

## **Review of Modeling and interpreting hydrological responses of sustainable urban drainage systems with explainable machine learning methods.**

The article by Yang and Chui shows the results of a study on the prediction of the hydrological response of sustainable urban drainage systems (SuDS) using Machine Learning algorithms.

Through an in-depth examination of both the manuscript and the authors' responses to other reviewers' comments, I was able to appreciate the effort made by the authors to adequately address the constructive comments of colleagues.

However, I believe that the article is not yet ready for publication and should be re-evaluated after further review.

I expect the Authors to further improve the article in order to overcome my main concerns:

- First, I believe that a Machine Learning-based approach is less suitable for addressing a SuDS problem than a physically based systemic approach. If it is true that in some cases the geometric and hydraulic characteristics of these systems are not well known, it is even more true that the presence of experimental field data on inflows and outflows represents a rare exception. For this reason, a physically based modeling is in most cases to be preferred and allows to address a wider variety of problems, while in this case a model based on Machine Learning algorithms would be limited to the specific case. Authors should give more convincing reasons for the choice of approach.
- The novelty of the work is not relevant from a methodological point of view. The XGBoost algorithm is widely used in literature, as well as the other tools (Nested cross-validation, Bayesian optimization, etc.) used in modeling. Authors should better highlight in the introduction section why this work would represent a significant upgrade over existing literature, worthy of publication in HESS.
- The article is very long, and in some places still unclear. It could certainly be shortened without compromising its contents. This is the most important aspect on which Authors should focus their efforts in order to obtain a more concise and clear manuscript.

Furthermore, I would ask the Authors to consider the following additional comments:

P3 L67 - "Machine learning methods, also referred to as data-driven modeling, predictive modeling, and statistical learning": These terms are not strictly equivalent.

P4 L126 – Section 2.1.1 Local and global methods: This section is very basic and not essential for the manuscript. Sections 2.1.1 and 2.1.2 should be merged and shortened.

P14 L183 – “The water levels are converted into discharge measurements using stage-discharge rating curves”: How were the curves obtained? Do they come from an appropriate calibration?

P16 L426 – “The feature engineering and XGBoost hyperparameters are automatically optimized using the Bayesian optimization”: Authors should report the optimal values of the hyperparameters, possibly in a table.

P18 L486 – Nash-Sutcliffe efficiency coefficient and coefficient of determination  $R^2$  are very similar metrics, it is not useful to consider both.

P28 L668 – Conclusions: This section should be much more concise and effective in summarizing the main findings of the study.