***We thank the reviewers for their insightful comments and constructive suggestions that have led to the improvement of our paper. Our responses to all the comments and suggestions are detailed below.***

***Reviewer #1:***

***This manuscript proposed an AI framework consisting of hyperparameter selection and training parts to improve training efficiency and reduce overfitting. And the case study for daily runoff prediction in the Maumee domain showed the good potential of this framework. In general, this manuscript is well-written, and the conclusion is reasonable. Therefore, I think the manuscript can be published in EGUsphere after minor revision.***

***(a) When discussing the hyperparameter SS, the paper mentioned that the daily EOF runoff is imbalanced data. Usually, this is very important for AI model training. Can authors explain how to deal with the imbalanced data? I recommend the authors preprocess the data to improve the availability of the data in model training if the authors did not deal with the imbalanced data.***

Thanks. We agreed with the reviewer that the imbalanced data poses great challenges to model training. There are several ways to improve the training quality for imbalanced data: 1) Select an effective machine learning (ML) algorithm that has built-in mechanisms to deal with imbalanced data. 2) Choose a good cross-validation strategy to ensure the training and test datasets follow a similar distribution of the target variable. 3) Set class weights on your target classes to give more weight to the minority class. In our study, as our focus is to identify the influential hyperparameters for the regression models that are trained to predict the magnitude of daily EOF runoff, we chose an effective ML algorithm, the Extreme Gradient Boosting (XGBoost) algorithm for model training; the XGBoost algorithm offers a range of hyperparameters that can give fine-grained control over the model training performance against imbalance data. For example, we used the Stratified K-Fold cross-validation to ensure the training and test datasets follow a similar distribution and defined a loss function that penalizes more the missing predictions of non-zero runoff events, that is, the minority class in this study (See Lines 206 – 209).

***(b) I suggest authors add a flowchart in the manuscript, which will be good for readers to understand the framework.***

Thanks. Figure 1 explains the different components of the methodological framework. To further clarify, we follow the suggestion by the reviewer to add some descriptions on the workflow as follows (See Lines 79 – 82):

We first choose a machine learning algorithm and its associated hyperparameters. Then, we feed the initial hyperparameters (1) to the hyperparameter selection (HS) module to determine the influential hyperparameters (2). Once initial values are assigned to the influential hyperparameters (4), we use the hyperparameter tuning (HT) module to identify their optimal values (3), which allows the algorithm to achieve the best performance in training. A case study is used to illustrate the workflow in detail.

***Reviewer #2:***

***The authors developed a generalizable framework for the improvement of the efficiency and effectiveness of model training and the reduction of model overfitting. This study makes attempt to predict daily runoff based on data-driven models. The two parts of proposed framework: hyperparameter selection and hyperparameter tuning are significant for machine learning. However, I suggest the authors should make more complete explanations on the results. I recommend the article for acceptance after minor revision.***

1. ***The data-driven model using the eXtreme Grandient Boosting. There are lots of the machine learning method and I suggest the authors explain the reason in the introduction part.***

   The framework is generally applicable to data-driven models using different machine learning algorithms, which need to be fine-tuned through hyperparameters. In this study, we chose to use the model using eXtreme Gradient Boosting algorithm (XGBoost), as it has been demonstrated to be effective for a wide range of regression and classification problems, such as the imbalanced data problem of runoff prediction in our case. We included an explanation of the choice of the XGBoost algorithm (See Lines 189 – 190).

2. ***I think the authors should make the simple introduction of study area (figure2) such as climate, soil.***

   Thanks for the suggestion. We included an introduction to the study site (See Section 2.3 Case Study, Lines 144 - 153).

3. ***Figure 5, It seems that the runoff training samples from July, 2012 to Dec.2013 is larger than runoff test samples from Jan.,2014 to Jan.,2016. What is the accuracy? If conversely, training samples are from Jan.,2014 to Jan.,2016 and test samples are from July, 2012 to Dec.2013.***

   Thanks. A training dataset is typically larger than a test dataset, as the purpose of training is to expose the model to more data to learn meaningful patterns from the data. If we do this reversely, we can expect to have worse test performance for the period from July 2012 to December 2013 for both the HT and HS-HT cases. Meanwhile, Figure 5 is intended to show that models preceded by Hyperparameter Selection (HS) and Hyperparameter Tuning (HT) approaches are less prone to overfitting than the case with the HT approach alone. When swapping the test dataset for training for both scenarios, the new results still support the conclusion, i.e., models are less prone to overfitting when using both the HS and HT approaches than that in the case with only the HT approach.

4. ***Figure 6(a), it seems that HS is better than HT when measured EOF has larger value. How about the performance of HS-HT when measured EOF has larger value?***

   Thanks. As shown by Figure 6(a), more blue dots are closer to the Y=X line as the measured EOF runoff values are greater than 10 mm/d, indicating that the model preceded by the HT approach can better predict larger EOF values than the case preceded by the HS approach. This is because the loss function, $R^2$ used for hyperparameter tuning penalizes more the missing predictions of large runoff values (See modifications, Lines 206 - 209). Similarly, the model preceded by the HS-HT approach performed the best to predict larger EOF values compared with both the HS and HT approaches alone, showing that the HT approach can be more effective in identifying the optimal hyperparameter values preceded by the HS approach, which gives the best performance on the test dataset (See Lines 316 – 320).

5. ***From Zenodo, we could find the input file and there are lots of inputs such as soil moisture for this study. I suggest the author give simple introduction about the data input in the Part 2 method.***

   Thanks for the suggestion. We included the data description in the Method section (See Section 2.3.1 Data Preparation) and the supporting information (See Table S2: Influential variables for the Maumee domain in Support Information).

6. ***From input data, soil temperature seems below 0 Celsius degree in winter. Is there some influences of soil frozen on runoff simulation based on the proposed framework?***

   Thanks. When predicting the edge-of-field (EOF) runoff, input variables were identified and selected from the previous study to predict the runoff for both winter and non-winter seasons. For example, the influences of frozen soil can be captured by two input variables in the Maumee domain, including ACSNOM (accumulated melting water out of snow bottom) and SOIL_T (soil Temperature). Please see the modifications in Section 2.3.1 Data Preparation.