Hydrol. Earth Syst. Sci. Discuss.,
doi:10.5194/hess-2016-247-AC3, 2016

Hydrology and Earth System Sciences Discussions
Open Access
EGU

HESSD

Interactive comment

Interactive comment on "Comparison of six different soft computing methods in modeling evaporation in different climates" *by* L. Wang et al.

L. Wang et al.

wang@cug.edu.cn

Received and published: 8 August 2016

Weakness

1. Language: needs substantial editing efforts to ensure consistency and readability via improving e.g., wording, sentence structure, paragraph connection and cohesion.

Reply: The language of the whole manuscript has been carefully revised and improved, thank you.

2. Methodology: Training dataset is select randomly, however without ensemble the randomness of the data selection is still weak. Cross-site validation is also necessary before concluding which model is the "best" one.

Reply: we further evaluated the applied models with full weather inputs by changing

Printer-friendly version

Discussion paper

training and testing period. Please see Table 14 for the results.

Major comments 1. Models used 50% of data for training and the rest for testing (randomly chosen) at each study site. However, random sampling is not repeated to ensure generality of the model testing results. With only one random sampling of training dataset, the trained model is potentially biased to that particular case. What will happen, if use the testing data for training and training data for testing? Will the model predictability retain? In order to show the generality of the model performance, one has to do multiple ensembles of training data sampling and present the "mean model".

Reply: as suggested models with full weather inputs were evaluated by changing training and testing period. Please see Table 14 for the results.

2. This study made lots of efforts on comparing different models and trying to find out the "best" one. The identification of the "best" model is based on within-site evaluation. In that sense, the significance of this study is highly limited. What happen if one would like to apply the "best" model to other sites with different climate, or do a regional modeling? This type of question could be answered by doing cross-site evaluation. For example, apply the trained model to other 7 sites and show the overall performance.

Reply: as suggested, the generalized MLP9 model was also tested at each site.

3. Results interpretation: Table, figures, and main text are heavily redundant, e.g., no need to repeat all the numbers (e.g., R2) in main text that have been included in table.

Reply: results section has been revised.

Specific comments

L24. The first sentence does not make too much sense, because this study focused on pan evaporation, which best inform water management such as agriculture. But for terrestrial ecological processes and regional climate change, evaporation is not as significant as it in agriculture. I would suggest the abstract starting with "Pan evaporation plays . . . in informing . . . ". And followed with another sentence highlighting the fact that one of the basic challenges is modeling . . ... L30. First time use Ep, better to define it earlier. L31. No need to list all the climate variables. Maybe: "We develop, train, and validate the eight models at various sites crossing a wide range of climate". L33. The first part of applications focused . . .. Remove this sentence, because the next sentence is actually presenting the accuracy comparison. L38. Generalized models were also developed and tested. . ... Remove this sentence. L42. BJ, CQ and HK station. Define the sites before use them. L42. Recommendation or major implication based on this study is needed to end the abstract. L48 "and air" -> "and air temperature"? L50, roles in . . . -> roles in informing water resources redistribution and irrigation system design. L55 is -> are L56 one of the, remove; aspects in the hydrological cycle, remove L57 to integrated -> for integrating L59 some, remove L63 for estimating Ep as a function of meteorological data -> to linking Ep to various meteorological drivers. L64. But some of . . . -> but the applications of these techniques are often limited by data availability and completeness. L68. What are conventional techniques, list a few. L75 at a site in hot and dry climate -> at a hot and dry site L76 is -> was L70 for example -> .For example, . . .and from L70 – L77 replace ";" with "." L110 provided an impetus -> impede L111. Which provided an impetus for . . ., remove. L112 "considering the importance of . . .or hydrological modeling", remove. L116 in modeling Ep. . . -> in Ep modeling with different combinations of climate inputs. L119 "using generalized . . . models" -> using eight different models. L129. MLPs are organized as hierarchical networks with several layers L133. its input -> input L142. but they do not use-> without using L143 The structure of, remove L148 Two types of neurons (S-summation and D-summation) are connected to patter layer unit L179 which is -> that is L183 which can be used for optimization problems, remove L193 which projects -> that projects L196 efficient is enough, redundant to say quick, converging to global optimum. L199 more simpler and more efficient L200 This issue is caused by-> , due to the reason that LSSVM solves linear equations instead of a quadratic programming problem in SVM. L204. This subject -> these models L212 consists ->consiting

Reply: All the corrections above have been made carefully.

L236 Why use monthly data? e.g., Goyal 2014 investigated various techniques to improve daily Ep in India. Is it possible to do daily at the eight selected sites as well?

Reply: we studied with monthly time scale following the related literature (Kisi 2015, Guven and Kisi 2013, Citakoglu et al. 2014, Tezel and Buyukyildiz 2016, Kisi 2009, Campos-Aranda 2004, Fennessey 2000, Savenije 1997, Francisco and Aranda 1997, Kim 2011, Alvarez et al. 2007, Almedeij 2012) which also used pan evaporation in monthly time scale.

References:

Kisi, O. (2015) Pan evaporation modeling using least square support vector machine, multivariate adaptive regression splines and M5 model tree. Journal of Hydrology 528, 312-320.

Guven, A. and Kisi, O. (2013) Monthly pan evaporation modeling using linear genetic programming. Journal of Hydrology 503, 178-185.

Citakoglu, H., Cobaner, M., Haktanir, T. and Kisi, O. (2014) Estimation of Monthly Mean Reference Evapotranspiration in Turkey. Water Resources Management 28(1), 99-113.

Tezel, G. and Buyukyildiz, M. (2016) Monthly evaporation forecasting using artificial neural networks and support vector machines. Theoretical and Applied Climatology 124(1-2), 69-80.

Kisi, O. (2009) Modeling monthly evaporation using two different neural computing techniques. Irrigation Science 27(5), 417-430.

Campos-Aranda, D.F. (2004) Estimation of monthly evaporation in a class-A pan in the Mexican Republic through temperature data. IngenieriaHidraulica En Mexico 19(4), 85-96.

Fennessey, N.M. (2000) Estimating average monthly lake evaporation in the northeast

United States. Journal of the American Water Resources Association 36(4), 759-769.

Savenije, H.H.G. (1997) Determination of evaporation from a catchment water balance at a monthly time scale. Hydrology and Earth System Sciences 1(1), 93-100.

Francisco, D. and Aranda, C. (1997) Estimate of the monthly evaporation from a simulation of operations of the El Cuchillo-Marte R. Gomez Dam system. IngenieriaHidraulica En Mexico 12(2), 49-56.

Kim, S. (2011) Nonlinear Hydrologic Modeling Using the Stochastic and Neural Networks Approach. Disaster Advances 4(1), 53-63.

Alvarez, V.M., Gonzalez-Real, M.M., Baille, A. and Martinez, J.M.M. (2007) A novel approach for estimating the pan coefficient of irrigation water reservoirs application to South Eastern Spain. Agricultural Water Management 92(1-2), 29-40.

Almedeij, J. (2012) Modeling Pan Evaporation for Kuwait by Multiple Linear Regression. Scientific World Journal.

L238 – 242. Site information are redundant, with figure 3. Or could be just put in a table.

Reply: It was corrected (see Table 1).

L242-258. Put the site description (lon, lat, alt, mean annual temperature, mean annual precipitation) in a table. In the text, highlight the most important fact, do not literally reiterate the site information.

Reply: It was corrected (see Table 1).

L280-286. Again, too many numbers in this section (which have already been shown in Table 1). Just illustrate the most important fact, e.g., which site has the highest Ep, why and how? L305. ANFIS, GP, . . ., and SS. -> six soft computing and two regression models L314-317 This study . . . in the application. Remove. Just present results and discussion, no need to reiterate what have been done. L329. Models with full weather data have the best accuracy. This seems obvious but indeed has significant implications. As using more and more predictor variables, the response variables could be commonly better predicted. However, the issue is the expenditure. In this case, is the data availability. Variables like air temperature is relatively easy to measure and important for evaporation, we definitely want to include it in ep modeling. However, is there a predictor variable that is relatively hard to measure (unavailable) but is “must be include” predictor variable? Is it necessary to use the full model for large-scale (regional) prediction? If so, is this conclusion also valid at other study sites?

Reply: model results for the 1st, 6th, 7th and 8th input combinations indicates that the soft computing models can be successfully used with local calibration (see Tables 4-11). Table 13 shows that the generalization of the soft computing models are possible in case of limited inputs. However, we have not investigated the accuracy of the generalized models with limited data for each station, separately because of the length of the paper. This may be done in another study.

L367. The best accuracy were generally obtained from five-input models and GNRR model perform better. It is excited to see that a certain model stand out. But it would be helpful, it one can go one step further and try to figure out the underlying reason why that model is “best”. What kind of feature of that model could possibly lead to the success?

Reply: MLP generally performed better than the other methods in estimating Ep. However, the accuracy of the other methods is also satisfactorily well. The advantages of each method were included in the methods section of the revised version.

L414. It is obvious that . . .. Throughout the paper, many places used this sentence structure “it is obvious that”. Try to avoid if necessary, because it imply that if one can not immediately understand the results then he is stupid. Furthermore, sometime the results are not that obvious and it is always authors’ responsibility to help the readers understand those results.

Reply: It was corrected.

Interactive comment on Hydrol. Earth Syst. Sci. Discuss., doi:10.5194/hess-2016-247, 2016.