Hydrol. Earth Syst. Sci. Discuss.,
https://doi.org/10.5194/hess-2017-331-AC1, 2018

Hydrology and
Earth System
Sciences
Discussions
Open Access

# *Interactive comment on* "Evaluation of ensemble precipitation forecasts generated through postprocessing in a Canadian catchment" *by* Sanjeev K. Jha et al.

**Sanjeev K. Jha et al.**

jha.sanj@gmail.com

Received and published: 9 January 2018

Reply to comments of Dr. Fabio Oriani:

We thank Dr. Fabio Oriani for his valuable comments. We would like to mention that we noticed discrepancy in reading data from NWP model outputs. Thanks to Dr. Vincent Fortin from the Environment and Climate Change Canada (ECCC) for pointing out our errors in reading in these datasets. The discrepancies were as follows:

The GDPS data obtained from ECCC provides precipitation forecasts as 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. In the previous analysis, we had used raw precipitation forecasts for each day, resulting in the increasing GDPS forecast bias previously seen on Figure 3.

In case of GEFS data obtained from NOAA, 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). In the previous analysis, we erroneously used summations of forecasts from 3 to 24 hours for days 1, 2 and 3.

In the revised version of the manuscript, modified results corresponding to raw and calibrated GEFS and GDPS are presented in Figures 2 to 7. Also we have modified the text accordingly.

GENERAL COMMENTS: The paper shows the application of a correction technique for daily numerical weather predictions (NWP) for short-term hydrological applications and early warning. The daily rainfall amount on a regular meso-scale (about 50-km) grid shows a consistent bias that is corrected by the proposed technique using a statistical approach based on the mapping of the joint probability distribution between the predicted and observed rainfall amount, in a multivariate normal framework (requiring the transformation of the variables).

The topic is very relevant to the scope of the journal, since the NWP approach is very common for meso-scale weather predictions, together with its problems of bias correction. The research work is not of extreme novelty, since the technique has been already proposed, but it is applied here for the first time on cold climates, bringing useful information for the practitioners. The results are clearly exposed and analyzed, showing a relevant improvement of the predictions brought the technique. I can recommend the paper for publication with some (minor) corrections.

The following main issues should be discussed: 1) Apparently, the results are shown for only 2 catchments over the 15 in the study area;

Response: Now we have provided results of all 15 subcatchments as supplementary figures.

2) The change of support between the observations (point measurement interpolation) and the predictions (25/50-km gridded data);

Response: In this study, we have considered 15 subcatchments, which are used by the Alberta River Forecast Center for hydrological prediction. For hydrological modeling, data are required at the centroid of a subcatchment. Therefore the average of observed precipitation at the centroid of a subcatchment is calculated. An area-weighted forecast is calculated for each subcatchment. In our analysis, we used observation and forecasts at the centroid of the subcatchment. Therefore there is no need to consider the change in support. If the information from observation and forecast were required at grids, then the change of support becomes a critical issue.

3) The authors mention the combined effects of snowmelt and precipitation as the cause of severe floods in the study area, but they don't investigate at all the hydrological response of the catchments, limiting the analysis to rainfall data. What is the effect of the correction on the hydrological response? This additional part would require some extra effort, but would be, in my opinion, a relevant improvement of the paper, constituting a truly novel step ahead in the research work.

Response: We agree with Dr. Oriani that investigation into the hydrological response of the catchments would be truly a novel step. In fact, the development of a hydrological model for Prairie region of Canada is underway and we are planning to explore the hydrological response using improved ensemble precipitation forecasts. However that

task is beyond the scope of current study and is left for future work.

4) you are using 3 years of data including an event with an ultra-centennial return time event (p.5 line 4). One can argue that it is a "lucky" training data set you are using, what is the sensibility of the technique to the lack of training data amount?

Response: In this study, we used daily precipitation over the period of 2013 to 2015 including the heavy precipitation causing the major flood of 2013 in Calgary, Alberta. The main reason for selecting small training data is limited availability of GDPS data from the Environment and Climate Change Canada (ECCC).

Like in other statistical techniques, the availability of more data including extreme events to infer the parameters of the RPP is desirable. This will result into forecasts that are consistent with the observations. Ongoing research lead by the second Author shows that rainfall post-processing parameters are sensitive when less than one year of data is used. The parameters are stabilised for data records greater than a year.

The following are mainly improvement suggestions and minor corrections.

SPECIFIC COMMENTS: P.4 line 10: it is not clear how the Schaake shuffle technique is applied (p.4 line 10). Please give more details about that step.

Response: We have now provided more details about the application of Schaake shuffle technique in the revised version of the manuscript (see Section 3.1).

Section 2.2 about the "statistical treatment of forecast" should be expanded a little to make it clearer: what is the aim? what is the verification score? you already have an ensemble of realizations, why not just validating each realization to obtain an ensemble of scores?

Response: The aim is to calculate uncertainty around raw and calibrated QPFs through a bootstrap procedure. The verification scores are defined in Section 3.2.

We do have ensemble of realizations from calibrated QPFs; however in order to calculate scores for raw QPF, we don't have ensembles. Since our study period is small (3 years only), there will be uncertainty associated with the evaluation scores. By applying bootstrap approach, we are accounting for the sampling uncertainty. Because of 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 have also added these explanations in the revised version of the manuscript (see Section 3.3).

How the space/time auto and cross correlation is preserved over the stochastic realization?

Response: We have presented spatial cross-correlation between subcatchments in the Supplementary Figure SF-3a and SF-3b for calibrated GEFS and GDPS forecasts respectively. By applying Schaake shuffle, the spatial correlation improved, sometimes better than the correlation obtained from observations.

Regarding the temporal correlation within a catchment, Lag-1 Kendall autocorrelation in all subcatchments is shown in the supplementary Figures SF-4a and SF-4b for calibrated GEFS and GDPS respectively. Figures show that the auto correlation was almost zero in the calibrated forecasts before applying Schaake shuffle. However, after the application of Schaake shuffle, the lag-1 correlation becomes closer to that of observation.

Section 2.3 I would add a brief indication about the climate class (e.g. refer to the koppen giger classification) and regime type of the river in the study zone. This section could be considered as an independent one and put before the Methodology (section 2).

Response: We have added the details of climate class and regime type of the river in the study area. 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 in Delavau et al. (2015) shows that the 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 Basin which flows eastward towards Canadian prairies. The combined basin area is approximately 101,720 km2 (AEP, 2017).

P.5 line 9: the "sub-catchment averaged observed precipitation" is in reality a IDW interpolation of punctual measurements, compared to mean values over 25/50 km2 areas. Is it eligible to compare these two types of data? There is a change in support, shouldn't the IDW interpolations be upscaled to the grid resolution of the predictions? For example, keeping the two different resolutions, the variance should be different, independently from the accuracy of the prediction. Also, the density of the rain-gauge network may allow or not a reasonable upscaling to the prediction grid. This is an important point on which the evaluation is very dependent.

Response: We addressed this point before in answer to General Comments (2) above. In summary, we are not comparing anything at a grid scale. It is worth mentioning that there is ongoing research by the second Author on this topic. The Rainfall Post-processing is being applied at a grid level and upscaling as suggested by Dr. Oriani is necessary. In the present study, we are focussing on the observation and forecasts at the centroid of subcatchments. Thus, there is no need to upscale and try to match the grid resolution of forecasts with observations.

As stated in the captions, the results shown in the figures concern catchments 10 and 11 only, what about the other catchments?

Response: To convey the main findings, we had provided results from some of the catchments. Following the suggestion on Dr. Oriani, we have now provided results of all 15 subcatchments as supplementary figures.

TECHNICAL CORRECTIONS: The form is accurate and concise. Minor corrections in the attached pdf.

The minor corrections are incorporated in the revised version of the manuscript.

POSSIBLE IMPROVEMENTS/OPEN QUESTIONS: Is the spread of the corrected projections ensemble realistic with respect to the observed data? A possible quantification, for example, could be based on the frequency of the observed data lying outside a certain confidence boundary of the ensemble. For example the 0.1-0.9 confidence boundary of the projection ensemble should contain 80% of the time the observed data in order to be reliable... in short-term risk assessment application this reliability measure can be relevant.

Response: We calculated the frequency of observations within [0.1,0.9] boundary of calibrated QPFs. The plot is included as Figure 6 in the revised version of the manuscript. In case of calibrated GEFS, the calculated frequency of observed data for lead time 1 to 5 days varies between 0.78 to 0.91. However, for calibrated GDPS, the frequency is 0.87 to 0.91.

As you said, topography plays a major role in the spatial rainfall distribution, how spatial non-stationarity is taken into account in the (raw QPF,observed) probabilistic relation? Is there a relation between the error in the predictions and the topography? And between correction and topography? Empirical joint pdfs between observed/predicted rainfall amount and elevation can be a good analysis tool to underline related strong points or pitfalls of the correction technique.

Response: 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 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.

Please also note the supplement to this comment: https://www.hydrol-earth-syst-sci-discuss.net/hess-2017-331/hess-2017-331-RC1- supplement.pdf

Response: The suggested changes are included in the revised version of the manuscript.

References Alberta Environment and Parks: http://www.environment.alberta.ca/apps/basins/ 2017. Delavau, C., Chun, K., Stadnyk, T., Birks, S., and Welker, J.: North American precipitation isotope ($\delta$18O) zones revealed in time series modeling across Canada and northern United States, Water Resources Research, 51, 1284-1299, 2015. 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. 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.

Please also note the supplement to this comment:
https://www.hydrol-earth-syst-sci-discuss.net/hess-2017-331/hess-2017-331-AC1-supplement.zip

---

Interactive comment on Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2017-331, 2017.