A Study of Rain Forecast Error Structure Based on Radar Observations over a Continental-Scale Spatial Domain

MAJID FEKRI

Department of Atmospheric and Oceanic Sciences, McGill University, Montréal, Québec, and Meteorological Research and Development, Pelmorex Media Inc., Oakville, Ontario, Canada

M. K. YAU

Department of Atmospheric and Oceanic Sciences, McGill University, Montréal, Québec, Canada

(Manuscript received 14 May 2015, in final form 10 May 2016)

ABSTRACT

This study examines the univariate error covariances of hourly rainfall accumulations using two different NWP models and a mosaic of radar reflectivity over a continental-scale domain. The study focuses on two main areas.

The focus of the first part of the paper is on the ensemble-based and the innovation-based error variance and correlation estimations. An ensemble of forecasts and a set of observations provide the basis for estimating the errors in two different ways. The results indicate that both ensemble- and innovation-based methods lead to comparable variance estimations, while the local error correlation estimates have larger differences due to the sensitivity of calculations to the gradient of the variance field.

The second part of the paper uses innovations for identifying the errors. The focus of this part is on a prognostic method for estimating the error statistics from the background based on the Bayesian inference technique. The case study shows that the predictive model produces a similar result regarding the magnitude and the dispersion of variance in comparison with the innovation and ensemble-based variances.

This study represents a step toward estimating local error variances and local error correlations to construct a nonhomogeneous and precipitation-dependent error covariance matrix of rainfall. These results will be used in a future paper in the design of a 2D-VAR Assimilation Method for Blending Extrapolated Radars (AMBER) with NWP precipitation forecast to form a precipitation nowcasting model.

1. Introduction

Knowledge of rain forecast errors and their spatial and temporal correlation plays an important role in data assimilation (DA) to improve precipitation forecasts (Macpherson et al. 2004; Sun 2005; Stephan et al. 2008; Caumont et al. 2010). This knowledge is particularly relevant in the construction of the error covariance matrix (ECM) in variational (VAR) methods (Chevallier et al. 2002; Moreau et al. 2003; Amerault and Zou 2006). However, the representations of ECM in rainy areas remain suboptimal because of the nonlinearity inherent in moist physical processes (Fillion and Errico 1997; Errico et al. 2003; Mahfouf and Bilodeau 2007) and the non-Gaussianity of error distributions (Errico et al. 2001, 2007; Michaelides et al. 2009). A better understanding of the statistical structure of the background errors would improve the specification of the ECM of rain.

A common question relating to the statistical structure of the background errors is “how to get precipitation-dependent statistics?” (Montmerle et al. 2010). There are various approaches to address this question. One approach is to model the error correlations by distinguishing different weather patterns (Monteiro and Berre 2010; Brousseau et al. 2011). Another approach is to distinguish the rain/no rain events and to use a binning method to obtain discrete sets of error statistics (Pagé et al. 2007; Caron and Fillion 2010; Montmerle and Berre 2010; Michel et al. 2011). Except for the latter, which presents multivariate background error covariances for hydrometeors, the above papers focus on modeling multivariate background error covariances in rainy areas for the
“classical” control variables like winds, temperature, surface pressure, and specific humidity. They did not compute the covariances for rain explicitly. Thus, the above works enhance the benefit of observations within precipitation but do not lead to analysis of precipitation.

Two methods are generally used to diagnose background error covariances. The first is from innovations, defined as the differences between background and observations in the observation space (Hollingsworth and Lönnberg 1986; Lonnberg and Hollingsworth 1986). The innovation-based error estimates depend both on the background errors and the observation errors, as well as on the forward observation operator that maps model variables into the observation space. In the event that the observation errors are less correlated compared to the background errors (as discussed in section 3b), the impact of the observation error can be separated in the innovation errors by applying proper spatial filtering.

A second method is to use an ensemble of perturbed forecasts to compute the statistics (Brousseau et al. 2011; Ménétrier et al. 2014). The ensemble-based method yields an approximate representation of background error in model space without a direct comparison with the observations. It should be noted that the estimates derived from either method are usually affected by noise due to sampling issues and require filtering before they can be compared or used in any DA scheme (Raynaud et al. 2009; Raynaud and Pannekoucke 2012; Ménétrier et al. 2015a,b).

The purpose of this paper is to present a study of the univariate precipitation-dependent rainfall forecast error statistics over a continental-scale spatial domain. We investigate the statistical structure of the errors using both the innovation and the ensemble methods and compare their differences. The work is motivated by the need for an efficient prognostic ECM to blend NWP model forecast with radar extrapolations in a precipitation nowcasting system called the Assimilation Method for Blending Extrapolated Radars (AMBER). AMBER (Fekri 2015) requires only the 2D univariate model error statistics to assimilate the radar extrapolations. Each short-term forecast of AMBER consists of several 2D VAR analyses that blend the Canadian Global Environment Model (GEM) NWP model forecast with radar extrapolation. A detailed description of AMBER will be presented in a future publication.

To obtain the ECM for precipitation, we first investigate the error characteristics and then present a method of diagnostic error estimation from the background. We combine the innovation-derived prior rules with forecast information as a proxy to estimate precipitation-dependent local background error variances and error correlations of NWP forecasts. To make sure that errors are not merely a product of internal biases of a single model, we study and compare the errors of two different sets of operational forecasts produced independently using different numerical models (GEM and WRF).

The rest of the paper is organized as follows. The second section describes the available datasets. The third section contains the results of error analysis on the three aspects of NWP forecast errors: bias, variance, and, correlation. The fourth section provides the methodology for parameterization and modeling of the errors. The fifth section offers a summary and discussion.

2. Datasets
   a. Radar mosaic
      The observation data were obtained from a multi-radar mosaic of reflectivity covering most regions of North America provided by the National Severe Storms Laboratory (NSSL) (Zhang et al. 2005). The process of producing mosaic from several radars involved the remapping of high-resolution radar radial scans to Cartesian coordinates and then to a continental grid. The process introduces some representation errors in the final analysis. We used constant-altitude plan position indicator (CAPPI) reflectivity maps at 2.5 km with 4-km horizontal grid spacing, similar to the dataset used in Berenguer et al. (2012). Every mosaic dataset includes a mask identifying the locations that are outside the radar coverage at every hour.
   b. GEM-based numerical forecast
      The hourly accumulation of precipitation was derived from a GEM model forecast (Côté et al. 1998). This model had been operational at the Canadian Meteorological Centre (CMC) with 15-km horizontal grid spacing and 58 variable vertical levels up to 10 hPa from 2004 to 2010. The forecasts were available twice a day at 0000 and 1200 UTC for 48 h, and we used the former forecasts at 0000 UTC. The assimilation process for this set of forecast includes satellite radiance, but no radar reflectivity. Convection schemes include the shallow convection scheme based on the Kuo-type closure and the deep convection scheme of Kain and Fritsch (Kain and Fritsch 1993). The Sundquist scheme was used for resolved clouds. The details are in Mailhot et al. (2006).
   c. WRF-based numerical forecast and ensemble
      The second set of numerical forecasts was the WRF Model run by the Center for Analysis and Prediction of Storms (CAPS) in the 2008 National Oceanic and
Atmospheric Administration (NOAA)/Hazardous Weather Testbed Spring Forecasting Experiment (Kain et al. 2010). Forecasts were run at a grid spacing of 4 km for 30 h starting at 0000 UTC of each day and continuing to 0600 UTC of the next day. We used 10 ensemble members in the ensemble-based calculations, and 1 of the forecasts in the innovation-based method. The forecast without the assimilation of radar data was used in the innovation analysis to avoid cross correlation between the observation and the background errors. This member was run with the Thompson microphysics scheme (Thompson et al. 2004) with two predicted moments for cloud ice similar to four other members. Boundary layer was treated using the Mellor–Yamada–Janjic (MYJ) scheme (Kong et al. 2008; Xue et al. 2008).

d. Domain of study

The analysis area is confined to the central and eastern continental United States from 110° to 78°W in longitudes and between 32° and 45°N in latitudes (Fig. 1). Some regions to the west of the domain do not have a constant radar coverage due to the interference of the mountainous areas. The areas without radar observations are masked and excluded from the analysis. We, therefore, remapped the data by linear interpolation to a common 16-km grid-spacing map over the domain of analysis. The data were available during three spring months in 2008, from 16 April to 6 June. There were 20 rainy days. We grouped each day into three 8-h cases of rain to account for the diurnal cycle effects. The cases from 0000 to 0800 UTC were used for presenting the spatial patterns of correlation and variance in section 3. We used all of the 480 h of precipitation to generate the innovation and variance sample pools in section 4.

Fig. 1. The domain of study in the central and eastern continental United States between 32° and 45°N in latitudes and from 110° to 78°W in longitudes.

3. Diagnosis of the rain forecast errors

a. Estimation of bias

The broad definition of bias includes any error that is systematic rather than random (Dee 2005). However, it is usually hard to isolate the systematic part of the errors. A common way of calculating bias is by taking the space average and the time average of the innovations, assuming that the observation has a relatively negligible bias (Dee and Da Silva 1998; Dee and Todling 2000). Areas with no radar coverage or no rain coverage were excluded from the average. The radar coverage mask was in the observation data and marked the areas without radar coverage. The rain coverage was defined as a nonzero value for precipitation in the observation or in either of the forecasts. The total number of grid points used for averaging is the same for GEM, WRF, and radar observations but changes from one hour to another depending on the coverage. After taking the spatial averages, each hour is considered as a sample of temporally distributed bias errors. Samples are binned into 0.01 mm h⁻¹ thresholds.

The distribution of the mean forecasts and observation values are shown in Fig. 2a. The histogram of binned GEM-based forecast shows a resemblance to that of observations while the distribution of WRF-based forecast tends to overestimate the frequency and number of stronger precipitations (over 0.3 mm h⁻¹ of mean precipitation) and underestimate the number of weaker rains (less than 0.1 mm h⁻¹).

The distributions of mean errors are illustrated in Fig. 2b for the GEM- and WRF-based datasets. The average error distributions are nonsymmetrical with a peak at the center. The mean, variance, skewness, and kurtosis of the error distributions are 0.001, 0.007, 2.089, and 1.246, respectively, for GEM-based forecast errors, and 0.033, 0.014, 0.463, and 2.116, respectively, for WRF-based forecast errors. A positive mean error indicates a total overestimation of rain in both forecasts. GEM has a lower bias error and a smaller variance. Skewness is a measure of departure from symmetrical distribution. The positive skewness of WRF-based forecasts corresponds to a longer extension to the right-hand side of the mean error distribution in Fig. 2b, and indicates a higher number of overestimation of rain. On the other hand, the negative skewness of GEM forecasts indicates an underestimation of high intensity rain events.

In summary, the GEM-based forecast has a smaller skewness and a more symmetrical distribution of errors. The kurtosis of a normal distribution is zero. Positive kurtosis values indicate heavier tails and higher peaks in the distribution. Kurtosis is used to describe “bimodality” of a distribution (Darlington 1970), or as a
measure of departure from the normal distribution (DeCarlo 1997). The kurtosis of the available forecast errors suggests that the GEM-based forecast has a more symmetrical and less peaked distribution of errors. The departure of errors from the normal distribution in precipitation is not unexpected (Errico et al. 2000; Fletcher and Zupanski 2006). However, the relative magnitude of such deviation between two forecasts can be used as a measure of their relative bias.

b. Estimation of the error variances

There are two pathways in calculating the background error variance: ensemble-based methods and innovation-based methods as summarized by Berre and Desroziers (2010). Ensemble-based variance estimates depend on the internal variability of the ensemble members about the ensemble mean. This method is used to derive space-dependent and time-dependent variances both at Météo-France (Berre et al. 2007) and at ECMWF (Bonavita et al. 2011). The main problem of the ensemble-derived variance field is the presence of unknown model errors. Another issue is the sampling noise due to the insufficient number of ensemble members. The sampling noise issue is related to the local background error correlation as demonstrated by Raynaud et al. (2008). One solution is to identify the noise covariance and to apply proper filtering methods before using the variance in data assimilation. Homogeneous spectral filtering (Raynaud et al. 2009) and heterogeneous diffusion algorithm (Raynaud and Pannekoucke 2012) are two methods of noise removal from the ensemble-based variance estimation. More recently, Ménétrier et al. (2015b) proposed a theoretical framework for obtaining the optimality criteria for filtering.

Innovation-based variances depend on the observation rather than the model-based ensembles. We use the difference between background and observations as an indicator of all background errors including the model errors. However, the accuracy of precipitation innovation can be affected by the correlated errors in radar observations.

To explain the basic assumptions about our use of the innovation, we start with the relation between the true precipitation values in model space $\mathbf{x}^t$, observation $\mathbf{x}^o$, and the error of the observation. They can be expressed as follows:

$$ H\mathbf{x}^t = \mathbf{x}^o - \mathbf{e}(\mathbf{x}^o), \tag{1} $$

where $H$ is the forward observation operator that maps model variables into observation space and can be nonlinear. First, we assume that $H$ can be represented by a linear matrix $\mathbf{H}$. Then, we introduce the forecast error $\mathbf{e}(\mathbf{x}^b) = \mathbf{x}^b - \mathbf{x}^t$ into (1):

$$ \mathbf{e}(\mathbf{x}^o) = \mathbf{x}^b + \mathbf{H}^{-1}[\mathbf{e}(\mathbf{x}^o) - \mathbf{x}^o]. \tag{2} $$

It is necessary to investigate the observation error in estimation of rainfall values from radar reflectivities $\mathbf{e}(\mathbf{x}^o)$. There are two main approaches in studying these errors. One is the comparison between radar data and another reference measurement such as rain gauge values (Lee and Zawadzki 2006; Bellon et al. 2007; Ciach et al. 2007). The other approach is based on the investigation of the most relevant error sources and their physical characteristics from model simulations or experimental data (Bellon et al. 2005; Lee and Zawadzki 2005). Berenguer and Zawadzki (2008) proposed a methodology for estimating the ECM of radar observations by examining different sources of errors. They concluded that two sources of error are dominant for
deriving rain rate from radar observations: range-dependent uncertainties induced by radar beam, and the uncertainty associated with the Z–R relation including variability of drop size distributions (DSDs). Later, they analyzed structure of these two sources of errors and found that the average decorrelation distance of range effect errors are within 15–20 km and that Z–R estimation error decorrelation lengths can extend out to 40 km (Berenguer and Zawadzki 2009). Decorrelation lengths are associated with exponential correlations and are larger than the Daley correlation length scales. Therefore, we conclude that the largest radar observation error correlation lengths are smaller than the average background error correlations in our study, and we define

\[ \tilde{x}^o = H^{-1}[x^o - \epsilon(x^o)], \]  

(3)

where \( \tilde{x}^o \) is the new observation vector after filtering features less than 16 km and mapping to the same grid as of the background. The innovations are directly calculated with respect to the filtered observations:

\[ \epsilon(x^o) = (x^b - \tilde{x}^o). \]  

(4)

Every innovation field only provides one sample in space and time, while a large number of error samples are required to calculate the variance. One solution is to assume ergodicity and use the time average of spatial variations in innovation as examined by Lindskog et al. (2006). Another solution is to apply a local homogeneity assumption and to take the spatial average of samples over finite areas to reduce the sampling noise issue (Berre et al. 2007). One aspect of the method suggested by Berre et al. (2007) and Berre and Desroziers (2010) is the possibility of comparison between the innovation-based and the ensemble-based variances for different observation types. There has not been enough research about this problem and specifically none could be found about the variance of precipitation errors. We use the spatial sampling method to derive innovation-based variance of precipitation forecast and compare our results with those of the ensemble-based method. Here, we are interested in a qualitative comparison and physical interpretation of the differences. Therefore, we chose to apply a simple spectral filter. The more sophisticated filtering methods listed above could be tested in the future to get more optimal results.

A time-averaged map of WRF ensemble-based and innovation-based standard deviations of hourly accumulations are shown in Fig. 3. The variances were calculated spatially, at every hour, using 10 ensemble members. One member of the ensemble (the one without perturbations and radar data assimilation) was compared with radar mosaic to form the innovation field. The innovation-based variance was calculated using a spatial sampling within a square of \( 80 \times 80 \) km\(^2\) around every grid point. At a grid spacing of 16 km, this translates into calculating the variance of a \( 5 \times 5 \) square that has exactly 25 samples at every point. Variance fields derived from both methods were then normalized on an hourly basis to have the same average magnitude equal to one. The square root of the normalized hourly variance fields were averaged over 160 h of precipitation from 16 April to 6 June 2008 and from 0000 to 0800 UTC. Most of the precipitation systems developed during these hours in the central regions of the domain. The resulting averaged standard deviation map of the ensemble-based method is presented in Fig. 3a and that of the innovation-based method in Fig. 3b. The blank area on the left side of Fig. 3b indicates the lack of radar coverage over the mountains given by the radar masks.

To remove the small-scale noise, we applied a homogeneous spectral filtering method based on the discrete cosine transform (DCT) in 2D (Denis et al. 2002). In the 2D-DCT filter, a transfer function was assigned to give weight 1 to wavenumbers that correspond to scales greater than 100 km, 0 to scales smaller than 50 km, and linear weight between 1 and 0 to scales in between. This transfer function was multiplied element by element onto the 2D spectral components. The resulting spectral field was then transformed by an inverse 2D-DCT to produce the filtered output in physical space.

Both ensemble-based and innovation-based variances seem to agree on the large-scale features but differ in location and intensity of the small-scale features. Some details with regards to local similarity and differences can be interpreted by the local variability of the rainfall. For example, we observe a local maximum precipitation variance region along the eastern edge of the Rocky Mountain ranges in the states of Wyoming and Colorado, where most of the rain systems initiate during that period of the day (0000–0008 UTC). The west–east elongated features in the center correspond to the eastward Lagrangian advection of the precipitating systems along the corridor through Nebraska and Iowa. Another local maximum variance region was seen in Oklahoma according to the ensemble-based method, but the innovation-based method estimated it to be farther north toward the Oklahoma and Kansas border. As suggested by Berre and Desroziers (2010), this type of verification can be used to evaluate the space-dependent variability of the ensemble. Spatial sampling of innovations provided a variance map comparable to that of the ensemble. Their similarity is plausibly due to their mutual dependence on the magnitude of the local rainfall events.
c. Estimation of the error correlations

Local error correlations of rainfall have a high variability in space and time that makes them a problematic part of modeling the precipitation-dependent covariances. Following Caron and Fillion (2010), precipitation-dependent geographical masks were used for describing such correlations (Michel et al. 2011; Montmerle 2012). One approach to estimating the nonhomogeneity of error correlations is the physical-space error approximation method (Pannekoucke and Massart 2008; Pannekoucke 2009; Pannekoucke et al. 2014). In this method, the heterogeneity is represented by specifying local correlation length scales on a gridpoint space. Combining this with the diffusion method produces a strong tool for modeling error correlations in physical space (Weaver and Courtier 2001; Weaver and Ricci 2003; Weaver and Mirouze 2013). The diffusion tensors are derived from the local Daley correlation length scales (Pereira and Berre 2006; Pannekoucke et al. 2008). We are interested in estimating these local error correlation length scales for precipitation. The Belo–Pereira–Berre (BPB) formula has a distinct practical importance in providing a heterogeneous formulation of the local error correlation length scales $L_D$ based on the variance estimation (Pereira and Berre 2006; Weaver and Mirouze 2013), and can be written as

$$L_D^2 = -\left. \frac{d^2 \rho}{dx^2} \right|_{x=0} = \text{Var}(\epsilon) \left\{ \text{Var} \left( \frac{\partial \epsilon}{\partial x} \right) - \left[ \frac{d \sqrt{\text{Var}(\epsilon)}}{dx} \right]^2 \right\}.$$  

(5)

From (5), Michel (2013) derived a more elegant formula that guarantees positive definiteness of the square length scale:

![Figure 3. WRF-based rain forecast error standard deviation map from (a) an ensemble of 10 members and (b) a spatial sampling of innovations.](image)
Here, $\epsilon$ is the error, $\eta = \epsilon / \sqrt{\text{Var}(\epsilon)}$ is the error normalized by its standard deviation, and $d/dx$ denotes derivation in any arbitrary axis of space $x$. The proof of (6) requires assuming a null error expectation $\langle \epsilon \rangle = 0$. This assumption is true when the error is defined as the deviation of ensembles from the mean. However, the innovation-based errors do not satisfy such condition as we will show in section 4. Therefore, we use (5) for calculating the innovation-based and (6) for calculating the ensemble-based error correlations. Even though (5) does not guarantee positive definiteness of the square length scale, we implemented the ensemble method with Michel formula and innovation method with BPB formula, and compared the results.

To calculate $L_D$ from the BPB formula, we needed the variance of error, the gradient of variance, and the variance of the gradient of error. Derivation of the local variance from the innovation-based method was discussed in section 3b. The gradient of variance was derived from the local standard deviation field by calculating the discretized differentiation in meridional and zonal directions. To obtain the variance of gradients of error, the gradient of rainfall field was first computed through discretized differentiation, and then its variance was obtained by applying a spatial sampling method similar to the innovation-based variance estimation. For the cross-correlation term, instead of the variance the covariance between the zonal and meridional gradient samples was calculated. After averaging the correlations over time, a 2D-DCT filter was applied to remove the noise. The filter was designed to eliminate features smaller than 100 km in size. Figures 4a and 4b show the

$$L_M^2 = \frac{1}{\text{Var}\left(\frac{d\eta}{dx}\right)}.$$  (6)
zonal and meridional components of the innovation-based error correlation length scales, respectively. Some areas on the west side of the domain did not have radar coverage to form the innovation. As a result, no correlation length was calculated for them.

To obtain the ensemble-based correlation length scales $\ell_M$ from (6), first the standard deviation field was formed. Second, each ensemble member was used to form a sample of the normalized error field $\eta$. Third, the discretized differentiation was performed in space on the $\eta$ of each member, and their variance was calculated. Fourth, the zonal, meridional, and cross components of the hourly $\ell_M$ estimations were averaged over 160 h of the forecast. Fifth, the same 2D-DCT filter was applied on the results. Finally, the ellipses of anisotropic correlations were calculated from the eigenvalue and the eigenvector of each local correlation matrices. Figures 5a and 5b show the zonal and meridional components of the ensemble-based local error correlation length scales, respectively.

Comparison of Figs. 4 and 5 indicates different outcomes for the error correlation estimations. The zonal correlation length scales in Figs. 4a and 5a are greater than those of the meridional correlations in Figs. 4b and 5b. This difference in magnitude could be due to the larger displacement errors that happen along the mostly west–east movement of the precipitating systems. Comparing Figs. 4a and 4b, there are some features in the zonal correlations (e.g., in Nebraska) that do not appear in the meridional correlations. On the other hand, the zonal and meridional components of the ensemble-based estimates in Figs. 5a and 5b have a similar pattern with different intensities. In general, the innovation-based correlations are smaller in size compared to that of the ensemble-based method. The gradients of innovation errors in the zonal and the
meridional directions depend on the irregular observation fields, while the gradient of the ensembles follows the internal variability of the models. As a result, the differences between the innovation-based (Fig. 4) and ensemble-based (Fig. 5) error correlation estimations are more remarkable than the variance estimations (Figs. 3a and 3b).

The anisotropy of the correlation function is obtained by calculating the Hessian matrix of the correlations at each point. In addition to the meridional and zonal components of the length scales, the cross components of length scales are needed. To obtain the cross components with respect to the $x$ and $y$ axis, it suffices to replace the $\text{Var}(\frac{d}{d x})$ term in (5) and (6) by $\text{Cov}(\frac{d}{d x}, \frac{d}{d y})$ and to replace the $(\frac{d}{d x})^2$ in (5) by $(\frac{d}{d x})(\frac{d}{d y})$ [see appendix B of Weaver and Mirozoue (2013)]. The meridional, zonal, and cross component of the Hessian form a symmetric $2 \times 2$ matrix with the inverse square of the zonal and meridional correlations length scales on the diagonal. The eigenvalues of this matrix are the inverse square of the major and minor axes of an ellipse that represents the local error correlations. The orientation of the ellipse is derived from the ratio of the normalized eigenvectors of the same matrix.

Figure 6 shows the local elliptical anisotropic correlations. Figures 6a and 6b present the innovation-based correlations derived from (5) and the ensemble-based correlations derived from (6), respectively. The ellipses are shown at regular intervals and are rescaled for better presentation. The colors indicate the geometric mean of the major and minor axes or radius of a circle that has the same area as of the local correlation ellipse. The orientation of ellipses and details of the local correlations show a noticeable variation between the two estimation methods. On average, the innovation-based correlation ellipses in Fig. 6a are smaller than that of the
ensemble-based ones in Fig. 6b. In the central region of the domain where the size of the ellipses are comparable between Figs. 6a and 6b, the degree of anisotropy or the elongation of ellipses is more prominent in innovation-based correlations than that of the ensemble-based ones.

4. Modeling errors from the background

In section 3, we provided a comparison between the ensemble-based and innovation-based variance and correlations. The similarity between the error variance estimations was an indication of the dependence of the errors on the magnitude of the forecasted rainfalls. In section 4, we investigate this dependence by forming a conditional variance of innovations and presenting a predictive model of rainfall error variances from the background.

a. Forming the conditional error distributions

Precipitating and nonprecipitating areas have markedly different background error statistics (Caron and Fillion 2010; Montmerle and Berre 2010). This difference implies that a rain-dependent and heterogeneous error estimation can better represent the true nature of the error structures. Rain-dependent binning appeared as a practical tool in the heterogeneous formulation of precipitation ECM (Michel et al. 2011). The remaining question is whether a discernible relation exists between precipitation forecast and its error, and whether this relation can be used to express error variances as a function of background precipitation (Fekri 2015).

We would like to study the best estimate of the conditional error $\varepsilon (\varepsilon | x^b)$ and the conditional probability distribution of the errors $\varepsilon$ given the background $x^b$, which in the case of minimum variance estimate are related as follows:

$$
\varepsilon (\varepsilon | x^b) = \int_{-\infty}^{\infty} \varepsilon P(\varepsilon | x^b) d\varepsilon. \quad (7)
$$

In general, the conditional PDF for two vectors $\varepsilon$ and $x^b$ is the probability density that the event $\varepsilon$ occurs, given that the event $x^b$ occurred. In Bayesian approach the posterior conditional PDF is defined as

$$
P(\varepsilon | x^b) = \frac{P(\varepsilon, x^b)}{P(x^b)}. \quad (8)
$$

Our method is similar to Bayesian inference in the sense that we formulate a probability model, estimate posterior distribution, and evaluate its ability to characterize the desired estimate (Wikle and Berliner 2007). The difference, however, is that instead of using assumptions about prior and marginal distributions, we describe a statistical model of the posterior distribution based on the innovation samples. The statistical model allows us to approximate the true distribution of the rainfall forecast errors without implying any other prior assumptions.
In the Bayesian frame of work, forecast, error, and variance are all considered random variables. Each innovation provides one sample of random variables at every grid space and time instance. Therefore, it is important to use a large volume of samples for describing probability distributions regarding the frequency of occurrence. We assume ergodicity in the sense that estimates of the moments of these random processes converge in probability toward the theoretical moments by increasing the number of available samples. By including the available dataset of 480 h of rainfall, and by considering a continental-scale grid at 16-km grid spacing, a large sample pool of innovations was formed from GEM-based forecasts as background. The number of available samples and binned rainfall events in each bin are in the order of 10^5. The less-intense rainfall events are exponentially more frequent than high-intensity events. Therefore, the bin thresholds were chosen to increase the sample size for higher intensities. The threshold of 1 mm h^{-1} was used to distinguish between rain and no-rain events. To have a finite range of the forecasts, and also to avoid outlier values or extreme events, we omitted the point values with more than 32 mm h^{-1} rainfalls. Figure 7 shows the distribution of forecast events computed every millimeter per hour and normalized by the number of available samples within a certain threshold. For each forecast \( x^b \), the innovation error \( \epsilon \), and spatially sampled innovation variance \( \text{Var}_s(\epsilon) \) were calculated. Accordingly, the conditional distribution of innovation (Fig. 8) and of spatial variance (Fig. 9) were obtained as a function of background \( x^b \). To avoid confusion, we keep the notation \( \sqrt{\text{Var}_s(\epsilon | x^b)} \) in referring to the local standard deviation of spatially sampled innovations as opposed to \( \sigma(\epsilon|x^b) \), which is the climatological standard deviation of the innovation distribution.

Figure 8 shows that the conditional distribution of rainfall innovations has a tail toward the negative values indicating the missed rainfalls. Other studies found similar error distribution for precipitation fields (Errico et al. 2000, 2001). Distributions in Fig. 8 are more likely to be log-normal than normal. In the case of nonnormal posterior distributions, the choice of the best estimator depends on the distribution. For example, Fletcher and Zupanski (2006) found that mode is the best estimator for a log-normal distribution. We obtain and compare four different estimators: mean, median, mode, and expectation.

Best estimates and standard deviations of forecast \( x^b \), innovation \( (\epsilon | x^b) \), and spatially sampled standard deviation \( \sqrt{\text{Var}_s(\epsilon | x^b)} \) are presented in Tables 1, 2, and 3, respectively. The mean and median minimize the squared error and absolute value of the loss function in Bayesian point estimation, respectively. The mode is the location of the maximum probability in the posterior distribution and coincides with the maximum-likelihood estimator in the case of a uniform prior probability. However, the exact location of the mode in a discretized sample distribution depends on the number of bins and their intervals. We provided the uncertainty measures of mode estimation in the tables.

![Figure 9](image_url)

**Figure 9.** Normalized distribution of the GEM-based spatially sampled standard deviations of errors conditioned by the background.

Table 1. Estimation of the background values within the binned thresholds. Mean, median, mode, and expectations of binned distributions and their standard deviations are presented.

| Binned forecast \( x^b \) (mm h^{-1}) | Estimation \( \epsilon|x^b \) | Mean | Median | Mode (±0.01) | Expectation (±0.005) | Std dev \( \sigma(x^b) \) |
|---------------------------------------|-------------------------------|------|--------|--------------|----------------------|-------------------|
| 1. \( <x^b \leq 2 \)                |                               | 1.435 | 1.402  | 1.01         | 1.425                | 0.286             |
| 2. \( <x^b \leq 4 \)                |                               | 2.834 | 2.760  | 2.01         | 2.814                | 0.564             |
| 4. \( <x^b \leq 8 \)                |                               | 5.425 | 5.197  | 4.01         | 5.385                | 1.065             |
| 8. \( <x^b \leq 16 \)               |                               | 10.60 | 10.05  | 8.01         | 10.522               | 2.093             |
| 16. \( <x^b \leq 32 \)              |                               | 20.13 | 18.98  | 16.01        | 19.975               | 3.759             |
Table 1 corresponds to the distributions of Fig. 7. It indicates that forecast event frequencies decrease monotonically with intensity. Therefore, the mode estimator simply refers to the lowest bound of intensity within the threshold of \( x_b \) events. Also, the median that is the midpoint of the probability distribution tends to be smaller than the mean. The expectation (first moment) is smaller but much closer to the mean.

Table 2 corresponds to the distributions of Fig. 8 and represents the conditional distribution of a posteriori error. The tail toward the left-hand side of the distributions is associated with underestimation of rain intensities or misses. Similarly, the distribution at the right-hand side demonstrates the frequency of the overestimation or false alarms. As can be seen in Fig. 8, distributions are asymmetrical, and the lower forecast intensity events have longer tails toward the left. In such cases, mean and expectation show a smaller estimation than mode and median at (1. \( < x_b \leq 2. \) ). The difference between estimators reduces at greater event intensities until (16. \( < x_b \leq 32. \) ) where the distribution becomes more normal.

Table 3 corresponds to the distributions of Fig. 9, representing the standard deviation of spatially sampled local variances. The distributions at lower intensity are similar to lognormal distributions and gradually become more normal at higher intensities. The estimators show a good agreement between mean and expectation while mode and median point to a smaller estimation.

### Table 2. Estimation of the expected conditional standard deviation of the error within the binned thresholds of the background.

| Binned forecast \( x_b \) (mm h\(^{-1}\)) | Mean | Median | Mode (±0.1) | Expectation (±0.05) | Std dev \( \sigma (\epsilon | x_b \) ) |
|---|---|---|---|---|---|
| 1. \( < x_b \leq 2. \) | 0.546 | 1.171 | 1.2 | 0.38 | 2.293 |
| 2. \( < x_b \leq 4. \) | 1.723 | 2.361 | 2.2 | 1.56 | 2.662 |
| 4. \( < x_b \leq 8. \) | 3.947 | 4.552 | 4.3 | 3.77 | 3.157 |
| 8. \( < x_b \leq 16. \) | 8.677 | 9.03 | 8.4 | 8.48 | 4.039 |
| 16. \( < x_b \leq 32. \) | 17.85 | 17.71 | 16.5 | 17.62 | 5.906 |

b. Using the Bayesian model to predict the errors

We postulated that a relation existed between the rain intensity in the background and its innovation error variances. It is now possible to investigate the correlation between background and innovation. Figure 10 summarizes the best estimates from Tables 1–3. Results suggest that a linear regression can be fitted to estimate the unobserved quantities. The fitted regression provides a linear statistical model:

\[
\epsilon (\epsilon | x_b) = a_1 x_b + b_1, \quad \sqrt{\text{Var}_x(\epsilon | x_b)} = a_2 x_b + b_2. \tag{9}
\]

The unobserved variables \( \epsilon (\epsilon | x_b) \) and \( \sqrt{\text{Var}_x(\epsilon | x_b)} \) are both modeled using linear functions of \( x_b \). Regarding quantifying the covariance between innovation-based variables and background, the slope of the regression lines in each case indicates the degree of covariance between that random variable and the background:

\[
a_1 = \frac{\text{Cov}(\epsilon, x_b)}{\text{Var}(x_b)}, \quad a_2 = \frac{\text{Cov}([\sqrt{\text{Var}_x(\epsilon)}], x_b)}{\text{Var}(x_b)}. \tag{10}
\]

The empirical values of \( a_1, b_1, a_2, \) and \( b_2 \) are presented in Table 4 for each estimator. The distributions and the choice of the estimator determine the linear constants. Therefore, every forecast will have its characteristic statistical model that depends on the climatology of innovations. All of the above analysis was repeated for 480 h of WRF-based forecast and final statistical parameters are presented in Table 5.

To obtain the variance of error, we adopt the law of total variance and use the conditional variances. The total variance of error is equal to the expected variance of the error at a given background \( x_b \) plus the variation in estimating the error for the same given background. Note that both \( \text{Var}(\epsilon | x_b) \) and \( \sqrt{\text{Var}_x(\epsilon | x_b)} \) are random variables that fluctuate with \( x_b \). By fitting a statistical

### Table 3. Estimation of the expected conditional standard deviation of the error within the binned thresholds of the background.

| Binned forecast \( x_b \) (mm h\(^{-1}\)) | Mean | Median | Mode (±0.08) | Expectation (±0.04) | Std dev \( \sigma [\sqrt{\text{Var}_x(\epsilon | x_b)} ] \) |
|---|---|---|---|---|---|
| 1. \( < x_b \leq 2. \) | 1.327 | 0.914 | 0.55 | 1.25 | 1.312 |
| 2. \( < x_b \leq 4. \) | 1.790 | 1.332 | 0.98 | 1.72 | 1.473 |
| 4. \( < x_b \leq 8. \) | 2.503 | 1.992 | 1.66 | 2.42 | 1.679 |
| 8. \( < x_b \leq 16. \) | 3.937 | 3.496 | 2.90 | 3.85 | 2.017 |
| 16. \( < x_b \leq 32. \) | 6.175 | 5.914 | 5.83 | 6.10 | 2.091 |
model, it becomes possible to predict both of these random variables as a function of the background $x^b$:

$$\text{Var}(\epsilon) = \epsilon \{ \text{Var}_{\epsilon}(\epsilon | x^b) \} + \text{Var}(\epsilon \{ \epsilon | x^b \}) . \tag{11}$$

We used linear regression for prediction of the conditional expectations. Therefore, we replaced the expected conditional variance by expected value of spatially sampled conditional variance and plugged in the constants of the linear model from Table 4 for GEM-based and from Table 5 for WRF-based innovations. The following equation is the predictive rainfall error variance estimator as a function of the background and relies on a posteriori distributions from a climatological sample pool:

$$\text{Var}(\epsilon) = (a_2 x^b + b_2)^2 + a_1^2 \text{Var}(x^b). \tag{12}$$

To show the application of this formula, we set up an experiment. At a given hour, we used the forecasts and observation to calculate the actual ensemble-based and innovation-based variances. Then, we used the predictive method in (12) to estimate the same variance field from a single background forecast. The results are filtered by 2D-DCT and shown in Fig. 11. Parameters used for prediction were $a_1 = 1.011$, $a_2 = 0.472$, and $b_2 = 0.131$ according to the mode estimator from Table 5 for WRF-based innovations. WRF-based forecast and observation were from the forecast and observation at 0300 UTC 18 April 2008. The variance predicted by the linear regression method shows good agreement with the variance fields of the ensembles and innovations.

In our next study, we blend the GEM-based rainfall forecast with radar extrapolations in a 2D-VAR scheme.

![Fig. 10](image-url) (left) Expected conditional error and (right) expected standard deviation of the conditional error as a function of background values from GEM-based forecasts. Different estimators are considered for both error and variance best estimates.

### Table 4. The statistical model of linear regression fitted to best estimates of a posteriori distributions of the GEM-based innovations.

|          | $\epsilon(\epsilon | x^b)$ | $\epsilon\{\text{Var}(\epsilon | x^b)\}$ |
|----------|-------------------------------|---------------------------------|
| Mean     | 0.928                         | 0.258                           |
| Median   | 0.942                         | 0.285                           |
| Mode     | 1.021                         | 0.348                           |
| Expectation | 0.931                      | 0.260                           |

### Table 5. The statistical model of linear regression fitted to best estimates of a posteriori distributions of the WRF-based innovations.

|          | $\epsilon(\epsilon | x^b)$ | $\epsilon\{\text{Var}(\epsilon | x^b)\}$ |
|----------|-------------------------------|---------------------------------|
| Mean     | 0.962                         | 0.281                           |
| Median   | 0.954                         | 0.330                           |
| Mode     | 1.011                         | 0.472                           |
| Expectation | 0.962                      | 0.282                           |
To do so, we need the error variance estimations in the absence of ensembles and before having observations. This method allows for a prognostic estimation of the error variance of rainfall.

5. Summary and conclusions

The description of rainfall ECM requires specifying the heterogeneous variance and local correlation length scales. The meaning of length scales depends on the definitions and assumptions about the correlation functions. We presented the derivation of local error correlation length scales from the BPB formula (Pereira and Berre 2006) for the innovation-based errors and the formula of Michel (2013) for the ensemble-based errors. The comparison between innovation-based and ensemble-based approaches showed an agreement on the average pattern of the variance field. However, the local
correlations had more contrast in their spatial patterns. The meridional and zonal components of the innovation-based correlations had more variations in space compared to those of the ensemble-based correlations. The sensitivity to the spatial gradients causes the error correlation estimations to vary significantly in the two estimation methods. It was also found that the length scale from the ensemble-based method is generally larger than that from the innovation-based method.

The prognostic variance estimation required an analysis of a large sample pool to obtain the error distribution. However, once the coefficients of the linear approximation are set, the background is enough to produce a low-cost error variance field. If the ensemble is not produced or the innovations are not available, this method can generate a good approximation of the error variances.

In conclusion, rainfall error covariance matrix requires special treatment regarding bias, variance, and correlation estimations. Results that we presented depend on the accuracy of the radar mosaic as the truth for verifications and formation of the innovation errors. Because of the specific settings of NWP forecasts and the limited length of the analyzed dataset, the intercomparison between GEM-based and WRF-based errors cannot be regarded as a definitive measure of accuracy or skill of these NWP models in general. Nevertheless, we relied on the innovations for error estimations because the NWP forecasts still have larger errors compared to radar observations.

In practical applications, a dynamic and heterogeneous error correlation matrix can be limited by conventional error correlation modeling methods due to the lack of flexibility, numerical instability, or heavy calculation costs. As further research suggestions, it is desired to investigate the practical aspects of error correlation modeling methods that are well suited for handling case-dependent anisotropic and heterogeneous error correlations of rainfall. This would be the subject of our companion study in a future paper.

Acknowledgments. This study was mainly done as part of Majid Fekri’s Ph.D. thesis work at McGill University, supported by the NSERC/Hydro-Québec industrial research program. We acknowledge the support and guidance of Luc Fillion from Environment Canada. We are also thankful to the Center for Analysis and Prediction of Storms, University of Oklahoma, for the ensemble forecast data.

REFERENCES


Fekri, K., 2015: Towards a better understanding of rain forecast error and skill. Department of Atmospheric and Oceanic Sciences, McGill University, 239 pp.


Observations of Clouds and Precipitation, Reading, United Kingdom, ECMWF–JCSDA, 121–132.


——, and S. Ricci, 2003: Constructing a background-error correlation model using generalized diffusion operators. Proc. ECMWF Seminar Series on “Recent Developments in Atmospheric and Ocean Data Assimilation,” Reading, United Kingdom, ECMWF, 8–12.

