

Interactive comment on “Bayesian performance evaluation of evapotranspiration models for an arid region in northwestern China” by Guoxiao Wei et al.

Anonymous Referee #1

Received and published: 16 November 2018

In this manuscript, the authors did a lot of work in using the Bayesian techniques to analyze ET models. This work could be interesting to the hydrological community. General comments: 1. Some efforts are needed to emphasize the significance of the work. For the objective (1), what is the purpose of selecting the best model using BME, to improve model prediction? If the purpose is to improve model prediction, did the authors try Bayesian model averaging? Based on the results, some models are underestimate, some models are overestimate, it is possible that model averaging could give a better prediction performance. For the objective (2), theoretically we know these statistics only measure model fit without considering model complexity, so they are not as robust as BME. And we know these statistics can be efficiently calculated, so

C1

there is no need for testing. Please justify the objective (2). I think, objective (3) is very meaningful, I would like to see more analysis on the model-data mismatch to improve model development and model performance. 2. In several places of the manuscript, the logic is not very clear. The English writing needs improvement.

Specific comments: 1. Abstract, I think including some insights obtained from the numerical experiments in the abstract would attract more audience and make this work more meaningful. 2. Line 35-36, the SW model performs best in this study area, but may not be the best in other areas. For example, in Li et al., (2013)'s study, PM performed better than SW in estimation of maize. Please justify the statement that SW should be the first choice for evaluating ET of spring maize in arid desert oasis areas. 3. Line 93-94, BME can be used to compare and select the best-performing model. This is well-known and not a hypothesis that needs to be determined. What do the authors mean by saying “an unbiased view”? 4. Line 95-97, the first part of the sentence says Bayesian applications have focused on comparison of alternative models, but the second part of the sentence says that little attention has been given to the Bayesian model comparison. The sentence is self-contradictory. Please clarify. 5. Line 277, for each chain? I thought you total have 40,000 samples from all chains. In addition, Line 848, from one chain? Please clarify. 6. Line 280-282, based on Figure 1, DREAM needs far less than 8000 generations to make the GR statistic smaller than 1.2. Also, based on Figure 1's x-axis scale, it is hard to tell “obviously” the chain converged after about 620 and 450 generations. 7. Figure 1. In Figure 1(b) the position of the dash line is not at 1.2. The position of the label (b) is not aligned well with the label (a). 8. Figure 2. If the authors cannot get more information from the CDFs than the histograms, I suggest deleting the CDFs which make Figure 2 busy and confusing. Also, I would like to see more discussion about Figure 2; what insights the authors can obtain from these plots? 9. Line 294-297, I found the discussion of the figure 2 is confusing. I think, the figure 2 says the histograms tend to concentrate in the upper bounds, not the lower bounds. Also, the authors should increase the upper limits of these parameters not decrease, because the histograms are concentrated in the upper bounds. 11. Line

C2

355-356, what do the author mean by saying “to sample groups of variable in turn”?

Technical corrections: 1. Line 29, obstained → obtained 2. Line 92, beed → been

Interactive comment on Hydrol. Earth Syst. Sci. Discuss., <https://doi.org/10.5194/hess-2018-430>, 2018.