Review of manuscript Lin et al., Symbolic regression-based regionalization of baseflow separation parameter using catchment-scale characteristics

I. Summary of the most important scientific findings

(1) First, I would like to express that I very much like the subject of the paper and the research questions that have been investigated. The regionalization of hydrological model parameters based on catchments attributes has been a long history in the PUB initiative, and, over the last decade with the developments of deep learning tools gained some new momentum.
(2) This paper addresses this area of research, by using symbolic regression methods and combination with genetic programing, to not only map the (transfer) relationship of model parameters with physio-geographic properties of the catchments, but also make these relations interpretable in terms of concrete mathematical expressions that are derived.
(3) In this way, the paper provides a valuable approach towards the development of "explainable AI" in hydrology and eater management.

II. Novelty and contributions of this paper

(4) In detail, the paper investigates how the single parameter N of the Smooth-Minima-approach to perform base flow separation can be estimated and regionalized from tracer-based measurements and physio-geographical data for 855 catchments in the US.
(5) Optimal algebraic equations and parameters were estimated using a genetic programming approach and are compared to standard literature formulations demonstrating superior performance.

III. Critical comments and recommendations to the editor

While I very much like the general scope of the paper as stated before, here are a number of critical questions, concerns or suggestions I would like to make and that are listed in the following:

(6) Why did you use an genetic programming approach? It is well known that this approach is limited to a few number of variables and levels of branches. You are generating a discrete optimization problem that is in my opinion very difficult to handle. In recent work by e.g. Feigl et al. (2020, cited by you) this is approached/solved by generation a grammar and "compressing" it with a VAE (and some constrains) into a latent space which is continuous and therefore usable for "gradient-based" search methods. Why didn't you use such an approach?
(7) You used N=5 (FD) and N=1.6*A^0.2 (FPL) as benchmark models taken as typical parameterizations from the literature. I think that is ok to see the limitations and the lower end of performance. But why didn't you use that equation structure (a*A^b) and looked ho well it worked in comparison?
(8) For such benchmark experiments, we often use a ML approach (e.g. Xgboost) in order to analyze the "maximum" possible performance given the calibration data and catchment properties, to see how well the CF approach does in comparison.

(9) While Appendix A is given some text book information on the principle of streamflow separation, I would be much more interested in how the separation was performed and especially what kind of uncertainties are involved.

Some technical questions of concern are the following:

(10) The equations in Table 2 look to me strongly depend on what units you are choosing for Ks and A. Also, in F2 and F3 you add A and Ks which makes no sense in terms of units! What are the impacts of different choices on the overall estimation process. Is there guidance needed for any other user application?

(11) Looking at Fig. 2, I find some information on the derived equations, but I do not see what kind of function space is actually generated to be used in the genetic programming approach – does the GE approach in your case for recursion? Please clarify.

Overall, I would recommend to the editor to accept the manuscript after some (major) revisions.

IV. Minor and specific comments

L95: Does the data set included nested catchments/information?

L117: Units of the variables in the equation?

L199: In Fig. 3 the individual (10) lines are hardly visible or distinguishable

L268: I am very surprised by the relatively small spread of Ks values that is observed in the catchments?

L336-340: Can you provide some reference for the statements.