Reply to Anonymous Referee #1

Thank you very much for your time and for your thoughtful and constructive review. The following are our point-by-point responses to your comments.

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.

Response: Thank you very much for your positive comments.

Proposed changes to manuscript: N/A

2. Full names should be shown for all abbreviations in their first occurrence in texts. For example, ERA in page 1, ECMWF in page 3, etc.

Response: Thank you for your carefulness.

Proposed changes to manuscript: We will show full names for all abbreviations in their first occurrence in the revised manuscript.

3. 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.

Response: Thank you for your suggestion. The major difficulties and challenges are high precision model input related to inflow of longer lead times and effective prediction model. This paper proposed a new hybrid inflow forecast framework with ERA-Interim (European Centre for Medium-Range Weather Forecasts (ECMWF) Re-Analysis Interim) data as input, adopting gradient boosting regression trees (GBRT) and the maximum information coefficient (MIC) was developed for multi-step ahead daily inflow forecasting. The proposed method overcomes the difficulties from three aspects. Firstly, the ERA-Interim dataset provides enough information for the framework to discover inflow for longer lead times. Secondly, MIC can identify effective feature subset from massive features that significantly affects inflow so that the framework can avoid over-fitting, distinguish key attributes with unimportant ones and provide a concise understanding of inflow. Lastly, the GBRT is a prediction model in the form of an ensemble of decision trees and has a strong ability to capture nonlinear relationships between input and output in long lead times more fully.

Proposed changes to manuscript: We will make careful modifications in Section "Abstract" and "Introduction" of the revised manuscript.

4. It is mentioned in page 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.

Response: Thank you for your careful review and suggestion. The ERA-Interim data are the result of assimilating observed data with forecast data, which has less error than observed data and forecast data (Balsamo et al., 2015). ERA-Interim shows the results of a global climate reanalysis from 1979 to date, which are produced by a fixed version of a NWP system (Dee et al., 2011). The fixed version ensures there are no spurious trends caused by an evolving NWP system. Therefore, meteorological reanalysis data satisfies the need for long sequences of consistent data and have been used for the prediction of wind speeds (Stopa and Cheung 2014) and solar radiation (Linares-Rodríguez, Ruiz-Arias et al. 2011, Ghimire, Deo et al. 2019). Meanwhile, ERA-Interim was proved to be one of the best reanalysis data describing atmospheric circulation and elements (Kishore et al., 2011).

Proposed changes to manuscript: More details about ERA-Interim data will be given in Section "Appendix" of the revised manuscript for readers to quickly catch the contribution.

5. It is mentioned in page 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.

Response: Thank you for your careful review and suggestion. The gradient boosting regression trees (GBRT) (Friedman 2001, Fienen, Nolan et al. 2018), a nonparametric machine learning method based on a boosting strategy and decision trees, was developed and had been used in traffic (Zhan, Zhang et al. 2019) and environmental (Wei, Meng et al. 2019) field and proved to alleviate the problems of being trapped by local minima, over-fitting problems and reduced generalizing performance.

Proposed changes to manuscript: More details about GBRT and other soft computing techniques will be given in Section 3 of the revised manuscript according to your suggestion.

6. It is mentioned in page 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.

Response: Thank you for your careful review and suggestion. The several studies had shown that artificial neural networks (ANN) (Rasouli et al., 2012; Cheng et al., 2015; El-Shafie and Noureldin, 2011; Chau, 2006; Ali Ghorbani et al., 2018) and support vector regression (SVR) (Tongal and Booij, 2018; Luo et al., 2019; Moazenzadeh et al., 2018) are the two powerful models for inflow predicting. They are widely used and very mature algorithms, which are scientific and reasonable compared with them.

Proposed changes to manuscript: More details about compared model will be given in Section "Introduction" of the revised manuscript.

7. It is mentioned in page 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.

Response: Thank you for your careful review and suggestion. The Xiaowan Hydropower Station in the lower reaches of the Lancang River, which is the longest river with most discard water in Yunnan Province, was chosen as the study site (as shown in Fig. 2). The Lancang River is approximately 2000 km long and has a drainage area of 113300 km^2 above the Xiaowan Hydropower Station. Thus, the Xiaowan Hydropower Station is the main control hydropower station in the Lancang River and it is very significant to adopte the Xiaowan Hydropower Station as the case study.

Proposed changes to manuscript: More details about case study will be given in Section 2.1 of the revised manuscript according to your suggestion.

8. It is mentioned in page 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.*

Response: Thank you for your careful review and suggestion. The maximal information coefficient (MIC) (Reshef et al., 2011) is a robust measure of the degree of correlation between two variables and has attracted a lot attention from academia (Zhao et al., 2013; Ge et al., 2016; Lyu et al., 2017; Sun et al., 2018), which can select effective input factors accurately and quickly.

Proposed changes to manuscript: More details about inputs selection will be given in Section 4.1 of the revised manuscript.

9. It is mentioned in page 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.

Response: Thank you for your careful review and suggestion. The autocorrelation function (ACF) measures the dependency or relationship of observed value with lagged observations of a considered variable. In a long-dependent series such as inflow time series, the ACF will decay slowly. The partial autocorrelation function (PACF) and cross-correlation function (CCF) are two other feasible alternatives. We use the PACF and CCF in these days for modeling, calculation and analysis in these days according to Referee(#3)'s suggestion. We agree to try PACF and CCF replace ACF to determine the model structure.

Proposed changes to manuscript: In the revised manuscript, PACF and CCF will be adopted to determining the model structures for inflow and rainfall, respectively. Hypothesis test is used to determine the significant relationships replacing user-defined threshold value.

10. It is mentioned in page 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.

Response: Thank you for your careful review and suggestion. The root mean square error (RMSE) and mean absolute error (MAE) are the most commonly used criteria to assess model performance (Luo et al., 2019; Chau, 2005; Chau, 2006). The Nash-Sutcliffe efficiency coefficient (NSE) (Nash and Sutcliffe, 1970) is commonly for evaluating the performance of hydrological models and it is one of the best performance metrics for reflecting the overall fit of a hydrograph. The Pearson correlation coefficient (CORR) is a good measurement of the average error. Peak flow criterion, degree of agreement and Kling-Gupta efficiency metrics are three other feasible alternatives.

Proposed changes to manuscript: Peak flow criterion, degree of agreement and Kling-Gupta efficiency metrics will be added to compare several model performances in Section 3.3 of the revised manuscript.

11. It is mentioned in page 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.

Response: Thank you for your careful review and suggestion. Grid search is considered as an effective parameter search method, which is widely used (Fienen et al., 2018). Two of other feasible alternatives are randomized search and Bayesian optimization. Bayesian optimization (Snoek et al., 2012) was employed to tune the hyperparameters of support vector regression (SVR) in this paper. We have performed some numerical experiments to compare grid search and randomized search and grid search can obtain more reasonable and stable hyperparameter combination.

Proposed changes to manuscript: More details about grid search will be given in Section 4.2 of the revised paper.

12. It is mentioned in page 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.

Response: Thank you for your careful review and suggestion. Same as question 11, grid search is considered as an effective parameter search method, which is widely used (Fienen et al., 2018). We have performed some numerical experiments to compare grid search and randomized search and grid search can obtain more reasonable and stable hyperparameter combination.

Proposed changes to manuscript: More details about grid search will be given in Section 4.2 of the revised paper.

13. It is mentioned in page 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.

Response: Thank you for your careful review and suggestion. Bayesian optimization (Snoek et al., 2012) is proved as an effective parameter search method, especially for wide domain space.

Proposed changes to manuscript: More details about Bayesian optimization will be given in Section 4.2 of the revised paper.

14. It is mentioned in page 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.

Response: Thank you for your careful review and suggestion. Python is an important tool for scientific computing and data analysis. It has many open source libraries and is easy to implement.

Proposed changes to manuscript: Brief introduction about Python will be given in Section 4.1 of the revised manuscript.

15. 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?

Response: Thank you for your careful review and suggestion. For ANN, A range of 2-20 neurons and four activation functions (Table 3) are selected by trail-and-error. The sensitivities of these parameters have been analyzed.

Proposed changes to manuscript: We will use more wide selected ranges of the model parameters in Section 4.2 of the revised manuscript.

16. 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.

Response: Thank you for your careful review and suggestion. The comparison of different models is based on the basic assumption that parameters are optimal. In the original manuscript, grid searching and Bayesian optimization were employed to tune the hyperparameters of model. Large models for each lead time were developed to find as possible as optimal parameters.

Proposed changes to manuscript: More details about the assumption of optimal parameters will be given in Section 4.2 of the revised paper.

17. The discussion section in the present form is relatively weak and should be strengthened with more details and justifications.

Response: Thank you for your careful review and suggestion. Peak flow criterion, degree of agreement and Kling-Gupta efficiency metrics will be added to compare several model performances and more details about the discussion of the obtained results will be discussed.

Proposed changes to manuscript: The discussion of the obtained results will be enriched in Section 4 of the revised manuscript.

18. 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: ïAn 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. ïAn Fotovatikhah, F., et al., "Survey of Computational Intelligence as ˇBasis to Big Flood Management: Challenges, research directions and Future Work," Engineering Applications of Computational Fluid Mechanics 12 (1): 411-437 2018. ïAn Mosavi, A., et al., "Flood Prediction Using Machine Learning Models: Literature Review," Water 10 (11): article no. 1536 2018. ïAn Moazenzadeh, R., et al., "Coupling a ˇfirefly algorithm with support vector regression to predict evaporation in northern Iran," Engineering Applications of Computational Fluid Mechanics 12 (1): 584-597 2018. ïAn

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. ïAn Chau, K.W., et al., "Use of Meta-Heuristic Techniques in Rainfall-Runoff Modelling" Water 9(3): article no. 186, 6p 2017.

Response: Thanks.

Proposed changes to manuscript: We have carefully looked up the mentioned literature, which has been cited in the paper, and we have also extensively looked up other literatures from HESS, JH and other relative journals, added some necessary literatures.

19. Some inconsistencies and minor errors that needed attention are: ïAn Replace "... was supply to depict…" with "…was supplied to depict ..." in line 86 of page 3 ïAn Replace "... into train set, validation set, and test set..." with "...into training set, validation set, and testing set..." in line 206-207 of page 7 ïAn Replace "... test set..." with "... testing set..." in line 209 of page 7 ïAn Replace "... more accuracy inflow forecasting..." with "... more accurate inflow forecasting ..." in line 283 of p10 ïAn Replace "... arisen in in areas..." with "... arisen in areas..." in line 288 of page 10 ïAn Replace "... for train, validation and test set ..." with "... for training, validation and testing set ..." in line 294 of page 10 ïAn Replace "... According to compare the forecasted results of …" with "... According to the comparison of forecasted results of ... " in line 330 of page 11

Response: Thank you for your careful review. We agree with all of the minor changes above, and we will go through carefully the manuscript to check and correct any errors.

Proposed changes to manuscript: Those typos have been corrected in the revision.

20. In the conclusion section, the limitations of this study, suggested improvements of this work and future directions should be highlighted.

Response: Thank you for your careful review and suggestion.

Proposed changes to manuscript: We have carefully checked the conclusion of the article and will add the limitations of this study, improvements of this work and future directions in Section 5 of the revised manuscript.