Interactive comment on "Exploring hydrologic post-processing of ensemble stream flow forecasts based on Affine kernel dressing and Nondominated sorting genetic algorithm II" *by* Jing Xu et al.

Jing Xu et al.

jing.xu.1@ulaval.ca

Received and published: 11 November 2020

[1]JingXu [1]FrançoisAnctil [2]Marie-AmélieBoucher

[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.

[Printer-friendly version](#)

[Discussion paper](#)

de l'Université, Sherbrooke, Québec, Canada

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

11

Interactive
comment

Reply on 'Referee comment on "Exploring hydrologic post-processing of ensemble stream flow forecasts based on Affine kernel dressing and Nondominated sorting genetic algorithm II"'

November 11, 2020

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 objective(s) of the research is(are) in my view not clearly stated, nor the intended contribution to the literature. Could the authors describe these?

Response : Many thanks for your comments. We rewrote the introduction for better clarifying our research objective. 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 : This perhaps also makes the literature review rather general, not zooming-in to identify a gap or under-represented aspects / applications of ensemble prediction, or a particular forecast challenge in the case study catchment.

Response : Thanks for your comments. We re-designed the introduction structure and added more literature review from the operational perspective. Operational forecasters are open to ensemble forecasting methods and products for assessing the flood in a probabilistic way. The main challenges for them 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 : As I understand, post-processing of the meteorological ensemble forecasts was not done. Could the authors comment in the paper on the performance of the meteorological ensemble forecasts and state the reason for not also applying meteorological post-processing?

Response : Thanks. We rewrote the introduction to give a better explanation: "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."

General question 4 : Based on Figure 4, presenting one forecast, I do not understand how it can be concluded that the members generated from the meteorological eps as forcing are not fully interchangeable, which is the basis of applying weights with NSGA-II. Perhaps that the video (I am sorry I could not find it, this is probably my omission) shows this, but this is not explicitly stated in the paper.

Response : Thanks! We will prepare and upload the animation (too large) that can reveal the phenomenon of not fully interchangeable of the meteorological forcing members for your consideration.

General question 5 : I do not understand why the authors choose to calibrate the post-processors on only one forecast horizon (day-4) and validate on the other horizons

(1-3, 5-7). Because of a generally present decrease of skill with increasing forecast horizon, usually a post-processor is calibrated for each forecast horizon separately. It also seems that the analysis period for which re-forecasts have been prepared has not been split in a calibration and validation period (or a leave-one-out approach). There may be good reasons for choosing this approach, e.g. stemming from catchment or application characteristics versus limited data availability, or as a research objective, but I missed the explanation in the paper. Could the authors perhaps explain the chosen calibration/validation approach?

Response : Thanks for you comments. 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. We will explain further the division of the calibration and validation period in the section 3.4.

General question 6 : Lastly, I would kindly encourage the authors to expand the presentation and interpretation / discussion of the results. For example, why present the 5 sub-catchments, what should we learn from the results? What about the inflows to the reservoirs? Why present the 5 single h-eps, what should we learn from the results? Why not assess the performance of the combined grand multi-model ensemble? How does the performance of the raw and post-processed forecasts compare with the performance of a reference forecast such as climatology or persistence (forecast skill)?

Response : Thanks for this suggestion. We will add further discussion 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. And we have already published an-

other article mainly focus on exploring "the hydrological post-processing of streamflow forecasts issued from multimodel ensemble prediction systems". Please check: Xu, $J.$, $Anctil$, $F.$ and $Boucher$, $M.A.$, 2019 $Hydrological$ $post-processing$ of $streamflow$ $forecasts$ $issued$ $from$ $multimodel$ $ensemble$ $prediction$ $systems$, $J.$ $Hydrol.$, 578, $p.124002$, $https://doi.org/10.1016/j.jhydrol.$ $2019.$ $124002.$ if this may interest you.

2 Detailed comments:

$Detailed$ $comments$ 1 : Introduction: Could you add explanation why AKD and NSGA-II have been chosen for this research? (line 51)

$Response$: Thank you for your comments. We rewrote the introduction and further explained the reason for selecting these two post-processing techniques.

$Detailed$ $comments$ 2 : Introduction, data description, and / or Results section: Could you comment on observational uncertainty?

$Response$: Thanks. We plan to add further description about the observational uncertainty in the section 2.

$Detailed$ $comments$ 3 : Line 92: Does the analysis of forecast performance take into account these different flood generating processes, and related seasonality? Would be interesting.

$Response$: Seasonality diversity analysis was not the the focus of our previous script.

But I will add further comparison and analysis in the result section (Section 4).

Detailed comments 4 : Line 99: In this section, kindly add some information on catchment response time to rainfall/snow melt, and travel time (routing), to inform us about potential forecast lead times without meteorological forecasts as forcing.

Response : Many thanks for you suggestions. We will add more detailed information on catchment response tie to rainfall/snow melt, and travel time.

Detailed comments 5 : Line 104: Could you mention why inflow to the reservoirs is not measured (for some reservoirs), and how the inflow time series have been constructed?

Response : Thanks. We will add further clarification about how the inflow times series have been constructed in Section 2.

Detailed comments 6 : Line 105/106: Could you briefly describe the observational network, and methods used to create sub-basin average precipitation? (to inform observational uncertainty, and perhaps a reason for not going for meteorological forecast post-processing).

Response : Thanks. We will add more detailed description about the observational network to inform the observation uncertainty.

Detailed comments 7 : Line 168: I think we are missing here, what the parameters are optimised on. From the later paragraph on Experimental set-up it seems that parameterization of AKD was done by minimising $MCRPS$.

Response : Yes, the parameterization of AKD was done by minimizing $MCRPS$. We will further clarify these relevant parameters in Section 3.1.

Detailed comments 8 : Lines 224-226: Kindly explain why these parameter values were chosen, and if a sensitivity analysis was done? Was the maximum evolution runs a result of a stopping criterion? If so, please mention this.

Response : Thank you for your comments. We will add further description about NSGA-II method in Section 3.2.

Detailed comments 9 : Lines 272-273: Please introduce the use of a moving window in Section 3.2, and expand explanation. Also the mentioning here of operational requirements is interesting and further explanation and discussion would be welcome.

Response : Thanks. We will give further explanation about why we chose using "moving window" in the experiments from the operational perspective.

Detailed comments 10 : Figure 5: The differences in bias and NSE over the range of the Pareto front are small. Please discuss. What weights are in the weight matrices of these solutions?

Response : Thanks. We will give further analysis about the weights obtained from the Pareto solutions.

Detailed comments 11 : Line 338: This is interesting. Could you discuss what could

be the reason? Something specific about Model M05?

Response : Thanks. We will add more analysis about the potentail reason why the predictive distributions of the kernel dressed ensemble are the most reliable for model M05 over almost all individual catchments.

3 Editorials:

Editorials 1 : Lines 25-27. Deterministic systems do not asses/quantify uncertainty, so the superiority question, I think, did not concern uncertainty quantification, that difference is simply a given. The superiority question concerned more the value when using the forecasts in decision making, and ensemble mean versus deterministic forecast performance.

Response : Many thanks for this comment. Yes, the main challenge for the operational forecasters are how to comprehensively quantify the predictive uncertainties from different sources as well as how to use the uncertainty information for better decision-making. The essential topic here is to bridge the gap between the "theory" (i.e., accuracy, reliability, etc.) and the "practice" (i.e., decision-driven trade-offs). We would like to emphasize this point in this paper.

Editorials 2 : Figure 1: Please indicate in the map more clearly the main river reach and flow direction.

Response : Thanks. We will refine the map in Figure 1.

Editorials 3 : Lines 82-84: Not clear from this sentence if in Section 4 the results are analysed for each model individually first (not taking into account model structure uncertainty), and then are considered and processed as a grand multi-model ensemble, which does take into account model structure uncertainty.

Response : Thanks. We will add further explanation here about how we consider the model structure uncertainty.

Editorials 4 : Lines 137-139: Consider to move up to Introduction for literature review, or down in the sections below. In these few introductory sentences to the methodology I would focus on announcing what was the general approach followed to reach the research objectives. After having introduced the overall methodology, going into the details of the two post-processing methods as of section 3.1 makes sense.

Response : Thank you for your suggestion. We plan to re-design the paper structure carefully.

Editorials 5 : Figure 3: Qobs is not output, so can be left out on the right. It is also not indicated that the output or final results concerns post-processed (interpreted) Qfcsts. The flowchart ending in only one set of post-processed forecasts is confusing, because up to now I was under the impression that AKD and NSGA-II would be used independently post-process and hence each method to result in a set of post-processed forecasts, after which the performance of each method will be analysed and compared.

Response : Thank you for proposing your concerns about the flowchart. We will refine

this flowchart.

Editorials 6 : Line 256: Spread Skill plots are announced, but later not presented.

Response : Thanks. The Spread Skill plot (SSP) were referd as "spread" in the corresponding results (i.e., Figure 8 and 10). We will rephrase the text to make it easier to track.

Editorials 7 : Figure 8: Presenting results in spider plots is a nice idea, but with the scores selected this does not work well, because some scores indicate a better performance with lower value ($RMSE$) while others the other way around (NSE), and some have a scale only to 1 (NSE) while others are not limited. This makes interpretation of the plots rather difficult.

Response : Thank you for your comment. We will re-plot those figures to fix the scale issue.

Editorials 8 : Lines 357-362: General/Literature - I suggest to delete or move to Introduction.

Response : Thanks. We will delete the general literature review here.

Editorials 9 : Lines 366-369: Consider to move to Introduction or Methodology.

Response : Thanks. We will move these descriptions here to the methodology section.

Printer-friendly version

Discussion paper

4   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) meteoro-

logical 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 MultiâĂŘObjective 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 rainfallâĂŤrunoff 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 mul-

tiobjective 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

HESSD

Interactive comment

Printer-friendly version

Discussion paper

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 hydroâĂŘmeteorological 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 stream-flow 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.