

Interactive comment on “A Hybridized NGBoost-XGBoost Framework for Robust Evaporation and Evapotranspiration Prediction” by Hakan Başağaoğlu et al.

Anonymous Referee #1

Received and published: 17 August 2020

In this work, the authors analyzed the relationship between potential evapotranspiration (ET_p), actual evapotranspiration (ET_a), and surface water evaporation, using data from multiple sources. Major comments:

1. Study objective #1 is not clear. Relationship between PET and Actual ET has been well studied in the literature. The authors should mention the previous works in this area. I found the literature survey is too cursory. Suggest that the authors move some materials from Section 2 to the Introduction. Even so, it's not clear to me from the Intro why the existing results are not sufficient, such that the authors need a sophisticated ML to revisit the problem. The motivation needs to be elaborated further in the Intro.

C1

2. Similarly, in Section 2 the motivation of using ML is not clear. What regression methods have been used before? Why existing methods are insufficient in terms of model performance? The authors need to touch on these aspects. Otherwise, the work seems to focus solely on a new ML algorithm without justification, and no baseline results (e.g., multivariate linear regression) were provided.

3. The ML pipeline is not clear. A diagram is needed to show inputs and output to the ML model. Around L185, the authors simply spelled out the inputs, without much reasoning. Why these features are selected? What is the lead time of prediction? The promise of ML is not so much for well gauged sites, but for sites with a lot of missing data.

4. Variable importance calculation is well established for tree-based method, which entails finding whether a variable is selected to split on during the tree building process, and how much the squared error (over all trees) is improved or reduced as a result. Why a new variable importance method is needed?

Minor comments:

In Abstract, the authors concluded “the deterministic analysis reveals that ET_p set the upper bound for ET_a”, isn't this expected?

Interactive comment on Hydrol. Earth Syst. Sci. Discuss., <https://doi.org/10.5194/hess-2020-260>, 2020.

C2