) of the short-term (1- to 2-day) ozone forecast. The Kalman filter predictor bias-corrected ensemble forecasts have better forecast skill than the raw forecasts for the locations and days used here. The corrected forecasts are improved for correlation, gross error, root mean square error, and unpaired peak prediction accuracy. KEK is the best and EK is the second best forecast overall when compared with the other 12 forecasts. The reason for the success of EK and KEK is that both the systematic and unsystematic errors are reduced, the first by Kalman filtering and the second by ensemble averaging.

Citation: Delle Monache, L., T. Nipen, X. Deng, Y. Zhou, and R. Stull (2006), Ozone ensemble forecasts: 2. A Kalman filter predictor bias correction, *J. Geophys. Res.*, *111*, D05308, doi:10.1029/2005JD006311.

1. Introduction

[2] The first part of this study [Delle Monache et al., 2006, hereinafter referred to as DM1] presented a new Ozone Ensemble Forecast System (OEFS), composed of 12 forecasts created using four different meteorological inputs and three different emission scenarios. The meteorological fields were obtained by running two mesoscale numerical weather prediction (NWP) models over two nested domains with 12 and 4 km horizontal grid spacing. The emission scenarios were a control run, a run with 50% more NO_x emissions, and a run with 50% less. The 12 combinations of the meteorological and emission fields were used to drive the U.S. Environmental Protection Agency (EPA) Models-3/Community Multiscale Air Quality

Model (CMAQ) Chemistry Transport Model (CTM) [Byun and Ching, 1999].

[3] This OEFS has been tested for the period 11–15 August 2004 using data from five stations across the Lower Fraser Valley (LFV), British Columbia (BC), Canada, a region where the ozone modeling is particularly challenging because of the complex coastal mountain setting. The main finding of DM1 is that, for the locations and days used to test this new OEFS, the ensemble mean is the most skillful forecast when tested against the observations, and compared to any other ensemble member.

[4] The results in DM1 show that all the forecasts have systematic errors (e.g., nighttime over prediction). This is a problem common to all CTMs [Russell and Dennis, 2000]. In this paper the Kalman filter predictor (KFP) postprocessing bias correction method [Bozic, 1994] has been applied to each ozone forecast (the 12 ensemble members and the ensemble mean) to improve the individual forecast skill for all sites where ozone observations are available. The KFP correction is an automatic postprocessing method that uses the recent past observations and forecasts to estimate the model bias in the future forecast, where bias here is defined as the “difference of the central location of the forecasts and the observations” [Jolliffe and Stephenson, 2003]. This estimate can then be used to correct the raw model prediction. It is a recursive, adaptive method that takes into

¹Atmospheric Science Programme, Department of Earth and Ocean Sciences, University of British Columbia, Vancouver, British Columbia, Canada.

²Now at Lawrence Livermore National Laboratory, Livermore, California, USA.

³Now at Meteorological Service of Canada, Environment Canada, Montreal, Quebec, Canada.

⁴Now at Meteorological Service of Canada, Environment Canada, Edmonton, Alberta, Canada.

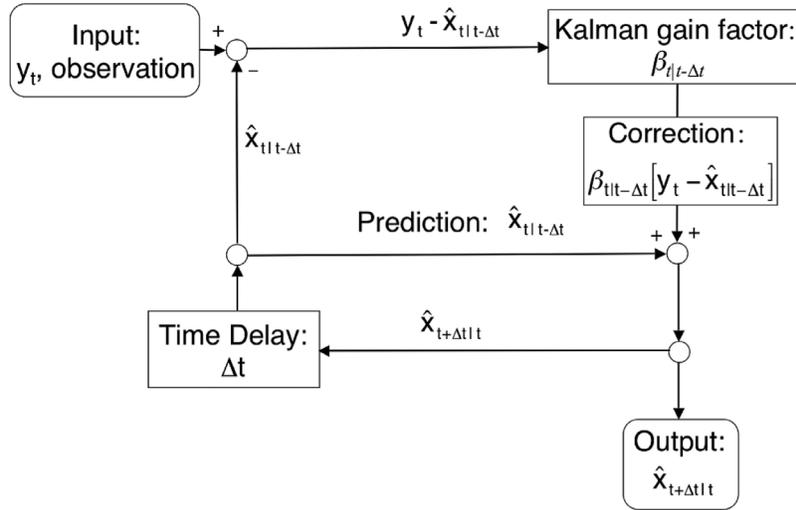


Figure 1. Flow diagram of the Kalman filter bias estimator. It uses a predictor corrector approach, starting with the previous estimate of the bias ($\hat{x}_{t|t-\Delta t}$) and correcting it by a fraction (β) of difference between the previous bias estimate and previous observed forecast error (y_t) to estimate the future bias ($\hat{x}_{t+\Delta t|t}$).

account the time variation of forecast error at a specific location.

[5] Details of the Kalman algorithm are given in section 2. Section 3 describes the experiment and methodology. In section 4, the performance of the raw (i.e., not corrected), the KFP bias-corrected forecasts, the ensemble mean of the KFP bias-corrected forecasts (EK, is a linear average of the KFP bias-corrected ensemble member predicted hourly concentrations), and the KFP bias-corrected EK (KEK) are compared using the same data set and statistical parameters as in DM1. Moreover, EK and KEK performances are compared with two other bias-correction methods; namely, the additive and multiplicative methods (section 5). In section 6 those results are discussed and conclusions are drawn.

2. Kalman Filter Predictor Bias Correction

[6] The Kalman filter (KF) is a recursive algorithm to estimate a signal from noisy measurements. It has been mainly used in data assimilation schemes to improve the accuracy of the initial conditions for both NWP [e.g., Burgers *et al.*, 1998; Hamill and Snyder, 2000; Houtekamer and Mitchell, 2001; Houtekamer *et al.*, 2005] and air quality (AQ) forecasts [e.g., van Loon *et al.*, 2000; Segers *et al.*, 2005]. The KF has also been used for NWP model forecasts as a predictor bias correction method during postprocessing of short-term weather forecasts [Homleid, 1995; Roeger *et al.*, 2003], an approach that is extended here for AQ forecasts (i.e., ozone).

[7] In a postprocessing predictor bias correction method, the information (i.e., recent past forecasts and observations) is used to revise the estimate of the current raw forecast. Previous bias values are used as input to KF. The filter estimates the systematic component of the forecast errors, or bias, which is often present in AQ forecasts as shown in DM1 and as reported in the literature [e.g., Russell and Dennis, 2000]. Once the future bias has been estimated, it can be removed from the forecast to produce an improved

forecast. Such a corrected forecast should be statistically more accurate in a least-squares sense.

[8] The KF models the true (unknown) forecast bias x_t at time t , by the previous true bias plus a white noise η term [Bozic, 1994]:

$$x_{t|t-\Delta t} = x_{t-\Delta t|t-2\Delta t} + \eta_{t-\Delta t} \quad (1)$$

where $\eta_{t-\Delta t}$ is assumed uncorrelated in time, and is normally distributed with zero-mean and variance σ_{η}^2 , Δt is a time lag (see Figure 1), and $t|t - \Delta t$ means that the value of the variable at time t depends on values at time $t - \Delta t$. Because of unresolved terrain features, numerical noise, lack of accuracy in the physical parameterizations, and errors in the observations themselves, the KF approach further assumes that the forecast error y_t (forecast minus observation at time t) is corrupted from truth by a random error term ε_t :

$$y_t = x_t + \varepsilon_t = x_{t-\Delta t} + \eta_{t-\Delta t} + \varepsilon_t \quad (2)$$

where ε_t is assumed uncorrelated in time and normally distributed with zero-mean and variance σ_{ε}^2 . Thus y_t includes both the systematic bias plus random errors.

[9] Kalman [1960] showed that the optimal recursive predictor of x_t (derived by minimizing the expected mean square error) can be written as a combination of the previous bias estimate and the previous forecast error:

$$\hat{x}_{t+\Delta t|t} = \hat{x}_{t|t-\Delta t} + \beta_{t|t-\Delta t}(y_t - \hat{x}_{t|t-\Delta t}) \quad (3)$$

where the hat (^) indicates the estimate.

[10] The weighting factor β , called Kalman gain, can be calculated from:

$$\beta_{t|t-\Delta t} = \frac{p_{t-\Delta t|t-2\Delta t} + \sigma_{\eta}^2}{(p_{t-\Delta t|t-2\Delta t} + \sigma_{\eta}^2 + \sigma_{\varepsilon}^2)} \quad (4)$$

where p is the expected mean square error, which can be computed as follows:

$$p_{t|t-\Delta t} = \left(p_{t-\Delta t|t-2\Delta t} + \sigma_{\eta}^2 \right) \left(1 - \beta_{t|t-\Delta t} \right) \quad (5)$$

[11] It can be shown [Dempster *et al.*, 1977] that the time series

$$z_t = y_{t+\Delta t} - y_t = \eta_t + \varepsilon_{t+\Delta t} - \varepsilon_t \quad (6)$$

has variance

$$\sigma_z^2 = \sigma_{\eta}^2 + 2\sigma_{\varepsilon}^2 \quad (7)$$

Assuming $r = \sigma_{\eta}^2 / \sigma_{\varepsilon}^2$, (7) become:

$$\sigma_z^2 = r\sigma_{\varepsilon}^2 + 2\sigma_{\varepsilon}^2 = (2 + r)\sigma_{\varepsilon}^2 \quad (8)$$

[12] σ_{ε}^2 (which is a time-varying quantity) can be estimated with the Kalman algorithm itself (i.e., by substituting \hat{x} with σ_{ε}^2 in equation (3)) in combination with (8). Further details on the filter implementation are given in Appendix A.

[13] Since here a time lag of $\Delta t = 24$ hours is used, today's forecast bias is estimated using yesterday's bias, which in turn was estimated using the day-before-yesterday's bias, and so on. Figure 1 shows the flow diagram of the KF algorithm. The difference between today's forecast error (y_t) and the portion of today's bias that was estimated yesterday ($\hat{x}_{t|t-\Delta t}$), is weighted by the Kalman gain to give the correction that was "learned" from previous errors. This correction is applied to yesterday's estimate of today's bias ($\hat{x}_{t|t-\Delta t}$) to produce today's estimate of the bias for tomorrow ($\hat{x}_{t+\Delta t|t}$). Thus real-time AQ forecasts are possible by taking the raw forecast from a model such as CMAQ, and correcting it with the bias forecast from KF.

[14] The KF algorithm will quickly and optimally converge (after few time step (Δt) iterations) for any reasonable initial estimate of p_0 and β_0 . However, the filter performance is sensitive to the ratio $\sigma_{\eta}^2 / \sigma_{\varepsilon}^2$. If the ratio is too high, the filter will put excessive confidence on the past forecasts, and will therefore fail to remove any error. On the other hand, if the ratio is too low, the filter will be unable to respond to changes in bias. Thus there exists an optimal value for the ratio that is given by the climatology of the forecast region, which can be estimated by evaluating the filter performance in different situations with different meteorology and different AQ scenarios (not only for AQ episodes).

[15] The data set presented in this study is not extended enough to compute an optimal ratio value that can also be used for a wide range of AQ scenarios (i.e., nonepisodic). A ratio value of 0.01 is used in this study. This is the value from previous studies where the KF was used to bias-correct weather forecasts in the steep mountains of BC, Canada [Roeger *et al.*, 2003], and close to the optimal value of 0.06 found by Homleid [1995]. With the availability of a longer data set (a full month or season), including both ozone forecasts and observations with a broader variability than just the AQ episode presented here, a different optimal value may result.

[16] A period of 2 days (9–10 August 2004) is used here in order to train the Kalman gain coefficients. Kalman corrections are then applied to the data for the subsequent 5 days (11–15 August 2004), during which time the KF continues training itself. Also, the filter algorithm is run on data for each hour of the day, using only values from previous days at the same hour of the day (corresponding to a $\Delta t = 24$ hours time delay in Figure 1). In this way, a given hour is corrected using only the past forecasts and observations at that same hour. This is to take into account the time-varying behavior the bias may have at different times of the day (e.g., different ozone reactions during daytime versus nighttime). Thus we compute and save different Kalman coefficients and variances for each hour of the day.

[17] When observations are missing for an hour, the filter uses the last known bias for that same hour from an earlier day. In some cases, however, the true bias changes considerably in such a time period, causing the algorithm to use incorrect, old values. This creates spikes in the Kalman coefficients that can be smoothed by applying the following low-pass filter twice:

$$x_t = \frac{1}{2}\hat{x}_t + \frac{1}{4}[\hat{x}_{t-1} + \hat{x}_{t+1}] \quad (9)$$

Since the bias correction is additive, the Kalman-filtered ozone concentrations were given a lower bound of 0 ppbv, in order to avoid negative forecast values.

[18] In summary, the Kalman filter predictor corrector approach is (1) linear, (2) adaptive, (3) recursive, and (4) optimal. Namely, it predicts the future bias as equal to the old bias plus uncertainty, but corrected by a linear function of the difference between the previous prediction and the verifying bias. Contrast this to a neural network approach, which is nonlinear [e.g., Cannon and Lord, 2000].

[19] Contrary to a neural network approach that requires a long training period and then behaves in a static manner, the KF approach adapts its coefficients during each time step. Advantages are a much shorter training period, and an ability to adapt to changing synoptic conditions, changing seasons, and even changing weather forecast models or AQ models. A disadvantage is that it is less likely to predict extreme bias events; namely, it is unable to anticipate a large bias when all biases for the past few days have been smaller.

[20] It is recursive because values of the KF coefficients at any one time step depend on the values at the previous time step. It is optimal in a least-square sense. Finally it is easy to implement and fast running on the computer, requiring storage of a handful of the KF coefficients for each AQ site for each forecast hour.

3. Method

3.1. Experiments

[21] Because each AQ ensemble member is a forecast based on a different meteorological model, different grid resolution, or different initial chemistry, it is anticipated that each forecast will have a different bias. Some of these biases could be quite large. Also, this bias could vary

depending on the hour of the day. To correct the individual AQ forecasts, we apply a separate Kalman filter for each ensemble member, for each hour. Individual Kalman-corrected AQ forecasts are denoted by K.

[22] Next, if we ensemble (E) average all of the Kalman-corrected (K) forecasts for any hour, then the result is denoted by EK. This ensemble average could have a small residual bias, because the bias corrections that were applied to the individual members were only estimates of future biases (as is the case for true AQ forecasts, not for ex-post-facto calculations of actual biases). Hence, as a final fine tuning, one can Kalman filter (K) the ensemble average (EK), with the result denoted by KEK.

[23] Experiments are performed here for the same suite of case study days, NWP models, and initial chemistry, as are described in DM1, but this study tests and compares the performance of the raw, K, EK, and KEK forecasts. During the 5-day period of 11–15 August 2004 used in this case study, there were typical conditions that lead to high ground-level ozone concentrations in the LVF. Those conditions are associated with a northward progressing low-level thermal trough from Washington State, associated with a stationary upper-level ridge situated across southern British Columbia, as described by *McKendry* [1994].

[24] The five AQ measurements sites for this study are in the complex terrain of the LFV, which is widest at its west terminus at the Georgia Strait. In the LFV sea breeze circulations, valley and slope flows exist, and with the addition of the photochemistry, ozone modeling becomes quite challenging in this area [*McKendry and Lundgren*, 2000].

[25] Roughly two million people in greater Vancouver live in this valley, causing significant anthropogenic emissions of NO_x that can mix with the volatile organic emissions from both anthropogenic sources and the surroundings evergreen forest. The Vancouver International Airport (CYVR) ozone monitoring site is at this western edge. The north and south walls of the valley are the steep Coast Range and Cascade Mountains. The valley width decreases considerably toward east, where the ozone site at the town of Hope is located in a very narrow, deep valley. See DM1 for a map and site details. KF postprocessing is particularly valuable at complex locations such as these, where both the NWP model and the AQ model can have difficulty.

3.2. Verification Statistics

[26] The skill of the 14 forecasts (12 ensemble members plus EK and KEK) have been measured using the same statistical parameters as defined in DM1: (1) Pearson product-moment coefficient of linear correlation (herein “correlation”):

$$\text{correlation}(station) = \frac{\sum_{t=1}^{N_{hour}} \{ [C_o(t, station) - \overline{C_o(station)}] [C_p(t, station) - \overline{C_p(station)}] \}}{\sqrt{\sum_{t=1}^{N_{hour}} [C_o(t, station) - \overline{C_o(station)}]^2 \sum_{t=1}^{N_{hour}} [C_p(t, station) - \overline{C_p(station)}]^2}} \quad (10)$$

(2) gross error (for hourly observed values of $\text{O}_3 > 30$ ppbv):

$$\text{gross error}(station) = \frac{1}{N_{hour}} \sum_{t=1}^{N_{hour}} \frac{|C_p(t, station) - C_o(t, station)|}{C_o(t, station)} \quad (11)$$

(3) root mean square error (RMSE):

$$\text{RMSE}(station) = \sqrt{\frac{1}{N_{hour}} \sum_{t=1}^{N_{hour}} [C_p(t, station) - C_o(t, station)]^2} \quad (12)$$

and (4) unpaired peak prediction accuracy (UPPA):

$$\text{UPPA}(station) = \frac{1}{N_{day}} \sum_{day=1}^{N_{day}} \frac{|C_p(day, station)_{\max} - C_o(day, station)_{\max}|}{C_o(day, station)_{\max}} \quad (13)$$

where N_{hour} is the number of 1-hour average concentrations over the 5-day period, N_{day} is the number of days, $C_o(t, station)$ is the 1-hour average observed concentration at a monitoring station for hour t , $C_p(t, station)$ is the 1-hour average predicted concentration at a monitoring station for hour t , $\overline{C_o(station)}$ is the average of 1-hour average observed concentrations at a monitoring station over the 5-day period, $\overline{C_p(station)}$ is the average of 1-hour average predicted concentrations at a monitoring station over the 5-day period, $C_o(day, station)_{\max}$ is the maximum 1-hour average observed concentration at a monitoring station over 1 day, and $C_p(day, station)_{\max}$ is the maximum 1-hour average predicted concentration at a monitoring station over 1 day. Predicted values also include EK and KEK.

[27] The gross error and UPPA are included in the U.S. EPA guidelines [*U. S. Environmental Protection Agency (U. S. EPA)*, 1991] to analyze historical ozone episodes using photochemical grid models. The EPA acceptable performance upper limit values are 35% for gross error, and $\pm 20\%$ for unpaired peak prediction accuracy. UPPA is computed here as an average (over the 5 days available) of the absolute value of the normalized difference between the predicted and observed maximum at each station (equation (13)). Thus UPPA is nonnegative; hence only the +20% acceptance performance upper limit is used in the next sections.

[28] The reasons for utilizing this set of statistics are as follows. We choose correlation to obtain an indirect indication of the phase differences between the predicted and measured ozone time series at a specific location. The closer the correlation is to one, the better is the correspondence of timing of ozone maxima and minima between the two signals.

[29] RMSE (measured in ppbv) gives important information about the skill in predicting the magnitude of ozone concentration, even though alone it does not draw a complete picture of a forecast value. It is very useful

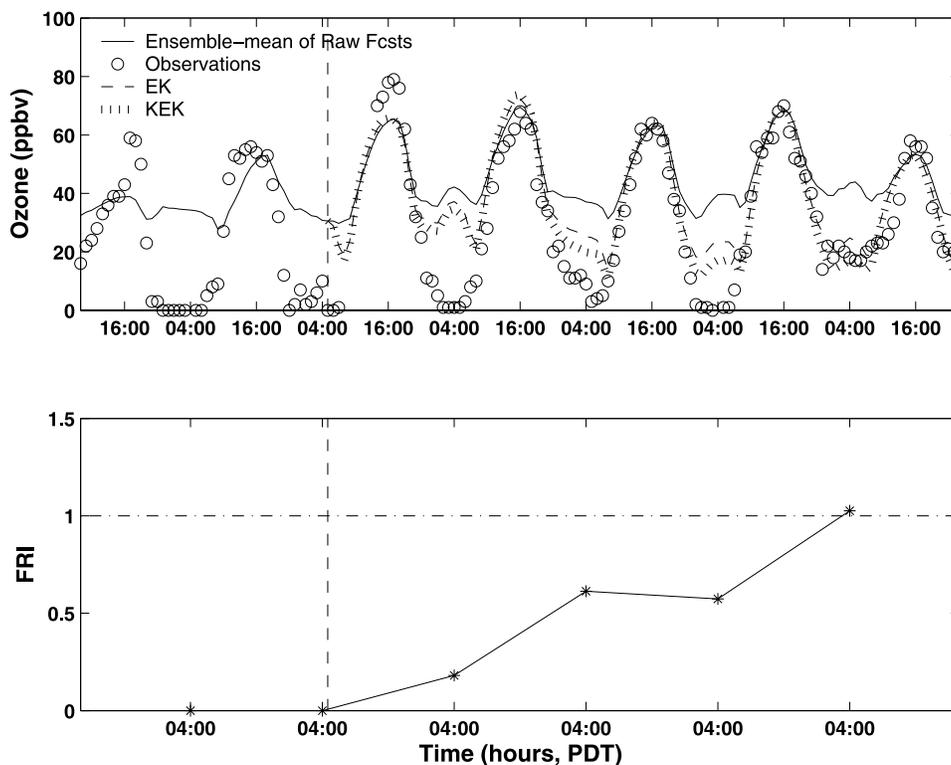


Figure 2. (top) Ozone ensemble mean forecasts and observations at Abbotsford for the 7-day period 9–15 August 2004. Continuous line is the raw ensemble mean, the dashed line represents the ensemble mean of the KFP bias-corrected forecasts (EK), and the dotted line represents the KFP bias-corrected EK (KEK). The circles are the observations. The vertical dashed line separates the training period (2 days, left) from the filter application (5 days, right). (bottom) Fractional Relative Improvement (FRI) at 4:00 am for each day. Vertical dashed line is as Figure 2 (top), and the dash-dotted line represents the optimal FRI value (one). Local Pacific Daylight Time (PDT) is UTC – 7 hours.

also for understanding the filter behavior, because it can be decomposed into systematic and unsystematic components as discussed in detail in section 4.3.

[30] The gross error statistic has been considered in this analysis because it is included in the U.S. EPA guidelines [U. S. EPA, 1991]. Also, being computed for hourly observed values of $O_3 > 30$ ppbv, it gives useful information about the forecast skill for higher concentration values, which are important for health-related issues. It gives information about the error magnitude (as RMSE), but as a portion of the observed ozone concentration (i.e., is measured in %).

[31] UPPA (%) is also used because it measures the ability of the forecasts to predict the ozone peak maximum on a given day. Traditionally, peak concentrations have been the main concern for the public health. However, in recent years over midlatitudes of the Northern Hemisphere, a rising trend for background ozone concentrations has been observed, while peak values have been steadily decreasing [Vingarzan, 2004].

4. Results

[32] Figure 2 shows a typical example of the KFP bias correction behavior. In Figure 2 (top), the time series include the observations (circles), the ensemble mean of the raw forecasts (solid line), EK (dashed line), and KEK

(dotted line), for the 7-day period of 9–15 August 2004, at Abbotsford. The first 2 days on the left side of the vertical dashed line represent the training period, when the coefficients start to be computed, but no correction is applied to the forecast.

[33] Even though the CMAQ model has been spun-up the 4 days before the start of training (i.e., in the period 5–8 August 2004), first day (9 August) still shows evidence that the forecast did not yet recover from the cold start. Therefore a longer CMAQ spin-up period would improve the filter performance as well.

[34] Nevertheless, KFP preserves the good performance of the raw ensemble mean for the peak concentration, except for the first day. The underestimated peak the first day is not adequately corrected by the KFP because the bias was much smaller for the previous (training) day. The overnight over prediction (that is indeed common to all the forecasts and the raw ensemble mean) is improved, with KEK closer to the observations than EK.

[35] Figure 2 (bottom) shows the behavior of the Fractional Relative Improvement (FRI), defined as follows:

$$FRI = \frac{|RawFcsts - KEK|}{|RawFcsts - Obs|} \quad (14)$$

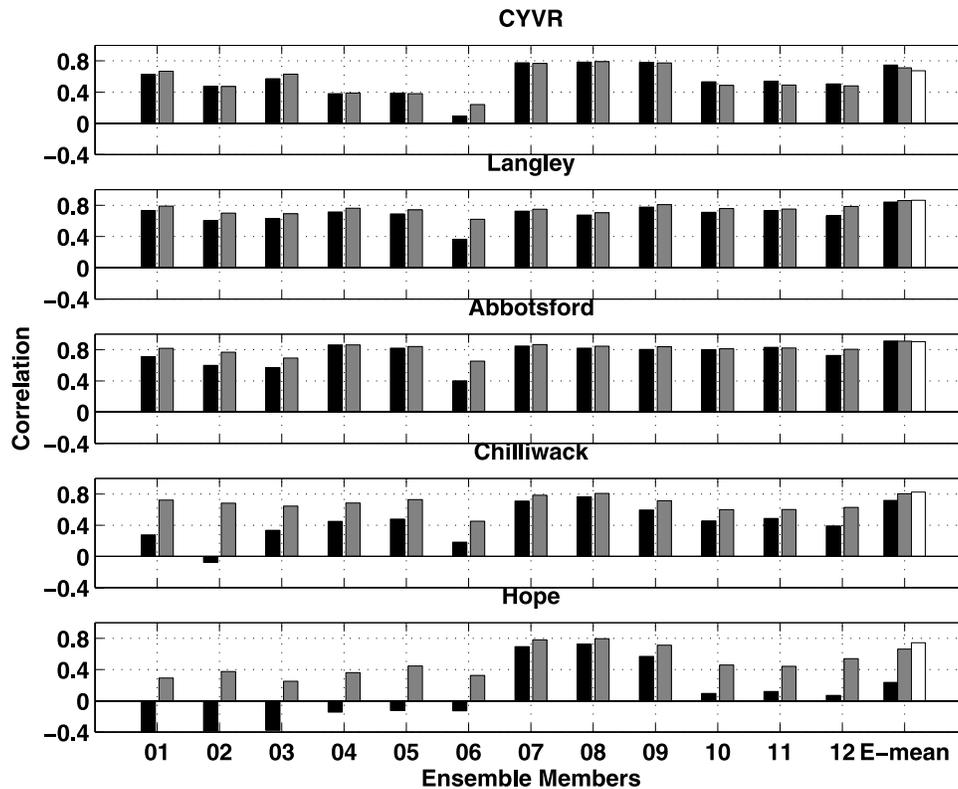


Figure 3. Correlation values between observed and predicted ozone 1-hour average concentrations are plotted for the 12-member Ozone Ensemble Forecast System (01, 02,..., 12) and the ensemble mean (E-mean). The solid bars are the values for the raw forecasts and raw ensemble mean, the shaded bars are the values for the Kalman filter predictor (KFP) bias-corrected forecasts and their ensemble mean (EK), and the open bar represents the KFP bias-corrected ensemble of the KFP members (KEK). Results are plotted at five stations (Vancouver International Airport (CYVR), Langley, Abbotsford, Chilliwack, and Hope), for the 5-day period 11–15 August 2004. Values are within the interval $[-1, 1]$, with correlation = 1 being the best possible value.

where $RawFcsts$ is the ensemble mean of the raw forecasts, and Obs is the observation. FRI is computed in Figure 2 each day at 4:00 am (PDT), when the nighttime over prediction is more evident. The fact that FRI, after the training period, almost steadily increases toward its optimal value (FRI = 1; i.e., when $KEK = Obs$) means that the filter, day after day, keeps learning about the over prediction at that hour, and progressively improves its performance. This also confirms what said in section 2, that the filter quickly and optimally converge after few time step iterations. It also means that, with a slightly longer training period, the results presented here could be improved, particularly for statistical parameters such as gross error and RMSE. The following subsections present and discuss the results by looking at correlation, gross error, RMSE, and UPPA.

4.1. Correlation

[36] Figure 3 shows the correlation results for the KFP bias-corrected 12 ensemble members and the ensemble mean for the 5-day period of 11–15 August 2004, at the five stations (CYVR, Langley, Abbotsford, Chilliwack and Hope). The solid bars are the values for the raw forecasts and raw ensemble-mean (as in Figure 6 of DM1), the shaded bars are the values for the KFP bias-corrected

forecasts and EK, while the open bars in the last column represent the KEK correlation values. There are improvements (higher correlation between forecast and observations) in most of the cases, except at CYVR where forecasts 10, 11 and 12 (MM5, 4 km) have slightly lower correlation after the KF. The EK improvements are up to a factor of six and they are larger for correlation values below 0.5. At Hope, six ensemble members have negative correlation before the KF bias correction, but have positive correlation (with values between 0.3 and 0.5) after the correction.

[37] The EK correlation is slightly worse (lower) than the raw ensemble mean at CYVR, slightly better at Abbotsford and Langley, better at Chilliwack, and significantly improved at Hope. The KEK correlation values are slightly worse than the EK values at CYVR and Abbotsford (but still very high correlation there), while they are better at the other stations. Notably, after the KFP bias correction, the correlation values of the forecasts are much more similar, meaning that the filter brings all of them closer to the same point — the observations.

[38] Table 1 shows for each station the ranking (from 1 to 14) of each ensemble member, EK, and KEK, where the best (highest) correlation value has a ranking of 1, and the worst (lowest) has 14. Forecast 08 has similar rankings when compared to EK, while forecasts 08 and 09 (MC2,

Table 1. Ranking of KFP Bias-Corrected 12 Ensemble Members (01, 02, . . . , 12), the Ensemble Mean of the KFP Bias-Corrected Forecasts (EK), and the KFP Bias-Corrected EK (KEK) at the Vancouver International Airport (CYVR), Langley, Abbotsford, Chilliwack and Hope Stations Tabulated for Correlation, Gross Error, Root Mean Square Error (RMSE), and Unpaired Peak Prediction Accuracy (UPPA)

	01	02	03	04	05	06	07	08	09	10	11	12	EK	KEK
<i>Correlation</i>														
CYVR	6	11	7	12	13	14	3	1	2	9	8	10	4	5
Langley	4	12	13	6	10	14	9	11	3	7	8	5	2	1
Abbotsford	9	12	13	4	6	14	3	5	7	10	8	11	1	2
Chilliwack	6	9	10	8	5	14	4	2	7	13	12	11	3	1
Hope	13	10	14	11	8	12	2	1	4	7	9	6	5	3
<i>Gross Error</i>														
CYVR	1	9	2	6	10	4	14	13	12	8	11	3	7	5
Langley	4	10	11	6	5	8	14	13	12	7	9	3	1	2
Abbotsford	4	6	12	3	5	11	13	14	10	7	9	8	2	1
Chilliwack	10	7	2	5	8	13	12	14	11	6	9	4	3	1
Hope	12	13	10	5	8	14	3	7	11	6	4	9	2	1
<i>RMSE</i>														
CYVR	2	7	1	9	11	4	14	12	13	8	10	6	5	3
Langley	2	11	6	8	9	4	14	12	13	7	10	5	3	1
Abbotsford	4	11	7	6	5	8	13	14	12	9	10	3	2	1
Chilliwack	14	6	9	7	2	10	8	13	4	11	12	5	3	1
Hope	12	7	13	14	9	10	3	4	2	8	11	6	5	1
<i>UPPA</i>														
CYVR	3	10	1	5	9	2	14	13	12	8	11	4	7	6
Langley	8	4	12	3	5	11	14	10	13	2	1	9	7	6
Abbotsford	8	10	13	2	4	11	12	14	9	5	6	7	1	3
Chilliwack	10	13	11	2	9	14	5	3	12	4	1	8	6	7
Hope	10	13	11	5	6	14	3	2	12	4	1	8	9	7

4 km) have a slightly worse performance. KEK rankings are the best when compared to any other forecast.

4.2. Gross Error

[39] The KFP bias-corrected forecasts have better (lower) gross error values than the raw forecasts, except at CYVR for forecasts 01 and 06 (Figure 4), with improvements roughly between 10 and 20%. KEK is always better than EK, which in turn is always better than the raw ensemble mean. The gross error computation (equation (11)) has a lower ozone concentration limit (observed 30 ppbv). Those improved gross error values after the KF correction means that the KFP bias correction is improving not only the forecast nighttime overprediction, but also efficiently remove bias throughout the time series, regardless the time of the day.

[40] Table 1 summarizes the rankings computed by looking at the gross error. KEK is clearly the best, while EK is the best when compared to the single deterministic forecasts. Here, as well as for the correlation (Table 1), the KFP forecast shows the same problem as the raw ones at CYVR, but not at Hope. The overall poor skill of the raw forecasts at CYVR and Hope are due to the fact that both stations are located in areas where all the individual ensemble members have difficulties, as explained in section 4.2 of DM1. The KFP is able to considerably improve the raw ensemble mean at Hope (where it was 4th), with EK being 2nd and KEK 1st. Moreover, both EK and KEK gross error are always well within the EPA acceptance limit (+35%).

4.3. RMSE

[41] The RMSE results are shown in Figure 5. With this parameter there is an improvement after the KFP bias

correction for all the forecasts, with values improved (decreased) up to 20–25%. The raw ensemble mean RMSE is considerably improved at each location, with further improvements (decreases) between 17 and 21% with EK, and between 29 and 36% with KEK. Table 1 shows the RMSE rankings. KEK is always the best except at CYVR where it is 3rd. EK is 3rd at Langley and Chilliwack, and second at Abbotsford, therefore it is the second best forecast when compared with the other 13.

[42] RMSE can be separated in different components. One decomposition was proposed by Willmott [1981]. First, an estimate of concentration $C^*(t, station)$ is defined as follows:

$$C^*(t, station) = a + bC_o(t, station) \quad (15)$$

where a and b are the least-square regression coefficients of $C_p(t, station)$ and $C_o(t, station)$ (the predicted and observed ozone concentrations, respectively, as defined in section 3.2). Then the following two quantities can be defined:

$$RMSE_s(station) = \sqrt{\frac{1}{N_{hour}} \sum_{t=1}^{N_{hour}} [C^*(t, station) - C_o(t, station)]^2} \quad (16)$$

$$RMSE_u(station) = \sqrt{\frac{1}{N_{hour}} \sum_{t=1}^{N_{hour}} [C^*(t, station) - C_p(t, station)]^2} \quad (17)$$

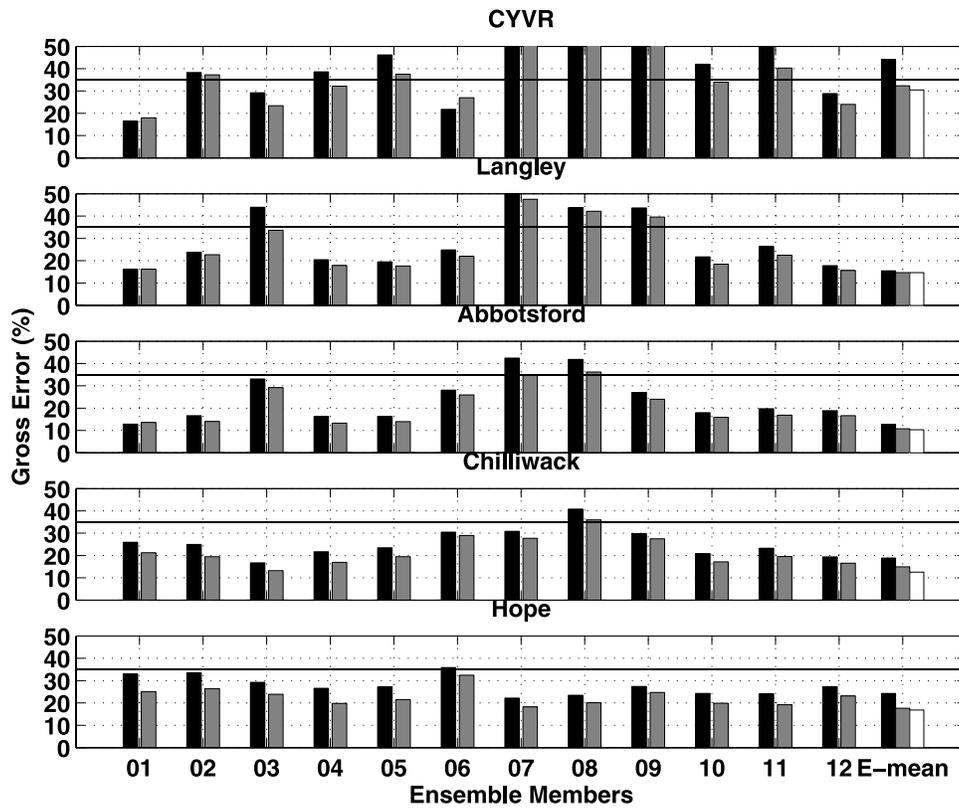


Figure 4. Similar to Figure 3 but for gross error values (%). The solid line is the EPA acceptance value (+35%). Values are within the interval $[0, +\infty)$, with a perfect forecast having gross error = 0.

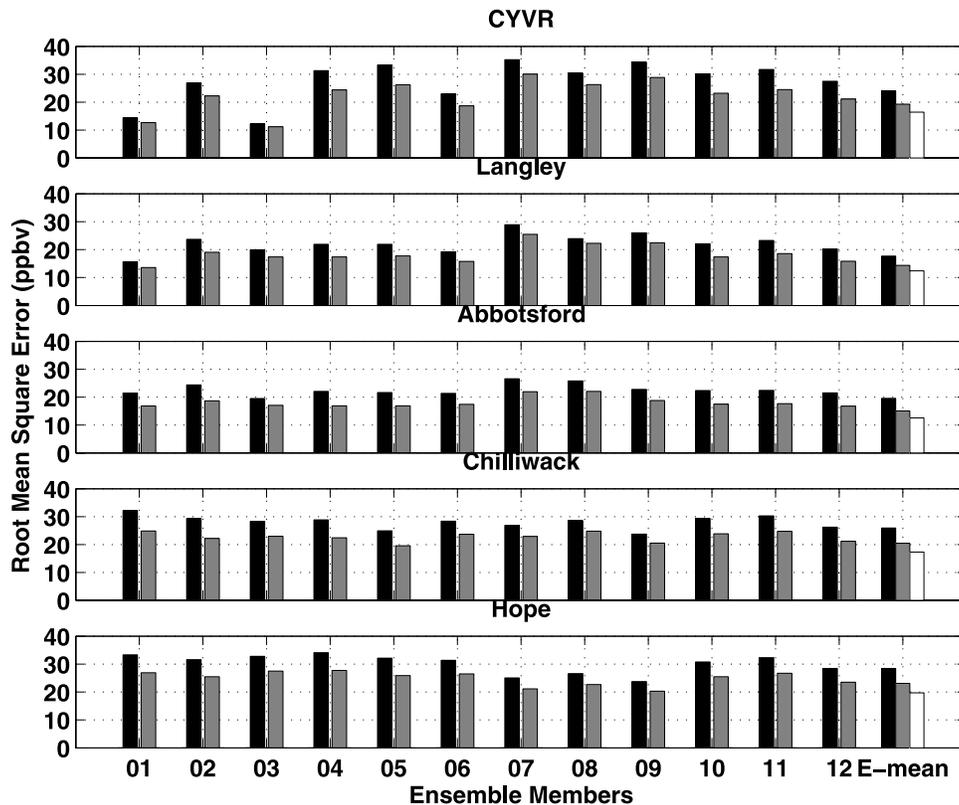


Figure 5. Similar to Figure 3 but for root mean square error (RMSE) values (ppbv). Values are within the interval $[0, +\infty)$, with a perfect forecast when RMSE = 0.

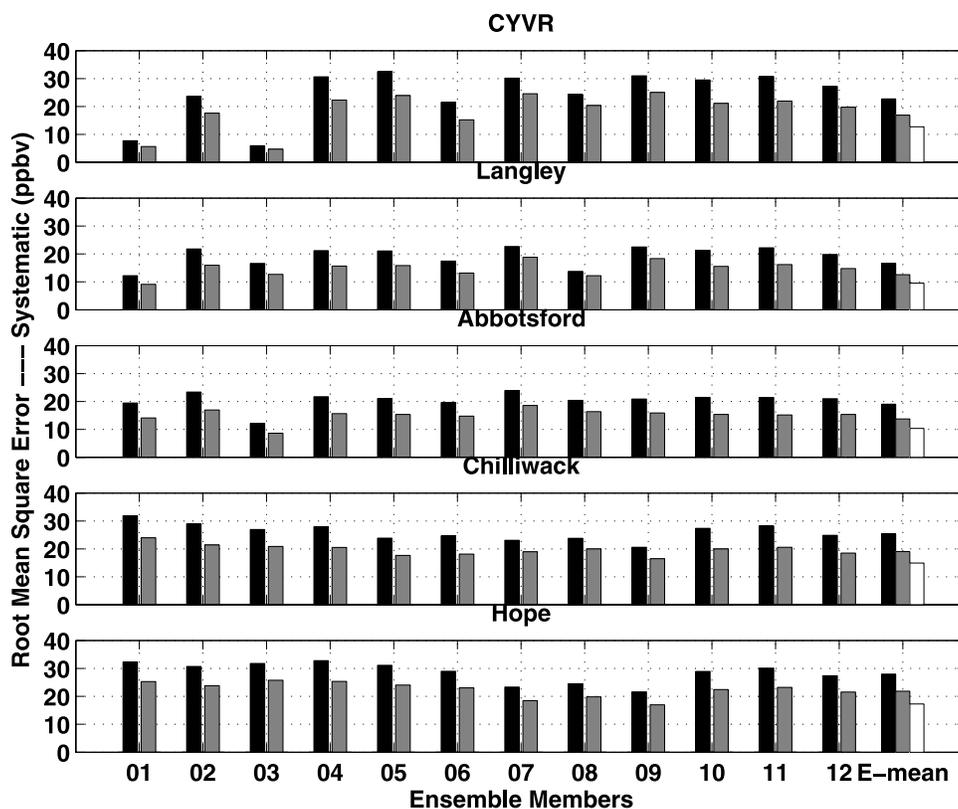


Figure 6. Similar to Figure 5 but for root mean square error (RMSE) systematic component values (ppbv). Values are within the interval $[0, +\infty)$, with a perfect forecast when $RMSE = 0$.

where $RMSE_s(station)$ is the RMSE systematic component, while $RMSE_u(station)$ is the unsystematic one. $RMSE_s$ indicates the portion of error that depends on errors in the model, while $RMSE_u$ depends on random errors, on errors resulting by a model skill deficiency in predicting a specific situation, and on initial condition errors. The following is an interesting relationship between RMSE and its components:

$$RMSE^2 = RMSE_s^2 + RMSE_u^2 \quad (18)$$

[43] The KF is expected to correct some of the systematic components of the errors (i.e., $RMSE_s$), while the unsystematic component ($RMSE_u$) on average (over the different forecasts) should be affected little by the filter correction. In fact, if $RMSE_u$ reflects errors introduced by model imperfections and initial condition errors, then it cannot be removed except by fundamental model improvements or improvements in initial conditions.

[44] Figure 6 shows the results for $RMSE_s$. The filter is correcting some of the forecast systematic errors, as expected, meaning that the algorithm is properly designed. There is an improvement even when the filter is applied twice (with KEK), meaning that successive applications of the filter correction will decrease further the systematic errors of all the forecasts.

[45] The 12-km runs (forecasts 01–06) have their highest systematic error at Hope. All these forecasts poorly reproduce the real topography at this location, and this leads to systematic misrepresentations of ozone temporal and spatial distribution. Conversely, the 4-km runs have their highest

systematic error at CYVR (in particular for MC2 driven runs, forecasts 07–09), where their ability to capture complex terrain more accurately than the 12-km runs is not an advantage, since at CYVR the terrain is flat.

[46] The results for $RMSE_u$ are shown in Figure 7. The filter does not decrease the unsystematic errors, and often increases them for this AQ episode. CYVR shows among the highest $RMSE_u$ values (particularly for MC2 driven runs, forecasts 01–03 and 07–09), indicating an intrinsic lack of predictive skill at this location. *Martilli and Steyn* [2004] discuss the effects of the superimposed valley, slope, and thermal flows over the LFV. Often the pollution plume is transported during night over the Georgia Strait waters, as a result of the combination of several transport processes. This makes it very challenging for the models to accurately predict the spatial and temporal evolution of ozone concentration near water locations, such as CYVR, where the overstrait pool of pollutants can be readvected over land during daytime sea breeze.

[47] For the ensemble mean, $RMSE_u$ keeps growing after successive filter applications, the opposite of what is observed for $RMSE_s$. This means that there is a finite upper limit on the number of useful corrections that can be obtained by successive KF applications. Here, for the ensemble mean, RMSE decreased until the fourth iteration, and considerably grew afterward (not shown).

4.4. UPPA

[48] Figure 8 shows the results for UPPA. There are improvements (values closer to zero) in the majority of

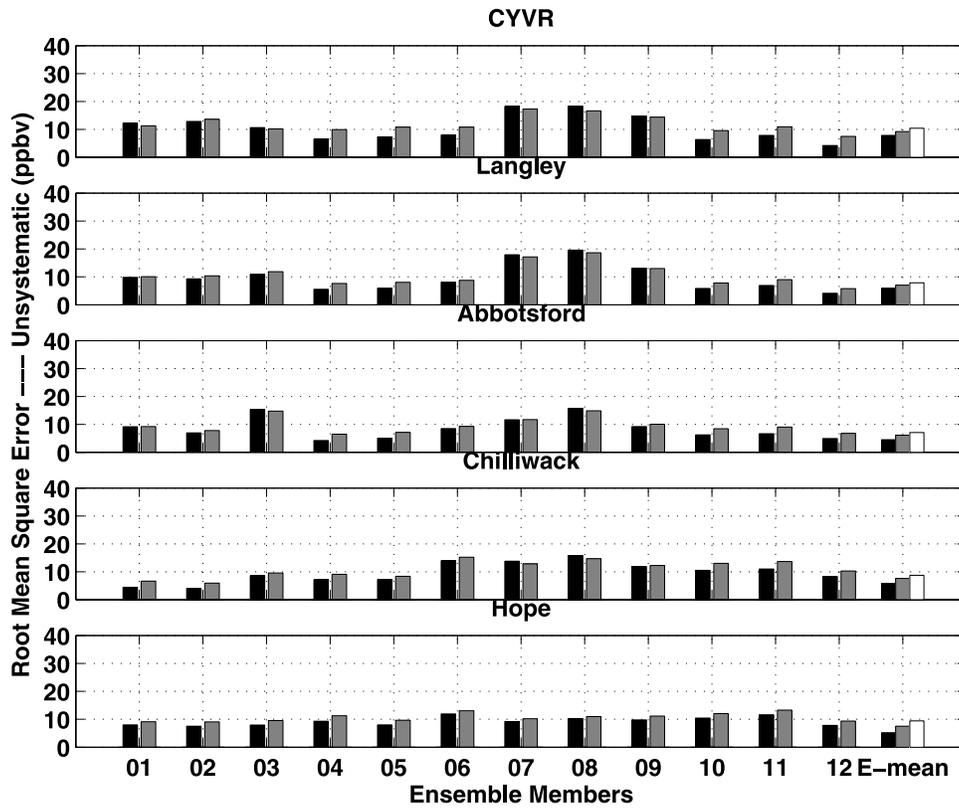


Figure 7. Similar to Figure 5 but for root mean square error (RMSE) unsystematic component values (ppbv). Values are within the interval $[0, +\infty)$, with a perfect forecast when RMSE = 0.

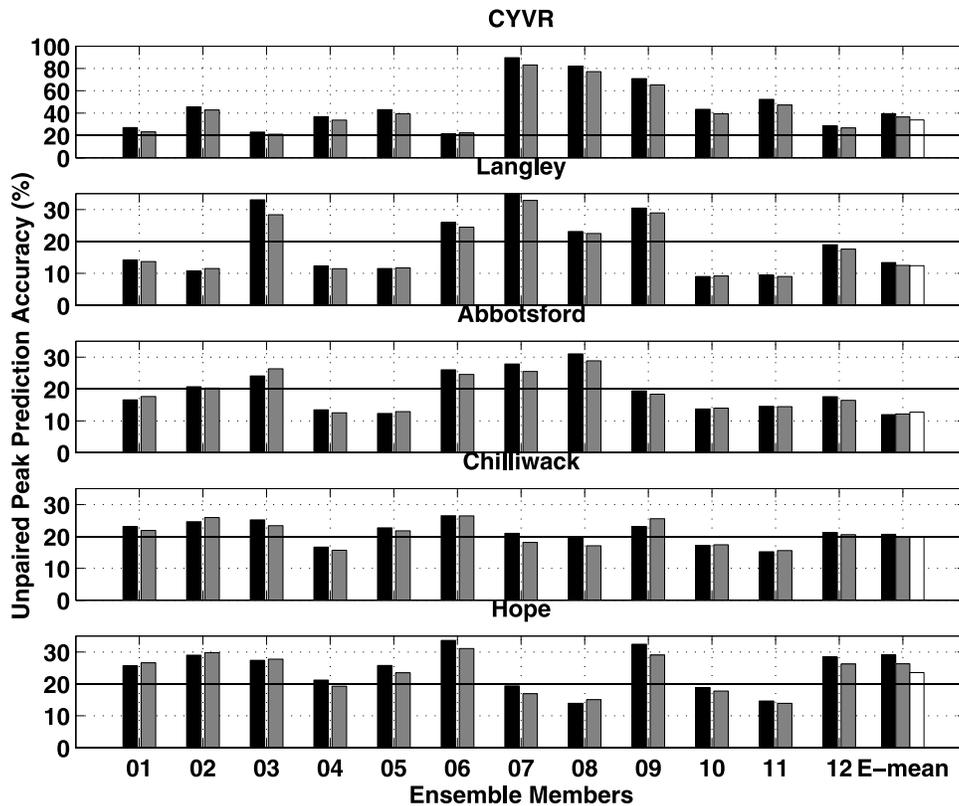


Figure 8. Similar to Figure 3 but for unpaired peak prediction accuracy (UPPA) values. The solid lines are the EPA acceptance values (+20%). Values are within the interval $[0, +\infty)$, with a perfect peak forecast when UPPA = 0.

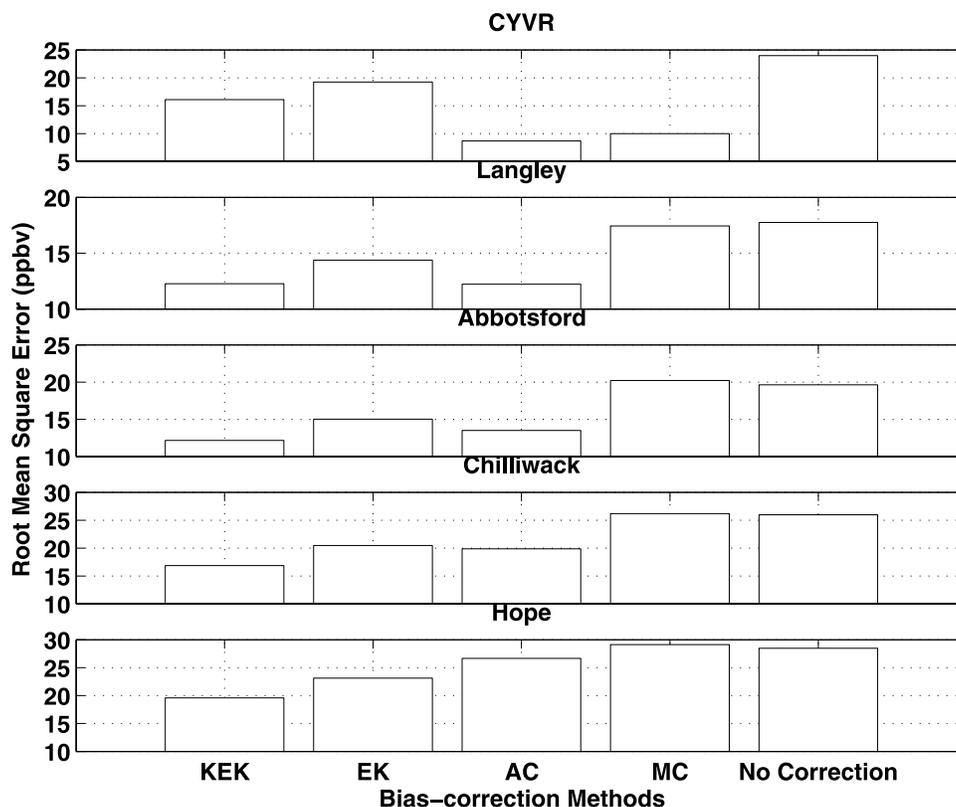


Figure 9. Root mean square error (RMSE) values (ppbv) are shown for four different bias correction methods applied to the ensemble mean. These methods are the Kalman filter predictor (KFP) bias-corrected ensemble mean of the KFP bias-corrected forecasts (KEK), the ensemble mean of the KFP bias-corrected forecasts (EK), the additive correction (AC), and the multiplicative correction (MC). The last values on the abscissa are for the raw ensemble mean with no corrections. Results are plotted at five stations (Vancouver International Airport (CYVR), Langley, Abbotsford, Chilliwack, and Hope), for the 5-day period 11–15 August 2004. Smaller values are better.

cases; however, in one, three, six, five and three cases out of 14 at CYVR, Langley, Abbotsford, Chilliwack and Hope, respectively, there is no improvement or the KF forecasts are slightly worse. The improvements of the UPPA KFP forecasts with respect to the raw forecasts are modest if compared with the improvements shown with the previous statistical parameters. EK is always better than the raw ensemble mean, except at Chilliwack, where it is slightly worse. The same can be said for KEK when compared to EK, with the larger improvements for both EK and KEK at Hope. EK and KEK have UPPA values within the EPA acceptance limit (+ 20%) at Langley, Abbotsford and Chilliwack, while it is close to this limit at Hope and above 30% at CYVR.

[49] UPPA is the only parameter where the ensemble mean does not have the best overall ranking, even after the forecasts are KFP bias corrected. Both EK and KEK have an average performance for UPPA, when compared with the other forecasts (Table 1).

5. Comparison With Other Bias Correction Methods

[50] Figure 9 shows the ensemble mean RMSE values for the five stations (CYVR, Langley, Abbotsford, Chilliwack and Hope), for the 5-day period 11–15 August 2004. On the

abscissa are KEK, EK, the additive bias correction (AC), the multiplicative bias correction (MC), and the raw ensemble mean for comparison purposes.

[51] The additive bias-corrected concentration is computed as follows:

$$C_{AC}(t, station) = C_p(t, station) - \frac{1}{N_{hour}} \sum_{t=1}^{N_{hour}} [C_p(t, station) - C_o(t, station)] \quad (19)$$

whereas the multiplicative bias-corrected concentration is given by

$$C_{MC}(t, station) = \frac{\sum_{t=1}^{N_{hour}} C_o(t, station)}{\sum_{t=1}^{N_{hour}} C_p(t, station)} C_p(t, station) \quad (20)$$

[52] Both AC and MC use observations throughout the experiment period, so the ozone time series corrected with these methods cannot be considered forecasts, since they cannot be computed in a predictor mode. Contrast this with both KEK and EK that are predictor postprocessing procedures of the forecasts, which use only observations available

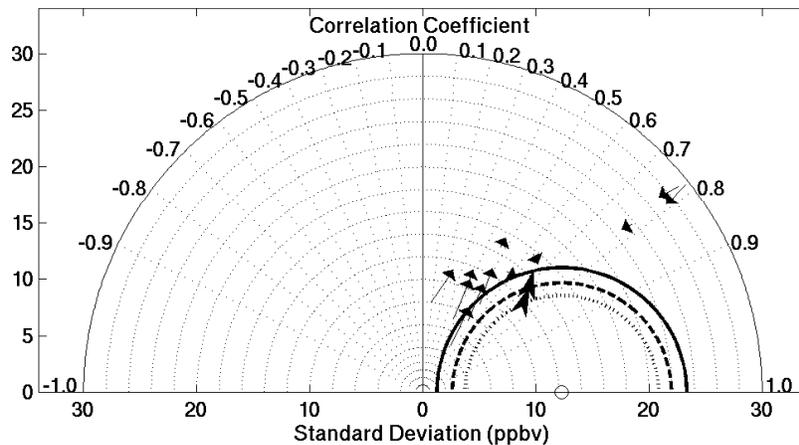


Figure 10. Taylor's diagram plotted for Vancouver International Airport (CYVR). The azimuthal position gives the correlation, while the radial distance from the origin is proportional to the standard deviation (ppbv). The smaller arrows represent the 12 ensemble members, and the bigger arrows (with different arrowhead) represent the ensemble mean of the Kalman filter predictor (KFP) bias-corrected forecasts (EK) and the KFP bias-corrected EK (KEK). Each arrow tail represents the forecast statistics of a raw forecast, and the arrowhead indicates KFP-corrected values. If the arrow points closer to the observation point (tiny circle) it means that the KFP is correcting the forecast in the right direction. The arrows representing EK and KEK are consecutive; that is, the EK arrowhead is also the KEK arrow tail, because EK is the raw version of KEK. The distance between the observation and a given point is proportional to the centered root mean square error (CRMSE) between the observation and the forecast. The three concentric lines centered over the point representing the observation indicate the CRMSE for the raw ensemble mean (dotted line), EK (thick dashed line), and KEK (thick solid line). If the line passing through the arrowhead is closer to the observation than the one passing through the tail, it means that the KFP is improving (reducing) the CRMSE.

before the time for which the forecast verify. In this sense, this is a stringent test for the KFP bias correction.

[53] Nevertheless, at every station (except CYVR) KEK is the best, while EK in general is better than MC, but has higher (worse) RMSE values than AC (except at Hope). Finally, at CYVR, KEK is third while EK is better only than the raw ensemble mean.

6. Discussion and Conclusions

[54] In summary, the Kalman filter predictor (KFP) bias-corrected forecasts and their ensemble mean have better forecast skill than the raw forecasts, for the locations and days used here to test their performance. The corrected forecasts are improved for correlation, gross error, root mean square error (RMSE), and unpaired peak prediction accuracy (UPPA), the latter being the statistical parameter showing the least pronounced improvement after the KFP bias correction. In general, the ensemble mean forecast benefits from the improvement of each single Kalman-corrected ensemble member. In fact, the ensemble mean of the KFP bias-corrected forecasts (EK) and the KFP bias-corrected EK (KEK) are the second best and the best forecasts overall when compared with the other 12 individual forecasts members and their raw ensemble mean. The results in section 4.3 showed also that only a limited number of successive KF application to the same forecast would result in an improvement.

[55] Those results indicate that the filter is improving the forecast timing of maxima and minima concentrations with

respect to the observations, because the correlation is closer to one. From the improved (decreased) RMSE and gross error values, we infer that the KF is improving the forecast accuracy in reproducing the magnitude of ozone concentrations. Better (closer to zero) UPPA and gross error values indicate that the filter is improving the forecast ability to capture rare (but important for health-related issues) events, such as the occurrence of ozone concentration peaks. Moreover, the KF reduced systematic errors such as can be induced by model error, as for example the poor representation of topographic complexity. Ensemble averaging tended to remove the unsystematic errors, as showed in DM1. This is why the combination of Kalman filtering and ensemble averaging results in the best forecasts; i.e., EK and KEK.

[56] EK and KEK performances have been compared also with the performances of two other bias correction (not in predictor mode) techniques, the additive bias correction (AC), the multiplicative bias correction (MC). At every station (except CYVR) KEK is the best, while EK is better than MC, but has higher (worse) RMSE values than AC (except at Hope). Finally, at CYVR, KEK is third while EK is better only than the raw ensemble mean.

[57] A concise way to summarize the results from section 4 is given in Figures 10–14. A Taylor's diagram [Taylor, 2001] is used to create a multistatistics plot of correlation, centered RMSE (CRMSE: RMSE computed after the overall bias is removed), and standard deviation. CRMSE is the distance on the diagram between the point representing the forecast and the one representing the observations. For each forecast (smaller arrows) and for EK and

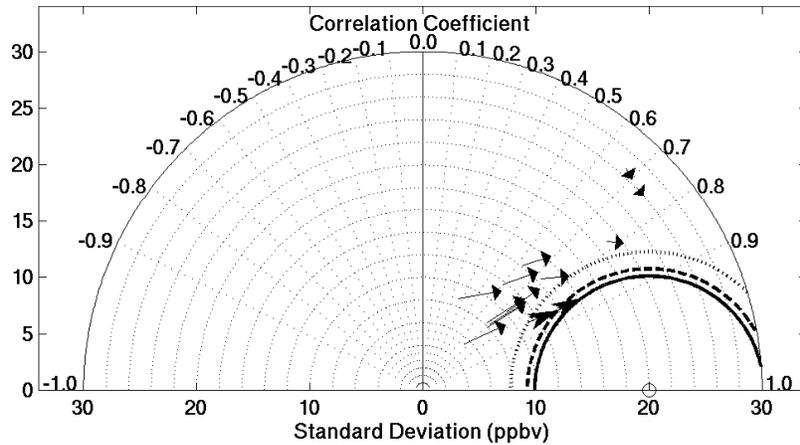


Figure 11. Taylor’s diagram for Langley (similar to Figure 8).

KEK (bigger arrows, with different arrowhead), the arrow tail gives the standard deviation and the correlation of a raw forecast, while the arrowhead represents the same values for the KFP bias-corrected version of the same forecast. If the arrow points toward to the observation (tiny circle) it means that the KFP is correcting the forecast statistically in the right direction. The arrows representing EK and KEK are consecutive; i.e., the EK arrowhead is also the KEK arrow tail, because EK is the raw version of KEK. The three concentric lines centered over the point representing the observation indicate the CRMSE for the raw ensemble mean (dotted line), EK (thick dashed line), and KEK (thick continuous line).

[58] At CYVR (Figure 10) the majority of arrows point away from the observation (including the arrows with different arrowhead for EK and KEK), indicating that the KFP in those cases degraded the raw forecasts. This is caused by the dominance of unsystematic errors at this location (as discussed in section 4.3), that prevent the filter to being able to do a successful correction.

[59] At Langley (Figure 11) the forecasts tend to be improved, as indicated by the arrows pointing closer to the observation. EK is better than the raw ensemble mean (which in turn is better than all the individual deterministic forecasts), since the thick dashed line passing through its

arrowhead is closer to the observations than the dotted line passing through the tail. KEK is the best being the closest to the observations (thick continuous line).

[60] The same conclusions can be drawn for Abbotsford (Figure 12), with even larger improvements after the correction. At this location, the forecast standard deviations after the correction are much more similar to the observation standard deviations (but the same can be said also at the other stations).

[61] Figure 13 shows the same diagram for Chilliwack. The forecasts are improved, since the arrows point toward the observations. At this location, EK is fourth best, while KEK is still the best.

[62] The results for Hope are shown in Figure 14. All the forecasts are improved, with EK and KEK being the third and fifth best, respectively. In this case (as well as for Chilliwack) the benefit of applying the KFP bias correction is even higher than at the other locations, demonstrating that the KF correction is particularly efficient if the raw forecast shows high systematic errors, as discussed in section 4.3. This is evident since the arrows are on average longer than at the other locations. At Hope, forecasts 07 and 08 are the first and second best forecasts (by comparison with Figure 19 in DM1), while they were among the worst at other locations, particularly at CYVR, Langley at Abbotsford.

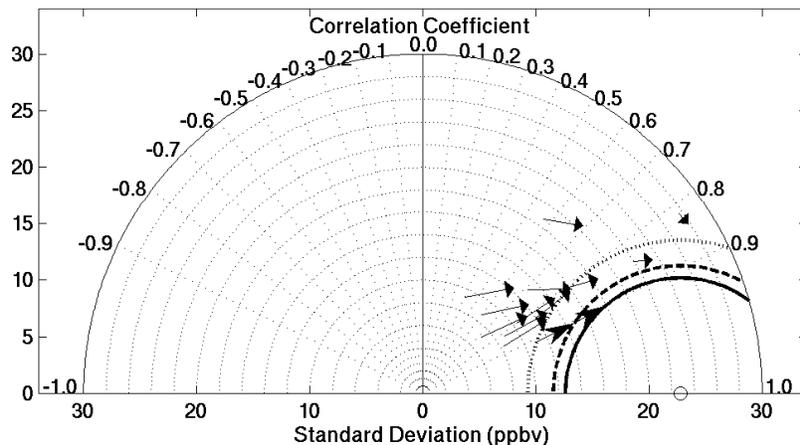


Figure 12. Taylor’s diagram for Abbotsford (similar to Figure 8).

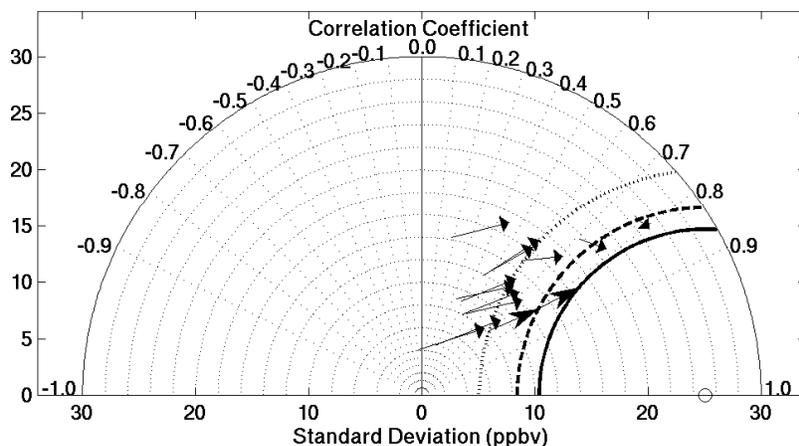


Figure 13. Taylor’s diagram for Chilliwack (similar to Figure 8).

[63] The KFP bias correction approach for the locations and days used in this study successfully remove the forecast bias. The filter is able to recognize systematic errors in the forecast, as for example the nighttime overprediction of ozone concentration induced by a poor representation of the nighttime boundary layer, or the errors at Chilliwack and Hope induced by the systematic misrepresentation of topographic complexity in the model. As a consequence of the improved nighttime over prediction, the ozone distribution low-concentration tail is better represented after the KF correction, resulting in forecasts having a variance that resembles more closely the observed variance, as discussed above.

[64] The experiments performed in this study suggest that better forecasts can be made with a longer KF training period (such as 5 days), and with a longer CMAQ model spin-up. Moreover, with the availability of a longer data set (a full month or season), including ozone forecasts and observations with a broader variability of low- and high-ozone events, an optimal value for the sigma ratio (as discussed in section 2) could be found.

[65] KEK, which combines the beneficial effects of ensemble averaging and KFP postprocessing, is overall the most skilful forecast for the locations and days tested here, where the ozone modeling is particular challenging because the complex coastal mountain setting. For this

reason the approach used here to improve ozone forecasts it might be equally successful when implemented in other regions with similar or less complex topographical settings.

[66] Finally, ensemble weather forecasts often provide information on the reliability of the forecast: if the ensemble members have a large spread (defined as the standard deviation of the ensemble members about the ensemble mean), this implies less confidence in the forecast. Perhaps a similar spread-skill relationship exists for air quality forecasts. However, in DM1, neither a correlation nor a relationship between the raw ensemble spread and the raw forecast error has been found. Similarly, a spread-skill relationship has not been found for the Kalman-filtered AQ forecasts in this study.

Appendix A

[67] Here a step-by-step description of the filter implementation is given. First, σ_ϵ^2 is estimated via the Kalman filter algorithm as follows (by applying equation (5)):

$$p_{t|t-\Delta t}^{\sigma_\epsilon^2} = \left(p_{t-\Delta t|t-2\Delta t}^{\sigma_\epsilon^2} + \sigma_{\sigma_\epsilon^2}^2 \right) \left(1 - \beta_{t|t-\Delta t}^{\sigma_\epsilon^2} \right)$$

where $p_{t|t-\Delta t}^{\sigma_\epsilon^2}$ is the expected mean square error in the σ_ϵ^2 estimate, $\sigma_{\sigma_\epsilon^2}^2$ is the variance of $\sigma_{\sigma_\epsilon^2}^2$, and $\beta_{t|t-\Delta t}^{\sigma_\epsilon^2}$ is the Kalman

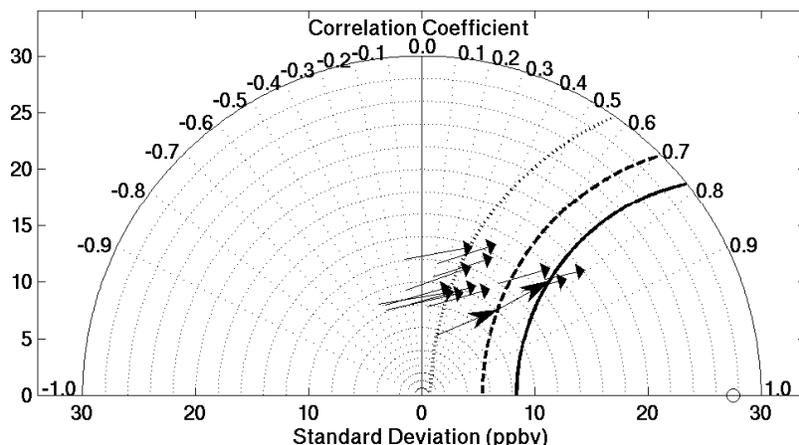


Figure 14. Taylor’s diagram for Hope (similar to Figure 8).

gain when the filter is used to estimate σ_{ε}^2 . Next, the new Kalman gain can be computed, similarly to equation (4):

$$\beta_{t+\Delta t|t}^{\sigma_{\varepsilon}^2} = \frac{P_{t|t-\Delta t}^{\sigma_{\varepsilon}^2} + \sigma_{\sigma_{\eta}}^2}{\left(P_{t|t-\Delta t}^{\sigma_{\varepsilon}^2} + \sigma_{\sigma_{\eta}}^2 + \sigma_{\sigma_{\varepsilon}}^2\right)}$$

where $\sigma_{\sigma_{\varepsilon}}^2$ is the variance of σ_{ε}^2 . Finally, σ_{ε}^2 can be estimated by combining equations (3) and (8):

$$\sigma_{\varepsilon,t+\Delta t|t}^2 = \sigma_{\varepsilon,t|t-\Delta t}^2 + \beta_{t|t-\Delta t}^{\sigma_{\varepsilon}^2} \left[\frac{(y_t - y_{t-\Delta t})^2}{2+r} - \sigma_{\varepsilon,t|t-\Delta t}^2 \right]$$

$\sigma_{\sigma_{\varepsilon}}^2$ and $\sigma_{\sigma_{\eta}}^2$ are assumed constant, with values of 1 and 0.0005, respectively, as determined from previous works [e.g., Roeger et al., 2003].

[68] Once σ_{ε}^2 is estimated, $\sigma_{\sigma_{\eta}}^2$ can be computed as $\sigma_{\sigma_{\eta}}^2 = r\sigma_{\varepsilon}^2$. Then, equations (5), (4) and (3) can be applied in sequence, resulting in the final estimate of the bias (\hat{x}). This process is iterated through different Δt , and for the first step, given initial values are used as discussed in section 2.

[69] **Acknowledgments.** We thank Miranda Holmes for the early development of the Kalman filter bias-corrected weather forecasts at the University of British Columbia that led to the tests here of the same approach to air quality forecasts. We also thank George Hicks, Henryk Modzelewski and Trina Cannon for maintaining the computing system used to perform the simulations presented here. Moreover, we thank Todd Plessel (of EPA) for providing very useful tools to handle Models-3 formatted data. We are grateful to RWDI for providing the emission inventory and the scripts to run SMOKE. Ken Stubbs and John Swalby (of the Greater Vancouver Regional District) graciously provided the ozone observation data. We are thankful to Bruce Thomson for carefully reviewing the paper. Grant support came from the Canadian Natural Science and Engineering Research Council, the BC Forest Investment Account, the British Columbia Ministry of Water Land and Air Protection, Environment Canada (Colin di Cenzo), and the Canadian Foundation for Climate and Atmospheric Science. Geophysical Disaster Computational Fluid Dynamics Center computers were used, funded by the Canadian Foundation for Innovation, the BC Knowledge Development Fund, and the University of British Columbia. Thanks are also due to two anonymous reviewers for their valuable comments and suggestions.

References

- Bozic, S. M. (1994), *Digital and Kalman Filtering*, 2nd ed., 160 pp., John Wiley, Hoboken, N. J.
- Burgers, G., P. J. van Leeuwen, and G. Evensen (1998), Analysis scheme in the ensemble Kalman filter, *Mon. Weather Rev.*, *126*, 1719–1724.
- Byun, D. W., and J. K. S. Ching (Eds.) (1999), Science algorithms of the EPA Models-3 Community Multiscale Air Quality (CMAQ) modeling system, *EPA/600/R-99/030*, Off. of Res. and Dev., U.S. Environ. Prot. Agency, Washington, D. C.
- Cannon, A. J., and E. R. Lord (2000), Forecasting summertime surface level ozone concentrations in the Lower Fraser Valley of British Columbia: An ensemble neural network approach, *J. Air Waste Manage. Assoc.*, *50*, 322–339.
- Delle Monache, L., X. Deng, Y. Zhou, and R. Stull (2006), Ozone ensemble forecasts: 1. A new ensemble design, *J. Geophys. Res.*, *111*, D05307, doi:10.1029/2005JD006310.
- Dempster, A., N. Laird, and D. Rubin (1977), Maximum likelihood from incomplete data via the EM algorithm, *J. R. Stat. Soc.*, *39*, 1–38.
- Hamill, T. M., and C. Snyder (2000), A hybrid ensemble Kalman filter-3D variational analysis scheme, *Mon. Weather Rev.*, *128*, 2905–2919.
- Homleid, M. (1995), Diurnal corrections of short-term surface temperature forecasts using Kalman filter, *Weather Forecasting*, *10*, 689–707.
- Houtekamer, P. L., and H. L. Mitchell (2001), A sequential ensemble Kalman filter for atmospheric data assimilation, *Mon. Weather Rev.*, *129*, 123–137.
- Houtekamer, P. L., H. L. Mitchell, G. Pellerin, M. Buehner, M. Charron, L. Spacek, and B. Hansen (2005), Atmospheric data assimilation with an Ensemble Kalman Filter: Results with real observations, *Mon. Weather Rev.*, *133*, 604–620.
- Jolliffe, I. T., and D. B. Stephenson (2003), *Forecast Verification: A Practitioner's Guide in Atmospheric Science*, 240 pp., John Wiley, Hoboken, N. J.
- Kalman, R. E. (1960), A new approach to linear filtering and prediction problems, *J. Basic Eng.*, *82*, 35–45.
- Martilli, A., and D. G. Steyn (2004), A numerical study of recirculation processes in the Lower Fraser Valley (British Columbia, Canada), paper presented at 27th NATO/CCMS Conference on Air Pollution Modeling 2004, NATO, Banff, Alberta, Canada.
- McKendry, I. G. (1994), Synoptic circulation and summertime ground-level ozone concentrations at Vancouver, British Columbia, *J. Appl. Meteorol.*, *33*, 627–641.
- McKendry, I. G., and J. Lundgren (2000), Tropospheric layering of ozone in regions of urbanized complex and/or coastal terrain: A review, *Prog. Phys. Geogr.*, *24*, 329–354.
- Roeger, C., R. B. Stull, D. McClung, J. Hacker, X. Deng, and H. Modzelewski (2003), Verification of mesoscale numerical weather forecast in mountainous terrain for application to avalanche prediction, *Weather Forecasting*, *18*, 1140–1160.
- Russell, A., and R. Dennis (2000), NARSTO critical review of photochemical models and modeling, *Atmos. Environ.*, *34*, 2283–2324.
- Segers, A. J., H. J. Eskes, R. J. van der A, R. F. van Oss, and P. F. J. van Velthoven (2005), Assimilation of GOME ozone profiles and a global chemistry-transport model using a Kalman filter with anisotropic covariance, *Q. J. R. Meteorol. Soc.*, *131*, 477–502.
- Taylor, K. E. (2001), Summarizing multiple aspects of model performance in a single diagram, *J. Geophys. Res.*, *106*, 7183–7192.
- U. S. Environmental Protection Agency (1991), Guideline for regulatory application of the Urban Airshed Model, *USEPA Rep. EPA-450/4-91-013*, Off. of Air Qual. Plann. and Stand., U. S. Environ. Prot. Agency, Research Triangle Park, N. C.
- van Loon, M., P. J. H. Builtjes, and A. J. Segers (2000), Data assimilation applied to LOTOS: First experiences, *Environ. Model. Software*, *15*, 603–609.
- Vingarzan, R. (2004), A review of surface ozone background levels and trends, *Atmos. Environ.*, *38*, 3431–3442.
- Willmott, C. J. (1981), On the validation of models, *Phys. Geogr.*, *2*, 184–194.
- L. Delle Monache, Lawrence Livermore National Laboratory, L-103, Livermore, CA 94550, USA. (ldm@llnl.gov)
- X. Deng, Meteorological Service of Canada, Environment Canada, Montreal, Quebec, Canada.
- T. Nipen and R. Stull, Department of Earth and Ocean Science, University of British Columbia, 6339 Stores Road, Vancouver, BC, Canada V6T 1Z4.
- Y. Zhou, Meteorological Service of Canada, Environment Canada, Edmonton, Alberta, Canada.