Hydrol. Earth Syst. Sci. Discuss.,
https://doi.org/10.5194/hess-2020-237-RC2, 2020

Hydrology and Earth System Sciences
Discussions

Interactive comment on "Technical note: Diagnostic efficiency – specific evaluation of model performance" *by* Robin Schwemmle et al.

Anonymous Referee #2

Received and published: 1 August 2020

The authors present an interesting technical note in which they link the idea of diagnostic model evaluation with that of efficiency metrics. They propose a new metric in which they integrate terms to assess constant, dynamic and timing errors. I like the idea and the paper, but I am unclear about the way this metric and its terms are formulated, and how they relate to previous work. Hopefully my comments below help the authors to strengthen their argument.

MAJOR COMMENTS

[1] I understand that the first term of their metric is the relative bias of the FDC. Why is this a more hydrologically relevant and insightful term than other bias estimates? Can you show evidence for this claim?

[2] Similarly, I would find it more informative if the authors were to compare their terms to the terms in KGE and the non-parametric version by Pool et al. (2018) to really understand the differences. Why are these more informative and can it be shown?

[3] Would it not be more informative if the different parameter sets in Figure 4 were to show that different errors dominate? Why do they all show essentially identical FDCs? Maybe use more varied examples?

[4] Is the main problem one of aggregation? And hence loss of information. See for example the separate use of KGE terms in Gudmundsson et al. (2012). Even your second term is more informative because it leads to less aggregation and loss of information. Is this the key?

[5] It would be good if the authors would clarify their assumptions better and discuss how these might relate to reality. For example, they assume that precipitation has a consistent input data error. Some previous studies suggest that such an input error varies significantly between rainfall events (e.g. Yatheendradas et al., 2008, WRR). Similarly, for the other errors. It would strengthen the study significantly if the authors where to review the literature thoroughly for studies that discuss how these different errors manifest themselves (the authors lines 61ff). The three assumptions made here are key to the paper, but they are currently not supported by literature. I am not arguing that the authors' assumptions are wrong (though I might disagree partially), but they need to show evidence why these assumptions are reasonable. How to assign these errors is key here, but it is also something many people have argued about before.

[6] There have been others who raised the question of benchmarks before. For example Jan Seibert (https://eprints.ncl.ac.uk/file_store/production/246998/A084BCF1-F4EA-4EDF-AE6D-9E85C27A9DC4.pdf or Seibert, 2001). It would be good if the authors would review the literature more thoroughly on this topic.

[7] Section 3.7 is difficult to follow. Maybe this can easier be summarized in a figure? I find these error combinations difficult to read and compare. Maybe another figure

instead of the table?

REFERENCES

Gudmundsson, L., T. Wagener, L. M. Tallaksen, and K. Engeland (2012), Evaluation of nine large-scale hydrological models with respect to the seasonal runoff climatology in Europe, Water Resour. Res., 48, W11504, doi:10.1029/2011WR010911.

Pool, S., Vis, M., & Seibert, J. (2018). Evaluating model performance: towards a non-parametric variant of the Kling-Gupta efficiency. Hydrological Sciences Journal, 63(13-14), 1941-1953.

Seibert J. 2001. On the need for benchmarks in hydrological modelling. Hydrological Processes 15 (6): 1063–1064 DOI: 10.1002/hyp.446

Yatheendradas, S., T. Wagener, H. Gupta, C. Unkrich, D. Goodrich, M. Schaffner, and A. Stewart (2008), Understandinguncertainty in distributed flash flood forecasting for semiarid regions, Water Resour. Res., 44, W05S19, doi:10.1029/2007WR005940.

Interactive comment on Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2020-237, 2020.