Hydrol. Earth Syst. Sci. Discuss.,
https://doi.org/10.5194/hess-2019-610-RC1, 2019

Hydrology and Earth System Sciences
Open Access
Discussions

Interactive comment on "Multi-step ahead daily inflow forecasting using ERA-Interim reanalysis dataset based on gradient boosting regression trees" *by* Shengli Liao et al.

Anonymous Referee #1

Received and published: 21 December 2019

1. The manuscript presents multi-step ahead daily inflow forecasting using ERA-Interim reanalysis dataset based on gradient boosting regression trees, which is interesting. It is relevant and within the scope of the journal. 2. However, the manuscript, in its present form, contains several weaknesses. Appropriate revisions to the following points should be undertaken in order to justify recommendation for publication. 3. Full names should be shown for all abbreviations in their first occurrence in texts. For example, ERA in p.1, ECMWF in p.3, etc. 4. For readers to quickly catch your contribution, it would be better to highlight major difficulties and challenges, and your original achievements to overcome them, in a clearer way in abstract and introduction. 5. It is mentioned in p.1 that ERA-Interim reanalysis data is adopted as input. What are

other feasible alternatives? What are the advantages of adopting this particular data over others in this case? How will this affect the results? The authors should provide more details on this. 6. It is mentioned in p.1 that gradient boosting regression tree is adopted as inflow forecast framework. What are the advantages of adopting this particular soft computing technique over others in this case? How will this affect the results? The authors should provide more details on this. 7. It is mentioned in p.1 that artificial neural networks, support vector regression and multiple linear regression models are adopted as benchmark for comparison. What are the other feasible alternatives? What are the advantages of adopting these particular models over others in this case? How will this affect the results? More details should be furnished. 8. It is mentioned in p.3 that the Xiaowan Hydropower Station is adopted as the case study. What are other feasible alternatives? What are the advantages of adopting this particular case study over others in this case? How will this affect the results? The authors should provide more details on this. 9. It is mentioned in p.4 that the maximum information coefficient is adopted to select inputs from 79 potential predictors from reanalysis data. What are the advantages of adopting this particular approach over others in this case? How will this affect the results? The authors should provide more details on this. 10. It is mentioned in p.4 that autocorrelation function is adopted to identify observed inflow and rainfall lags. What are other feasible alternatives? What are the advantages of adopting this particular approach over others in this case? How will this affect the results? The authors should provide more details on this. 11. It is mentioned in p.6 that four evaluation criteria are adopted to evaluate the performance of the models. What are the other feasible alternatives? What are the advantages of adopting these particular evaluation criteria over others in this case? How will this affect the results? More details should be furnished. 12. It is mentioned in p.7 that a grid search algorithm is adopted to optimization model parameters. What are other feasible alternatives? What are the advantages of adopting this particular algorithm over others in this case? How will this affect the results? The authors should provide more details on this. 13. It is mentioned in p.9 that grid searching is adopted to tune the hyperparameters of GBRT,

GBRT-MIC, ANN-MIC. What are other feasible alternatives? What are the advantages of adopting this particular approach over others in this case? How will this affect the results? The authors should provide more details on this. 14. It is mentioned in p.9 that Bayesian optimization (Snoek et al., 2012) is adopted to tune the hyperparameters of SVR-MIC. What are other feasible alternatives? What are the advantages of adopting this particular approach over others in this case? How will this affect the results? The authors should provide more details on this. 15. It is mentioned in p.9 that Python is adopted to perform all computations. What are other feasible alternatives? What are the advantages of adopting this particular software over others in this case? How will this affect the results? The authors should provide more details on this. 16. Some key parameters are not mentioned. The rationale on the choice of the particular set of parameters should be explained with more details. Have the authors experimented with other sets of values? What are the sensitivities of these parameters on the results? 17. Some assumptions are stated in various sections. Justifications should be provided on these assumptions. Evaluation on how they will affect the results should be made. 18. The discussion section in the present form is relatively weak and should be strengthened with more details and justifications. 19. Moreover, the manuscript could be substantially improved by relying and citing more on recent literatures about contemporary real-life case studies of soft computing techniques in hydrological prediction such as the followings: ïĄň Yaseen, Z.M., et al., "An enhanced extreme learning machine model for river flow forecasting: state-of-the-art, practical applications in water resource engineering area and future research direction," Journal of Hydrology 569: 387-408 2019. ïĄň Fotovatikhah, F., et al., "Survey of Computational Intelligence as Basis to Big Flood Management: Challenges, research directions and Future Work," Engineering Applications of Computational Fluid Mechanics 12 (1): 411-437 2018. ïĄň Mosavi, A., et al., "Flood Prediction Using Machine Learning Models: Literature Review," Water 10 (11): article no. 1536 2018. ïĄň Moazenzadeh, R., et al., "Coupling a firefly algorithm with support vector regression to predict evaporation in northern Iran," Engineering Applications of Computational Fluid Mechanics 12 (1): 584-597 2018.

ïĄň Ghorbani, M.A., et al., "Forecasting pan evaporation with an integrated Artificial Neural Network Quantum-behaved Particle Swarm Optimization model: a case study in Talesh, Northern Iran," Engineering Applications of Computational Fluid Mechanics 12 (1): 724-737 2018. ïĄň Chau, K.W., et al., "Use of Meta-Heuristic Techniques in Rainfall-Runoff Modelling" Water 9(3): article no. 186, 6p 2017. 20. Some inconsistencies and minor errors that needed attention are: ïĄň Replace "...was supply to depict..." with "...was supplied to depict..." in line 86 of p.3 ïĄň Replace "...into train set, validation set, and test set..." with "...into training set, validation set, and testing set..." in lines 206-207 of p.7 ïĄň Replace "...test set..." with "...testing set..." in line 209 of p.7 ïĄň Replace "...more accuracy inflow forecasting..." with "...more accurate inflow forecasting..." in line 283 of p.10 ïĄň Replace "...arisen in in areas..." with "...arisen in areas..." in line 288 of p.10 ïĄň Replace "...for train, validation and test set..." with "...for training, validation and testing set..." in line 294 of p.10 ïĄň Replace "...According to compare the forecasted results of..." with "...According to the comparison of forecasted results of..." in line 330 of p.11 21. In the conclusion section, the limitations of this study, suggested improvements of this work and future directions should be highlighted.

Interactive comment on Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2019-610, 2019.