Hydrology and Earth System Sciences Discussions

Hydrol. Earth Syst. Sci. Discuss., referee comment RC1
https://doi.org/10.5194/hess-2021-539-RC1, 2021

Comment on hess-2021-539

Anonymous Referee #1

Referee comment on "Evaluating the impact of post-processing medium-range ensemble streamflow forecasts from the European Flood Awareness System" by Gwyneth Matthews et al., Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2021-539-RC1, 2021

General comments:

The paper is very well written and very detailed. The topic of the study is certainly of great interest to the forecasting community. The combination of a hydrological uncertainty processor (MCP) and EMOS with the help of a Kalman filter approach is quite novel and attractive. The discussion of the results is, however, a bit too long and could be summarized more concisely focusing on floods aspects. Therefore, I suggest that it is worth publishing after some minor revision.

There are many EFAS paper available now and all describe the EFAS system. This could be shortened and only the differences to the operational settings of EFAS should be explained, in particular how the reforecasts are used. The biggest challenge in using reforecasts is the reduced number of ensemble members (11 instead of 51). This small number of members causes difficulties in computing the CRPS for making a fair comparison with a CRPS derived from the PDFs of the post-processed forecasts (see for example Zamo and Naveau, 2018). However, this problem is not mentioned and the presented results of the CRPSS should be treated with care.

A most recent paper from Skoven, et al. (2021) has a similar topic about evaluating the post-processing methods for EFAS (EMOS and the application of transformations like NQT). Therefore, the differences and novelty of this study should be stressed clearly and discussed in more detail. One difference is, apart from using the reforecasts and MCP, that the EMOS correction parameters are lead-time invariant. This could be stressed more clearly right from the beginning. For me, this lead-time invariance is a big drawback of the proposed method and rather problematic for deriving the total predictive uncertainty, which forces the Kalman filter sometimes to give more wait to the hydrological uncertainties.

Although the analysis of the different aspects like catchment size, elevation, regulation, length of period, are very interesting, it maybe could be shortened focusing on floods, which is the main topic of EFAS. Also, the detailed analysis of the Nash-Sutcliffe (KGE) is maybe too long, since the results of the post-processing methods are probabilistic and the KGE reduces the information content to the mean (median) of the Ensembles (or PDF). Therefore, I would suggest that the emphasis of the verification should be the CRPS.

Specific comments

Page 5: Highlighting the differences between the operational setting of EFAS and the setting used in this study should be sufficient. More details can be found in many other papers. However, the calibration period of the LISFLOOD model is missing. Is there an overlap between the period for calibrating the parameters of the hydrological model and the historical period p for the off-line calibration?

In Figure 1, the index of the parameters μ and Σ is ψ, whereas in the caption you use the index Φ.

You mention several times (e.g. line 172, page 7) that the minimum for the off-line calibration is 2 years. However, for the fitting of the GPD you use 1000 values (page 9). So you will need more than 2 years?

I would suggest to include a list of nomenclature, so you avoid repetitive descriptions like the tilde for the physical space (line 182, 204, 220, 340) and the definition of the timestep notation introduced on page 7 (lines 176-177).

On page 9, line 235, you write that the location parameter a is used for defining the breakpoint, but in Figure 2 the shape parameter c is used as breakpoint.

On page 9, line 251, you write about consistence between the 2 distribution. What does it mean? How do you check this?

On page 10, line 257 the concentrated likelihood method is mentioned without any further explanation what this method does. Maybe some more details would be helpful. Also, it is not clear for me how the GPD is weighted (line 261-262) .

Whereas the description of the linear approximation (page 11 – line 270 – 280) is maybe not necessary.

On page 13, it is not clear for me why you have observations (line 336) and water balance (line 337) for the period until t, but the forecasts (line 339) until t-1 ?

Line 349: …using a MCP method …

Page 14, line 372: in the recent period …

Page 16, line 411: you write that a set of forecasts are used estimate the two spread correction parameters. How did you choose the size of these sets of forecasts ?

In the Figure 3 you write CRPS besides the legend bar, but it should be CRPSS?

On page 19, line 483 you write that only 11 reforecasts are available (I suppose this 11 comes from 40 days/7day x 2). The number 11 could be misleading, since it happens to coincide with the number of members mentioned in the next sentence (line 485). Why do you fix it to 40 days? Since there is this a discrepancy between the operational setting and this analysis anyhow, you could set q to a longer period to include more reforecasts (e.g q=70 ~ 20 reforecasts).

At line 486 I don’t understand why the mean discharge value is predicted for the previous 6 hours?

The difference between the raw and the post-processed forecasts in Fig. 11a (mentioned on page 35, line 862) is very difficult to see and almost not visible.

Line 880: …uncertainties show a small increase

On page 41 the paragraph from line 1004 – 1010 can be removed

Line 1021 ..greater than..

I have some doubts about your suggestion that very short periods are sufficient (line

1045): the chance that such a short period will show the variability of the discharge needed for applying the NQT is rather small and the fitting of the GPD almost impossible. Consequently the back-transformation of the variables from the Normal space will always produce poor and very unreliable results for floods.

The citation of Coccia (line 1135) is incomplete. Also, the term “Multi-Temporal” in combination with the MCP (MT-MCP) is mentioned only in line 153-154 and in the conclusions (line 1054), but is not explained.

Zamo, M., Naveau, P. Estimation of the Continuous Ranked Probability Score with Limited Information and Applications to Ensemble Weather Forecasts. *Math Geosci* **50,** 209–234 (2018). https://doi.org/10.1007/s11004-017-9709-7)

Skøien, J. O., Bogner, K., Salamon, P., & Wetterhall, F. (2021). On the Implementation of Postprocessing of Runoff Forecast Ensembles, Journal of Hydrometeorology, 22(10), 2731-2749