Hydrol. Earth Syst. Sci. Discuss.,
https://doi.org/10.5194/hess-2019-650-RC2, 2020

Hydrology and Earth System Sciences
Discussions
Open Access
EGU

Interactive comment on "Predicting tile drainage discharge using machine learning algorithms" *by* Saghar Khodadad Motarjemi et al.

Anonymous Referee #2

Received and published: 6 February 2020

The manuscript describe a study about modelling tile-drainage discharge aggregated on annual level on different sites in Denmark, by using machine learning algorithms, in particular Random Forest and Cubist. The overall presentation of the study is fairly concise. However, the manuscript lacks more detailed explanation on motivation of such study and final conclusions on applicability of the results. The latter is most probably the case due to the miss-conception of the validation process. More details on the study and manuscript's sections are given below.

The study itself has been thought systematically on how to approach the modelling phase. However, some phases were misconducted. First of all, the time scale of the study is considered to be annual in regard with the output, which is not clearly specified how then the input has been encoded/aggregated, knowing the fact that meteo data

Printer-friendly version

Discussion paper

are available on daily bases. Next problematic approach is using mechanistic models to encode/represent the input in the modelling process. Such case is with meteorological data that are run through water balanced model EVACROP. Finally, after performing the cross validation, the study does not extract any new knowledge, rather discuss differences in cross-validation techniques - which clearly does not fit the scope of the journal. To this end, I would rather say that finding out that percolation (Db) is most important attribute upon running huge machinery is not an added value, as that fact is proven by theory and more specific by correlation of both variables discharge and percolation, which is obvious from Figure 7. Rather more interesting contribution would be to see which of precipitation and/or evapotranspiration is more significant in combination with different/specific landscape and soil characteristics. Similarly, second most importantly identified covariate - elevation - is pretty difficult to be simply explained as cause for discharge. The small range of values with pretty small sample size cause a behaviour as a clustering bias, especially if experimental sites are uniformly (equidistantly) distributed along the given range. So instead of discovering more interesting patterns, those are replaces with single covariate that encapsulate different processes under the hood. Therefore, I would rather see what is happening if this covariate is removed.

Regarding the manuscript, the sections introduction and data are well described. The methodology and validation part is also fairly good described, except the part for how the importance of covariates is performed - especially knowing the fact that RF is not that open model so to be able to easily extract the most important covariates. Results and discussion section are lacking more details and focus on actual findings and less (or at least not that dominantly) on performance from different validation schemas. Validation schemas are well defined, and in discussion difference in performance of the models should be discussed - talking of which, spatial bias is not mention upon introduction. Such given discussion sounds more of evaluating three different validation schemas, rather than discussion of new findings in the domain of hydrology.

Printer-friendly version

Discussion paper

Finally, few sentences need to be strongly changed as they are not true:

l.20: "This work opens up for a better understanding of the dynamics of tile drainage discharge and proves that machine-learning techniques can perform as predictive models in this specific concept." - too optimistic conclusion without good ground for such claim.

l.229: "The proposed tile-drainage discharge predictive model is not dependent on the climatic and constantly measured data and makes it possible to use different geographical properties as predictive parameters." - this is absolutely not true as percolation is derived from a model that uses at input precipitation and evapotranspiration data.

Interactive comment on Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2019-650, 2020.

Printer-friendly version

Discussion paper