A framework for ensemble streamflow forecast using improved post processed precipitation forecasts

Sanjeev Kumar Jha
Assistant Professor
Earth and Environmental Sciences
Indian Institute of Science Education and Research Bhopal
Email: sanjeevj@iiserb.ac.in

Outline

1. Motivation

   FloodNet Project in Canada

   Visits to the Canadian Flood Forecast Centres (FFCs)

2. Identified challenges/needs of the forecast centres
3. Rainfall post-processing (RPP) technique
4. Application of RPP on the watersheds in Alberta and British Columbia
5. Ongoing development

FloodNet motivation

- Floods are a major concern in Canada
- Need for improved flood forecasting methods
- Opportunity for universities to work closely with flood forecast centres and conservation authorities

FloodNet Project 3.1

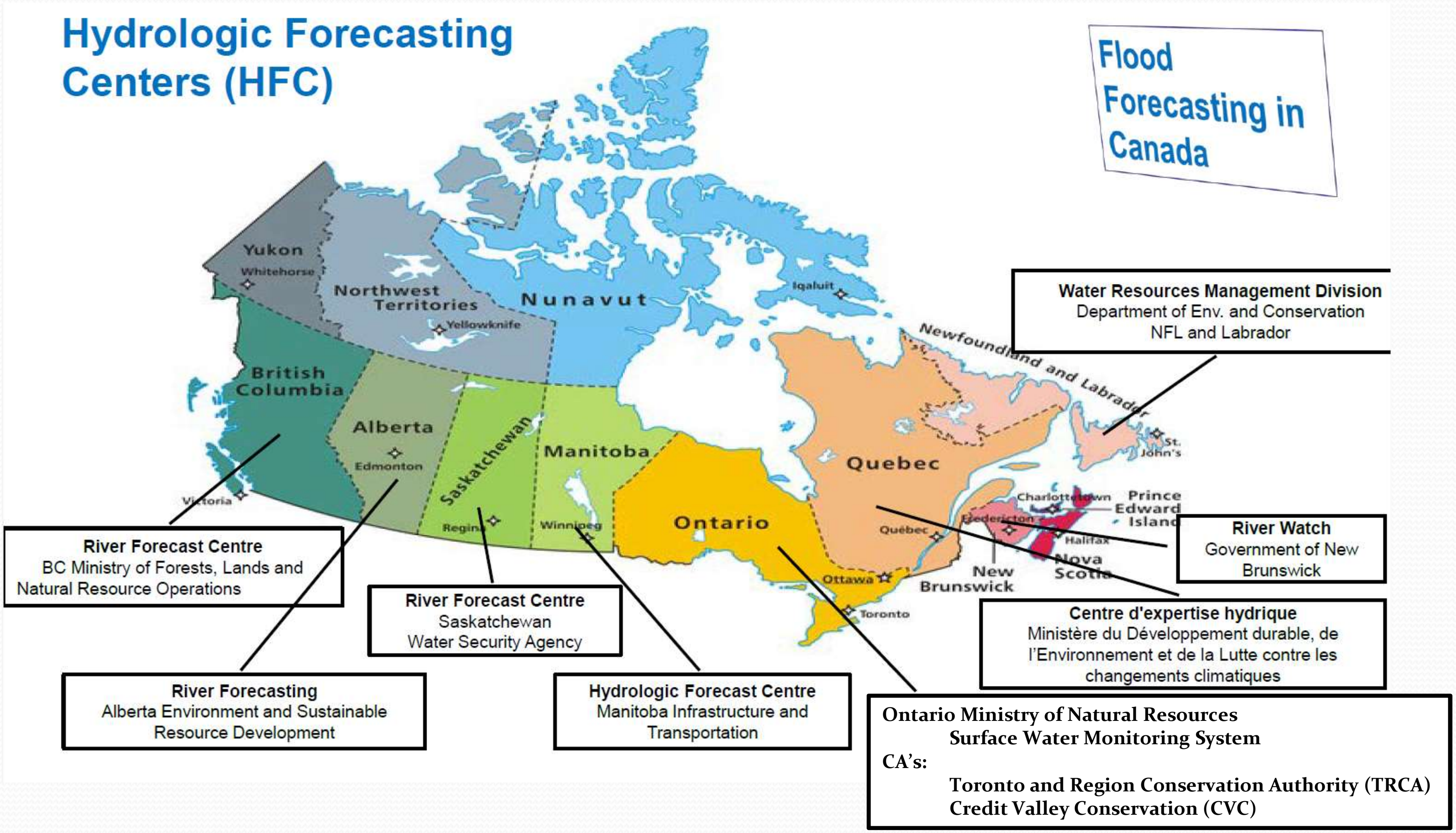
Objective

- Review flood forecasting systems currently implemented by Canadian provinces and evaluate their performance

Proposed research

- Collect information about existing flood forecasting techniques and challenges across the country
  - Methods (hydrologic models, statistical models, etc.)
  - Challenges (e.g. antecedent soil moisture, urban/rural lands)
  - Inputs/data (collection, storage, management)
  - Communication (internal/external)
- Select a number of case studies
- Conduct research to address the challenges
- Develop tools to implement the proposed research into operation

Site visits



Hydrologic Forecasting Centers (HFC)
Flood Forecasting in Canada
Yukon
Whitehorse
Northwest Territories
Yellowknife
Nunavut
Iqaluit
British Columbia
Victoria
Alberta
Edmonton
Saskatchewan
Regina
Manitoba
Winnipeg
Ontario
Ottawa
Toronto
Quebec
Québec
Newfoundland and Labrador
St. John's
Prince Edward Island
Charlottetown
Fredericton
Halifax
New Brunswick
Nova Scotia
Water Resources Management Division
Department of Env. and Conservation
NFL and Labrador
River Forecast Centre
BC Ministry of Forests, Lands and Natural Resource Operations
River Forecast Centre
Saskatchewan Water Security Agency
River Forecasting
Alberta Environment and Sustainable Resource Development
Hydrologic Forecast Centre
Manitoba Infrastructure and Transportation
River Watch
Government of New Brunswick
Centre d'expertise hydrique
Ministère du Développement durable, de l'Environnement et de la Lutte contre les changements climatiques
Ontario Ministry of Natural Resources
Surface Water Monitoring System
CA's:
Toronto and Region Conservation Authority (TRCA)
Credit Valley Conservation (CVC)

Research questions identified

- **Data sources, collection and processing**
  - ***Uncertainty in precipitation forecast obtained from Numerical Weather Prediction models***
  - The accuracy of streamflow forecast at upstream locations in neighbouring provinces or US states
  - Determining antecedent soil moisture
  - Estimating snow-water equivalent

- **Hydrologic and hydraulic modeling**
  - Hydrologic modelling of the Prairie region, characterized by a high percentage of non-contributing areas due to potholes (western provinces)
  - Presence of urban and rural areas in the same watershed (mostly eastern provinces)
  - Consideration of regulated flow in hydrologic modelling
  - Selection of the appropriate modelling system
  - Need for an automated and integrated real-time forecast system

Evaluation of post-processed ensemble precipitation forecasts in Canadian catchments

Raw Precipitation Forecast

from Numerical Weather Prediction model (NWP) has limitations due to difficulties in perturbing initial conditions and physical parametrization of the NWP model.

Rainfall Post Processing

generates ensemble forecasts by exploiting the joint relationship between observations and raw NWP forecast.

Rainfall Runoff Model

can now use 1000 generated ensemble forecasts from Rainfall Post-processing approach

Water Volume forecast

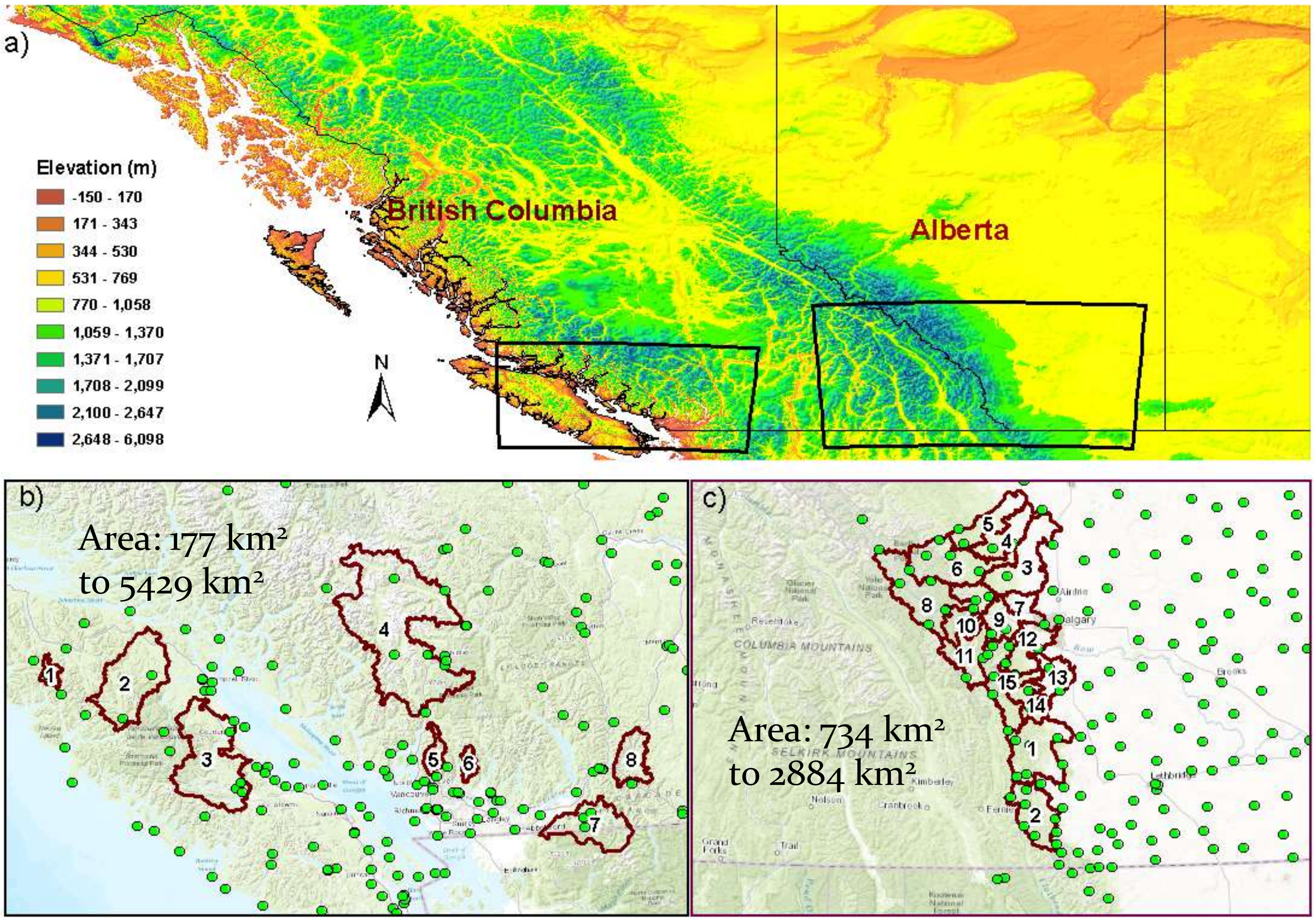
can have estimate of likelihood of occurrence of events instead of a single estimate. Highly beneficial for water managers in managing water structures, issuing flood/drought warnings etc.

The aims of this study are to:

(a) evaluate the performance of Rainfall post-processing approach (RPP), developed in Australia, in improving cold regions precipitation forecasts, and

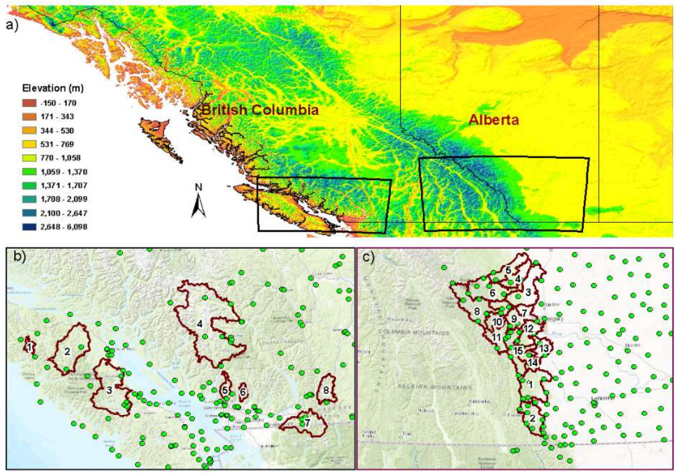
(b) compare the ensembles generated from applying RPP to the deterministic raw precipitation forecasts obtained from two NWP models (GEFS from NCEP, and GDPS from ECCC).

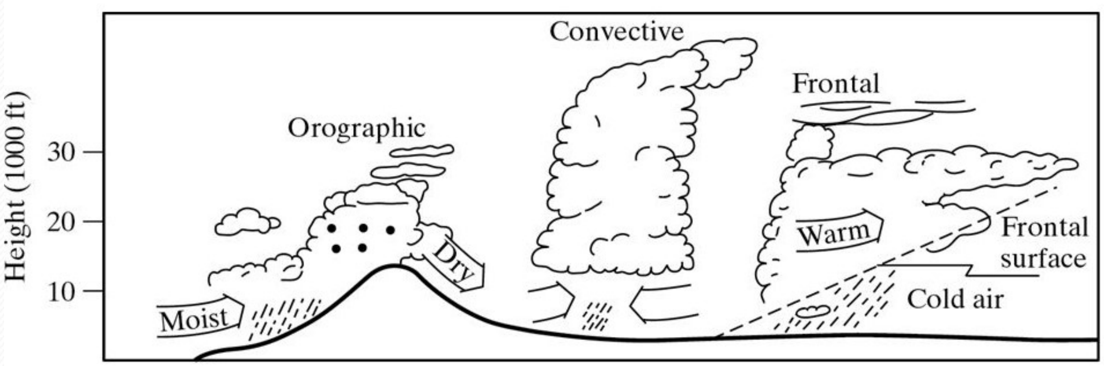
Study area



a)
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
N
British Columbia
Alberta
b)
Area: 177 km²
to 5429 km²
c)
Area: 734 km²
to 2884 km²
COLUMBIA MOUNTAINS
SELKIRK MOUNTAINS
Calgary
Lethbridge
Kimberley
Cranbrook
Nelson
Trail
Vancouver

Study area



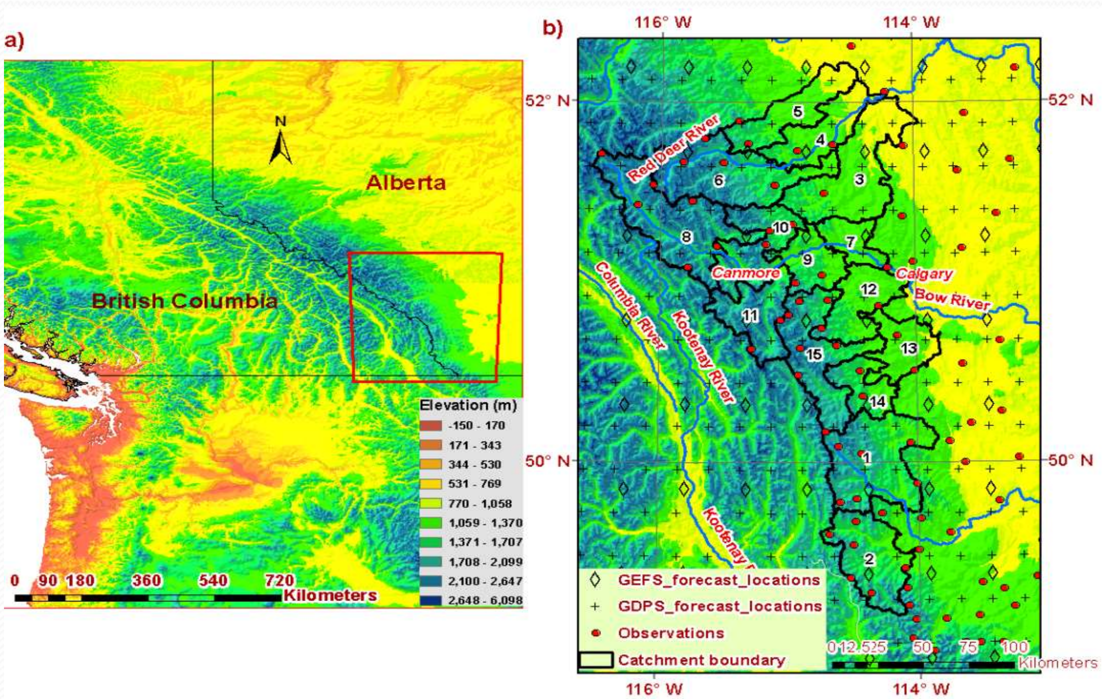


- NWP output - errors present in both initial conditions and numerical models
- Coupled atmospheric/land/ocean dynamical system is chaotic
- Forecasts can have negative impact on users
- Issues of spatial and temporal resolution
- For decision making, expected forecast errors need to be considered carefully

Source: Slideshare, google images

Data sources

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



GEFS: Global ensemble forecast system

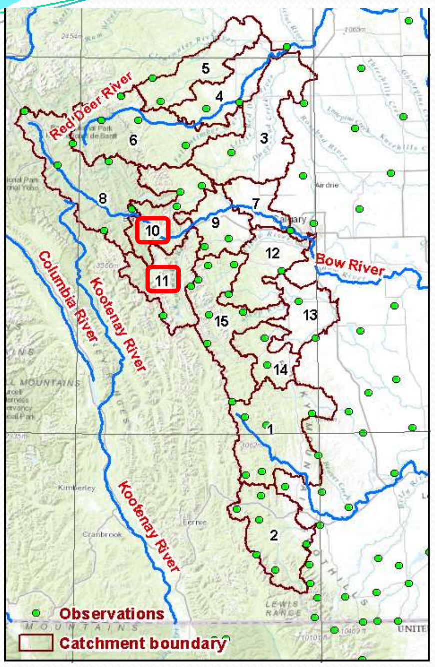
GDPS: Global deterministic prediction system

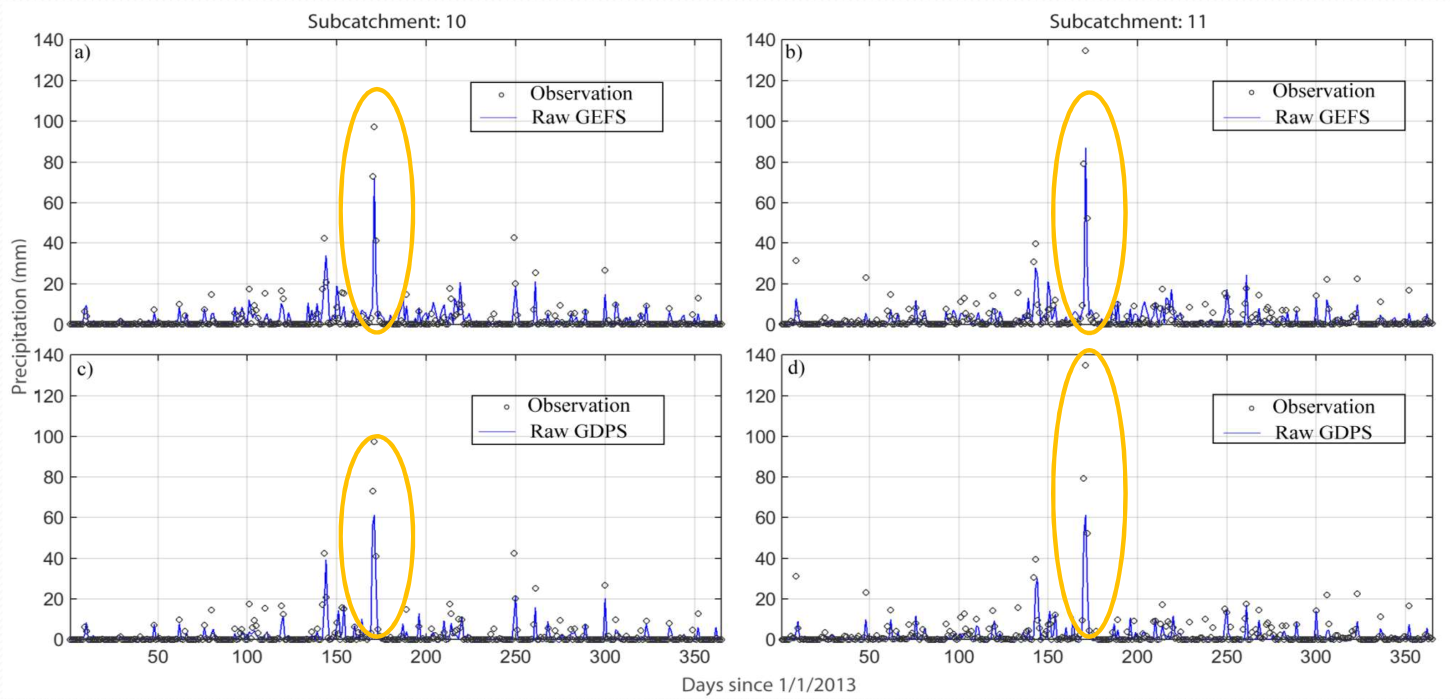
NCEP: National Centers for Environmental Protection

ECCC: Environment and Climate Change Canada

UTC: Universal Time Coordinated

Raw precipitation forecast





The comparison of weighted-area raw forecasts and subcatchment-averaged observed precipitation in subareas 10 and 11 for GEFS and GDPS with a lead-time of one day for 2013.

Figure shows that there is always a bias between raw forecasts and observations.

Methodology

The Rainfall Post Processing (RPP) relates raw quantitative precipitation forecasts (QPF) and observed precipitation using a Bayesian joint probability (BJP) modeling approach, followed by the Schaake shuffle.

(i) Transform the raw QPF (x) and observed precipitation (y) as:

$$\hat{x} = \frac{1}{b_x}\ln[\sinh(a_x + b_x x)]; \quad \hat{y} = \frac{1}{b_y}\ln\left[\sinh\left(a_y + b_y y\right)\right]$$

(ii) Transformed variables follow a bivariate normal distribution, $p(\hat{x}, \hat{y}) \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$.

BJP model parameters are $\boldsymbol{\theta} = [a_x, b_x, \mu_{\hat{x}}, \sigma_{\hat{x}}, a_y, b_y, \mu_{\hat{y}}, \sigma_{\hat{y}}, \rho_{\hat{x}\hat{y}}]$.

(iii) Model the joint distribution of $\hat{x}$ and $\hat{y}$. Use the shuffled complex evolution algorithm to infer parameters of the marginal distribution of each $\hat{x}$ - $\hat{y}$ pairs. Infer all parameters ($\boldsymbol{\theta}$) using a leave-one-month-out-cross-validation approach on the historical data.

(iv) Estimate a forecast from the conditional bivariate normal distribution

$$f(\hat{y}|\hat{x}) \sim N(\mu_{\hat{y}|\hat{x}}, \sigma^2_{\hat{y}|\hat{x}}) \text{ with } \mu_{\hat{y}|\hat{x}} = \mu_{\hat{y}} + \rho_{\hat{x}\hat{y}}\sigma_{\hat{y}}\frac{(\hat{x}-\mu_{\hat{x}})}{\sigma_{\hat{x}}}, \text{ and } \sigma^2_{\hat{y}|\hat{x}} = \sigma^2_{\hat{y}}\,(1-\rho^2_{\hat{x}\hat{y}})$$

(v) Obtain the forecast value in an untransformed space

$$y_f = \frac{1}{b_y}\{archsinh[\exp(b_y\hat{y}_f)] - a_y\}$$

(vi) BJP modeling approach is applied to each lead time and each forecast locations separately, therefore the Schaake shuffle is required to derive appropriate space-time correlations.

Improving precipitation forecasts by generating ensembles through postprocessing

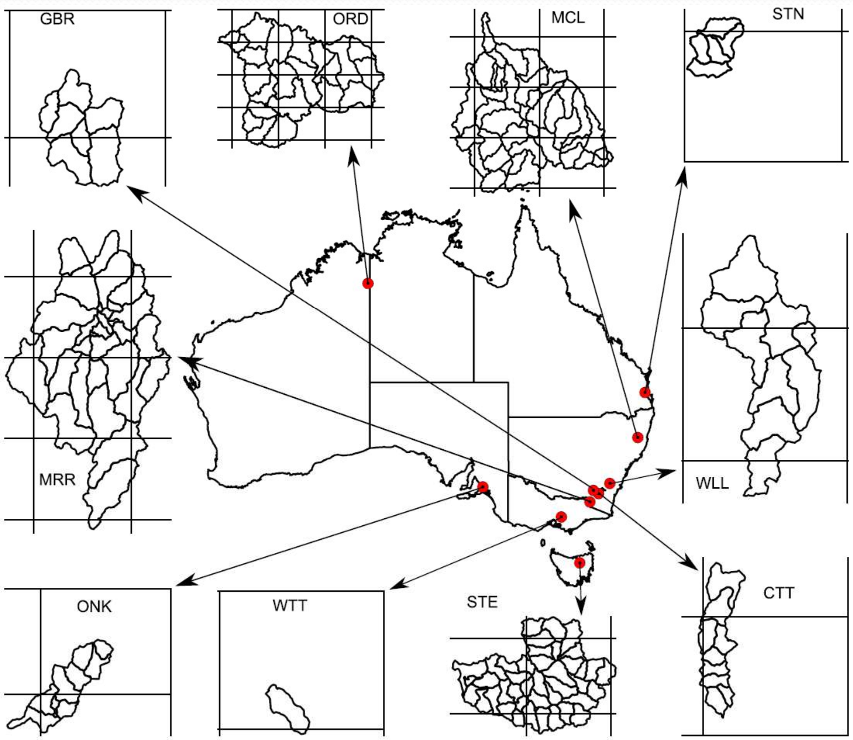


FIG. 1. Location of catchments used in this study. Descriptions of the catchments are given in Table 1. Subarea delineations and NWP

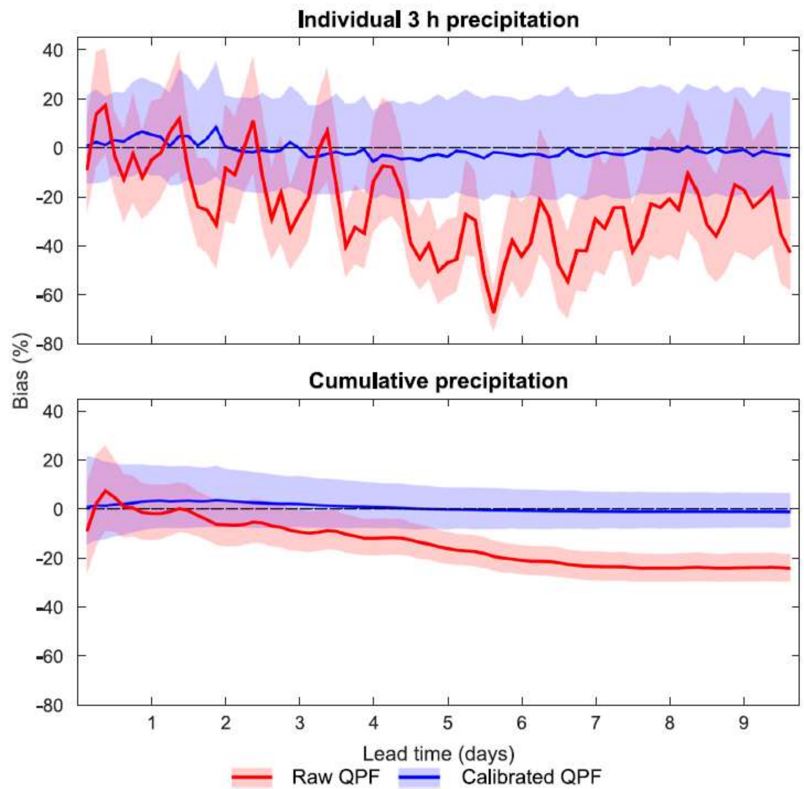


FIG. 5. Catchment average bias (%) in the raw QPF and the calibrated QPF as function of lead time in the South Esk catchment. (top) The individual 3-h precipitation and (bottom) for the cumulative precipitation. The shaded bands correspond to the [0.05, 0.95] confidence intervals of sampling uncertainty in the raw QPF (light red color) and the calibrated QPF (light blue color).

Evaluation statistics

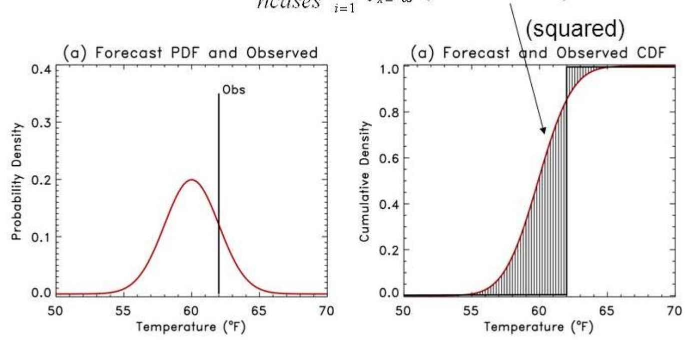
$$Bias = \frac{\sum_{1}^{t} z_f - \sum_{1}^{t} z_o}{\sum_{1}^{t} z_o} \cdot 100$$

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

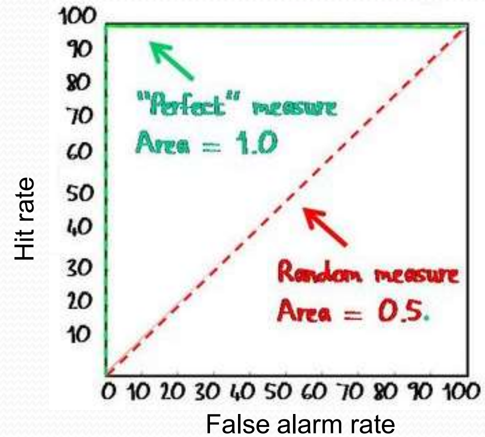
Continuous rank probability score

- Let $F_i^f(x)$ be the forecast probability CDF for the ith forecast case.
- Let $F_i^o(x)$ be the observed probability CDF (Heaviside function).

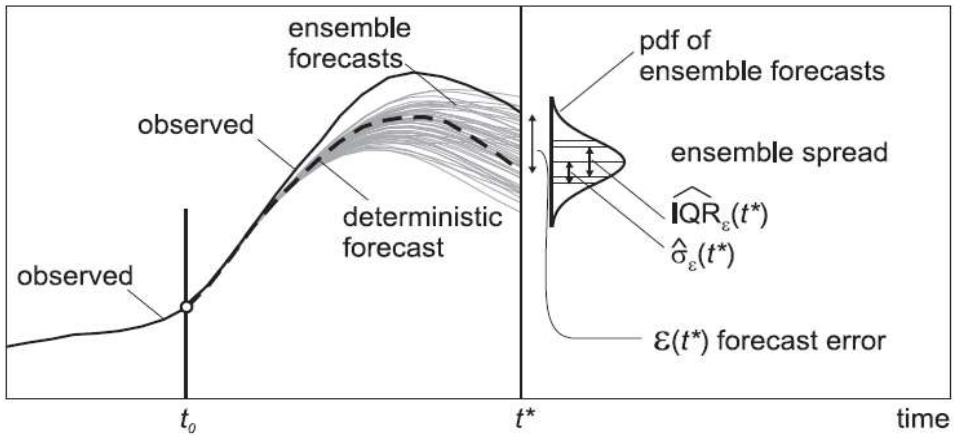
$$CRPS(forecast) = \frac{1}{ncases}\sum_{i=1}^{ncases}\int_{x=-\infty}^{x=-\infty}\left(F_i^f(x) - F_i^o(x)\right)^2 dx$$



Receiver operating characteristic

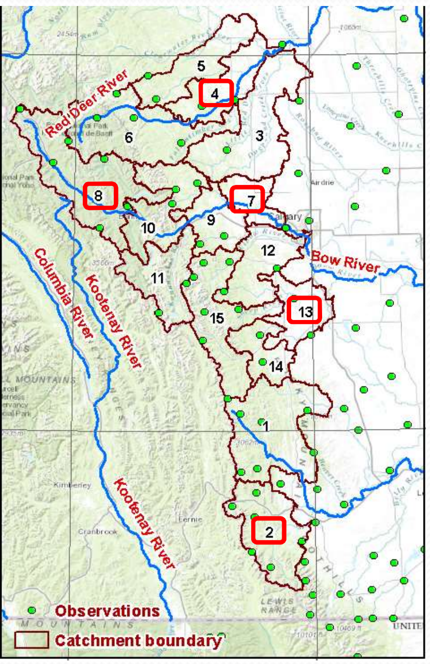


Spread-skill plot

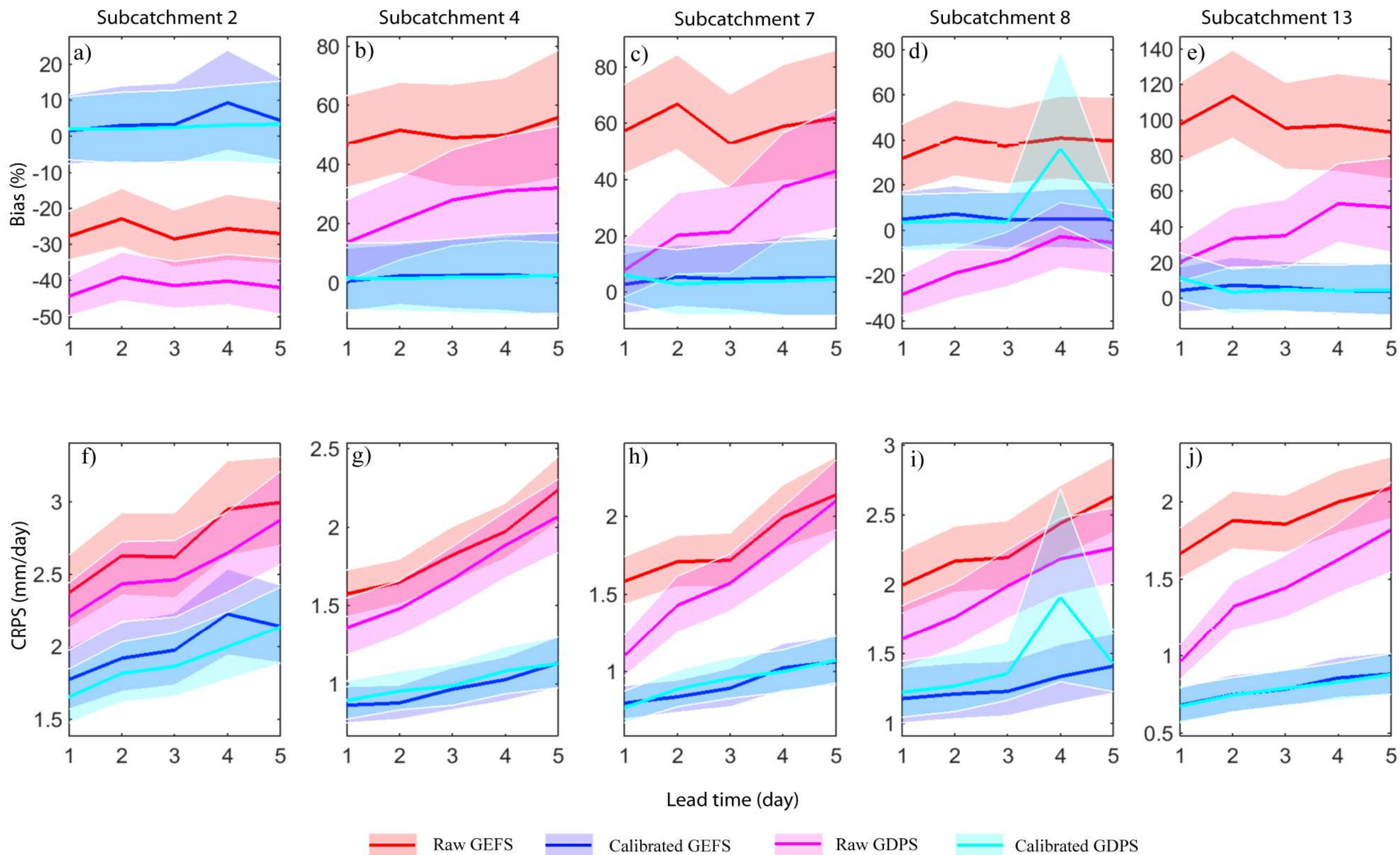


Source: pinimg.com; Nester et al. (2012), WRR; online presentation of Tom Hill, NOAA

Results

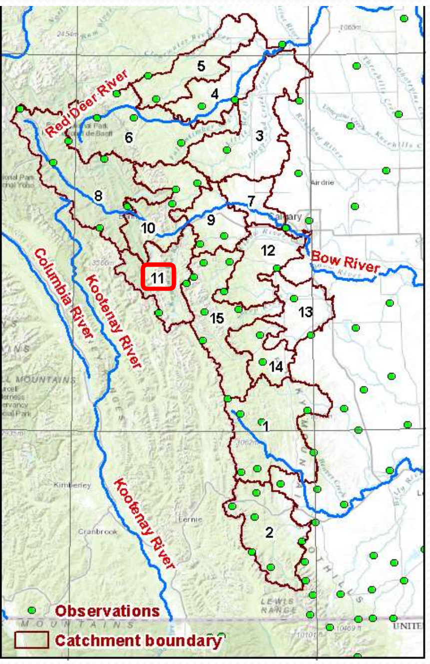


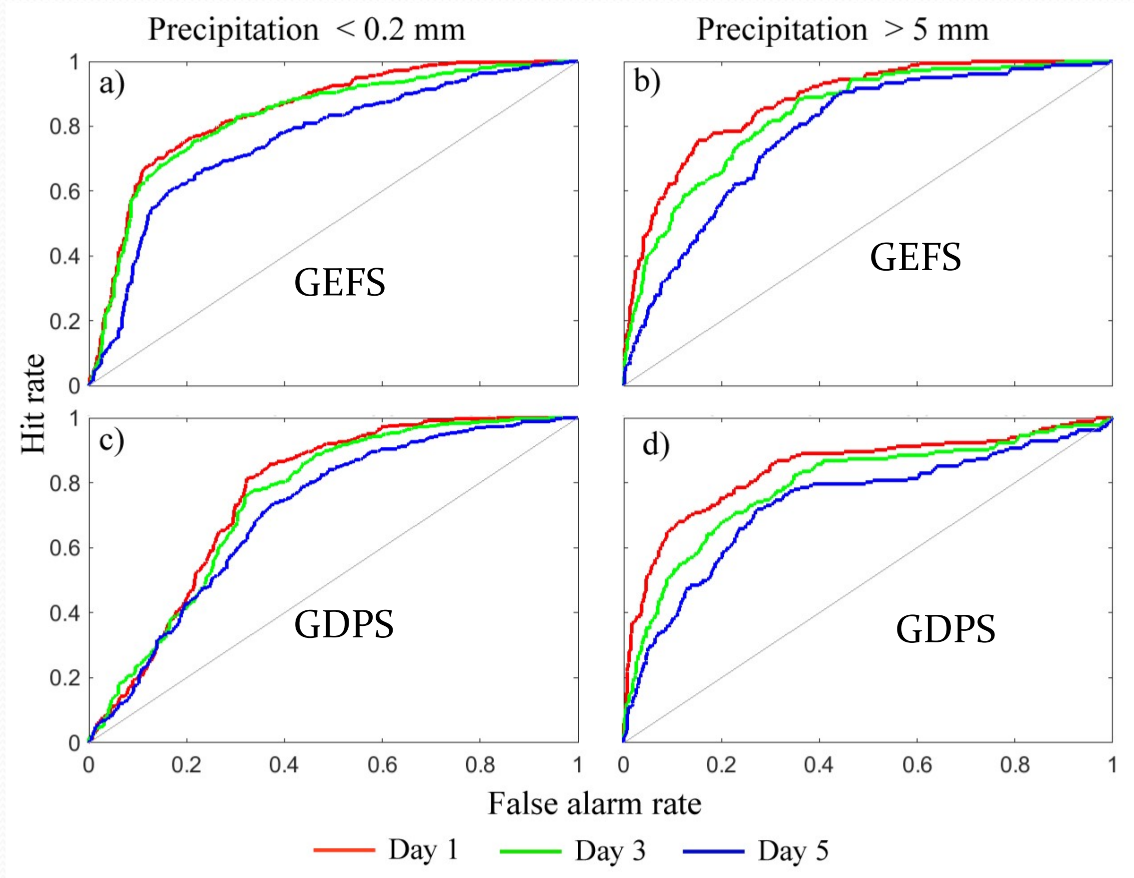
Red Deer River
Bow River
Columbia River
Kootenay River
Kootenay River
5
4
6
3
8
7
10
9
12
11
13
15
14
1
2
Observations
Catchment boundary



Subcatchment 2
Subcatchment 4
Subcatchment 7
Subcatchment 8
Subcatchment 13
a)
b)
c)
d)
e)
f)
g)
h)
i)
j)
Bias (%)
CRPS (mm/day)
Lead time (day)
Raw GEFS
Calibrated GEFS
Raw GDPS
Calibrated GDPS

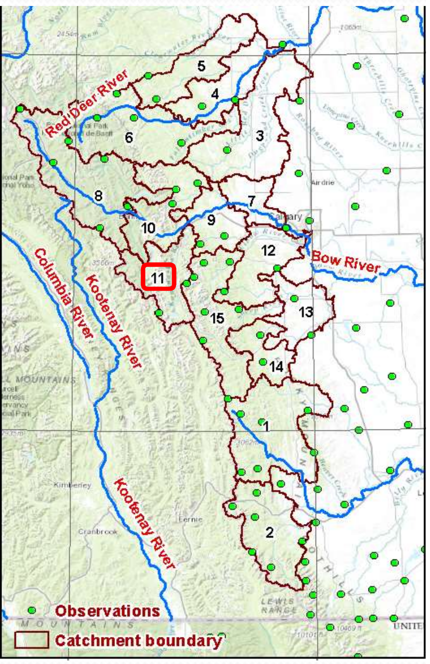
Results contd…

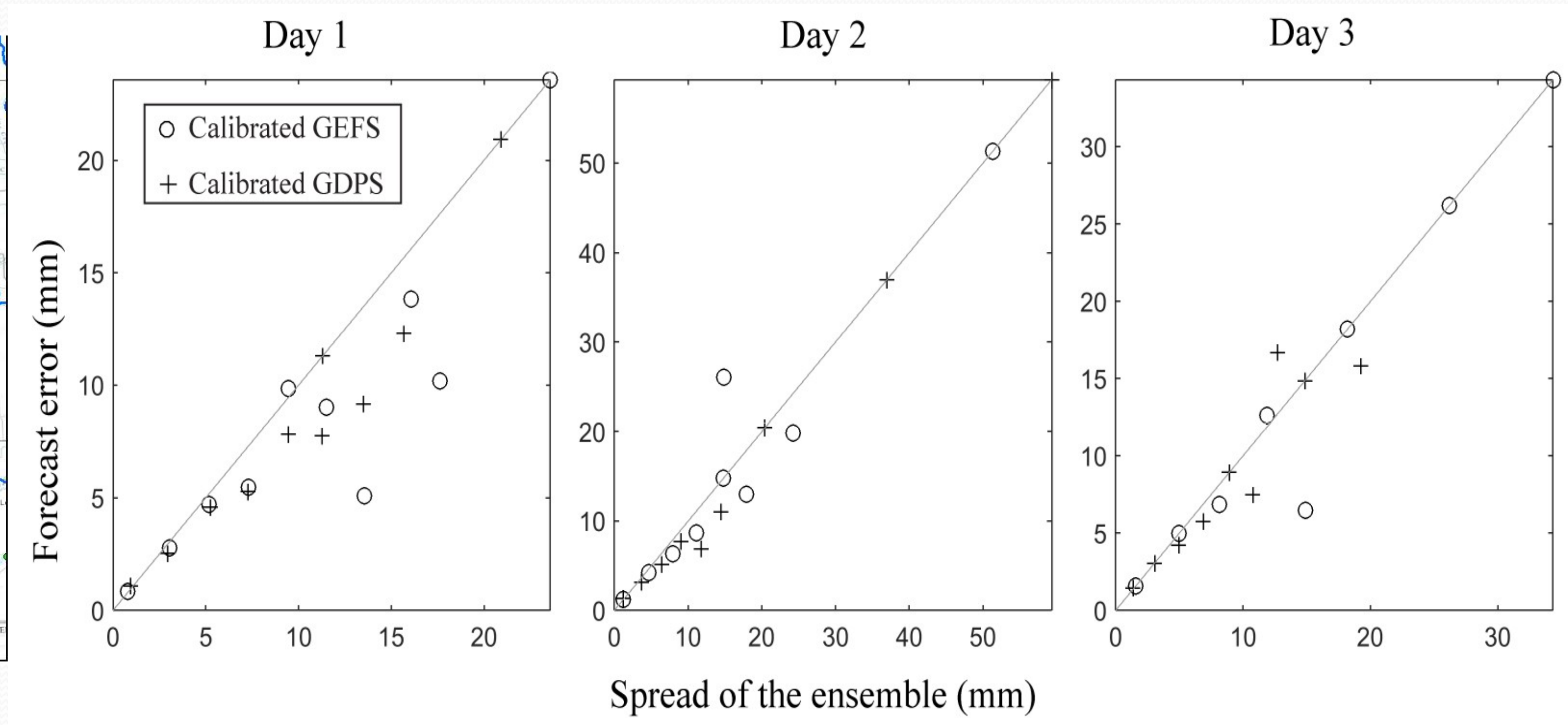




Given a precipitation threshold, hit rate refers to probability of forecasts that detected events smaller or larger in magnitude than the threshold.

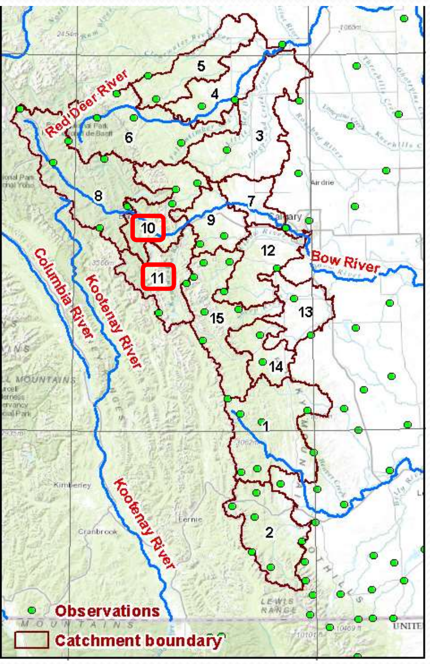
Results contd…

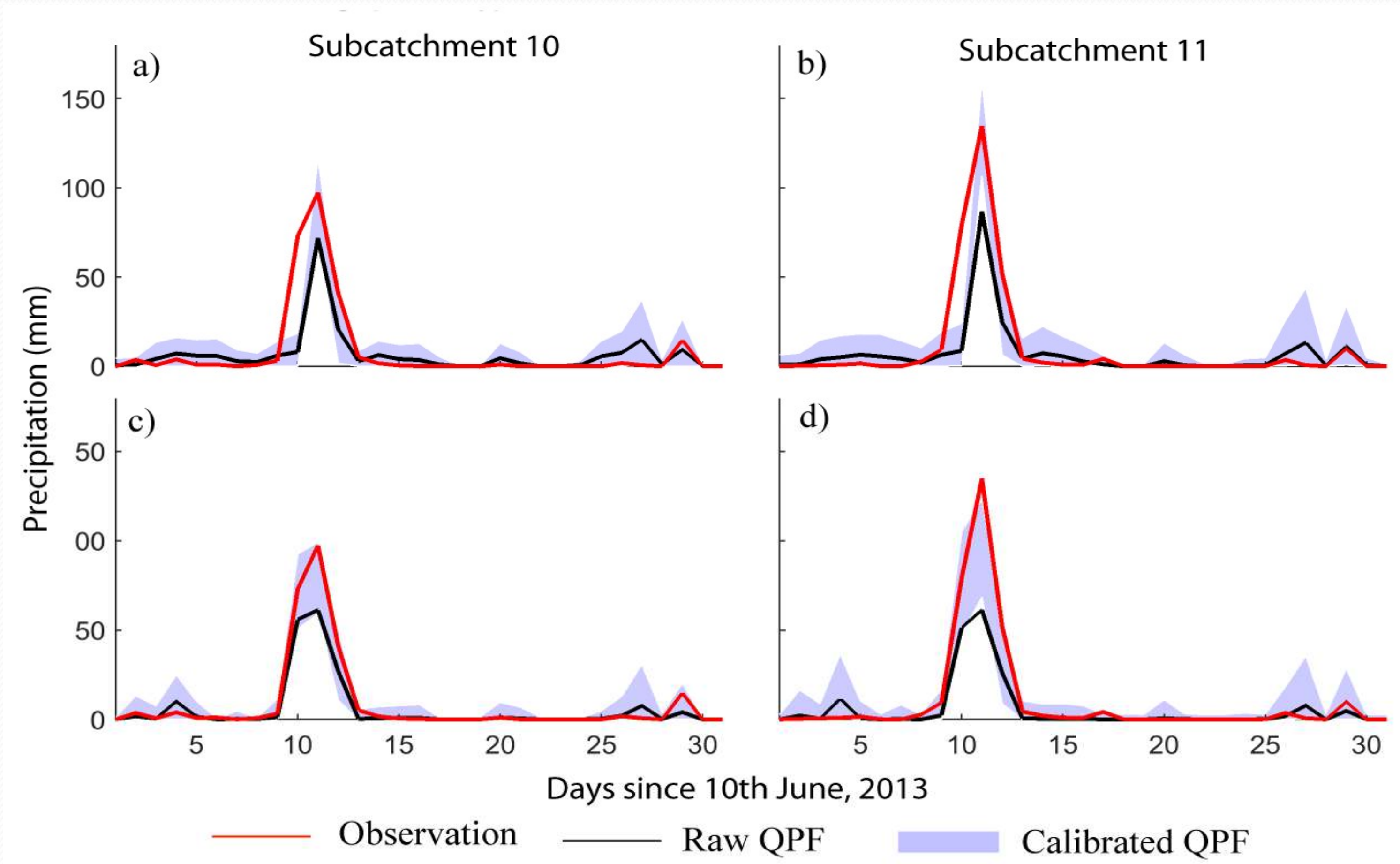




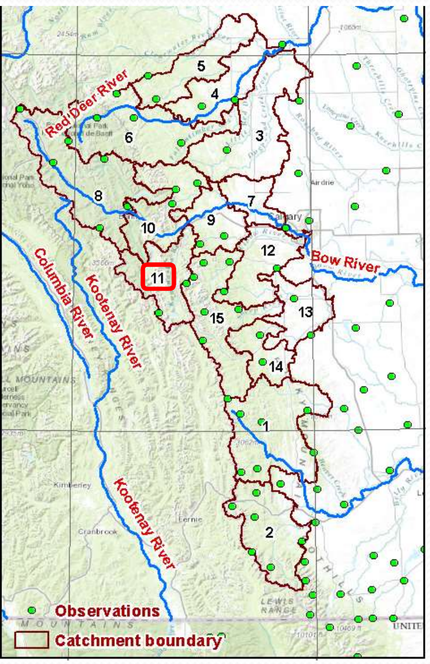
The ensemble spread is defined as mean absolute difference between the ensemble members and the mean.
The absolute difference between the observation and the ensemble mean is defined as forecast error.

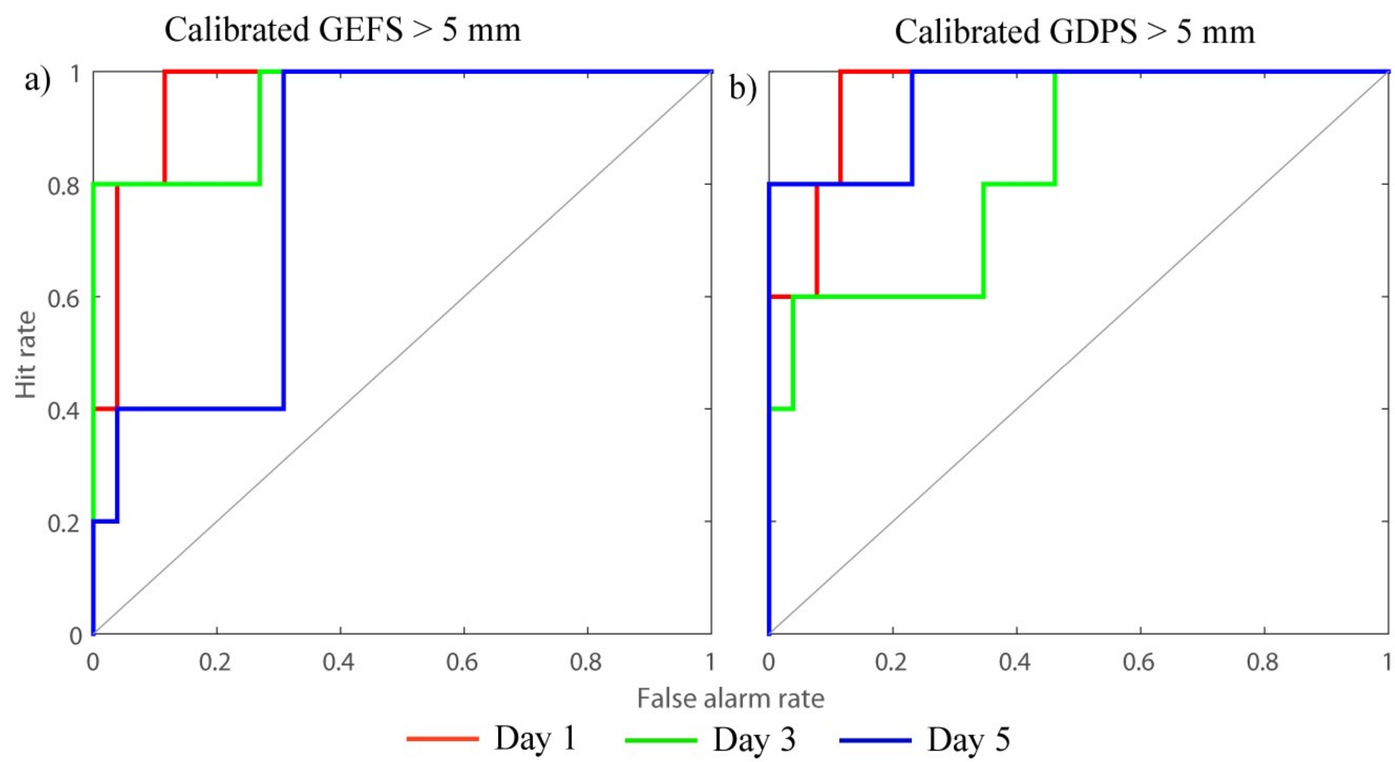
Results contd…





Results contd…





Forecasts during 10/6/2013 to 10/7/2013 are considered.

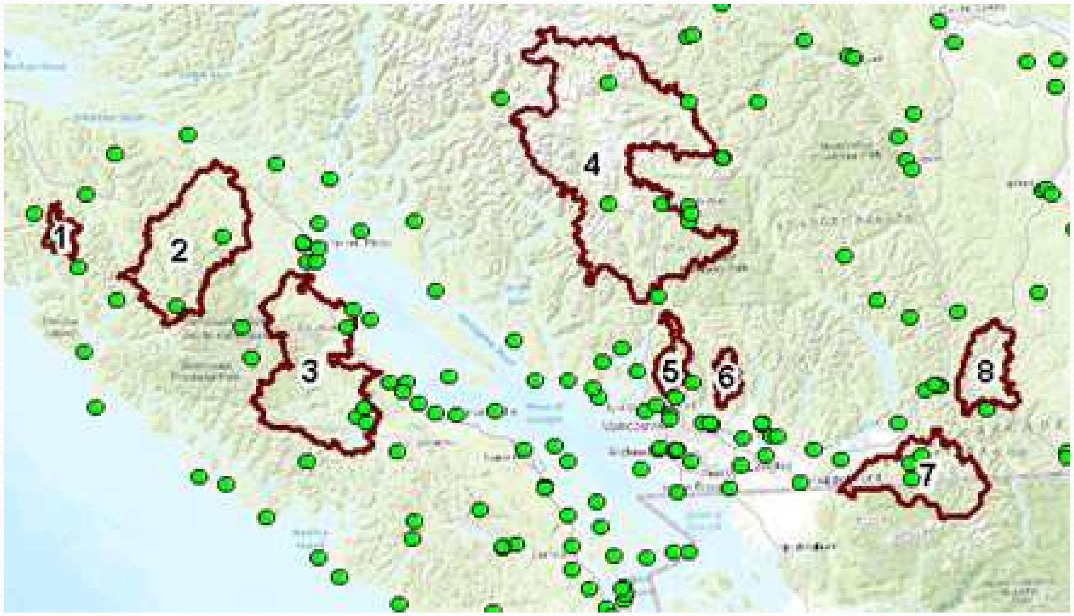
Conclusion:

- Post-processed forecasts are demonstrated to have low bias, and higher accuracy for each lead-time in 15 subareas covering a range of topographical conditions from hills to plains.
- Post-processed forecast ensembles are able to capture peak precipitation events, which caused a major flood event in the study area.
- Raw QPF need to be carefully examined before using in streamflow forecasting.

Ongoing work:

- Apply RPP to other Canadian catchments under different climatic conditions
- Investigate into the influence of density of rain gauges
- Use a gridded reanalysis type data, e.g., CaPA as a substitute for observations.
- Uncertainty in streamflow forecasts using post-processed precipitation forecasts in a hydrologic model.

Ongoing work



Ongoing development at IISER Bhopal

- Development of SWAT model for river basins in MP
- NCMWRF agreed to share 35 km resolution NWP model output
- IIITM provided medium-range 100 km resolution NWP output
- Observation data collected from MPWRD
- Gridded precipitation data from IMD

Acknowledgments

- Alberta and British Columbia River Forecast Centres provided rain gauge data and shape files of catchments.
- Environment Canada provided GDPS forecast data
- CSIRO, Australia to share Rainfall Post-processing tool
- Dr. Gary Bates at NOAA – for his support related to GEFS data
- Late Dr. Peter Rasmussen – for leading FloodNet 3.1 project till January 2017

Thank you!

Contact: sanjeevj@iiserb.ac.in

Schaake shuffle

Our application of the Schaake shuffle is briefly described here.

1. For a given forecast date, an observation sample (date and amount of data) of the same size as that of the ensemble is selected from the historical observation period;
2. The observation sample data for each lead time are ranked. Similarly, the data from the forecast ensemble for each lead time are ranked;
3. A date from the observation sample is 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.

The above procedure is extended for both temporal and spatial correlation in this study.