Supplement of Hydrol. Earth Syst. Sci., 26, 197–220, 2022
https://doi.org/10.5194/hess-26-197-2022-supplement

Hydrology and Earth System Sciences
Open Access
EGU

Supplement of

Choosing between post-processing precipitation forecasts or chaining several uncertainty quantification tools in hydrological forecasting systems

Emixi Sthefany Valdez et al.

Correspondence to: Emixi Sthefany Valdez (emixi-sthefany.valdez-medina.1@ulaval.ca)

1 Hydrological models evaluation

In this document, we provide information on how the models represent the precipitation-streamflow response of the catchments and a description of the metrics used to evaluate performance. The evaluation is based on volumetric efficiency (VE), KGEm, and relative bias (BIAS).

1.1 Evaluation criteria

1.1.1 Volumetric efficiency (VE)

We use the volumetric efficiency (VE, Criss and Winston, 2008) to evaluate and compare the performance of the seven hydrological models. The VE represents the fractional volumetric difference between the simulated and observed streamflows. It ranges from 0 to 1. A perfect value of 1 indicates that the volume of water predicted by the model matches the observed volume. This criterion gives the same weight to any flow range (e.g., slow recession and rapid rising flow), relaxing the constraint of model residuals heteroscedasticity.

$$VE = 1 - \frac{\sum_{k=1}^{N} |Q_{sim}(k) - Q_{obs}(k)|}{\sum_{k=1}^{N} Q_{obs}(k)} \tag{1}$$

where ($Q_{sim}(k)$ and $Q_{obs}(k)$) is the k^{th} of N pairs of simulated and observed streamflows.

1.1.2 Modified Kling-Gupta efficiency

The modified Kling-Gupta efficiency (KGEm, Kling et al., 2012) was used as objective function to identify the optimal set of parameters, but also as criteria to evaluate the quality of model outputs in the calibration and validation periods. It assesses the shape, timing, water balance, and variability of discharge time series. KGEm values range between $-\infty$ and 1 (positively oriented), the latter is the perfect value.

$$KGE = \sqrt{(r-1)^2 + (\beta-1)^2 + (\gamma-1)^2} \tag{2}$$

$$\beta = \frac{\mu_{sim}}{\mu_{obs}} \tag{3}$$

$$\gamma = \frac{CV_{sim}}{CV_{sim}} = \frac{\sigma_{sim}/\mu_{sim}}{\sigma_{obs}/\mu_{obs}} \tag{4}$$

where r is the correlation coefficient reflecting the linear relationship between simulation (the ensemble mean in forecasting mode) and observation. μ_{sim} (μobs) and σ_{sim} (σ_{obs}) are the mean and the standard deviations of the simulated (observed) time series. CV is the coefficient of variation.

1.1.3 Relative Bias (BIAS)

The relative bias (BIAS) is used to measure the overall unconditional bias (systematic errors) of the forecasts (Anctil and Ramos, 2018). Mathematically, it is defined as the ratio between the mean of the ensemble average and the mean observation. The BIAS is sensitive to the direction of errors: values higher (lower) than 1 indicate an overall overestimation (underestimation) of the observed values.

$$BIAS = \frac{\sum_{k=1}^{N} Fct_{avg}(k)}{\sum_{k=1}^{N} Obs(k)} \tag{5}$$

where ($Fct_{avg}(k)$, $Obs(k)$) is the k^{th} of N pairs of deterministic forecasts and observations.

1.2 Performance of the hydrological models

Figure S1 ilustrates the performance of the seven hydrological models and the multimodel in calibration (1997-2007) and validation (2008-2016). The boxplots correspond to calibration (red), validation without DA (light blue), and validation with DA (dark blue). Each boxplot summarizes the distribution (minimum, quantiles 0.25, 0.5, and 0.75, and maximum) of the 30 catchments. Multimodel performance is estimated from the mean of each model simulation.

As Figure S1 shows, no model is systematically better than the others for the different criteria and over the catchments, which is to be expected since no single model excels in all situations. Overall, models present a better performance in calibration, as also generally expected. When using EnKF DA, some models have slightly better or similar VE and KGEm values in validation (e.g., HBV, HACRES, PDM, and SACRAMENTO). However, the EnKF DA does not display a uniform influence on the models. Models CEQUEAU, MORDOR, and XINANJIANG exhibit lower transposability; that is, the parameters estimated in calibration are less suitable for the validation period. This highlights the multimodel value since its performance is similar to that of the best performing models.

The fraction of simulated and observed volumes differ on average by 3 % in calibration and 18 % in validation, but improves to 8 % when activating the EnKF DA in validation (Fig. S1, VE). BIAS reveals that this difference is usually an overestimation and confirms the effectiveness of the EnKF, especially to deal with transposability issues between calibration and validation. Concerning the seven models, we obtained an average KGEm of 0.83 in calibration, 0.64 in validation without EnKF, and 0.82 in validation with EnKF, over the 30 catchments. While in the multimodel, these values were equal to 0.88, 0.73 and 0.87, respectively.

Figure S2 presents the interannual hydrographs of the catchments showing the best and the worst performance values spanning over the calibration and validation periods. The hydrographs reveal that overestimation occurs mostly in spring, during snowmelt. Overall, the ensemble of hydrological models reproduces the hydrographs well.

Figure S3 ilustrates the performance of the seven hydrological models for the different catchment groups. The Medium group has the best performance in the calibration period for the three criteria. In contrast, the Larger group has a better performance in validation, both with and without DA EnKF.

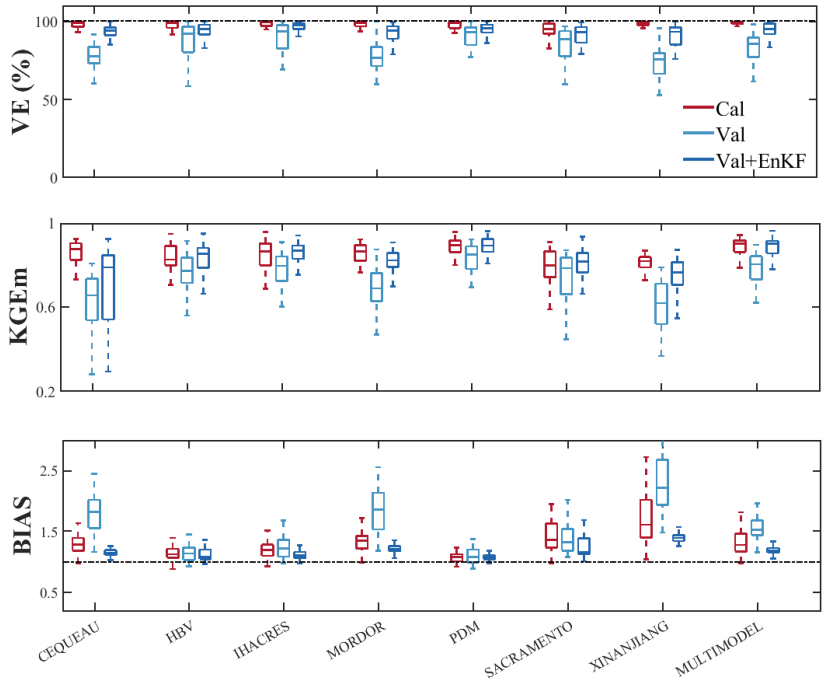

Figure S1. Performance of the seven hydrological models and the multimodel in terms of volumetric efficiency (VE), KGEm, and relative bias (BIAS) in calibration (1997-2007; red), validation (2008-2016) without EnKF DA (light blue), and with EnKF DA (dark blue). Boxplots represent the distribution of the scores over 30 catchments. The scores exclude wintertime (December-March) when river ice interferes with streamflow measurements.

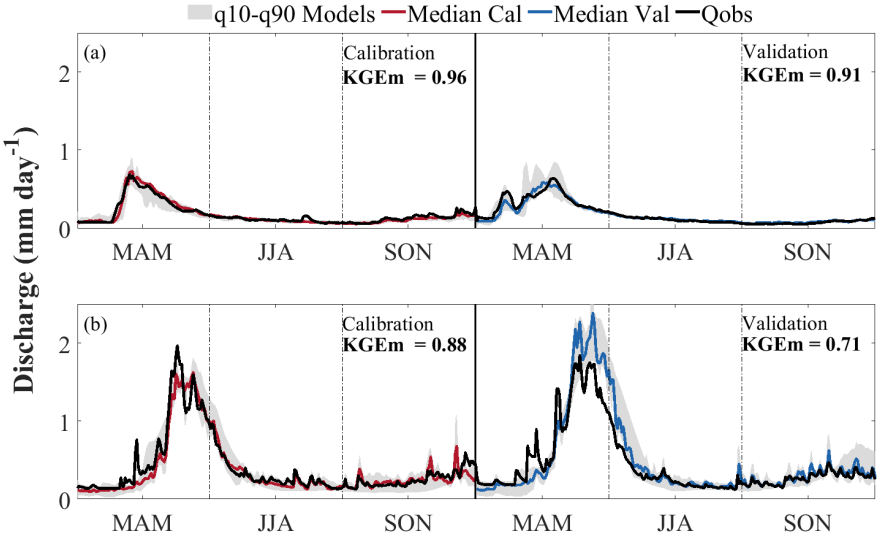

Figure S2. Simulated and observed (black) interannual streamflow of the catchments with the best (a) and the worst (b) performance in calibration (1997-2007; red) and validation (2008-2016; blue). The grey shaded area represents the variability (80 % interval) of the seven hydrological models. MAM: March-April-May; JJA: June-July-August; SON: September-October-November.

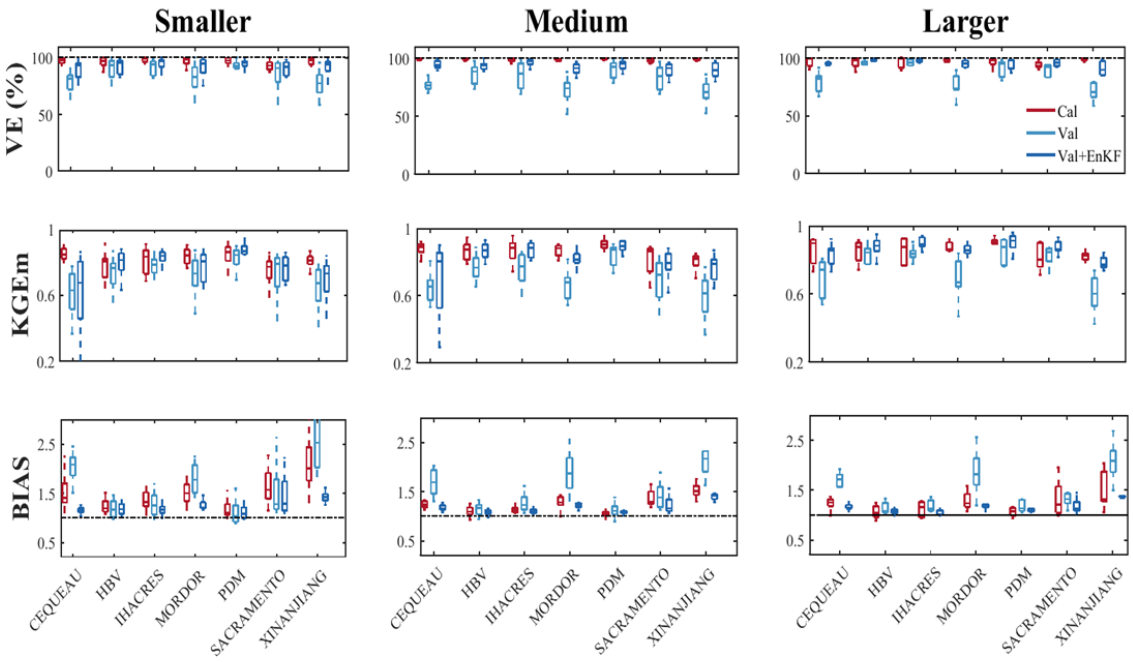

Figure S3. Performance of the seven hydrological models in terms of volumetric efficiency (VE), KGEm, and relative bias (BIAS) in calibration (1997-2007; red), validation (2008-2016) without EnKF DA (light blue), and with EnKF DA (dark blue). Boxplots represent the distribution of the scores over the catchments in each group: Smaller, Medium and Larger. The scores exclude wintertime (December-March).

References

Anctil, F., Ramos, M.H., 2018. Verification Metrics for Hydrological Ensemble Forecasts, in: Handbook of Hydrometeorological Ensemble Forecasting. Springer Berlin Heidelberg, pp. 1–30.

Criss, R.E., Winston, W.E., 2008. Do nash values have value? discussion and alternate proposals. Hydrological Processes 22, 2723–2725. doi:https://doi.org/10.1002/hyp.7072, arXiv:https://onlinelibrary.wiley.com/doi/pdf/10.1002/hyp.7072.

Kling, H., Fuchs, M., Paulin, M., 2012. Runoff conditions in the upper danube basin under an ensemble of climate change scenarios. Journal of Hydrology 424, 264–277. doi:https://doi.org/10.1016/j.jhydrol.2012.01.011.