Hydrol. Earth Syst. Sci. Discuss.,
https://doi.org/10.5194/hess-2020-224-RC2, 2020

Hydrology and Earth System Sciences
Discussions

Interactive comment on "Comparative analysis of Kernel-based versus BFGS-ANN and deep learning methods in monthly reference evaporation estimation" *by* Mohammad Taghi Sattari et al.

Anonymous Referee #2

Received and published: 24 September 2020

The study estimated monthly reference evapotranspiration (ET0) using four different machine learning techniques, including Gaussian process regression (GPR), support vector regression (SVR), long short-term memory (LSTM), and artificial neural network with the training function of Broyden-Fletcher-Goldfarb-Shanno quasi-Newton (BFGS-ANN). To obtain the best modeling performance, three different kernel functions for both GPR and SVM, and ten different combinations as inputs for all the models proposed were evaluated, respectively. LSTM method is currently an extensively used method in literature to address nonlinear regression problems in a wide range of ap-

plications. LSTM was compared with three conventional approaches (ANN, SVM and GPR), which provides a good and new insight to the existing studies. Regrettably, these models were not well investigated in terms of their generalization ability and computational efficiency. Moreover, the manuscript was not well-written, and its shortcomings can be found in each section. Substantial language improvements should be also made. Therefore, the manuscript needs major revisions before I can recommend it for publication.

Major Comments: 1). Sections Introduction and Methods were not well-written, as well as the organization and design of figures and tables. 2). The dataset was split into two parts for training and testing. However, the results of all figures and tables were only shown in the testing period. 3). As we all know, machine learning is being widely used for addressing many issues, mainly including classification and regression. This study was conducted for regression and aimed at modeling and predicting monthly ET0. I don't know why the descriptions related to classification and classifier were frequently shown.

Introduction 1). For the first paragraph, is it a popular science article? Or suggest deleting this paragraph. 2). Lines 41-52: Some classical previous studies and reviews should be cited for support these descriptions. Besides, it is well known that many physical and empirical models as common methods have been widely used to estimate ET0. Suggest pointing out their advantages and disadvantages, and give some reasons why artificial intelligence (AI) techniques were adopted as alternative tools for this work. 3). As shown in Lines 53-119, so many previous studies (18) of ET0 estimation using different artificial intelligence models were reviewed monotonously. It is utterly pointless. Why did you carry out this study? It should be supported by more sound reasons. Suggest focusing on reviewing some extensively methods (e.g., ANN, SVM, GRNN) for ET0 prediction, and point out their advantages and disadvantages when estimating ET0 in terms of their performance and computational efficiency. For example, both ANN and SVM methods have received a great deal of attention in the

last decade and have been extensively utilized in diverse fields. Nevertheless, these two approaches still have some shortcomings, which have been revealed by previous studies. In general, the ability of ANN method is limited by several disadvantages, such as slow learning speed, over-fitting and local minima. Additionally, it is also relatively difficult to determine some key parameters, such as training function and activation function. SVM also exists several drawbacks, such as high memory requirement and a large amount of computing time during learning process. In order to overcome the disadvantages of these two approaches, many new modeling techniques have been proposed in recent years. For instance, two state-of-the-art machine learning techniques, namely LSTM and GPR, are widely utilized in the hydrologic time series modeling and forecasting. To the best of our knowledge, however, there have been very few attempts to test the practicability and ability of these two advanced approaches (LSTM and GPR) for ET0 modeling and prediction. 4). Regarding the last paragraph, the comparison of different kernel functions for SVM and GPR models, was designed as one goal of this study. Why did you attempt to compare these kernel functions? This aim should be supported by more sound reasons. To the best of my knowledge, many similar studies have been reported, which should reviewed before this paragraph. 5). For ANN model, training function plays an important role in its generalization performance. To my knowledge, a number of training functions (>10) can be used as alternative inner functions, such as conjugate gradient algorithms, gradient descent methods, quasi-Newton methods, Bayesian regulation backpropagation and one step secant backpropagation. The effects of these training functions on ANN have been reported frequently in in diverse fields. These related studies should be reviewed for offering more sound reasons for this paper. More importantly, in this study, why was BFGS selected as training function for ANN model? In order to better check the performance of these training functions, more training functions also can be adopted and compared with BFGS algorithm in this work. Materials and methods 1). Check the titles of “2 Material and method” and “3 Methods”. 2). For Table1, to better compare and evaluate the performance of the used models, the statistics of the data should be divided according to training and testing

periods.

Methods 1). For each method used in this paper, many irrelevant descriptions and inessential details should be omitted. More rigorous and precise description about the principle of the method used in this study should be given. Furthermore, some important and classical papers should be cited. 2). For each method, please point out some special inner functions and parameters of the developed models. Because different functions and parameters have great effects on the generalization of those models. Taking ANN method for example, its generalization performance is generally dependent on many factors, mainly including topological structure of network and relevant parameters (e.g., learning rate, regularization factor and momentum factor) and functions (e.g., learning, activation and training algorithms). In this study, apart from training algorithm, the remaining features above-mentioned were determined by the trial and error method. 3). Suggest adding some descriptions about the used toolbox, package or software for each method.

Results 1). The descriptions of all the tables and figures were so simple and monotonous. 2). As the title of this section is shown, more discussion should be given about this study.

Conclusion 1). In this study, ET0 and its related meteorological data at a time scale of month were gathered from one weather station. Results showed that all the proposed models did a good job in simulating monthly ET0. Nevertheless, these machine learning methods are likely to be questioned in that the intrinsic mechanisms of these well-trained black box models remain poorly described or understood. To a certain degree, this limitation decreases the reliability of these techniques. 2). In the follow-up work, the performance of the GPR and LSTM models for the present study should be further evaluated at finer time scales, such as daily. Moreover, more weather stations or regions should be taken into consideration.

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

224, 2020.