Evaluation of ensemble precipitation forecasts generated through postprocessing in a Canadian catchment

Sanjeev K. Jha[1+], D. L. Shrestha[2], T. Stadnyk[1], P. Coulibaly[3]

[1]Department of Civil Engineering, University of Manitoba, Winnipeg, R3T 5V6, Canada
[2]Commonwealth Science and Industrial Research Organization, Clayton South Victoria, 3169, Australia
[3]Department of Civil Engineering, McMaster University, Hamilton, L8S 4L7, Canada
[+] Now at Indian Institute of Science Education and Research Bhopal, Madhya Pradesh, 462066, India

Correspondence to: Sanjeev K. Jha (jha.sanj@gmail.com)

Formatted: Superscript

Abstract. Flooding in Canada is often caused by heavy rainfall during the snowmelt period. Hydrologic forecast centers rely on precipitation forecasts obtained from numerical weather prediction (NWP) models to enforce hydrological models for streamflow forecasting. The uncertainties in raw quantitative precipitation forecasts (QPFs) are enhanced by physiography and orography effect over diverse landscape, particularly in the western catchments of Canada. A Bayesian post-processing approach called rainfall-post processing (RPP), developed in Australia (Robertson et al., 2013;Shrestha et al., 2015), has been applied to assess its forecast performance in a Canadian catchment. Raw QPFs obtained from two sources, Global ensemble forecasting system (GEFS) Reforecast 2 project from National Centers for Environmental Protection (NCEP), and Global deterministic forecast system (GDPS) from Environment and Climate Change Canada (ECCC) are used in this study. The study period from Jan 2013 to Dec 2015 covered a major flood event in Calgary, Alberta, Canada. Post-processed results show that the RPP is able to remove the bias, and reduce the errors ~~continuous ranked probability score~~ of both GEFS and GDPS forecasts. Ensembles generated from the RPP ~~better depict~~ reliably quantify the forecast uncertainty.

1 Introduction

Quantitative precipitation forecasts (QPFs) obtained from ~~Numerical~~ numerical ~~Weather~~ weather ~~Prediction~~ prediction (NWP) models are one of the main inputs to hydrological models when used for streamflow forecasting (Ahmed et al., 2014;Coulibaly, 2003;Cuo et al., 2011;Liu and Coulibaly, 2011). A deterministic forecast, representing a single state of the weather, is unreliable due to known errors associated with approximate simulation of atmospheric processes, and in defining initial conditions for a NWP model (Palmer et al., 2005). A single estimate of streamflow using a poor or high quality precipitation forecast would ~~will~~ have significant impact on decision -support, such as management of water structures, issuing warnings of pending floods or droughts, scheduling reservoir operations, etc. In recent years, there is growing interest in moving toward probabilistic forecasts, suitable for estimating the likelihood of occurrence of any future meteorological event, thus allowing water managers and emergency agencies to prepare for the risks involved during low or

high flow events, at least a few days or weeks in advance (Palmer, 2002). The precipitation forecasts, however, are constrained by major limitations surrounding the technical difficulties and computational requirements involved in perturbing initial conditions and physical parametrization of the NWP model. The QPFs, ensemble or deterministic, have to be post-processed prior to use as reliable estimates of any observations (e.g., streamflow).

In the last decade, several post-processing methods reliant on statistical models have been proposed. The basic idea is to develop a statistical model by exploiting the joint relationship between observations and NWP forecasts, estimate the model parameters using past data, and reproduce post-processed ensemble forecasts of the future (Roulin and Vannitsem, 2012;Robertson et al., 2013;Chen et al., 2014;Khajehei, 2015;Shrestha et al., 2015;Khajehei and Moradkhani, 2017;Schaake et al., 2007;Wu et al., 2011;Tao et al., 2014). The range of complexity in the post processing approaches vary from regression-based approaches to parametric or non-parametric models based on the meteorological variables (wind speed, temperature, precipitation etc.) and specific applications. Precipitation is known to have complex spatial structure and behavior (Jha et al., 2015a;Jha et al., 2015b). Thus it is much more difficult to forecast than other atmospheric variables because of nonlinearities and the sensitive processes involved in its generation (Bardossy and Plate, 1992;Jha et al., 2013). From the perspective of a hydrologic forecast center, the post-processing approach should be effective while involving few parameters. For instance, the United States National Weather Service River Forecast System has been using an Ensemble Pre-p~~P~~rocessing (EPP) technique that constructs ensemble forecasts through the Bayesian Forecasting System by correlating normal quantile transform of QPFs and observations (Wu et al., 2011). In order to instill space-time variability of precipitation forecast in ensemble, the post-processed forecast ensemble is reordered based on historical data using the Schaake-Shuffle procedure (Clark et al., 2004;Schaake et al., 2007). This pre-processing technique requires a long historical hindcast database as it relates single NWP forecasts to corresponding observations. In Australia, Robertson et al. (2013) developed a Bayesian post-processing approach called Rainfall Post-processing (RPP) to generate ~~sub daily~~ precipitation ensemble forecasts. The approach was based on combining Bayesian Joint Probability (BJP) approach of Wang et al. (2009) and Wang and Robertson (2011), along with the Schaake-Shuffle procedure (Clark et al., 2004). In contrast to EPP, the RPP approach of Robertson et al. (2013) has been described with few parameters and it can better deal ~~with the transformation and~~ zero value problems in NWP forecasts (Tao et al., 2014) and observations. The RPP approach has been successfully applied to remove rainfall forecast bias and quantify forecast uncertainty from NWP models in Australian catchments (Bennett et al., 2014;Shrestha et al., 2015).

Recent developments in post-processing techniques and advantage of generating ensembles, and thus estimating uncertainty, are well established in the literature. In an operational context, however, forecast centres in Canada tend to use deterministic forecasts in hydrologic models to obtain streamflow forecasts. The main reasons for this are the higher spatial and temporal resolution of the deterministic forecasts over the ensemble QPFs, and the associated computational complexities in dealing with ensemble members. The added advantage of using ensemble forecasts over deterministic forecasts have been addressed

in many previous studies ~~(~~e.g., (Abaza et al., 2013;Boucher et al., 2011)~~)~~. When the computational facilities are available, using a set of QPFs obtained from different NWP models run by different agencies (such as the European Centre for Medium-range Weather Forecasts (ECMWF), The Japan Meteorological Agency (JMA), The National Center for Environmental Prediction (NCEP), The Canadian Meteorological Center (CMC), etc.) seem to be a preferred choice (Ye et al., 2016;Zsótér et al., 2016;Qu et al., 2017;Hamill, 2012).

The aims of this study are to: (a) evaluate the performance of RPP in improving cold regions precipitation forecasts; (b) compare the ensembles generated from applying RPP to the deterministic QPFs obtained from GEFS and GDPS (referred to as calibrated QPF); and, (c) investigate forecast performance during an extreme precipitation event like that of 2013 in Alberta, Canada. The methodology and description of the study area and datasets are presented in Section 2. Section 3 presents results, followed by discussion and conclusion in Section 4.

2.~~3~~ Study area and datasets

The selected study area is southern Alberta, located in western Canada (Figure 1a). The Rocky Mountains are located at the Southern boundary with the USA and Western boundary with British Columbia, with the Canadian Prairie region extending toward the south-eastern portion of the province. Topography plays a major role in generating cyclonic weather systems common to Alberta. The Oldman, Bow and Red Deer ~~r~~River basins, all located at the foothills of the Canadian Rocky mountain range, are subjected to extreme precipitation events. In June 2013 a major flood occurred in this region resulting from the combined effect of heavy rainfall during mountain snowpack melt over partially frozen ground (Pomeroy et al., 2016;Teufel et al., 2016). Some river basins received 1.5 times 1:100 year rainfall, estimated to be 250 mm rain in 24 hours. The 2013 flood affected most of Southern Alberta from Canmore to Calgary and beyond, causing evacuation of around 100,000 people and a reported cost of damage of infrastructure exceeding $6CAD billion (Milrad et al., 2015). The spatial distribution of convective precipitation and orography make it difficult for any NWP model to successfully predict the summertime convective precipitation in Alberta. The NWP forecasts at the time predicted less (about half of the actual amount) rainfall during this event (AMEC, 2014).

The dataset used in this study consists of observed and forecast daily precipitation over for the period of 2013 to 2015, including the heavy precipitation event causing the major flood of 2013. Observed data were obtained from the Environment and Climate Change Canada (ECCC) precipitation gauges. Two precipitation forecasts were obtained: the Global Ensemble Forecast System (GEFS) Reforecast Version 2 data from National Centers for Environmental Protection (NCEP), and Global deterministic forecast system (GDPS) from ECCC. A description of both forecasts is presented in Table 1. The spatial resolution of GEFS ~~and GDPS~~ forecasts ~~are~~is: [0.47$^{0}_{\Delta}$ latitude, 0.47$^{0}_{\Delta}$ longitude] ~~and~~ and GDPS forecasts are ~~[0.24$^{0}_{\Delta}$approximately 50 and 25 km, respectively latitude, 0.24$^{0}_{\Delta}$ longitude] respectively.~~ ~~In the case of GEFS, there are~~

Formatted: Normal, Don't adjust space between Latin and Asian text, Don't adjust space between Asian text and numbers

Formatted: Superscript

Formatted: Superscript

Formatted: Superscript

contains eleven forecast members including one control run and ten ensembles. ~~–~~The control run uses the same model physics but without perturbing the analysis or model. The ensembles are obtained by perturbing the initial conditions slightly (WMO, 2012). ~~All forecast members are equally likely, while the ensemble mean is expected to have better skill than any individual member (Personal correspondence with Dr. Gary Bates at NOAA)~~. The forecast is available at 00 UTC at an interval of three hours for the first three days and then six hourly up to eight days. It is worth pointing here that in the GEFS data, the forecasts at hours 3, 9, 15, and 21 are three hour accumulations, whereas the forecasts at 6, 12, 18, and 24 hours are six hours accumulations for forecasts valid for days 1 to 3. In order to obtain a 24-hour (daily) forecast for days 1, 2, and 3, we need to consider the summation of forecasts valid at hours 6, 12, 18, and 24 for a given day. For days 4 and 5, forecasts are only available for 6, 12, 18, and 24 hours (i.e., there is no forecast for the 3-hour accumulation). The control run of GEFS for a period of three years (01/01/2013 to 31/12/2015) with lead-time of 5 days are used in the present analysis.

GDPS ~~P~~precipitation forecasts from the Canadian NWP model, Global Environmental Multi-scale model (GEM) are obtained from ECCC~~, by request~~. ~~For the global NWP, the operational meteorological prediction system of ECCC relies on an ensemble Kalman filter based on a data assimilation technique that produces 20 ensemble members at a spatial resolution of 100 km, while the output from operational deterministic forecast system (~~GDPS~~)~~ forecasts ~~is~~are available at approximately ~~25 km~~[0.24^0 latitude, 0.24^0 longitude] spatial resolution at an interval of three hours for forecast lead times up to and beyond two weeks. Precipitation forecasts are accumulations from the start to the forecast period. To obtain a forecast for a specific day, let's say day 2, the precipitation forecast at the end of day 1 has to be subtracted from the precipitation forecast at the end of day 2. Three years of continuous GDPS forecasts from 01/01/2013 to 31/12/2015 with lead-times of 5 days at 00 UTC are used in the present analysis.

There are three major rivers passing through the study area: Bow ~~River~~, Oldman ~~River~~, and Red Deer Rivers (Figure 1b). ~~Peel et al. (2007) All the three river basins are part of the South Saskatchewan River Basin which flows eastward towards Canadian prairies. The combined watershed area is approximately 101,720 km^2 (AEP, 2017).~~ Based on the world map of Peel et al. (2007), the climate of the study area is classified as warm summer humid continental. The Köppen-Geiger classification system presented for Canada in Delavau et al. (2015) shows that ~~the~~our study area falls within the KPN42 (Dfb – snow, fully humid precipitation, warm summer), KPN43 (Dfc – snow, fully humid precipitation, cool summer) and KPN 62 (ET polar tundra). All the three river basins are part of the South Saskatchewan River ~~B~~basin which flows eastward towards Canadian prairies. The combined basin area is approximately 101,720 km^2 (AEP, 2017). For the purpose of hydrological prediction, the River Forecast Centre in Alberta uses fifteen subcatchments (marked with numbers 1 to 15 in Figure 1b) to delineate the study area, with drainage areas as indicated in Table 2.

The distribution of precipitation gauges and forecast locations are uneven in the various subcatchments (Figure 1b). For hydrological modeling purposes, ~~data are required at the centroid of a~~average precipitation over each subcatchment~~,~~

Field Code Changed

Formatted: Font: 10 pt, Font color: Auto

Formatted: Font: 10 pt, Font color: Auto

Formatted: Font: 10 pt, Font color: Auto

~~therefore the average of observed precipitation at the centroid of a subcatchment~~ is calculated using an inverse-distance weighting (IDW) method (Shepard, 1968) considering four neighbouring gauges. Subcatchment 2 receive~~d~~s the highest subcatchment-averaged annual precipitation, while subcatchment 13 receive~~d~~s the lowest average annual precipitation during ~~all~~ the three year study~~s~~ period (Table 2). In each subcatchment, an area-weighted forecast is calculated by considering the portion of the forecast grid that overlaps with the subcatchment.

Figure 2 shows the comparison of ~~weighted area~~ raw QPFs and ~~subcatchment averaged~~ observed precipitation in subcatchments 10 and 11 for GEFS and GDPS with a lead-time of one day for 2013. The large peak observed ~~peaks in May large peak observed~~ (Figures 2a to 2d) ~~indicates~~is the result of a ~~the~~ major precipitation event responsible for severe flooding in Alberta in ~~May~~June 2013. Figure 2 indicates that there is ~~always~~ a substantial bias between the raw QPFs and observations. Raw QPFs from GEFS ~~seems to forecast this peak precipitation value quite~~reasonably well~~, while~~and ~~GDPS~~ consistently underestimate~~s~~ peak events and~~s as well as~~ medium precipitation amounts, which is of concern to hydrologic forecast centres predicting streamflow peak volume and timing~~, however, in subcatchment 11, it shows an underestimation of precipitation amount. There is a relatively larger bias observed in the raw QPFs from GDPS (Figures 2c and 2d).~~

3~~2~~ Methodology

2~~3~~.1 Post-processing approach

We use the RPP to post-process the precipitation forecasts. The RPP was developed by Robertson et al. (2013) and successfully applied to a range of Australian catchments (Shrestha et al., 2015). Detailed descriptions of the RPP can be found in above references, here we present a brief overview of the method.

The input to the post processing approach is observation (z_O) and raw QPF (z_{rf}). A log-sinh transformation is applied to both observations and raw precipitation forecast:

$$\hat{z}_O = \frac{1}{\beta_O} \ln(\sinh(\alpha_O + \beta_O z_O)) \quad (1)$$

$$\hat{z}_{rf} = \frac{1}{\beta_f} \ln(\sinh(\alpha_f + \beta_f z_{rf})) \quad (2)$$

where $\hat{z}_O$ and $\hat{z}_{rf}$ are transformed observation and raw forecast; α_O and β_O are parameters used in the transformation of z_O; α_f and β_f are parameters used in transformation of z_{rf}. It is assumed that the transformed variables ($\hat{z}_O$ and $\hat{z}_{rf}$) follow a bivariate normal distribution $p(\hat{z}_O, \hat{z}_{rf}) \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, in which $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ are defined:

$$\boldsymbol{\mu} = \begin{bmatrix} \mu_{\hat{z}_O} \\ \mu_{\hat{z}_{rf}} \end{bmatrix} \text{ and} \tag{3}$$

$$\boldsymbol{\Sigma} = \begin{bmatrix} \sigma^2_{\hat{z}_O} & \rho_{\hat{z}_O \hat{z}_{rf}} \sigma_{\hat{z}_O} \sigma_{\hat{z}_{rf}} \\ \rho_{\hat{z}_O \hat{z}_{rf}} \sigma_{\hat{z}_O} \sigma_{\hat{z}_{rf}} & \sigma^2_{\hat{z}_{rf}} \end{bmatrix} \tag{4}$$

where $\mu_{\hat{z}_O}$ and $\sigma_{\hat{z}_O}$ represent the mean and standard deviation of $\hat{z}_O$ respectively; $\mu_{\hat{z}_{rf}}$ and $\sigma_{\hat{z}_{rf}}$ represent the mean and standard deviation of $\hat{z}_{rf}$ respectively; $\rho_{\hat{z}_O \hat{z}_{rf}}$ is the correlation coefficient between $\hat{z}_O$ and $\hat{z}_{rf}$. Thus, there are nine parameters ($\alpha_O, \beta_O, \mu_{\hat{z}_O}, \sigma_{\hat{z}_O}, \alpha_f, \beta_f, \mu_{\hat{z}_{rf}}, \sigma_{\hat{z}_{rf}}, \rho_{\hat{z}_O \hat{z}_{rf}}$) to model the joint distribution of raw QPF and observation.

We infer a single set of model parameters that maximizes the likelihood of posterior parameter distribution using the shuffled complex evolution algorithm (Duan et al., 1994). All model parameters are reparametrized to ease the parameter inference. Once the parameters are inferred, the forecast is estimated using the ~~parameters and~~ bivariate normal distribution conditioned on the raw forecast. The random sampling from the conditional distribution generates the ensemble of forecasts. The forecast values are then transformed to the original space using inverse of Equations (1) and (2).

Since the forecasts are generated at each location for each lead-time separately, the space-time correlation in the ensemble members will be unrealistic. The Schaake shuffle (Clark et al., 2004) is then applied ~~at the same time step as the forecast~~ to adjust the space-time correlations in the ensemble similar to what was observed in the historically observed data. The Schaake shuffle calculates the rank in the observed data and ~~instills~~preserves the same rank in the sorted, new ensemble forecast. ~~The~~Our application of the Schaake shuffle ~~for instilling temporal correlation~~ is briefly described here.

1. For a given forecast date, an observation sample (date and amount of data) of the same size as ~~of~~ that of the ensemble is selected from the historical observation period~~.~~;
2. The observation sample data for each lead time ~~is~~are ranked. Similarly, the data from the forecast ensemble for each lead time ~~is~~are ranked~~.~~;
3. A date from the observation sample is ~~selected~~ randomly selected and the ranks of the observation data for the selected date for all lead times are identified~~.~~;
4. For a given lead time, we select the forecast (from the forecast ensemble) that has same rank as that of the selected observation~~.~~;
5. In order to construct an ensemble trace across all lead times, step 3 is repeated for all lead times~~.~~; and
6. Steps 3 to 5 are repeated as many times as the size of ensembles. ~~for ensemble size times.~~

The above procedure is extended ~~to instill~~for both temporal and spatial correlation in this study. ~~The Schaake shuffle calculates the rank in the observed data and instills the same rank in the sorted new ensemble forecast. This procedure is repeated to obtain a realistic space-time correlation in the forecasts. The forecast values are then transformed to the original space using inverse of Equations (1) and (2).~~

Formatted: Numbered + Level: 1 + Numbering Style: 1, 2, 3, … + Start at: 1 + Alignment: Left + Aligned at: 0.63 cm + Indent at: 1.27 cm

Formatted: Normal

An important feature of the RPP is the treatment of (near) zero precipitation values, which are treated as censored data ~~(i.e. having values less than or equal to (near) zero)~~ in the parameter inference. This enables the use of the continuous bivariate normal distribution for a problem which is otherwise solved using a mixed discrete-continuous probability distribution (e.g. Wu et al. (2011)).

~~**2.2 Statistical treatment of forecasts**~~

~~**Because of the short record of data, few extreme events or outlier may significantly affect the verification scores. Therefore it is desirable to understand the effect of the sampling variability on the verification scores. Accounting for sampling variability in calculations of verification scores adds confidence that results are robust and likely to apply under operational conditions (Shrestha et al., 2015). We calculate sampling uncertainty around raw and calibrated QPFs through a bootstrap procedure (e.g. Shrestha et al. (2013)). The first 1095 pairs (three years of data) of forecast observation are sampled from the original forecast observation pairs, with replacement and verification scores calculated. These steps are repeated 5000 times to obtain a distribution of the verification score, from which 5% and 95% confidence interval are calculated.**~~ **3.2~~~~1~~ Verification of the post-processed forecasts**

We assess the processed forecast in terms of deterministic metrics, such as percent bias defined as percent deviation from the observations (Bias (%)), mean absolute error (MAE in mm), and a probabilistic metric, continuous rank probability score (CRPS) at each site for the forecast period (t). The percent bias is estimated:

$$Bias = \frac{\sum_{1}^{t} z_f - \sum_{1}^{t} z_o}{\sum_{1}^{t} z_o} \cdot 100 \quad (5)$$

where z_f could be either raw (z_{rf}) or post-processed forecast (z_{pf}), and z_O represents observation.

MAE measures the closeness of the forecasts and observations over the forecast period.

$$MAE = \frac{1}{t}\sum_{1}^{t}\left|z_f - z_o\right| \quad (6)$$

In case of ensemble forecast, we use the mean value of forecasts in the calculation of both bias and MAE. Low value of both bias and MAE indicate that the forecasts are closer to observations. Percent bias close to zero indicates that forecasts are unbiased.

The CRPS is a probabilistic measure to relate the cumulative distribution of the forecasts and the observations:

$$CRPS = \int_{-\infty}^{\infty} (F_{z_f}(t) - F_{z_o}(t))^2 \, dt \quad (7)$$

where F_{z_f} is the cumulative distribution function of ensemble forecasts; F_{z_o} is the cumulative distribution function of observation, which turns out to be a Heaviside function (= 1 for values greater than a value, otherwise 0). In case of deterministic forecast, CRPS reduces to MAE. The forecast is considered better when the CRPS values are close to zero.

The relative operating characteristic (ROC) curves are used to assess the forecast's ability to discriminate precipitation events in terms of hit rate and false alarm rate. Given a precipitation threshold, hit rate refers to probability of forecasts that detected events smaller or larger in magnitude than the threshold; false alarm rate refers to probability of erroneous forecasts or false detection (Atger, 2004;Golding, 2000). If the ROC curves (plot between hit rate versus false alarm rate) approach the top left corner of the plot, the forecast is considered to have greater ability to discriminate precipitation events. The discrimination ability of the forecast is considered low when the ROC curves are close to the diagonal.

To compare the spread in forecast ensembles against the observations, we perform spread-skill analysis by plotting ensemble spread versus forecast error (e.g., Nester et al. (2012)). The ensemble spread is defined as mean absolute difference between the ensemble members and the mean. The absolute difference between the observation and the ensemble mean is defined as forecast error. For each lead time in the cross-validation period, we compute the ensemble spread and forecast error for 1000 ensemble forecasts, sort them in increasing order, group the values in 10 classes, and calculate an average spread and error in each class.

3.3 Statistical treatment of forecasts

Because of the short record of data (3 years only), few extreme events may significantly affect the verification scores. Therefore it is desirable to understand the effect of the sampling variability on the verification scores. Accounting for sampling variability in calculationsing ofthe verification scores adds confidence that results are robust and likely to apply under operational conditions (Shrestha et al., 2015). We calculate sampling uncertainty through a bootstrap procedure (e.g. Shrestha et al. (2013)). The first 1095 pairs (three years of data) of forecast-observation are sampled with replacement from the original forecast-observation pairs, with replacement and verification scores ~~(discussed in Section 3.1 below)~~ calculated (discussed in Section 3.1 below). These steps are repeated 5000 times to obtain a distribution of the verification score, from which 5% and 95% confidence interval are calculated.

~~**2.3 Study area and datasets**~~

~~The selected study area is southern Alberta, located in western Canada (Figure 1a). The Rocky Mountains are located at the Southern boundary with the USA and Western boundary with British Columbia, with the Canadian Prairie region extending~~

toward the south eastern portion of the province. Topography plays a major role in generating cyclonic weather systems common to Alberta. The Oldman, Bow and Red Deer river basins, all located at the foothills of the Canadian Rocky mountain range, are subjected to extreme precipitation events. In June 2013 a major flood occurred in this region resulting from the combined effect of heavy rainfall during mountain snowpack melt over partially frozen ground (Pomeroy et al., 2016;Teufel et al., 2016). Some river basins received 1.5 times 1:100 year rainfall, estimated to be 250 mm rain in 24 hours. The flood affected most of Southern Alberta from Canmore to Calgary and beyond, causing evacuation of around 100,000 people and a reported cost of damage of infrastructure exceeding $6CAD billion (Milrad et al., 2015). The spatial distribution of convective precipitation and orography make it difficult for any NWP model to successfully predict the summertime convective precipitation in Alberta. The NWP forecasts at the time predicted less (about half of the actual amount) rainfall during this event (AMEC, 2014).

The dataset used in this study consists of observed and forecast daily precipitation over for the period of 2013 to 2015, including the heavy precipitation event causing the major flood of 2013. Observed data were obtained from the Environment and Climate Change Canada (ECCC) precipitation gauges. Two precipitation forecasts were obtained: the Global Ensemble Forecast System (GEFS) Reforecast Version 2 data from National Centers for Environmental Protection (NCEP), and Global deterministic forecast system (GDPS) from ECCC. A description of both forecasts is presented in Table 1. The spatial resolution of GEFS and GDPS forecasts are approximately 50 and 25 km, respectively. In the case of GEFS, there are eleven forecast members including one control run and ten ensembles. All forecast members are equally likely, while the ensemble mean is expected to have better skill than any individual member (Personal correspondence with Dr. Cary Bates at NOAA). The forecast is available at 00 UTC at an interval of three hours for the first three days and then six hourly up to eight days. The control run of GEFS for a period of three years (01/01/2013 to 31/12/2015) with lead time of 5 days are used in the present analysis.

Precipitation forecasts from the Canadian NWP model, Global Environmental Multi-scale model (GEM) are obtained from ECCC, by request. For the global NWP, the operational meteorological prediction system of ECCC relies on an ensemble Kalman filter based on a data assimilation technique that produces 20 ensemble members at a spatial resolution of 100 km, while the output from operational deterministic forecast system (GDPS) is available at 25 km resolution for forecast lead times up to and beyond two weeks. Three years of continuous GDPS forecasts from 01/01/2013 to 31/12/2015 with lead time of 5 days at 00 UTC are used in the present analysis.

There are three major rivers passing through the study area: Bow River, Oldman River, and Red Deer River (Figure 1b). For the purpose of hydrological prediction, the River Forecast Centre in Alberta uses fifteen subcatchments (marked with numbers 1 to 15 in Figure 1b) to delineate the study area, with drainage areas as indicated in Table 2.

~~The distribution of precipitation gauges and forecast locations are uneven in the various subcatchments (Figure 1b). For hydrological modeling purposes, data are required at the centroid of a subcatchment, therefore the average of observed precipitation at the centroid of a subcatchment is calculated using an inverse distance weighting (IDW) (Shepard, 1968) considering four neighboring gauges. Subcatchment 2 received the highest subcatchment averaged annual precipitation while subcatchment 13 received lowest average annual precipitation during all the three years period (Table 2). In each subcatchment, an area weighted forecast is calculated by considering the portion of the forecast grid that overlaps with the subcatchment.~~

~~Figure 2 shows the comparison of weighted area raw QPFs and subcatchment averaged observed precipitation in subcatchments 10 and 11 for GEFS and GDPS with a lead time of one day for 2013. The large peak observed (Figures 2a to 2d) indicates the major precipitation event responsible for severe flooding in Alberta in May 2013. Figure 2 indicates there is always a bias between the raw QPFs and observations. Raw QPFs from GEFS seems to forecast this peak precipitation value quite well, however, in subcatchment 11, it shows an underestimation of precipitation amount. There is a relatively larger bias observed in the raw QPFs from GDPS (Figures 2c and 2d).~~

~~3.4~~ ~~2.4~~ Experimental set up

Post-processing is applied to precipitation forecasts in 15 subcatchments, making use of the subcatchment-averaged precipitation forecast data for the total study duration (i.e., 2013 to 2015 for GEFS and GDPS), for each day~~s~~ of forecast at 00 UTC up to ~~the~~ a lead-time of 5 days. We ~~applied~~ apply a leave-one-month~~-~~out cross validation approach. The simulation runs in two modes: inference and forecast. In the inference mode, for example, ~~and in the case of GEFS data:~~ 36 months of precipitation forecast and observations are used to estimate the model parameters. Once parameters are estimated, the simulation runs in forecast mode to generate 1000 ensembles (or realizations) of precipitation forecast for the month that was left out of the calibration. The process is repeated 36 times ~~(i.e., three years of data)~~ to generate forecasts for 2013 to 2015. ~~In the case of GDPS data, the same procedure is applied.~~

~~3~~4 Results

~~3.1 Verification of the post-processed forecasts~~

~~We assess the processed forecast in terms of deterministic metrics, such as percent bias defined as percent deviation from the observations (Bias (%)), mean absolute error (MAE in mm), and a probabilistic metric, continuous rank probability score (CRPS) at each site for the forecast period (t). The percent bias is estimated:~~

$$Bias = \frac{\sum_{1}^{t} z_f - \sum_{1}^{t} z_o}{\sum_{1}^{t} z_o} \cdot 100 \quad (5)$$

where z_f could be either raw (z_{rf}) or post processed forecast (z_{pf}), and z_O represents observation.

MAE measures the closeness of the forecasts and observations over the forecast period.

$$MAE = \frac{1}{t}\sum_{1}^{t}\left|z_f - z_o\right| \quad (6)$$

In case of ensemble forecast, we use the mean value of forecasts in the calculation of both bias and MAE. Low value of both bias and MAE indicate that the forecasts are closer to observations. Percent bias close to zero indicates that forecasts are unbiased.

The CRPS is a probabilistic measure to relate the cumulative distribution of the forecasts and the observations:

$$CRPS = \int_{-\infty}^{\infty} (F_{z_f}(t) - F_{z_o}(t))^2 \, dt \quad (7)$$

where F_{z_f} is the cumulative distribution function of ensemble forecasts; F_{z_o} is the cumulative distribution function of observation, which turns out to be a Heaviside function (= 1 for values greater than a value, otherwise 0). In case of deterministic forecast, CRPS reduces to MAE. The forecast is considered better when the CRPS values are close to zero.

The relative operating characteristic (ROC) curves are used to assess the forecast's ability to discriminate precipitation events in terms of hit rate and false alarm rate. Given a precipitation threshold, hit rate refers to probability of forecasts that detected events smaller or larger in magnitude than the threshold; false alarm rate refers to probability of erroneous forecasts or false detection (Atger, 2004;Golding, 2000). If the ROC curves (plot between hit rate versus false alarm rate) approach the top left corner of the plot, the forecast is considered to have greater ability to discriminate precipitation events. The discrimination ability of the forecast is considered low when the ROC curves are close to the diagonal.

To compare the spread in forecast ensembles against the observations, we perform spread-skill analysis by plotting ensemble spread versus forecast error (e.g., Nester et al. (2012)). The ensemble spread is defined as mean absolute difference between the ensemble members and the mean. The absolute difference between the observation and the ensemble mean is defined as forecast error. For each lead time in the cross-validation period, we compute the ensemble spread and forecast error for 1000

~~ensemble forecasts, sort them in increasing order, group the values in 10 classes, and calculate an average spread and error in each class.~~

~~4~~3.12 Evaluation of calibrated QPFs

Figures 3 presents the percent bias and CRPS in five subcatchments for both GEFS and GDPS forecasts~~, based on their unique characteristics~~. Out of 15 subcatchments considered in this study, the maximum and minimum total annual subcatchment-averaged precipitation ~~respectively~~ for years 2013 to 2015 occurred in subcatchments 2 and 13, respectively (see Table 2); subcatchments 7 and 8 have minimum and maximum size, respectively. The four selected subcatchments covered the middle and southern portions of the study area, therefore we include subcatchment 4 to facilitate discussion on the performance of calibrated QPFs in the northern portion of the study area ~~as well~~. The percent bias ~~plot of~~for all 15 subcatchments ~~are~~is pr~~esented~~ovided in ~~the~~ Supplementary Figure SF-1(a).

Based on visual inspection of bias plots (Figure 3a to 3e), the bias in the calibrated QPF is close to zero in almost all five subcatchments (except at lead time 4 in the subcatchment 8) ~~in almost all five subcatchments~~, indicating that the RPP is able to reduce the bias in the raw QPF. As shown in Supplementary Figure SF-1(b), the calibrated RPP doesn't capture a peak precipitation event at lead time 4 in the subcatchment 8 which resulted into large bias. The anomaly can be attributed to the fact that the subcatchment 8 has largest area and only few observation stations lie inside the subcatchment. In case of GEFS forecasts for the lead-time of one day, the raw QPFs ~~has~~ have an average bias ranging from ~~5~~ -30 % (in subcatchment 2) to around ~~2~~100 % (in subcatchment 13). In all ~~the~~ subcatchments, the bias ~~is very high for the lead-time of up~~increases slightly from day 1 to day 2, then drops in ~~to 3~~ day~~s~~ 3 and afterwards either increases or remain almost constant ~~(except subcatchment 2), and drops afterwards to a lower value for higher lead-times~~. ~~In subcatchment 2, the bias is close to zero in the raw QPF for the lead-time up to three days, and around -40% during lead-times from 4 to 5 days. Higher~~Increase in bias in the first ~~three~~ two days lead-time can be attributed to the 'spin-up' of the NWP model. Spin-up is expected to influence only the first day or two~~, however, and extending the same reasoning for larger biases up to 3 days is counter-intuitive~~. In the case of GDPS, the bias in the raw QPF is close to zero (except in subcatchments 2 and 8, where bias is negative) for the 1-day lead time, but the bias increases up to as high as ~~700~~50% (subcatchment 13) for ~~the~~ a lead-time of 5 days. The 5% and 95% confidence interval around the raw QPF also increases slightly with lead-time indicating that the forecast for lead-time of 1-~~-~~day will have higher confidence (hence a narrow shaded area), than for the latter days, which is intuitive. ~~The bias of calibrated QPFs is close to zero in all the subcatchments (except at lead-time 4 in the subcatchment 8), demonstrating the efficacy of RPP technique.~~ It may be argued that the variations in bias in different subcatchments can be attributed to topography and physiographic characteristics. It is worth pointing out that in this study, we are not considering spatial non-stationarity because the goal is to set up a simple Bayesian model that relates the subcatchment precipitation forecasts and the observations. Accounting for the topography and elevation in the probabilistic model increases the complexity

Formatted: List Paragraph, Adjust space between Latin and Asian text, Adjust space between Asian text and numbers

Formatted: Font: 10 pt, Not Italic, Font color: Auto

significantly and it is unlikely that the forecast performance will increase given the length of data used to infer the model parameters. Thus, we are not concerned with linking topography and corrections in the forecasts.

Formatted: Font color: Auto

The subcatchment-averaged CRPS of raw and calibrated QPFs ~~is~~ are shown on Figures 3f to 3j. It is worth mentioning that for the deterministic forecast, the CRPS reduces to mean-absolute error (MAE), thus the plots for raw QPF (Figures 3f to 3j) show MAE. For simplicity, we therefore refer to MAE of raw QPF as its CRPS. The CRPS estimated on the calibrated QPFs are based on 1000 ensembles generated from the RPP approach. In the case of GEFS and similar to the bias plots (Figures 3a to 3e), we notice that ~~e drastic change in~~ the CRPS first increases then decreases and then ~~keep~~ increas~~inges~~~~pattern~~ again after a lead-time of 4 days in the raw QPF. The CRPS based on the calibrated QPFs, however, consistently increases (except for subcatchment 2) indicating that as the lead-time increases, deviation between forecasts and observation will be larger. In case of GDPS, the CRPS of raw QPF ~~is around~~varies between 1 to 2.2 mm/day for lead-time of one day, almost linearly increasing up to 3~~10~~ mm/day for forecast lead-times of 5 days. The RPP approach reduces the CRPS ~~significantly~~significantly~~to a nearly constant value (around 1 to 2.5 mm/day)~~ for each lead-time in all subcatchments (except at a lead time 4 in the subcatchment 8)~~a~~. ~~Despite the size and hydrological characteristics across the subcatchments,~~ Overall the calibrated QPF has lower CRPS than the raw QPF, which demonstrates that the RPP is able to improve the raw QPF across all lead times. The CRPS of other subcatchments are presented in ~~the~~ Supplementary Figure SF-1(b).

To assess the calibrated forecasts' ability to discriminate between low ~~precipitation events~~ (<0.2 mm) and high precipitation events (>5 mm) for all lead times, Figure 4 presents the ROC curves for years 2013 to 2015. We only present results for lead-times of 1, 3 and 5 days for calibrated GEFS and GDPS forecasts for subcatchment 11. The results of other subcatchments are presented in ~~the~~ Supplementary Figures SF-2(a) to 2(d). For GEFS (Figures 4a and 4b), the ROC curves for days 1, 3 and 5 increasingly move away from the top left corner of the plot, suggesting that forecasts for shorter lead times have slightly higher discriminative ability than those for longer lead times. GDPS shows similar ~~behavior~~behaviour (Figures 4c and 4d), indicating that f~~F~~orecasts at longer lead times are less ~~skillful~~skilful than those at shorter lead times. ~~The spread in ROC curves is larger for GDPS than GEFS, however, indicating that the skill of GDPS forecasts decreases significantly for longer lead times. Another major difference between calibrated GEFS and GDPS forecasts is that none of the ROC curves approach the diagonal for GEFS forecasts, suggesting that calibrated forecasts are always skillful; while ROC curves are very close to the diagonal for longer lead times for GDPS.~~ Both GEFS and GDPS forecasts for a lead time of 1 day suggest that the forecast discrimination is stronger for higher rainfall events (> 5 mm) where ROC curves are closer to left corner of the plot (Figures 4c and 4d) than for smaller precipitation events (< 0.2 mm).

Figure 5 indicates the forecast error versus spread of the ensembles for calibrated GEFS and GDPS forecasts with lead times of 1, 3 and 5 days for subcatchment 11. For days 3 and 5, m~~M~~ost of the points seem to fall on the diagonal (1:1 line),

suggesting good agreement between the forecast error and the spread across all the lead times. For day 1, the deviation of the points from the diagonal (1:1 line) is higher indicating larger bias for day 1 compared to days 3 and 5. ~~This indicates the calibrated GEFS and GDPS forecasts are unbiased at all lead times.~~ To explore it further, we calculated the frequency of observed data lying within 10-90% confidence boundary of calibrated QPFs. Figure 6 shows that in case of calibrated GEFS, the calculated frequency of observed data for lead time of ~~one~~1 to ~~five~~5 days varies between 0.78 to 0.88. However, for calibrated GDPS, the frequency lies between 0.87 to 0.9. Figure 6 demonstrates that as the lead-time increases, the frequency of observed data lying within the [0.1-0.9] confidence boundary is higher.

4.2~~3~~.3 Performance of calibrated QPFs during an extreme event

As mentioned in Section 2.~~3~~, a severe flood event occurred from 20th to 24th of June, 2013 in Calgary (located near the outlet of subcatchment 7, see Figure 1b). Therefore we examine subcatchment-averaged precipitation obtained from raw and calibrated QPFs against observed data. From the historical observed data, we notice that most peak precipitation events tend to occur~~red~~ over the mountains (i.e., in subcatchments 10 and 11). To consider both the peak precipitation event responsible for triggering the 2013 flood, and also the series of smaller precipitation events before and after the peak event, we select a one month period from 10th June to 10th July 2013. Results for the 1-day lead-time ~~-~~in subcatchments 10 and 11 (Figure 6) relative to observed data ~~-~~suggest ~~that~~ there were series of high precipitation events on day 10, 11 and 12~~. Compared to these high precipitation events, there was~~ with almost negligible precipitation on the remaining days relative to these peak events (with the exception of some small events on days 26 and 29). In both subcatchments 10 and 11, raw GEFS forecasts show significantly less precipitation compared to the observations ~~on~~ from days 10 to 12~~. There is, however, a close match between the raw forecast and observed precipitation on days 11 and 12~~ (see Figures ~~6a~~ 7a and ~~6b~~7b). On the remaining days, raw GEFS consistently forecasts~~-a~~ higher magnitudes of precipitation ~~than~~ relative to the observations. The raw GDPS forecast also shows significantly lower magnitudes of precipitation relative to observed during the peak~~entire~~ event (days 10 to 12; Figures ~~6c~~ 7c and Figure ~~6d~~7d). The GDPS forecast shows overprediction of a smaller event on day 26 and underprediction on day 29. For the remaining days, the raw GDPS forecast closely matches observed precipitation. The shaded area for the calibrated QPF in the case of both GEFS and GDPS indicates the range of precipitation forecasts obtained from 1000 ensemble forecast members. In both subcatchments 10 and 11, the catchment-averaged calibrated QPFs (shaded area) is able to capture peak precipitation and the smaller events (except for day 10 in calibrated GEFS).

We ~~also~~ have also evaluated the ability of the calibrated QPFs to discriminate between events and non-events for large rainfall events (>5 mm) from 10th June to 10th July 2013. The ROC curves for lead times of 1, 3, and 5 days for both calibrated GEFS and GDPS in subcatchment 11 (Figure ~~7~~8) indicate that~~, similar to Figure 4,~~ the calibrated GEFS (lead

times of 1 and 3 days) and calibrated GDPS (lead time of 1 and 5 days) have a greater ability to discriminate between events and non-events.

45 Discussion and conclusion

Based on the results presented, the RPP shows promising performance for catchments in cold and snowy climates, such as that in Western Canada. Bias free precipitation is a vital component, among other inputs, for improved streamflow forecasts from hydrological models. For raw GEFS and GDPS, the RPP approach was able to reduce the bias in the calibrated QPFs close to zero. ~~It was expected that the bias would increase with increasing lead time given forecasts for longer lead times were less certain than those for the shorter lead times.~~ The bias calculated from raw GEFS forecasts show almost similar bias, with slight variations, from lead time of one to five days. ~~Counter intuitive behavior was observed in the bias calculated from raw GEFS forecasts, where higher bias was computed in the first three days lead time, and then reduced to a constant value for the forecast following that. There was no known reason for this behaviour, even in consultation with the data providers (personal correspondence with Dr. Gary Bates and Dr. Tom Hamill at NOAA). To the best of our knowledge, GEFS data has not previously been used or tested specifically in Western Canadian catchments.~~ The GDPS forecast, however, showed an expected trend of increasing bias with increasing lead time. The advantage of applying the RPP approach was that, irrespective of the nature of the inherent bias in the raw forecasts, the calibrated QPFs were bias-free.

The calibrated QPFs have significantly reduced CRPS values in all subcatchments in both GEFS and GDPS forecasts. Furthermore, the ensembles produced from the deterministic QPF were mostly able to capture the peak precipitation events within the study area (i.e., June 2013). It is noted that in the absence of ensembles, a hydrological model would take only the raw QPF, and would therefore not forecast the resulting streamflow correctly during a major flood event. Ensemble precipitation forecasts would enable uncertainty bands to be produced around the forecast streamflow simulated from a hydrological model, thus increasing the chance of properly assessing the associated risks associated with sudden, high precipitation events.

ROC curves for calibrated QPFs showed that GEFS forecasts have greater ability to discriminate between events and non-events for both low and high precipitation across all lead times. The discrimination ability of GDPS forecasts, however, reduces significantly with increasing lead-time.

In conclusion, this study assessed the performance of a post-processing approach, RPP, developed in Australia to a catchment in Alberta, Canada. The RPP approach was applied on two sets of raw forecasts, GEFS and GDPS, obtained from two different NWP models for common periods in 2013 and 2015. In each case, 1000 post-processed forecast ensembles were created. Post-processed forecasts were demonstrated to have low bias and higher accuracy for each lead-time in 15 subcatchments covering a range of topographical conditions, from mountains to western plains, inducing different

precipitation mechanisms. Unlike raw forecasts, the post-processed forecast ensembles are able to capture peak precipitation events, which resulted in a major flood event in 2013 within the study area. Future work will involve applying RPP to other Canadian catchments, under different climatic conditions such as coastal, plains, and lake-dominated Boreal Shield, among others. The influence of the density of rain gauges, and perhaps use of a gridded reanalysis product for the observation dataset, are left for future investigations. The authors aim to test the post-processed precipitation forecasts for streamflow forecasting in different Canadian catchments as future work.

Acknowledgements

This work was supported by the Natural Science and Engineering Research Council of Canada (NSERC) and a post-doctoral fellowship under the NSERC Canadian FloodNet. The first author would also like to acknowledge the support through the Initiation grant from the Indian Institute of Science Education and Research Bhopal. We would like to extend our sincere honour to Late Dr. Peter Rasmussen who was the FloodNet Theme 3-1 project leader and actively participated in the initial discussions. Dr. Rasmussen passed away on 2nd January, 2017 and is greatly missed by all of us. Alberta Forecast Centre provided the rain gauge data and the shape files of the catchment. Environment and Climate Change Canada provided the GDPS forecast data. We would also like to thank Dr. Gary Bates at NOAA for promptly responding to our queries on GEFS data. Special thanks to Dr. Vincent Fortin from ECCC for helping us ~~in~~to properly understand~~ing~~ the raw NWP model output~~arrangement of data from NWP models~~.

References

Abaza, M., Anctil, F., Fortin, V., and Turcotte, R.: A comparison of the Canadian global and regional meteorological ensemble prediction systems for short-term hydrological forecasting, Monthly Weather Review, 141, 3462-3476, 2013.

Alberta Environment and Parks: http://www.environment.alberta.ca/apps/basins/default.aspx?Basin=8, 2017.

Ahmed, S., Coulibaly, P., and Tsanis, I.: Improved Spring Peak-Flow Forecasting Using Ensemble Meteorological Predictions, Journal of Hydrologic Engineering, 20, 04014044, 2014.

Weather forecast review project for operational open-water river forecasting: http://aep.alberta.ca/water/programs-and-services/flood-mitigation/documents/weather-forecast-review-project-for-operational-open-water-river-forecasting.pdf, 2014.

Atger, F.: Estimation of the reliability of ensemble-based probabilistic forecasts, Quarterly Journal of the Royal Meteorological Society, 130, 627-646, 2004.

Bardossy, A., and Plate, E. J.: Space-time model for daily rainfall using atmospheric circulation patterns, Water Resources Research, 28, 1247-1259, 1992.

Bennett, J. C., Robertson, D. E., Shrestha, D. L., Wang, Q., Enever, D., Hapuarachchi, P., and Tuteja, N. K.: A System for Continuous Hydrological Ensemble Forecasting (SCHEF) to lead times of 9days, Journal of Hydrology, 519, 2832-2846, 2014.

Boucher, M.-A., Anctil, F., Perreault, L., and Tremblay, D.: A comparison between ensemble and deterministic hydrological forecasts in an operational context, Advances in Geosciences, 29, 85-94, 2011.

Chen, J., Brissette, F. P., and Li, Z.: Postprocessing of ensemble weather forecasts using a stochastic weather generator, Monthly Weather Review, 142, 1106-1124, 2014.

Clark, M., Gangopadhyay, S., Hay, L., Rajagopalan, B., and Wilby, R.: The Schaake shuffle: A method for reconstructing space–time variability in forecasted precipitation and temperature fields, Journal of Hydrometeorology, 5, 243-262, 2004.

Coulibaly, P.: Impact of meteorological predictions on real-time spring flow forecasting, Hydrological Processes, 17, 3791-3801, 2003.
Cuo, L., Pagano, T. C., and Wang, Q.: A review of quantitative precipitation forecasts and their use in short-to medium-range streamflow forecasting, Journal of Hydrometeorology, 12, 713-728, 2011.
Delavau, C., Chun, K., Stadnyk, T., Birks, S., and Welker, J.: North American precipitation isotope (δ18O) zones revealed in time series modeling across Canada and northern United States, Water Resources Research, 51, 1284-1299, 2015.
Duan, Q., Sorooshian, S., and Gupta, V. K.: Optimal use of the SCE-UA global optimization method for calibrating watershed models, J. Hydrol., 158, 265-284, 1994.
Golding, B.: Quantitative precipitation forecasting in the UK, Journal of Hydrology, 239, 286-305, 2000.
Hamill, T. M.: Verification of TIGGE multimodel and ECMWF reforecast-calibrated probabilistic precipitation forecasts over the contiguous United States, Monthly Weather Review, 140, 2232-2252, 2012.
Jha, S. K., Mariethoz, G., Evans, J. P., and McCabe, M. F.: Demonstration of a geostatistical approach to physically consistent downscaling of climate modeling simulations, Water Resources Research, 49, 245-259, 2013.
Jha, S. K., Mariethoz, G., Evans, J., McCabe, M. F., and Sharma, A.: A space and time scale-dependent nonlinear geostatistical approach for downscaling daily precipitation and temperature, Water Resources Research, 51, 6244-6261, 2015a.
Jha, S. K., Zhao, H., Woldemeskel, F. M., and Sivakumar, B.: Network theory and spatial rainfall connections: An interpretation, Journal of Hydrology, 527, 13-19, 2015b.
Khajehei, S.: A multivariate modeling approach for generating ensemble climatology forcing for hydrologic applications, 2015.
Khajehei, S., and Moradkhani, H.: Towards an improved ensemble precipitation forecast: A probabilistic post-processing approach, Journal of Hydrology, 546, 476-489, 2017.
Liu, X., and Coulibaly, P.: Downscaling ensemble weather predictions for improved week-2 hydrologic forecasting, Journal of Hydrometeorology, 12, 1564-1580, 2011.
Milrad, S. M., Gyakum, J. R., and Atallah, E. H.: A meteorological analysis of the 2013 Alberta flood: antecedent large-scale flow pattern and synoptic–dynamic characteristics, Monthly Weather Review, 143, 2817-2841, 2015.
Nester, T., Komma, J., Viglione, A., and Blöschl, G.: Flood forecast errors and ensemble spread—A case study, Water Resources Research, 48, 2012.
Palmer, T.: The economic value of ensemble forecasts as a tool for risk assessment: From days to decades, Quarterly Journal of the Royal Meteorological Society, 128, 747-774, 2002.
Palmer, T., Shutts, G., Hagedorn, R., Doblas-Reyes, F., Jung, T., and Leutbecher, M.: Representing model uncertainty in weather and climate prediction, Annu. Rev. Earth Planet. Sci., 33, 163-193, 2005.
Peel, M. C., Finlayson, B. L., and McMahon, T. A.: Updated world map of the Köppen-Geiger climate classification, Hydrology and earth system sciences discussions, 4, 439-473, 2007.
Pomeroy, J. W., Stewart, R. E., and Whitfield, P. H.: The 2013 flood event in the South Saskatchewan and Elk River basins: causes, assessment and damages, Canadian Water Resources Journal/Revue canadienne des ressources hydriques, 41, 105-117, 2016.
Qu, B., Zhang, X., Pappenberger, F., Zhang, T., and Fang, Y.: Multi-Model Grand Ensemble Hydrologic Forecasting in the Fu River Basin Using Bayesian Model Averaging, Water, 9, 74, 2017.
Robertson, D., Shrestha, D., and Wang, Q.: Post-processing rainfall forecasts from numerical weather prediction models for short-term streamflow forecasting, Hydrology and Earth System Sciences, 17, 3587, 2013.
Roulin, E., and Vannitsem, S.: Postprocessing of ensemble precipitation predictions with extended logistic regression based on hindcasts, Monthly Weather Review, 140, 874-888, 2012.
Schaake, J., Demargne, J., Hartman, R., Mullusky, M., Welles, E., Wu, L., Herr, H., Fan, X., and Seo, D.: Precipitation and temperature ensemble forecasts from single-value forecasts, Hydrology and Earth System Sciences Discussions, 4, 655-717, 2007.
Shepard, D.: A two-dimensional interpolation function for irregularly-spaced data, Proceedings of the 1968 23rd ACM national conference, 1968, 517-524,

Shrestha, D., Robertson, D., Wang, Q., Pagano, T., and Hapuarachchi, H.: Evaluation of numerical weather prediction model precipitation forecasts for short-term streamflow forecasting purpose, Hydrology and Earth System Sciences, 17, 1913-1931, 2013.

Shrestha, D. L., Robertson, D. E., Bennett, J. C., and Wang, Q.: Improving precipitation forecasts by generating ensembles through postprocessing, Monthly Weather Review, 143, 3642-3663, 2015.

Tao, Y., Duan, Q., Ye, A., Gong, W., Di, Z., Xiao, M., and Hsu, K.: An evaluation of post-processed TIGGE multimodel ensemble precipitation forecast in the Huai river basin, Journal of Hydrology, 519, 2890-2905, 2014.

Teufel, B., Diro, G., Whan, K., Milrad, S., Jeong, D., Ganji, A., Huziy, O., Winger, K., Gyakum, J., and de Elia, R.: Investigation of the 2013 Alberta flood from weather and climate perspectives, Climate Dynamics, 1-19, 2016.

Wang, Q., Robertson, D., and Chiew, F.: A Bayesian joint probability modeling approach for seasonal forecasting of streamflows at multiple sites, Water Resources Research, 45, 2009.

Wang, Q., and Robertson, D.: Multisite probabilistic forecasting of seasonal flows for streams with zero value occurrences, Water Resources Research, 47, 2011.

WMO: Guidelines on ensemble prediction systems and forecasting. World Meteorological Organization, Geneva, Switzerland, Geneva, Switzerland, 2012.

Wu, L., Seo, D.-J., Demargne, J., Brown, J. D., Cong, S., and Schaake, J.: Generation of ensemble precipitation forecast from single-valued quantitative precipitation forecast for hydrologic ensemble prediction, Journal of Hydrology, 399, 281-298, 2011.

Ye, J., Shao, Y., and Li, Z.: Flood Forecasting Based on TIGGE Precipitation Ensemble Forecast, Advances in Meteorology, 2016, 2016.

Zsótér, E., Pappenberger, F., Smith, P., Emerton, R. E., Dutra, E., Wetterhall, F., Richardson, D., Bogner, K., and Balsamo, G.: Building a Multimodel Flood Prediction System with the TIGGE Archive, Journal of Hydrometeorology, 17, 2923-2940, 2016.

List of Tables:

List of Figures:

Formatted: Font: 10 pt, Not Bold

Formatted: Justified

the range of values obtained from 1000 post-processed ensembles, (a) and (b) show results of calibrated GEFS, and (c) and (d) show results of calibrated GDPS.

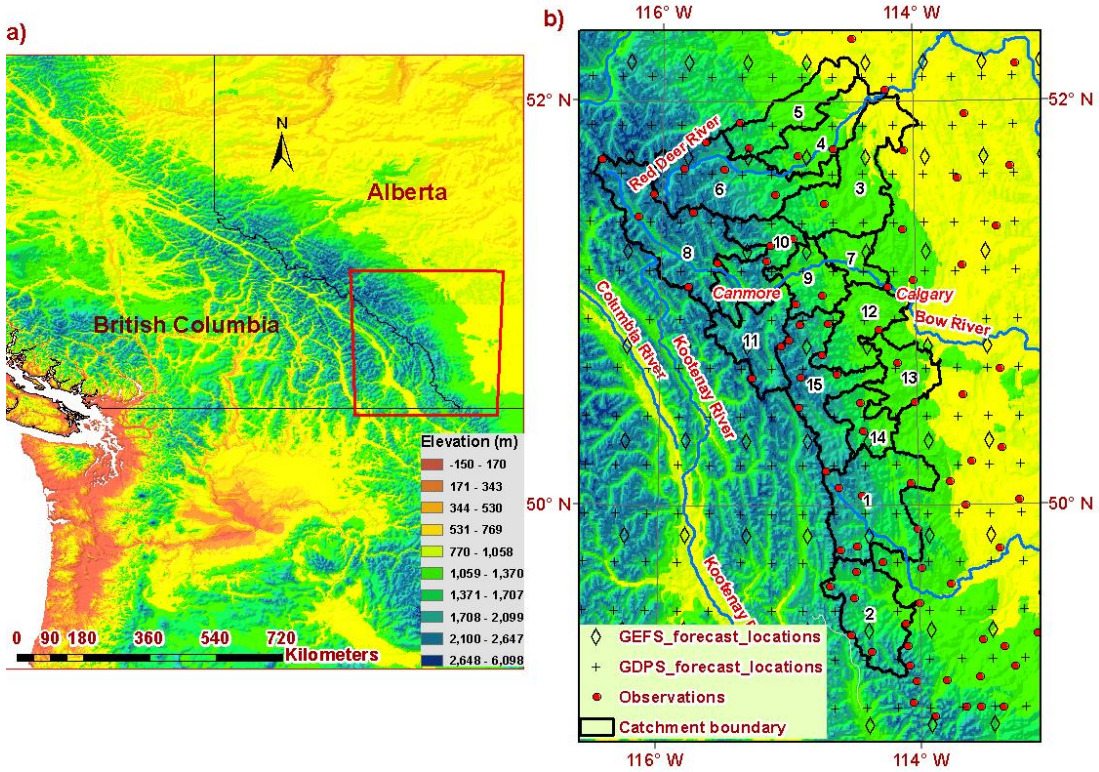
Figure 8~~7~~. Relative operating characteristic (ROC) curve at lead times of 1, 3, and 5 days for calibrated QPFs for precipitation events greater than 5 mm for subcatchment 11 during 10/6/2013 to 10/7/2013, with (a) and (b) showing ROC curves of calibrated GEFS and GDPS, respectively.

Table 1. Description of precipitation forecasts

Data Source	NWP name	Variable	Ensembles/ Deterministic	Time period	Daily /Subdaily	Lead time (days)	Spatial resolution (km)	Forecast hour
NCEP	GEFS	Precipitation	Control run	2013-2015	Daily	5 days	50 km	00 UTC
ECCC	GDPS	Precipitation	Deterministic	2013-2015	Daily	5 days	25 km	00 UTC

Table 2. Description of subcatchments in the study area

Subcatchment ID	Name	Area (km^2)	Subcatchment-averaged total annual precipitation (mm)		
			Year 2013	*Year 2014*	*Year 2015*
1	Up Oldman Willow	2664	808	699	514
2	Crows nest Castle	1848	***1206***	***1196***	***909***
3	Little Red Deer	2574	608	587	458
4	Mid Red Deer	1398	614	549	458
5	James Raven	1464	715	655	541
6	Up Red Deer	2723	897	662	571
7	Low Bow Local Bearspaw	***734***	628	533	409
8	Up Bow Banff Cascade	***2884***	663	806	659
9	Mid Bow Local Ghost Jumping pound	1063	871	689	540
10	Canmore Ghost Waiparous	1642	900	723	544
11	Spray Kananaskis	1445	1136	983	821
12	Fish Threepoint Low Elbow	1405	646	516	487
13	Low Sheep Highwood	1111	***502***	***440***	***371***
14	Trap Peki Stim	890	587	691	559
15	Up Highwood Sheep Elbow	2153	1062	784	693


a)
N
Alberta
British Columbia
Elevation (m)
-150 - 170
171 - 343
344 - 530
531 - 769
770 - 1,058
1,059 - 1,370
1,371 - 1,707
1,708 - 2,099
2,100 - 2,647
2,648 - 6,098
0 90 180 360 540 720
Kilometers
b)
116° W
114° W
52° N
50° N
Red Deer River
Canmore
Calgary
Bow River
Columbia River
Kootenay River
GEFS_forecast_locations
GDPS_forecast_locations
Observations
Catchment boundary

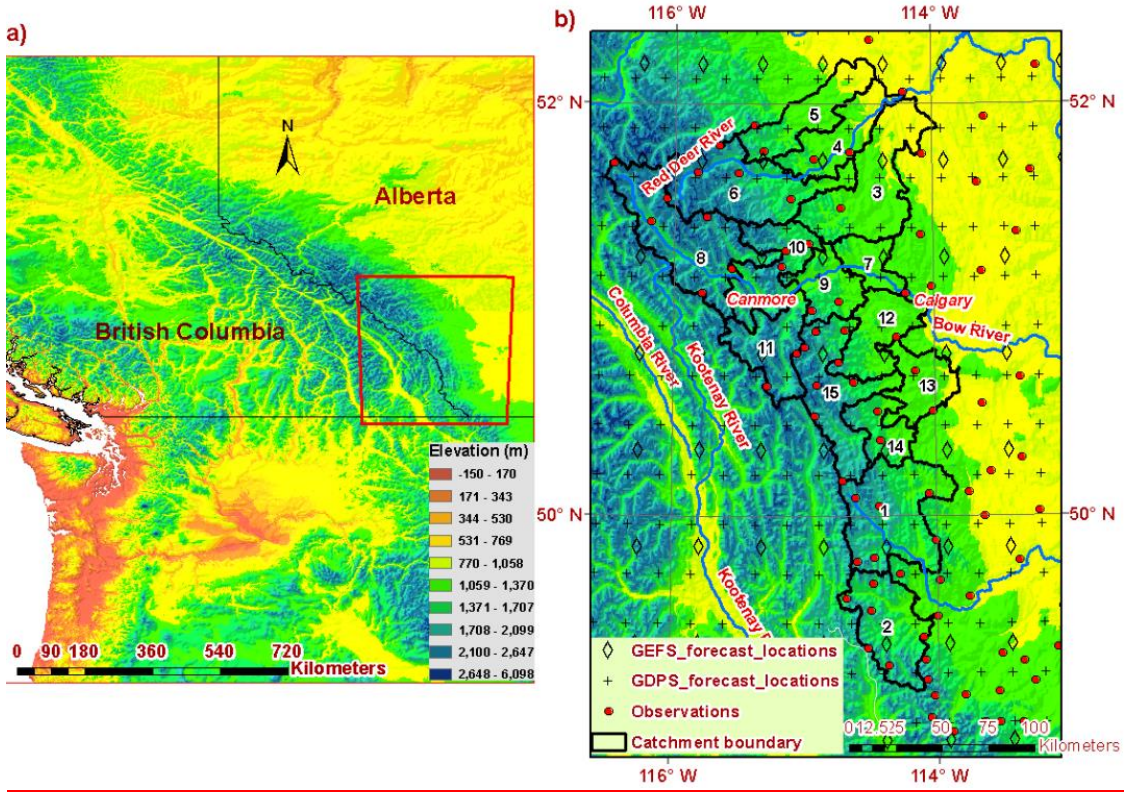
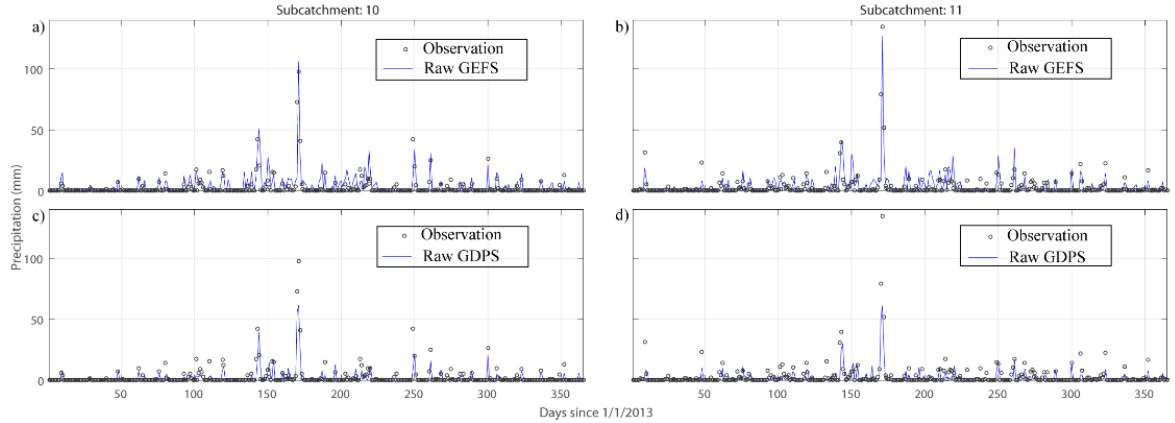


Figure 1: Location of the study basin in Alberta, Canada; showing (a) topography, a major driver of different precipitation mechanism; and (b) the study area with locations of observed and forecast data.


Subcatchment: 10
Subcatchment: 11
a)
b)
c)
d)
Observation
Raw GEFS
Raw GDPS
Precipitation (mm)
Days since 1/1/2013

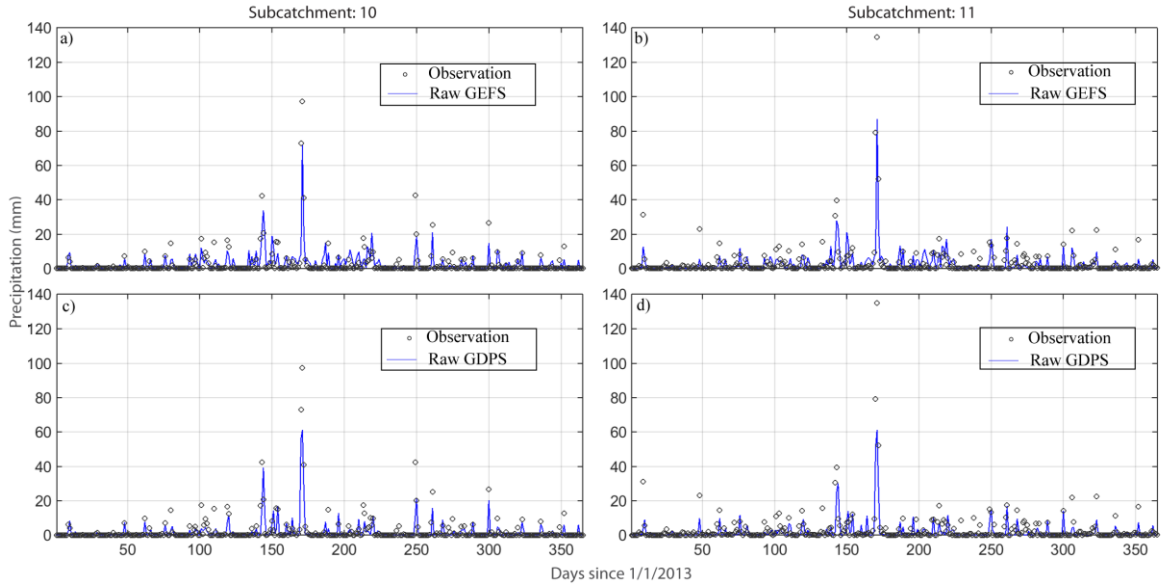
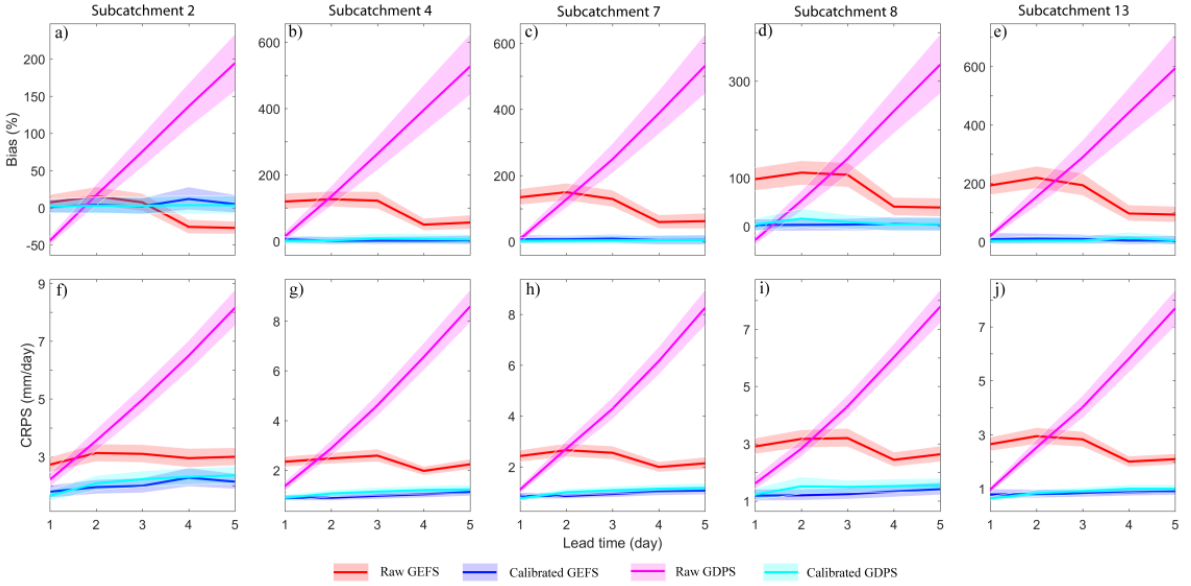

Subcatchment: 10
Subcatchment: 11
a)
b)
c)
d)
Observation
Raw GEFS
Raw GDPS
Precipitation (mm)
Days since 1/1/2013

Figure 2: Comparison of weighted-area raw QPF with subcatchment-averaged observations for the year 2013 in subcatchments 10 and 11. Raw GEFS are plotted in (a) and (b), while (c) and (d) show raw GDPS, along with observations.



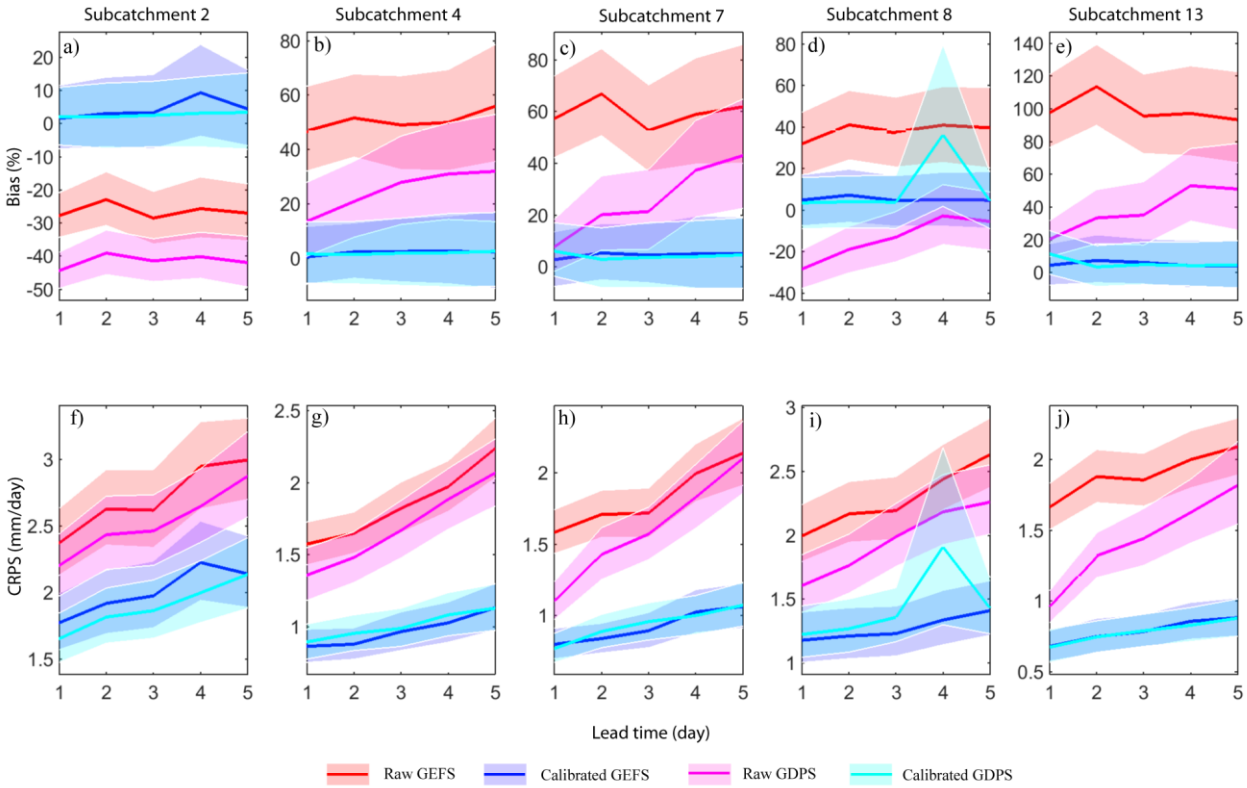
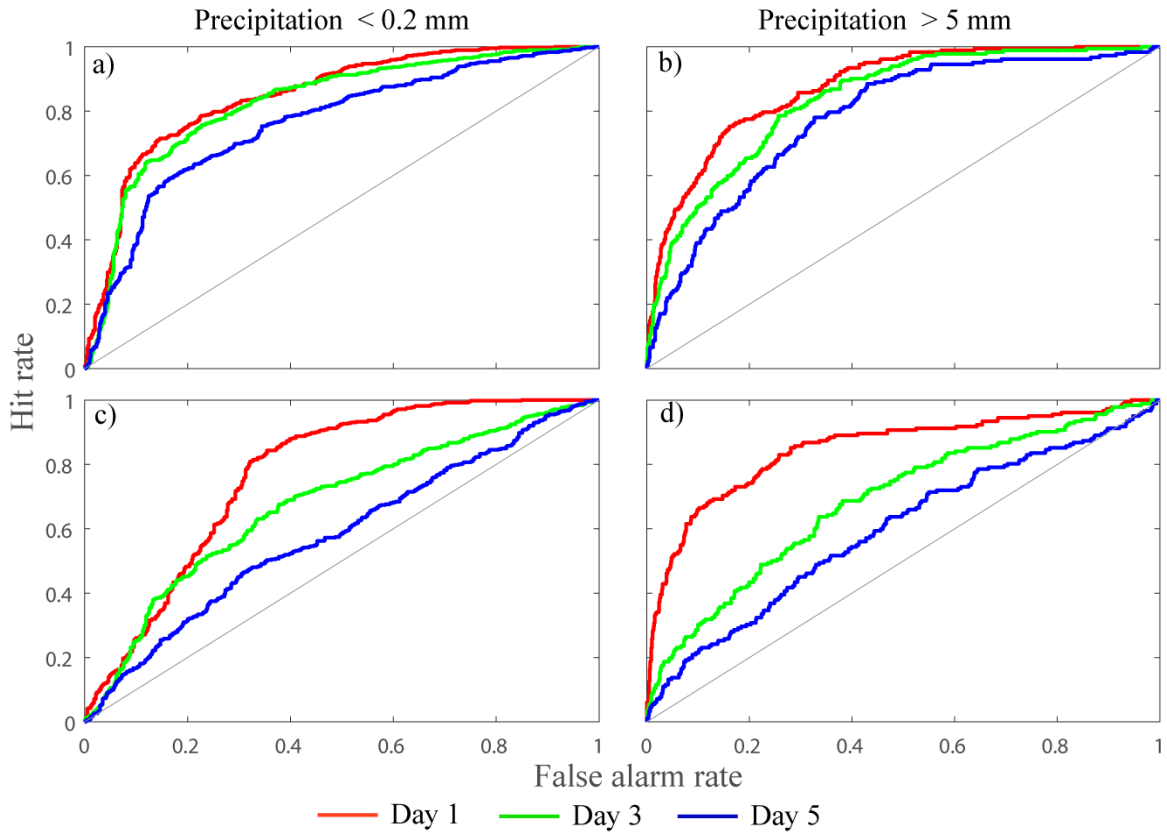


Figure 3: Subcatchment-averaged bias (%) in the raw QPFs and calibrated QPFs for individual daily forecasts as a function of lead-time for subcatchments 2, 4, 7, 8 and 13 ((a) to (e), respectively); (f) to (j) show subcatchment-averaged CRPS (mm/day) in the raw QPFs and calibrated QPFs for daily precipitation as a function of lead-time. The shaded region represents 5% and 95% confidence intervals generated using a bootstrap approach. Note the different scales on the vertical axes.


Precipitation < 0.2 mm
Precipitation > 5 mm
a)
b)
c)
d)
Hit rate
False alarm rate
0
0.2
0.4
0.6
0.8
1
Day 1
Day 3
Day 5

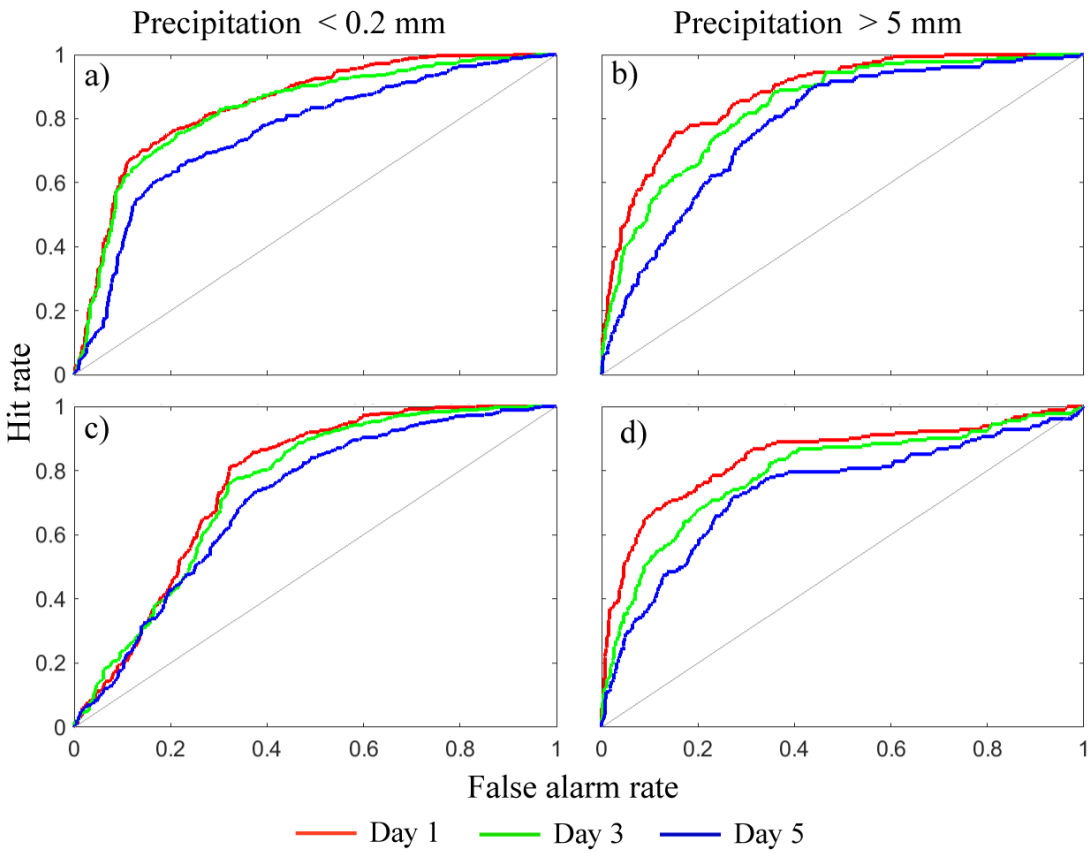
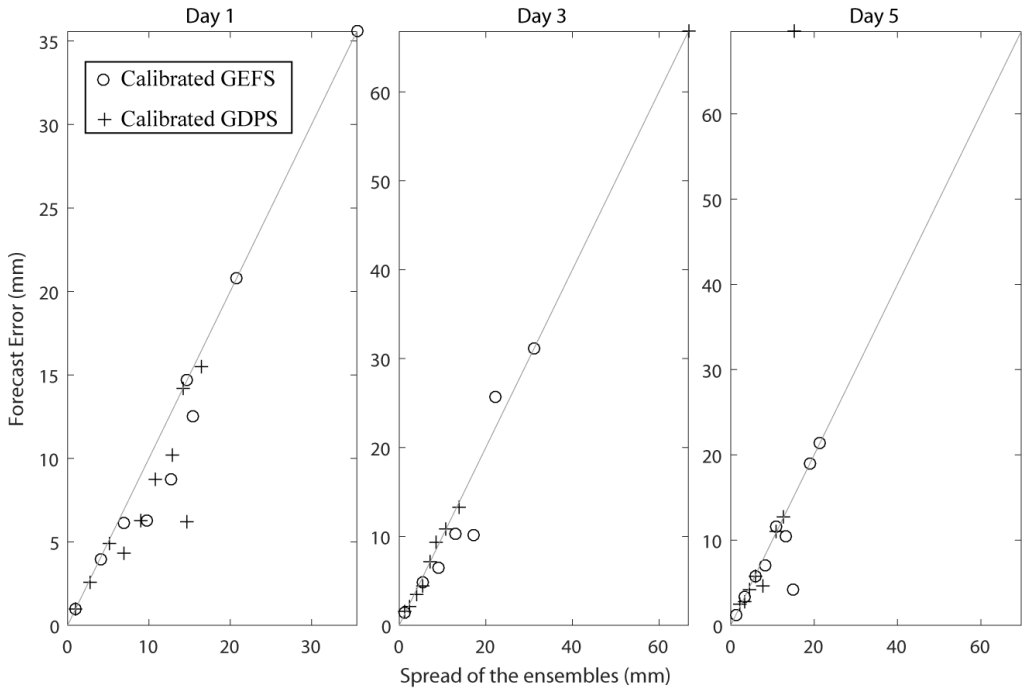


Figure 4: Relative operating characteristic (ROC) curve at lead times of 1, 3, and 5 days for calibrated QPFs for events of precipitation less than 0.2 mm and events greater than 5 mm for subcatchment 11. (a) and (b) show ROC curves of calibrated GEFS; (c) and (d) show ROC curves of calibrated GDPS. In the calculation of ROC, the daily data from 2013 to 2015 are used.


Day 1
Day 3
Day 5
Calibrated GEFS
Calibrated GDPS
Forecast Error (mm)
Spread of the ensembles (mm)

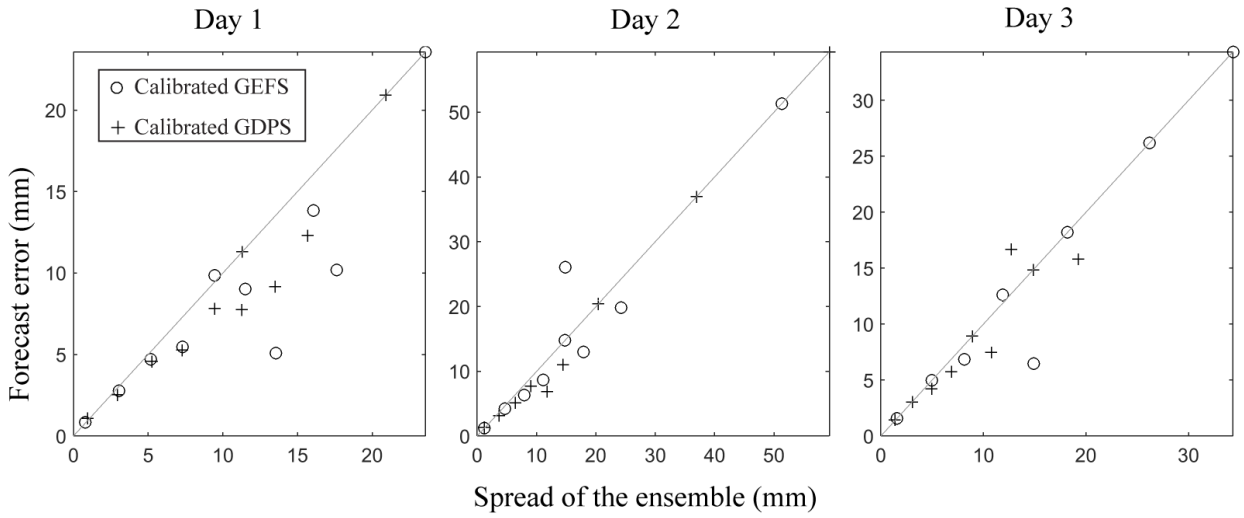


Figure 5. Scatterplots of forecast error versus spread for the 1000 ensembles of calibrated QPFs for lead times of 1, 3, and 5 days for subcatchment 11.

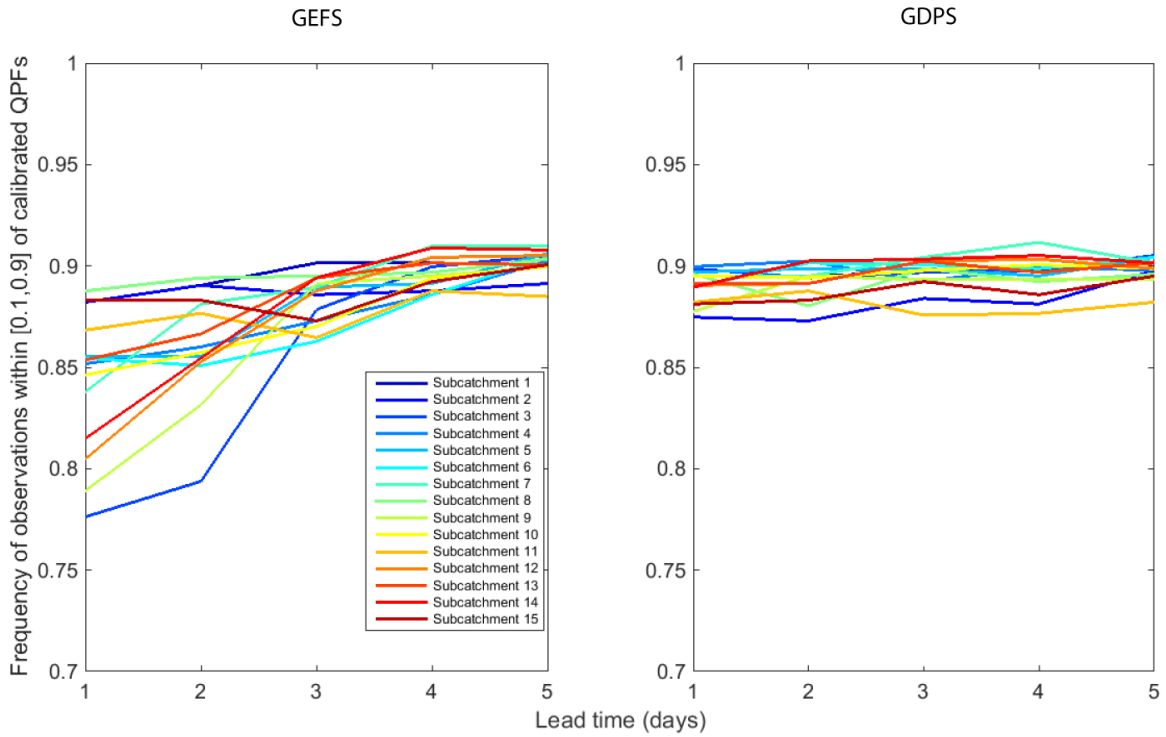
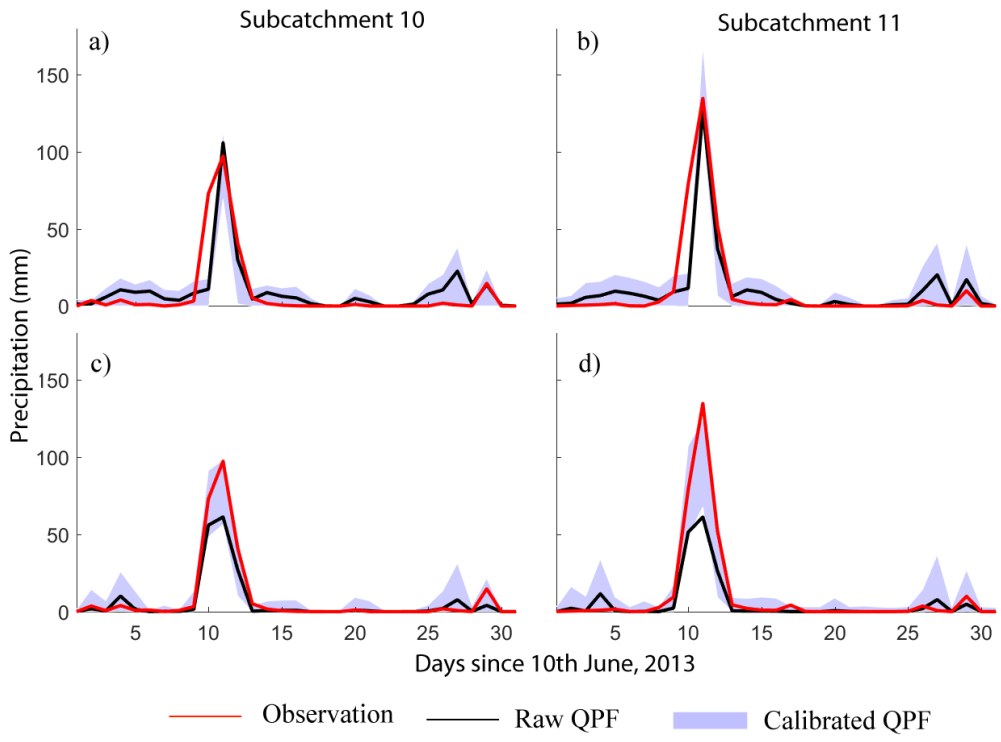


Figure 6: Frequency of observations lying within 10 and 90 percentile of calibrated GEFS and calibrated GDPS.


Subcatchment 10
Subcatchment 11
a)
b)
c)
d)
Precipitation (mm)
150
100
50
0
5
10
15
20
25
30
Days since 10th June, 2013
Observation
Raw QPF
Calibrated QPF

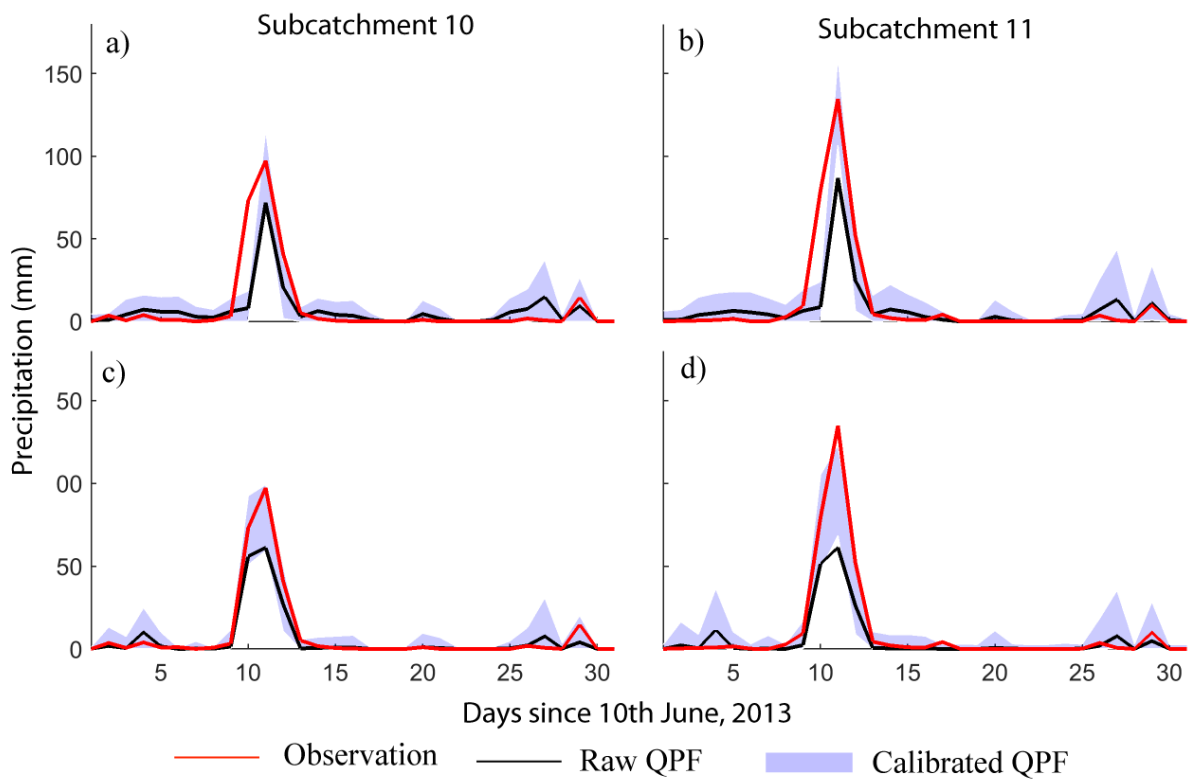
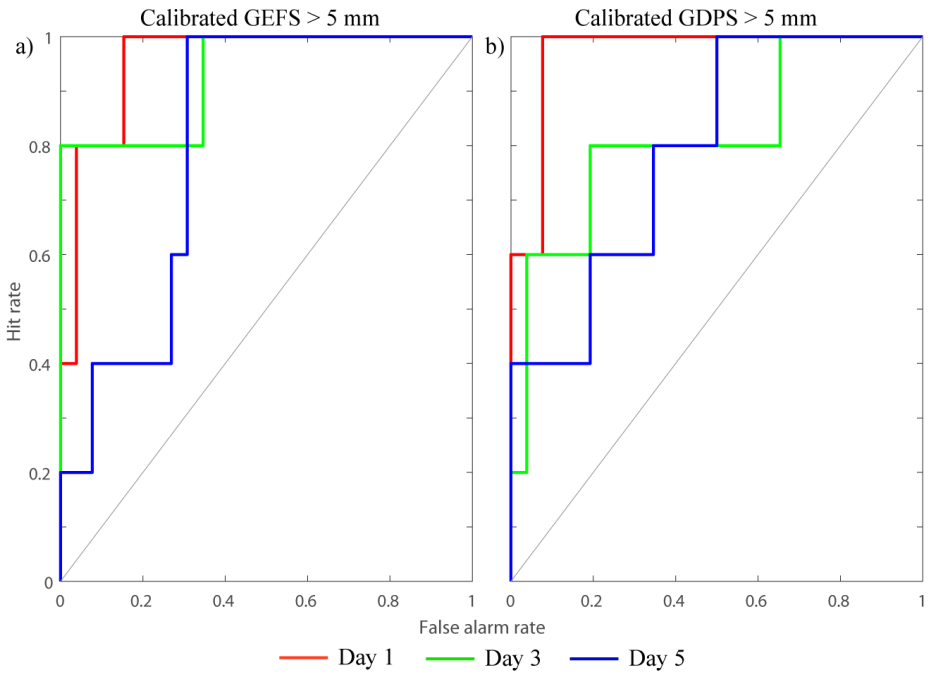


Figure 67: Comparison of time series of precipitation obtained from subcatchment-averaged raw QPF, subcatchment-averaged observations, and subcatchment-averaged calibrated QPFs in subcatchments 10 and 11. The shaded area represents the range of values obtained from 1000 post-processed ensembles, (a) and (b) show results of calibrated GEFS, and (c) and (d) show results of calibrated GDPS.


Calibrated GEFS > 5 mm
Calibrated GDPS > 5 mm
a)
b)
Hit rate
False alarm rate
0
0.2
0.4
0.6
0.8
1
Day 1
Day 3
Day 5

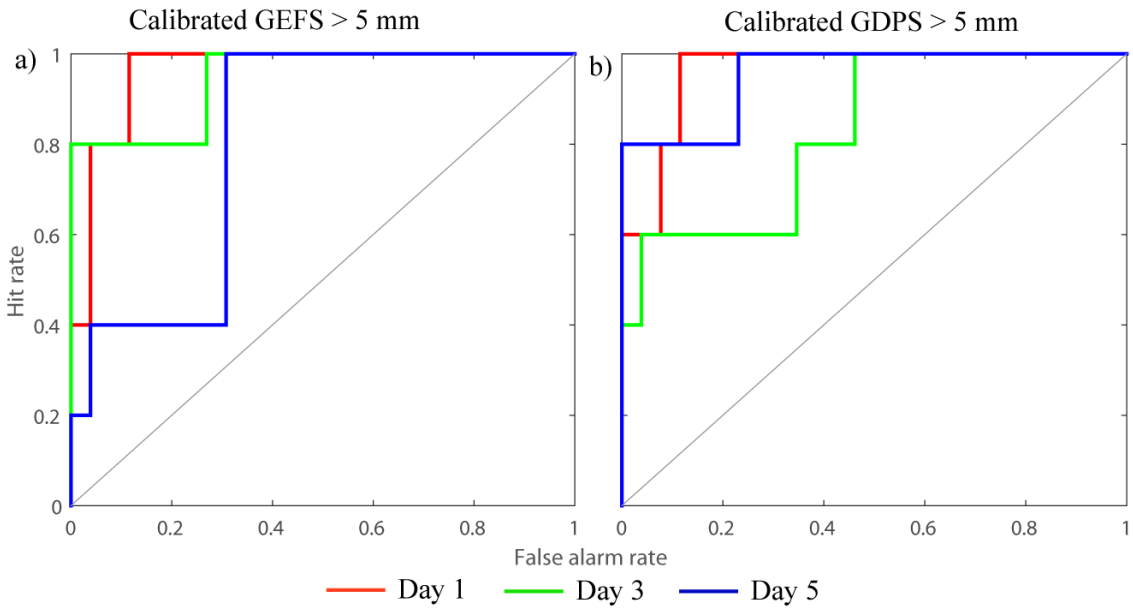


Figure ~~8~~7: Relative operating characteristic (ROC) curve at lead times of 1, 3, and 5 days for calibrated QPFs for precipitation events greater than 5 mm for subcatchment 11 during 10/6/2013 to 10/7/2013, with (a) and (b) showing ROC curves of calibrated GEFS and GDPS, respectively.

Formatted: Centered