Hydrol. Earth Syst. Sci. Discuss.,
https://doi.org/10.5194/hess-2020-594-RC2, 2021

Hydrology and Earth System Sciences
Discussions

Interactive comment on "Design flood estimation for global river networks based on machine learning models" *by* Gang Zhao et al.

Anonymous Referee #2

Received and published: 22 January 2021

Overall Comments: The paper proposes a machine learning based approach to estimate design floods globally. It includes three stages. First is using Anderson-Darling test and Bayesian MCMC method to choose suitable distribution and estimate parameters. Then at-site frequency curve will be sure. Second is clustering these stations into subgroups by a L-means model based on 12 globally available catchment descriptors. Third is developing a regression model in each subgroup for regional design flood estimation using the same descriptors. 11793 stations' data is used to predict regional flood and a support vector machine regression provide the highest prediction performance with root mean square normalized error of 0.708 for 100-year return period flood estimation and relative mean relative biases of all climate types being less than 20%. This paper proposes a large-scale regional flood estimation method by machine

learning which covers 11793 stations globally. The method performance is also satisfactory compared with previous work. However, there are still some shortcomings. Some explanations should be complemented and the negative value of RBIAS should be analyzed more. Specific Comments: 1) Page 6: "... These explanatory factors can be grouped into 135 four categories as follows: ...". A correlation analysis of all factors can be done to make sure they have weak correlation with each other. 2) Page 9: "... The adopted Anderson-Darling test and Bayesian MCMC method are briefly described as follows. ...". A brief introduction of distributions is necessary. For example, Pearson type three distribution is used widely in China. 3) Page 11: "... The adopted Bayesian MCMC method was proposed by Reis and Stedinger (2005) and is reported to provide better parameter estimates than the MOM and MLE approaches in some studies ...". L-moment method is a valid method on estimating the parameters. Bayesian MCMC method should compare with it. 4) Page 11: "... The detail of the Bayesian MCMC method is comprehensively described in the research of Reis and Stedinger (2005) ...". Although its numerical method will be complicated, a brief explanation is still essential. 5) Page 13: "... SVM regression has shown advantages in solving complicated non-linear problems in the field of hydrology ...". As a major method of this article, the introduction of SVM may be too simple. More detailed description can be added. 6) Page 13: "... RF regression is a representative type of ensemble machine learning model ...". Math is the best language of science. Several mathematical formulas of RF will help readers to understand it abstractly. 7) Page 18: "... Figure 7 (a) Factor importance evaluated by RF model and (b) the impact of catchment descriptors for regression ...". Figure 7 (a) Factor importance evaluated by RF model and (b) the impact of catchment descriptors for regression 8) Page 19: "... The negative value of RBIAS reflected some overestimation which mainly occurred due to low discharge in small catchments ...". It is normal to underestimate 100-year return period floods, but why all the RBIAS indexes are negative? RBIAS index is just a relative index so the absolute value of discharge should not take much effect on the index. Please analyze more about it.

Interactive comment on Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2020-594, 2020.