Review Report

Journal:	Hydrology and Earth System Sciences
Manuscript's Ref.:	hess-2020-460
Title:	Modeling and interpreting hydrological responses of sustainable urban drainage systems with explainable machine learning methods
Authors:	Yang Yang, Ting Fong May Chui
Reviewer:	Georgia Papacharalampous
Date agreed:	2020-10-09
Date submitted:	2020-11-05
Recommendation:	Minor revisions

Summary

The paper focuses on the predictive modeling of sustainable drainage systems (SuDS) at fine temporal scales using boosting (Friedman 2001). Several boosting variants are formed and exploited in two case studies, while comparisons with the linear regression algorithm and the Storm Water Management Model (SWMM; Rossman 2015) are also provided. Furthermore, the SHapley Additive exPlanations (SHAP) method (Lundberg and Lee 2017) is used to explain the contribution of each variable (else referred to as "feature") to the issued predictions, thereby facilitating interpretability to some extent.

General comments

In general, I find that the manuscript is well-formulated and -written, and I think that the work done so far (including the release of the `R` codes at GitHub) should be appreciated. Nonetheless, I also think that there is some room for improvement before publication.

I recommend minor revisions. My comments are given right below.

Comments

(1) To my view, the following clarification is required: Which are the similarities and differences between basic variable importance measures (available in the `xgboost` `R` package) and the SHAP methodology (available in the `SHAPforxgboost` `R` package)?

(2) Since interpretability is one of the main themes of the present work, I feel that a comparison (direct or indirect, depending on the answer to comment #1) between basic variable importance measures and the SHAP methodology is currently missing from the manuscript and should be necessarily made for both case studies. New computations are needed for this comment to be fully addressed (independently of the answer to comment #1); however, these computations will only require the `xgboost` `R` package (which is already used in the paper).

(3) In light of comments #1 and #2, other hydrological studies using boosting or random forests while also emphasizing on interpretability (by using variable importance measures) could be discussed (in comparison to the present study) somewhere in the manuscript. What is the added value of the present work with respect to such existing works?

(4) In the "Introduction" section, it is written that "only a few studies adopted machine learning methods to investigate the hydrological processes of SuDS", with the studies by Eric et al. (2015), Khan et al. (2013), Li et al. (2019), and Yang and Chui (2019) being discussed as examples of such studies. Since such studies are quite close to the present work, more of them could be reported (provided that they exist).

(5) In the same section, it is also written that "modeling the responses of SuDS at fine temporal scales requires high-dimensional hydrometeorological time series to be used as input, which is difficult in machine learning". Could this sentence be further elaborated? I would say that the opposite holds, i.e., that machine learning methods are ideal for handling high-dimensional hydrometeorological time series.

(6) The reader could also be referred to several specialized books (e.g., Hastie et al. 2009; James et al. 2013; Witten et al. 2007), for further information on the machine learning (or statistical learning) methods used in the paper.

(7) Another concern of mine is related to the small number of real-world cases examined in the paper. I think that the application of the proposed procedures to large real-word datasets (comprising hundreds of cases) should be addressed at least with extensive relevant discussions in the manuscript (e.g., future research recommendations). (Currently, it is only suggested using "the SHAP method in more case studies"). To my view, these extensive discussions are important, especially given that (i) there are studies in the hydro-meteorological literature validating their

models using big datasets, and (ii) the first aim of the paper is to "evaluate the usefulness of machine learning methods in predicting the hydrological responses of SuDS at fine temporal scales". The necessity of evaluating machine learning methods using big datasets is extensively discussed by Boulesteix et al. (2018).

(8) In the "Conclusions" section it is written that "the proposed model training methods are semi-automatic, requiring minimal user input". It would be useful to discuss (somewhere in the paper) which parts of the proposed methods are not (fully) automatic, and how one could overcome this limitation to allow large-scale (even global-scale) investigations (see also comment #7).

(9) Currently, the use of the `xgboost` and `SHAPforxgboost` `R` packages is reported in the manuscript. To my view, all utilized software packages (which, of course, at the moment can be found online at https://github.com/stsfk/explainable_ml_hydro, since the `R` code has been made available) should necessarily be reported and cited in the paper.

(10) Finally, the manuscript is not typo-free at the moment. Particular attention should be placed on the mathematical notations. For instance, the transpose operator should not be written in italics (therefore, T should be replaced with T) and the vectors should be bolded (therefore, $X_{t-m,t}$ should be replaced with $\boldsymbol{X}_{t-m,t}$).

References

Boulesteix AL, Binder H, Abrahamowicz M, Sauerbrei W, for the Simulation Panel of the STRATOS Initiative (2018) On the necessity and design of studies comparing statistical methods. Biometrical Journal 60(1):216–218. doi:10.1002/bimj.201700129

Hastie T, Tibshirani R, Friedman JH (2009) The elements of statistical learning: Data mining, inference and prediction, second edition. Springer, New York. doi:10.1007/978-0-387-84858-7

James G, Witten D, Hastie T, Tibshirani R (2013) An introduction to statistical learning. Springer, New York. doi:10.1007/978-1-4614-7138-7

Witten IH, Frank E, Hall MA, Pal CJ (2017) Data Mining: Practical machine learning tools and techniques, fourth edition. Elsevier Inc. ISBN:978-0-12-804291-5