Reply on "Exploring hydrologic post-processing of ensemble streamflow forecasts based on Affine kernel dressing and Nondominated sorting genetic algorithm II"

Jing Xu[1], François Anctil[1], and Marie-Amélie Boucher[2]

[1]Department of Civil and Water Engineering, Université Laval, 1065 avenue de la Médecine, Québec, Québec, Canada;
[2]Department of Civil and Building Engineering, Université de Sherbrooke, 2500 Boul. de l'Université, Sherbrooke, Québec, Canada

Correspondence: Jing Xu (jing.xu.1@ulaval.ca)

Dear Prof. Solomatine and reviewers:

Many thanks for your review comments that we received with respect to our paper. Those valuable comments have significantly enhanced our paper. We have carefully considered and addressed the reviewers' comments and suggestions, which will lead to significant revisions in many parts of the paper. Particularly, we rewrote the introduction section attached at the end of this view letter. Below we hereby provide our point by point responses to each of the reviewer's comments.

1 General questions and remarks:

General question 1 : The aim of the paper should be more clearly stated already (and earlier) in the Introduction.My impression is that we discover the aim of the study while reading the methods and results (for instance, line 262). I also struggled to find out what the novelty of the paper is, with regards to other existing post-processing techniques in the literature. What is the additional (scientific or operational) value of the paper?

Response : Many thanks for your valuable comments. We rewrote the introduction for better clarifying our research aim. Particularly, the novelty of this paper is to emphasize that in the practice, not only quantifying comprehensively, but also communicating the predictive uncertainties in probabilistic forecasts effectively will become an more essential topic progressively. And compared to the conventional post-processing methods, such as Affine kernel dressing (AKD), how the multi objective genetic algorithm (i.e., here, NSGA-II) can open up the opportunities to improve the forecast quality in harmony with the forecasting aims and the specific needs of end-users.

General question 2 : Concerning the Introduction, I found it very difficult to follow the argumentation, since I could not see the direct links between paragraphs, and, most importantly, why the authors were raising, and long discussing, the issue of "sources of uncertainty": if a statistical post-processor is going to be applied, what difference does it make if one, previously, in the raw ensemble, quantified all sources of uncertainty, or, for instance, all but one source of uncertainty? Wouldn't the post-processor work equally well if we had 50 ensemble members from each hydrological model instead of 50x50 members?

Response : Operational forecasters are open to ensemble forecasting methods and products for assessing the flood in a probabilistic way. Their main concerns are how to comprehensively quantify the predictive uncertainties from different sources as well as how to use the uncertainty information for better decision-making. We rewrote the introduction to build stronger and more logical between paragraphs. In addition to clarifying the different sources of uncertainty in the hydrometeorological forecast chain, we explored the possibility of using NSGA-II for better fitting the end-user's specific needs.

General question 3 : Also in the Introduction, overall, I think the key concepts are not introduced very clearly and just loosely thrown in the sentences. For a reader not used to the techniques, it becomes uncomprehensive. For instance, the whole paragraph on lines 49-65 reads very confusing to me. We read about "bias-corrected ensemble member", "normally distributed data", "predictive weights", or "other dressing parameters", without much explanation about what these terms mean.

Response : Thank you for your comments. We rewrote the introduction to give a better explanation of these terms.

General question 4 :Then from line 66 onwards, it is not clear why it is novel to apply NSGA-II and compare it to a kernel-based dressing method. What are the advantages of using NSGA-II? Line 70: what "different conceptualizations" are we talking about? Line 74: what do you mean by "credibility"?

Response : Thanks! "different conceptualizations" refers that the mechanisms of these two statistical post-processing methods (i.e., kernel-based dressing method and NSGA-II) are different. While the term of "credibility" means "reliability". The two techniques share one similarity from another perspective, which is they can estimate the probability density directly from the data (i.e., ensemble forecast) without assuming any particular underlying distribution. The advantages of using NSGA-II is to offer the flexibility to improve the forecast quality in harmony with the forecasting aims and the specific needs of end-users

General question 5 : I think the authors should completely re-write the Introduction, and think about better presenting the literature, the novel aspect of the paper and the questions the paper wants to answer (i.e., its aim). Some review of the literature presented in the "methods" section 3.2 (page 9, lines 179-204) should go to the Introduction to better explain the reader why using NSGA-II could be considered a novel aspect in this paper.

Response : Thanks. We rewrote the introduction to emphasize the novel aspect of this study.

General question 6 : The paper investigates post-processing of ensemble forecasts based on 5 hydrological models and 5 sub-catchments in Canada. However, there is nothing in the paper that discusses the differences in performance among models and sub-catchments? What drives a better/worse performance of the post-processors used in the study? I missed some reflexions about this issue, which would certainly increase the value of the paper. Without this reflexion, and without aggregated (averages) results, I do not understand very well the usefulness of carrying out the study over 5 models and 5 sub-catchments. What does this diversity of applications bring to the analysis?

$Response$: Many thanks for you suggestions. We will add more analysis and discussion for comparing the forecast performance among different models and sub-catchments. The emphasis of this study is to highlight that NSGA-II not only improves the forecast performance compared to conventional post-processing methods but also enhance the predictive uncertainty communication by setting multiple specific objective functions from scratch.

$General\ question\ 7$: I found the distinction between training and validation datasets and criteria very confusing. For instance, we present $MCRPS$ as a validation criteria (section 3.3), but it is then said it is used in calibration (line 269). It is also not clear to me why we do not have a calibration for each lead time. What is the impact of using one unique lead time for calibration?

$Response$: Thanks for you comments. We will modify the criteria for calibration and keep the $MCRPS$ as the verifying score. Besides, the skill of flood forecasts fades away with increasing lead time. The target ensemble has a horizon that extends from day 1 to 7. The 4-day-ahead ensemble forecasts issued from each single-model H-EPSs and their corresponding observations are chosen as a training dataset, since it locates in the middle of the forecast horizon as a compromise.

$General\ question\ 8$: Much of the justification for the selection of the study area comes from its operational role in reservoir management. However, the post-processing application presented in the paper is based on a "non-operational" context: the parameters of the post-processor are calibrated over the entire data available (not over a split sample) for a given lead time (4 days) and validated over different lead times. Operationally, though, a forecaster would have to calibrate the post-processor over a long series of past pairs of forecasts and observations, and apply it to a different set of real-time forecast (for which the observations are not yet available). What are the implications of the method proposed for an operational service? Would the operational service be fine with a post-processing that is optimized for a 4-day lead time? Is that the lead-time that most count for the service when forecasting over these catchments? Maybe some lines of discussion would be interesting in the final section of the paper.

$Response$: Thanks for you valuable suggestions. We will add more discussion and reflexions in the final section to explore what potential benefits will post-processing techniques will bring to the operational services.

2 Specific questions and remarks:

Specific question 1 : lines 23-24: these sentences are not very clear to me.

Response : Thanks. We will rewrote the abstract as well to make it more clear.

Specific question 2 : line 38: what are the three main sources mentioned?

Response : These different sources of uncertainty are related to deficiencies in the: (1) meteorological forcing; (2) mis-specified hydrologic initial and boundary conditions; (3) inherent hydrologic model structure errors, and biased estimated parameters.

Specific question 3 : line 47-48: what are the implications of autocorrelation in the post-processing? Besides, aren't meteorological forecasts also auto-correlated? Why is it specifically a problem to hydrological forecasts?

Response : Yes, the autocorrelation is the problem for both meteorological forecasts and hydrological forecasts. We deleted this description in the updated version of introduction.

Specific question 4 : Fig. 2: I understand these are daily streamflow (it is written: mm/day) averaged over each month, and not monthly streamflows. Is that so? The caption should state the period over which the averages were obtained.

Response : Yes, these are daily streamflow (mm/day) averaged over each month. We will modify the caption. Thanks.

Specific question 5 : Table 1: I do not understand the data on reservoir area: why it is important to this paper? Furthermore, I do not understand all these physical and climatic data provided: if the results are not going to be interpreted according to the characteristics of the catchments, why are these characteristics presented in the table? In what do they influence the results?

Response : Thanks. We will modify this table to keep the useful characteristics especially for this study.

Specific question 6 : Line 122: why have you chosen 5 models and why not work with the 20 models? If this is a matter of computational time, could you explain it to the reader? How long it takes to post-process one single model H-EPS?

Response : All of these 5 models are lumped models. They are representative of the 20 models. It is more of a layout concern rather than computation time.

Specific question 7 : Line 136: I am used to forecast post-processing, but not with the term “ensemble interpretation method” or “interpreted ensemble (line 157). I would be happy with more explanations here.

Response : Thanks. We will add more useful explanations here.

Specific question 8 : Line 147: correct English

Response : We will rephrase this content.

Specific question 9 : Line 169: what is this rule of thumb? Please, clarify.

Response : Thanks. We will add more explanation about this parameter h_S (Silverman, 1986) in the methodology section.

Specific question 10 : Equations, overall: it seems to me that not all terms are always defined, explained after the equations where they are presented. r1, r2, s1, s2, etc. zi is lower case in equation 10 but upper case in equation 11. a is alpha (line 169)? Please, check the equations and the way terms are presented.

Response : Thanks. We will check all those equations and define all related terms.

Specific question 11 : Line 175: “Eq. (6) can be further defined”, should maybe be replaced to “can be re-written”

Response : Thanks. We will rephrase it as "can be re-written".

Specific question 12 : Line 205: X_t was already defined in line 143. Please, check.

Response : Thanks. Yes, it is. We will remove the description of X_t in Line 205.

Specific question 13 : Line 191, 192: I tend not to agree with the authors. I think “accuracy” is what is first of all searched when issuing a forecast at a given day for a short lead time such as 7 days. This is specially the case for flood events, for instance. Please, explain your arguments.

Response : Thanks. Yes, probabilistic forecasts must be, first of all, accurate. We will rephrase this paragraph.

$Specific\,question$ 14 : Lines 195-196: not very clear to me. Hydrologists may rely on NSE, but for simulations (long time series), not necessarily for forecasters. Please, clarify.

$Response$: Thanks. We will add more clarification here.

$Specific\,question$ 15 : Line 200: I do not understand “elitist”. Please, clarify.

$Response$: The Nondominated sorting genetic algorithm II (NSGA-II; Deb et al. (2002)) is admitted as a fast and elitist multiobjective genetic algorithm, adopted for searching for the Pareto solution set. I will add more description about the "elitism" of NGSA-II in section 3.2.

$Specific\,question$ 16 : Line 215: the concept of crowding distance was not clear to me. Please, clarify.

$Response$: Thanks. I will add more description about the "crowding distance" of NGSA-II in section 3.2.

$Specific\,question$ 17 : line 235: was the $MCRPS$ calculated using empirical distributions or a fitted theoretical distribution? Please, clarify.

$Response$: Thanks. We will clarify this in the section 3.3.

$Specific\,question$ 18 : line 238: why do you need both, MAE and MSE?

$Response$: Thanks. We will only keep MSE later.

$Specific\,question$ 19 : line 248-249: I do not understand why the Taylor diagram is mentioned here. Did you use it? How? Can you explain it?

$Response$: Thanks. We will remove this short description of the Taylor diagram.

$Specific\,question$ 20 : Figure 3: where do we find “w” in the text (output of NSGA-II in the figure)?

$Response$:Thanks. We will redraw this flowchart and provide more details.

$Specific\ question$ 21 : lines 287-288: it is not unexpected that forecast performance decreases with lead time. I do no understand why it is “revealed” here. Please, clarify.

$Response$:Thanks. We will rephrase this sentence.

$Specific\ question$ 22 : lines 293-294: check for the English language.

$Response$: Thanks. We will check the English here.

$Specific\ question$ 23 : line 324: delete “In the meanwhile,”

$Response$: Thanks. We will delete "In the meanwhile,".

$Specific\ question$ 24 : line 335: I understand that “error growth” is usually depicted as an increase in spread with lead time and decrease in accuracy. Why should it be maintained for a single model H-EPS if the post-processor was calibrated for 4 days of lead time only and applied to other lead times? Please, clarify.

$Response$: Thanks. The target ensemble has a horizon that extends from day 1 to 7. The 4-day-ahead ensemble forecasts issued from each single-model H-EPSs and their corresponding observations are chosen as a training dataset, since it locates in the middle of the forecast horizon as a compromise. We will add more clarification here.

$Specific\ question$ 25 : Figure 9 is not explained in the text (notably the number of lines in each graph). Also, why AKD seems to work well with M05?

$Response$: We will add more explanations in text about Figure 9.

$Specific\ question$ 26 : line 343-344: not clear; please, revise it.

$Response$: Thanks. We will rephrase this paragraph.

$Specific\ question$ 27 : line 345: figure 10 shows much more than spread. Please, clarify when presenting (fully) the figure.

$Response$: Thanks. We will modify the description and analysis for Figure 10.

Specific question 28 : Figure 10 is very difficult to read. It is not clear (BW print) which graph is AKD, which is NSGA-II. We can barely see what is inside the figure. I think it needs to be re-designed.

Response : We will enlarge the figure size and add more analysis corresponding to Figure 10.

Specific question 29 : Overall, terminology could be uniformed (ex., use of AKD).

Response : Thanks. We will uniform the terminology in the whole paper.

Specific question 30 : It is a pity that the paper does not have a discussion section. I would suggest the authors to introduce one, commenting further the results obtained, comparing post-processing performance among catchments (i.e., geographic location) and hydrological models in a summarized way. This piece of work is missing in the paper and would better justify the use of several catchments and models in the analysis.

Response : Thanks for this suggestion. We will add further discussion section about: (1) comparing post-processing performance among catchments hydrological models in a summarized way; (2) highlight the novelty and potential benefits that post-processing techniques may bring to the operational services.

3 Introduction

Hydrologic forecasting is crucial for flood warning and mitigation (e.g., Shim and Fontane, 2002; Cheng and Chau, 2004), water supply operation and reservoir management (e.g., Datta and Burges, 1984; Coulibaly et al., 2000; Boucher et al., 2011), navigation, and other related activities. Sufficient risk awareness, enhanced disaster preparedness in the flood mitigation measures, and strengthened early warning systems are crucial in reducing the weather-related event losses. Hydrologic models are typically driven by dynamic meteorological models in order to issue forecasts over a medium range horizon of 2 to 15 days (Cloke and Pappenberger, 2009). This kind of coupled hydrometeorologic forecasting systems are admitted as effective tools to issue longer lead times. Inherent in the coupled hydrometeorologic forecasting systems, some predictive uncertainties are then inevitable given the limits of knowledge and available information (Ajami et al., 2007). In fact, those uncertainties occur all along the different steps of the hydrometeorological modeling chain (e.g., Liu and Gupta, 2007; Beven and Binley, 2014). These different sources of uncertainty are related to deficiencies in the meteorological forcing, mis-specified hydrologic initial and boundary conditions, inherent hydrologic model structure errors, and biased estimated parameters (e.g., Vrugt and Robinson, 2007; Ajami et al., 2007; Salamon and Feyen, 2010; Thiboult et al., 2016). Among most cases, a single deterministic forecasts turns out to be way more insufficient.

Many substantive theories have been proposed in order to quantify and reduce the different sources of cascading forecast uncertainties and to add good values to flood forecasting and warning. Among them, the superiority of ensemble forecasting systems in quantifying the propagation of predictive uncertainties (over deterministic systems) is now well established (e.g., Cloke and Pappenberger, 2009; Palmer, 2002; Seo et al., 2006; Velázquez et al., 2009; Abaza et al., 2013; Wetterhall et al., 2013; Madadgar et al., 2014). Numerous challenges have been well tackled, for example: (1) meteorological ensemble prediction systems (M-EPSs) (e.g., Palmer, 1993; Houtekamer et al., 1996; Toth and Kalnay, 1997) are refined and operated worldwide by national agencies such as the European Centre for Medium-Range Weather Forecasts (ECMWF), the National Center for Environmental Prediction (NCEP), the Meteorological Service of Canada (MSC), and more; (2) the forecast accuracy is highly improved by adopting higher resolution data collection and assimilation. Sequential data assimilation techniques, such as the particle filter (e.g., Moradkhani et al., 2012; Thirel et al., 2013) and the ensemble Kalman filter (e.g., Evensen , 1994; Reichle et al., 2002; Moradkhani et al, 2005; McMillan et al., 2013) provide an ensemble of possible re-initializations of the initial conditions, expressed in the hydrologic model as state variables, such as soil moisture, groundwater level and so on; (3) forecasting skills of the coupled hydrometeorologic forecasting systems are also improved by tracking predictive errors using the full uncertainty analysis. Multimodel schemes were proposed to increase performance and decipher structural uncertainty (e.g., Duan et al., 2007; Fisher et al., 2008; Weigel et al., 2008; Najafi et al., 2011; Velázquez et al., 2011; Marty et al., 2015; Mockler et al., 2016). Thiboult et al. (2016) compared many H-EPS, accounting for the three main sources of uncertainties located along the hydrometeorological modeling chain. They pointed out that EnKF probabilistic data assimilation provided most of the dispersion for the early forecasting horizons but failed in maintaining its effectiveness with increasing lead times. A multimodel scheme allowed sharper and more reliable ensemble predictions over a longer forecast horizon; (4) statistical hydrologic post-processing component is added in the H-EPS for rectifying biases and dispersion errors (i.e., too narrow/too large) are numerous, as reviewed by Li et al. (2017). It is noteworthy that many hydrologic variables, such as discharge, follow a skewed distribution (i.e., low probability associated to the highest streamflow values), which complicates the task. Usually, in a hydrologic ensemble prediction system (H-EPS) framework (e.g., Schaake et al., 2007; Cloke and Pappenberger, 2009; Velázquez et al., 2009; Boucher et al., 2012; Abaza et al., 2017), the post-processing procedure over the atmospheric input ensemble is often referred as pre-processing, while post-processing aims at improving the hydrologic ensemble forecasting outputs.

However, another challenge still remains: how to improve the human interpretation of probabilistic forecasts and the communication of integrated ensemble forecast products to end-users (e.g., operational hydrologists, water managers, local conservation authorities, stakeholders and other relevant decision makers). This step is considered to be the key part of facilitating the implementation of H-EPS in real-time operational forecasting effectively. Buizza et al. (2007) emphasized that both functional and technical qualities are supposed to be assessed for evaluating the overall forecast value of a hydrometeorologic forecasts. Ramos et al. (2010) further noted that the best way to communicate probabilistic forecast and interpret its usefulness should be in harmony with the goals of the forecasting system and the specific needs of end-users. She also demonstrated the main achievements from two studies obtained from a Member States workshop (Thielen et al., 2005) role-play game and another survey to explore the users' risk perception of forecasting uncertainties and how they dealt with uncertain forecasts for decision-making. The results revealed that there is still space for enhancing the forecasters' knowledge and experience on bridge the community gap between predictive uncertainties quantification and effective decision-making.

Hence, in practice, which forecast quality impacts a given decision the most? Different end-users share their unique requirements: Crochemore et al. (2017) produced the seasonal streamflow forecasting by conditioning climatology with precipitations indices (SPI3). Forecast reliability, sharpness (i.e., spread), overall performance and low-flow event detection were verified to assess the conditioning impact. In some cases, the reliability and sharpness could be improved simultaneously while more often, there was a trade-off between them. Another IMPREX project conduct an optimization for the reservoir-based hydropower production to explore the relationship between the forecast quality and economic values. They found that an over-estimation comes along with more penalization. In the operational filed, not only quantifying, but also communicating the predictive uncertainties in probabilistic forecasts will become an more essential topic progressively.

The study is a contribution to probe this topic by exploring hydrological post-processing of ensemble streamflow forecasts based on Affine kernel dressing and Non-dominated sorting genetic algorithm II. The mechanisms of these two statistical post-processing methods are completely different, however, they share one similarity from another perspective, which is they can estimate the probability density directly from the data (i.e., ensemble forecast) without assuming any particular underlying distribution. As a more conventional method, Silverman (1986) firstly proposed the kernel density smoothing method to estimate the distribution from the data by centering a kernel function K that determines the shape of a probability distribution (kernel) fitted around every data point (i.e., the bias-corrected ensemble member). The smooth kernel estimate is then the sum of those kernels. As for the choice of bandwidth h of each dressing kernel, Silverman's rule of thumb finds an optimal h by assuming that the data is normally distributed. Improvements to the original idea were soon to follow. For instance, the improved Sheather Jones (ISJ) algorithm is more suitable and robust with respect to multimodality (Wand and Jones, 1994). Roulston and Smith (2003) rely on the series of "best forecasts" (i.e., best-member dressing) to compute the kernel bandwidth. Wang and Bishop (2005) as well as Fortin et al. (2006) further improved the best member method. The later advocated that the more extreme ensemble members are more likely to be the best member of raw under-dispersive forecasts, while the central members tend to be more "precise" for over-dispersive ensemble. They proposed the idea that different predictive weights should be set over each ensemble member, given each member's rank within the ensemble. Instead of standard dressing kernels that act on individual ensemble members, Bröcker and Smith (2008) proposed the affine kernel dressing (AKD) by assuming an affine mapping between ensemble and observation over the entire ensemble. The mapping parameters are determined from the training data simultaneously with the other dressing parameters. They approximate the distribution of the observation given the ensemble.

While the other post-processor of Non-dominated sorting genetic algorithm II (NSGA-II) open up the opportunity of improving the forecast quality in harmony with the forecasting aims and the specific needs of end-users. Given the single-model H-EPSs studied here, the hydrologic ensemble is generated by activating two forecasting tools: the ensemble weather forecast and the EnKF. Henceforth, enhancing the H-EPS forecasting skill by assigning different credibility to ensemble members becomes preferred than reducing the number of members. Multiple objective functions (i.e., here, verifying scores) for evaluating the forecasting performances of the H-EPS are selected to guide the optimization process. The expected output is a group of solutions, also known as Pareto fronts, that can give the trade-offs between different objectives. Other post-processing techniques, like the Non-dominated sorting genetic algorithm II (NSGA-II), are now common (e.g., Liong et al., 2001; De Vos and Rientjes, 2007; Confesor and Whittaker, 2007). Such techniques are conceptually linked to the multiobjective parameter calibration of hydrologic models using Pareto approaches. Indeed, formulating a model structure or representing the hydrologic processes using a unique global optimal parameter set proves to be very subjective. Multiple optimal parameter sets exist with satisfying behavior given the different conceptualizations, albeit not identical Beven and Binley (1992). For example, Brochero et al. (2013) utilized the Pareto fronts generated with NSGA-II for selecting the "best" ensemble from a hydrologic forecasting model with a pool of 800 streamflow predictors, in order to reduce the H-EPS complexity.

In this study, the daily streamflow ensemble forecasts issued from five single-model H-EPSs over the Gatineau River (Province of Québec, Canada) are post-processed. Details about the study area, hydrologic models, and hydrometeorologic data are described in Section 2. Section 3 explains the methodology and training strategy of Affine kernel dressing (AKD) and Non-dominated sorting genetic algorithm II (NSGA-II) methods, in parallel with the scoring rules that evaluate the performance of the forecasts. Specific concepts associated with those scores are also introduced in this section. Predictive distribution estimation based on the five single-model H-EPSs configurations, which lack accounting for the model structure uncertainty, is presented in Section 4. The comparison of both statistical post-processing methods in improving the forecasting quality as well as enhancing the uncertainty communication are discussed and analyzed as well. Conclusion follows in Section 5.

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, Mon. Weather. Rev., 141(10), 3462-3476, https://doi.org/10.1175/MWR-D-12-00206.1, 2013.

Abaza, M., Anctil, F., Fortin, V., and Perreault, L.: Hydrological Evaluation of the Canadian Meteorological Ensemble Reforecast Product, Atmos.Ocean., 55(3), 195-211, https://doi.org/10.1080/07055900.2017.1341384, 2017.

Ajami, N.K., Duan, Q., and Sorooshian, S.: An integrated hydrologic Bayesian multimodel combination framework: Confronting input, parameter, and model structural uncertainty in hydrologic prediction, Water. Resour. Res., 43(1), 1-19, https://doi.org/10.1029/2005WR004745, 2007.

Bergström, S. and Forsman, A.: Development of a conceptual deterministic rainfall-runoff model, Nordic. Hydr., 4, 147-170. https://doi.org/10.2166/nh.1973.013, 1973.

Beven, K. and Binley, A.: GLUE: 20 years on, Hydrol. Process., 28(24), 5897-5918, https://doi.org/10.1002/hyp.10082, 2014.

Beven, K. and Binley, A.: The future of distributed models: Model calibration and uncertainty prediction, Hydrol. Process., 6(3), 279-298, https://doi.org/10.1002/hyp.3360060305, 1992.

Buizza, R. Asensio, H. Balint, G. Bartholmes J, et al.: EURORISK/PREVIEW report on the technical quality, functional quality and forecast value of meteorological and hydrological forecasts, ECMWF Technical Memorandum, ECMWF Research Department: Shinfield Park, Reading, United Kingdom, 516, 1-21, http://www.ecmwf.int/publications/, 2007.

Burnash, R.J.C., Ferral, R.L., and McGuire, R.A.: A generalized streamflow simulation system: conceptual modeling for digital computers, Technical Report, Joint Federal and State River Forecast Center, US National Weather Service and California Department of Water Resources, Sacramento, 204 pp., 1973.

Brochero, D., Gagné, C., and Anctil, F.: Evolutionary multiobjective optimization for selecting members of an ensemble streamflow forecasting model. Proceeding of the fifteenth annual conference on Genetic and evolutionary computation conference-GECCO, New York, United States, 6 July 2013, 13, 1221-1228, https://doi.org/10.1145/2463372.2463538, 2013.

Bröcker, J. and Smith, L.A.: From ensemble forecasts to predictive distribution functions, Tellus. A., 60(4), 663-678, https://doi.org/10.1111/j.1600-0870.2007.00333.x, 2008.

Boucher, M.A., Anctil, F., Perreault, L., and Tremblay, D.: A comparison between ensemble and deterministic hydrological forecasts in an operational context, Adv. Geosci., 29, 85-94, https://doi.org/ 10.5194/adgeo-29-85-2011, 2011.

Boucher, M.A., Tremblay, D., Delorme, L., Perreault, L., and Anctil, F.: Hydro-economic assessment of hydrological forecasting systems, J. Hudrol., 416, 133-144, https://doi.org/10.1016/j.jhydrol.2011.11.042, 2012.

Cheng, C.T. and Chau, K.W.: Flood control management system for reservoirs, Environ. Modell. Softw., 19(12), 1141-1150, https://doi.org/10.1016/j.envsoft.2003.12.004, 2004.

Cloke, H.L. and Pappenberger, F.: Ensemble flood forecasting: A review, J. Hydrol., 375, 613-626, https://doi.org/10.1016/j.jhydrol.2009.06.005, 2009.

Confesor, Jr.R.B. and Whittaker, G.W., 2007. Automatic Calibration of Hydrologic Models With Multi-Objective Evolutionary Algorithm and Pareto Optimization 1, JAWRA Journal of the American Water Resources Association, 43(4), 981-989, https://doi.org/10.1111/j.1752-1688.2007.00080.x, 2007.

Coulibaly, P., Anctil, F., and Bobée, B.: Daily reservoir inflow forecasting using artificial neural networks with stopped training approach, J. Hydrol., 230, 244-257, https://doi.org/10.1016/S0022-1694(00)00214-6, 2000.

Crochemore, L., Ramos, M.H., Pappenberger, F., and Perrin, C.: Seasonal streamflow forecasting by conditioning climatology with precipitation indices, Hydrol.Earth Syst. Sci., 21, 1573-1591, https://doi.org/10.5194/hess-21-1573-2017, 2017.

De Vos, N.J. and Rientjes, T.H.M.: Multi-objective performance comparison of an artificial neural network and a conceptual rainfall—runoff model, Hydrolog. Sci. J., 52(3), 397-413, https://doi.org/10.1029/2007WR006734, 2007.

Datta, B. and Burges, S.J.: Short-term, single, multiple-purpose reservoir operation: importance of loss functions and forecast errors, Water. Resour. Res., 20, 1167-1176, https://doi.org/10.1029/WR020i009p01167, 1984.

Deb, K., Pratap, A., Agarwal, S., and Meyarivan, T.A.M.T.: A fast and elitist multiobjective genetic algorithm: NSGA-II. IEEE. T. Evolut. Comput., 6(2), 182-197, https://doi.org/10.1109/4235.996017, 2002.

Duan, Q., Ajami, N.K., Gao, X., Sorooshian, S.: Multi-model ensemble hydrologic prediction using Bayesian model averaging, Adv. Water. Resour., 1371-1386, https://doi.org/10.1016/j.advwatres.2006.11.014, 2007.

Evensen, G.: Inverse Methods and Data Assimilation in Nonlinear Ocean Models, Physica. D., 77, 108-129, https://doi.org/10.1016/0167-2789(94)90130-9http://dx.doi.org/10.1016/0167-2789(94)90130-9, 1994.

Fraley, C., Raftery, A.E., and Gneiting, T.: Calibrating Multimodel Forecast Ensembles with Exchangeable and Missing Members Using Bayesian Model Averaging, Mon. Weather. Rev., 138, 190-202, https://doi.org/10.1175/2009MWR3046.1, 2010.

Fisher, J.B., Tu, K.P., and Baldocchi, D.D.: Global estimates of the land-atmosphere water flux based on monthly AVHRR and ISLSCP-II data, validated at 16 FLUXNET sites, Remote Sens. Environ., 112, 901-919, https://doi.org/10.1016/j.rse.2007.06.025, 2008.

Fortin, V., Abaza, M., Anctil, F., and Turcotte, R.: Why Should Ensemble Spread Match the RMSE of the Ensemble Mean? J. Hydrometeorol., 15, 1708-1713, https://doi.org/10.1175/JHM-D-14-0008.1, 2014.

Fortin, V., Favre, A.C., Saïd, M., 2006. Probabilistic forecasting from ensemble prediction systems: Improving upon the best-member method by using a different weight and dressing kernel for each member, Q. J. R. Meteorol Soc., 132, 1349-1369, https://doi.org/10.1256/qj.05.167, 2006.

Gaborit, É., Anctil, F., Fortin V., and Pelletier, G.: On the reliability of spatially disaggregated global ensemble rainfall forecasts, Hydrol. Process., 27(1), 45-56. https://doi.org/10.1002/hyp, 2013.

Gneiting, T. and Raftery, A.E.: Strictly proper scoring rules, prediction, and estimation, J. Am. Stat. Assoc., 102, 359-378. https://doi.org/10.1198/016214506000001437, 2007.

Gupta, H.V., Kling, H., Yilmaz, K.K., and Martinez, G.F.: Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modeling, J. Hydrol., 377, 80-91, https://doi.org/10.1016/j.jhydrol.2009.08.003, 2009.

Hersbach, H., 2000. Decomposition of the Continuous Ranked Probability Score for Ensemble Prediction Systems,Wea. Fore., 15, 559-570, https://doi.org/10.1175/1520-0434(2000)015<0559:DOTCRP>2.0.CO;2, 2000.

Houtekamer, P.L., Lefaivre, L., Derome, J., and Ritchie, H.: A system simulation approach to ensemble prediction, Mon. Weater. Rev., 124, 1225-1242, https://doi.org/10.1175/1520-0493(1996)124<1225:ASSATE>2.0.CO;2, 1996.

Jakeman, A.J., Littlewood, I.G., and Whitehead, P.G.: Computation of the instantaneous unit hydrograph and identifiable component flows with application to two small upland catchments, J. Hydrol., 117, 275-300, https://doi.org/10.1016/0022-1694(90)90097-H, 1990.

Jewson, S.: Comparing the ensemble mean and the ensemble standard deviation as inputs for probabilistic medium–range temperature forecasts, arXiv preprint physics, arXiv [preprint], arXiv:0310059, 13 October 2003.

Klemeš, V.: Operational testing of hydrological simulation models, Hydrolog. Sci. J., 31(1), 13-24, https://doi.org/10.1080/02626668609491024, 1986.

Kling, H., Fuchs, M., and Paulin, M.: Runoff conditions in the upper Danube basin under an ensemble of climate change scenarios, J. Hydrol., 424, 264-277, https://doi.org/10.1016/j.jhydrol.2012.01.011, 2012

Kulturel-Konak, S., Smith, A.E., and Norman, B.A.: Multi-objective search using a multinomial probability mass function, Eur. J. Oper. Res., 169(3), 918-931, https://doi.org/10.1016/j.ejor.2004.08.026, 2006.

Kottek, M., Grieser, J., Beck, C., Rudolf, B., and Rubel, F.: World Map of the Köppen-Geiger climate classification updated, Meteorol. Z., 15(3), 259-263. https://doi.org/10.1127/0941-2948/2006/0130, 2006.

Li, W., Duan, Q., Miao, C., Ye, A., Gong, W., and Di, Z.: A review on statistical postprocessing methods for hydrometeorological ensemble forecasting, Wires. Water., 4(6), e1246. https://doi.org/10.1002/wat2.1246, 2017.

Liong, S.Y., Khu, S.T. and Chan, W.T.: Derivation of Pareto front with genetic algorithm and neural network, J. Hydrol. Eng., 6(1), 52-61, https://doi.org/10.1061/(ASCE)1084-0699(2001)6:1(52), 2001.

Liu, Y. and Gupta, H. V.: Uncertainty in hydrologic modeling: Toward an integrated data assimilation framework, Water. Resour. Res., 43, 1-18, https://doi.org/10.1029/2006WR005756, 2007.

Madadgar, S., Moradkhani, H., and Garen, D.: Towards improved post-processing of hydrologic forecast ensembles, Hydrol. Process., 28(1), 104-122, https://doi.org/10.1002/hyp.9562, 2014.

Marty, R., Fortin, V., Kuswanto, H., Favre, A.C., and Parent, E.: Combining the bayesian processor of output with bayesian model averaging for reliable ensemble forecasting, J. R. Stat. Soc. C-Appl., 64, 75-92, https://doi.org/10.1111/rssc.12062, 2015.

Matheson, J.E. and Winkler, R.L.: Scoring Rules for Continuous Probability Distributions. Management Science, Manage. Sci., 22(10), 1087-1096, https://doi.org/10.1287/mnsc.22.10.1087, 1976.

McMillan, H.K., Hreinsson, E.O., Clark, M.P., Singh, S.K., Zammit, C., and Uddstrom, M.J.: Operational hydrological data assimilation with the recursive ensemble Kalman filter, Hydrol. Earth Syst. Sci., 17, 21-38, https://doi.org/10.5194/hess-17-21-2013, 2013.

Mockler, E.M., O'Loughlin, F.E., and Bruen, M.: Understanding hydrological flow paths in conceptual catchment models using uncertainty and sensitivity analysis, Comput Geosci., 90, 66-77, https://doi.org/10.1016/j.cageo.2015.08.015, 2016.

Moradkhani, H., Dechant, C.M., and Sorooshian, S.: Evolution of ensemble data assimilation for uncertainty quantification using the particle filter-Markov chain Monte Carlo method, Water Resour. Res., 48(12), 121-134, https://doi.org/10.1029/2012WR012144, 2012.

Moradkhani, H., Sorooshian, S., Gupta, H.V., and Houser, P.R.: Dual state-parameter estimation of hydrological models using ensemble Kalman filter, Adv. Water. Resour., 28(2), 135-147, https://doi.org/10.1016/j.advwatres.2004.09.002, 2005.

Movahedinia, F.: Assessing hydro-climatic uncertainties on hydropower generation, Université Laval, Québec city, 7 pp, https://corpus.ulaval.ca/jspui/handle/20.500.11794/25294, 2014.

Najafi, M.R., Moradkhani, H., and Jung, I.W.: Assessing the uncertainties of hydrologic model selection in climate change impact studies, Hydrol. Process., 25, 2814-2826, https://doi.org/10.1002/hyp.8043, 2011.

Nash, J.E. and Sutcliffe, I.: River flow forecasting through conceptual models. Part 1-A discussion of principles. J. Hydrol., 10(3), 282–290, https://doi.org/10.1016/0022-1694(70)90255-6, 1970.

Oudin, L., Michel, C., and Anctil, F.: Which potential evapotranspiration input for a lumped rainfall-runoff model? Part 1 - Can rainfall-runoff models effectively handle detailed potential evapotranspiration inputs? J. Hydrol., 303, 275–289, https://doi.org/10.1016/j.jhydrol.2004.08.025, 2005.

Palmer, T.N.: The economic value of ensemble forecasts as a tool for risk assessment: From days to decades, Q. J. R. Meteorol. Soc., 128, 747-774, https://doi.org/10.1256/0035900021643593, 2002.

Palmer, T.N.: Extended-Range Atmospheric Prediction and the Lorenz Model, Bull. Am. Meteorol. Soc., 74, 49-65, https://doi.org/10.1175/1520-0477(1993)074<0049:ERAPAT>2.0.CO;2, 1993.

Ramos, M.H., Mathevet, T., Thielen, J. and Pappenberger, F.: Communicating uncertainty in hydro-meteorological forecasts: mission impossible? Meteorol. Appl., 17(2), pp.223-235, https://doi.org/10.1002/met.202, 2010.

Pappenberger, F., Beven, K.J., Hunter, N.M., Bates, P.D., Gouweleeuw, B.T., Thielen, J., and de Roo, A.P.J.: Cascading model uncertainty from medium range weather forecasts (10 days) through a rainfall-runoff model to flood inundation predictions within the Euro- pean Flood Forecasting System (EFFS), Hydrol. Earth. Syst. Sc., 9, 381–393, https://hal.archives-ouvertes.fr/hal-00304846, 2005.

Perrin, C.: Vers une amélioration d'un modèle global pluie-débit, PhD diss., Institut National Polytechnique de Grenoble-INPG, Grenoble, 287 pp, 2000.

Perrin, C., Michel, C., and Andréassian, V.: Improvement of a parsimonious model for streamflow simulation, J. Hydrol., 279, 275-289, https://doi.org/10.1016/S0022-1694(03)00225-7, 2003.

Reichle, R., McLaughlin, D.B., and Entekhabi, D.: Hydrologic data assimilation with the ensemble Kalman filter, Mon. Weather Rev., 130, 103-114, https://doi.org/10.1175/1520-0493(2002)130<0103:HDAWTE>2.0.CO;2, 2002.

Roulston, M.S. and Smith, L.A.: Combining dynamical and statistical ensembles. Tellus. A., 55, 16-30, https://doi.org/10.3402/tellusa.v55i1.12082, 2003.

Salamon, P. and Feyen, L.: Disentangling uncertainties in distributed hydrological modeling using multiplicative error models and sequential data assimilation, Water Resour. Res., 46(12), 1-20, https://doi.org/10.1029/2009WR009022, 2010.

Schaffer, J.: Multiple Objective Optimization with Vector Evaluated Genetic Algorithms, Proceedings of the First International Conference on Genetic Algortithms, Lawrence Erlbaum Associates. Inc., 93-100, July 1985.

Seiller, G., Roy, R., and Anctil, F.: Influence of three common calibration metrics on the diagnosis of climate change impacts on water resources, J. Hydrol., 547, 280-295, https://doi.org/10.1016/j.jhydrol.2017.02.004, 2017.

Seiller, G., Anctil, F., and Perrin, C.: Multimodel evaluation of twenty lumped hydrological models under contrasted climate conditions, Hydrol. Earth. Syst. Sc., 16, 1171-1189, https://doi.org/DOI : 10.5194/hess-1116-1171-2012, 2012.

Seo, D., Herr, H.D., and Schaake, J.C.: A statistical post-processor for accounting of hydrologic uncertainty in short-range ensemble streamflow prediction, Hydrol. Earth. Syst. Sc., 3(4), 1987-2035, https://hal.archives-ouvertes.fr/hal-00298744, 2006.

Shim, K.C., Fontane, D.G., and Labadie, J.W.: Spatial Decision Support System for Integrated River Basin Flood Control, J. Water. Res. Pl-Asce., 128, 190-201, https://doi.org/10.1061/(ASCE)0733-9496(2002)128:3(190), 2002.

Schaake, J., Demargne, J., Hartman, R., Mullusky, M., Welles, E., Wu, L., Fan, X., and Seo, D.J.: Precipitation and temperature ensemble forecasts from single-value forecasts, Hydrol. Earth. Syst. Sci. Discuss., 4(2), 655-717. https://hal.archives-ouvertes.fr/hal-00298823/document, 2007.

Silverman, B.W.: Density estimation for statistics and data analysis, CRC press, London, 26, 1986.

Stanski, H.R., Wilson, L.J., and Burrows, W.R.: Survey of common verification methods in meteorology, World Weather Watch Tech. Report 8, WMO/TD, 358, 114 pp, 1989.

Sloughter, J.M.L., Raftery, A.E., Gneiting, T., and Fraley, C.: Probabilistic Quantitative Precipitation Forecasting Using Bayesian Model Averaging, Mon. Weather. Rev., 135, 3209-3220. https://doi.org/10.1175/MWR3441.1, 2007.

Thiboult, A., Anctil, F., Boucher, M.A.: Accounting for three sources of uncertainty in ensemble hydrological forecasting, Hydrol. Earth. Syst. Sc., 20(5), 1809-1825, https://doi.org/10.5194/hess-20-1809-2016, 2016.

Thiboult, A., Seiller, G., Poncelet, C., and Anctil, F.: The HOOPLA toolbox: a HydrOlOgical Prediction LAboratory to explore ensemble rainfall-runoff modeling, arXiv [preprint], Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2020-6, 28 January 2020.

Thielen, J., Ramos,M.H., Bartholmes, J., De Roo, A., Cloke, H., Pappenberger, F., and Demeritt, D.: Summary report of the 1st EFAS workshop on the use of Ensemble Prediction System in flood forecasting, European Report EUR, Ispra., 22118. http://floods.jrc.ec.europa.eu/efas-documents, 2005.

Thirel, G., Salamon, P., Burek, P., and Kalas, M.: Assimilation of MODIS snow cover area data in a distributed hydrological model using the particle filter, Remote. Sens., 5, 5825-5850. https://doi.org/10.3390/rs5115825, 2013.

Thornthwaite, C.W. and Mather, J.R.: The water balance, Publications in Climatology, 8, 1-104, 1955.

Toth, Z., Kalnay, E.: Ensemble Forecasting at NCEP and the Breeding Method. Mon. Weather. Rev., 125, 3297-3319, https://doi.org/10.1175/1520-0493(1997)125<3297:EFANAT>2.0.CO;2, 1997.

Vrugt, J.A., and Robinson, B.A.: Treatment of uncertainty using ensemble methods: Comparison of sequential data assimilation and Bayesian model averaging, Water Resour. Res., 43(1), 1-15, https://doi.org/10.1029/2005WR004838, 2007.

Valéry, A., Andréassian, V., and Perrin, C.: As simple as possible but not simpler': What is useful in a temperature-based snow-accounting routine? Part 2 – Sensitivity analysis of the Cemaneige snow accounting routine on 380 catchments, J. Hydrol., 517: 1176-1187, https://doi.org/10.1016/j.jhydrol.2014.04.058, 2014.

Velázquez, J.A., Anctil, F., Ramos, M.H., and Perrin, C.: Can a multi-model approach improve hydrological ensemble forecasting? A study on 29 French catchments using 16 hydrological model structures, Adv. Geosci., 29, 33-42, https://doi.org/10.5194/adgeo-29-33-2011, 2011.

Velázquez, J.A., Petit, T., Lavoie, A., Boucher, M.A., Turcotte, R., Fortin, V., and Anctil. F.: An evaluation of the Canadian Global Meteorological Ensemble Prediction System for short-term hydrological forecasting, Hydrol. Earth. Syst. Sci., 13, 2221-2231, https://doi.org/10.5194/hess-13-2221-2009, 2009.

Wand, M.P. and Jones, M.C.: Kernel smoothing, CRC press, vol. 60, 1 December 1994.

Wang, X. and Bishop, C.H.: Improvement of ensemble reliability with a new dressing kernel, 131(607), 965-986, https://doi.org/10.1256/qj.04.120, 2005.

Weigel, A. P., Liniger, M., and Appenzeller, C.: Can multi-model combination really enhance the prediction skill of probabilistic ensemble forecasts? Q. J. Roy. Meteorol. Soc., 134, 241–260, https://doi.org/10.1002/qj.210, 2008.

Wetterhall, F., Pappenberger, F., Alfieri, L., Cloke, H.L., Thielen-del Pozo, J., Balabanova, S., Danhelka, J., Vogelbacher, A., Salamon, P., Carrasco, I. and Corzo-Toscano, M.: HESS Opinions "forecaster priorities for improving probabilistic flood forecasts, Hydrol. Earth. Syst. Sc., 17(11), 4389-4399, 2013.

Wilks, D.S.: On the Reliability of the Rank Histogram, Mon. Weather. Rev., 139, 311-316, https://doi.org/10.1175/2010MWR3446.1, 2011.

Wilks, D.S.: Smoothing forecast ensembles with fitted probability distributions. Q. J. Roy. Meteorol. Soc., 128, 2821–2836, https://doi.org/10.1256/qj.01.215, 2002.