Interactive comment on "*Machine learning methods for stream water temperature prediction*" by Feigl et al.

Pr. Salim Heddam
heddamsalim@yahoo.fr.
https://orcid.org/0000-0002-8055-8463

General Comments

In the present study, the authors applied and compared six machines learning (ML) models for predicting river water temperature: XGBoost, FNN, RF, two deep learning models and the step-wise linear regression. Results obtained using the proposed ML were compared to those obtained using the air2stream and the linear regression (LR). The proposed models were developed and compared using data collected at ten Austria catchments. A standard modelling approach has been adopted for this study based on linking a set of input variables to one output variable. Mean, maximal and minimal air temperatures, precipitation and global solar radiation were selected as the most relevant regressors and used as input variables for predicting water temperature. Overall the paper is very interesting, well-structured, easy to ready, and written in a scientifically sound manner. Although modelling reviver water temperature is broadly discussed in the literature, and a large amount of work has been done in recent years, admittedly with some important conclusive answers, the present work has a large potential to expand significantly our knowledge in this subject. The approach proposed in the present work owes its originality from several points, including: (*i*) new introduced ML models belonging to different categories and having different modelling strategies, (*ii*) the use of different climatic variables as input instead of what is already done in the literature (i.e., only air temperature and discharge), and (*iii*) the use of the Bayesian optimization method (BO) is innovative and the authors have developed a detail proposal. The most important finding of the present study is that the ML models were more accurate than the air2stream and the LR models, and the BO can help in improving significantly the models accuracies. In addition, the performances of the ML models varied from one catchment to another and in overall, worked equally with slightly difference. I have a number of concerns related to the paper, to my opinion needs to be clarified by the authors.

Major's Comments

1. The comparisons of models results with in situ measured data using only errors metrics is insufficient and does not help in providing robust conclusions regarding models accuracies, robustness and fitting capabilities. Specifically, using several kinds of goodness-of-fit indicators should be more useful: the coefficient of determination (R^2), the Nash-Sutcliffe efficiency (NSE), and the index of agreement *d*, are highly recommended for hydrological models evaluation (Legates and McCabe 1999; Moriasi et al. 2007; Harmel and Smith 2007; Gupta 1998, 2008; Krause et al. 2005).
2. Models structures need to be clarified. In Lines 173-175, the authors argued that including the lag of all variables for 4 previous days can help in improving models accuracies according to Webb et al. (2003). First, using only 4 previous lag should be justified, on which basis it was selected (i.e., cross-correlation analysis can be helpful for answering this question)? Second, according to Webb et al. (2003), adopting the previous lag as input variables can be useful on only hourly data scenario. Therefore, a comparison between models with and without lag data may be a good option.

Specific Comments

1. The introduction is not deeply written and in some cases need improvement. Specifically, the proposed ML reported in the literature should be presented, discussed, and the strength and weakness of each one would be more useful and effective if they are highlighted. Using lumped references do not help in understanding the mains contribution of the work.
2. Research gap. What are the mains contributions of the present study in comparison to what is already done? What does it add to existing literature?
3. Lines 47 to 50, from Austria to characteristics. To our opinion this paragraph is more suitable to be moved to section 2.1.
4. Line 79: *''To the author's knowledge, RF has not been applied for river water temperature prediction yet''*. This statement is incorrect. The RF was recently reported as a powerful tool for predicting river water temperature (Heddam et al. 2020).
5. Models comparison using cross-station scenarios can help in providing more conclusions, and a clear idea about models capabilities outside of their own catchment area: models calibration using data from on station and validated for other stations (i.e., see Zhu and Heddam 2019).

References

Gupta, H.V., Sorooshian, S., Yapo, P.O., (1998). Toward improved calibration of hydrological models: multiple and non-commensurable measures of information. Water Resour. Res. 34, 751-763.

Gupta, H.V., Wagener, T., Liu, Y., (2008). Reconciling theory with observations: elements of a diagnostic approach to model evaluation. Hydrol. Proc. 22, 3802-3813.

Harmel, R.D., & Smith, P.K. (2007). Consideration of measurement uncertainty in the evaluation of goodness-of-fit in hydrologic and water quality modeling. *Journal of Hydrology*, *337*(3-4), 326-336.

Heddam, S., Ptak, M., & Zhu, S. (2020). Modelling of daily lake surface water temperature from air temperature: Extremely randomized trees (ERT) versus Air2Water, MARS, M5Tree, RF and MLPNN. *Journal of Hydrology*, *588*, 125130.

Krause, P., Boyle, D., Bse, F., (2005). Comparison of different efficiency criteria for hydrological model assessment. Adv. Geosci. 5, 89-97.

Legates, D.R., & McCabe Jr, G.J. (1999). Evaluating the use of "goodness-of-fit" measures in hydrologic and hydroclimatic model validation. *Water resources research*, *35*(1), 233-241.

Moriasi, D.N., Arnold, J.G., Van Liew, M.W., Bingner, R.L., Harmel, R.D., & Veith, T.L. (2007). Model evaluation guidelines for systematic quantification of accuracy in watershed simulations. *Transactions of the ASABE*, *50*(3), 885-900.

Webb, B.W., Clack, P.D., Walling, D.E. (2003). Water-air temperature relationships in a Devon river system and the role of flow, Hydrological Processes, 17, 3069-3084.

Zhu, S., & Heddam, S. (2019). Modelling of maximum daily water temperature for streams: Optimally pruned extreme learning machine (OPELM) versus radial basis function neural networks (RBFNN). *Environmental Processes*, *6*(3), 789-804.