Geoscientific
Model Development
Discussions

Interactive comment on "HydroMix v1.0: a new Bayesian mixing framework for attributing uncertain hydrological sources" by Harsh Beria et al.

David Ham

david.ham@imperial.ac.uk

I am writing as an executive editor of GMD to highlight two issues with the code availability section which needs to be remedied in the revised manuscript.

Code and data only on GitHub

Thank you for providing a reference to the code and some of the data used in the experiments presented in your manuscript. There are two problems with providing this data via GitHub. The first is that a reader cannot identify the exact version of the code

that was used in the paper (for example, you may fix bugs or add features in the future). The second issue is that projects sometimes change the revision control system they use, or the hosting (the project might move to GitLab, for example). The solution to both of these issues is to provide a reference to a persistent archive of the exact version of the code that was used in the manuscript. This reference can, and should, be in addition to the GitHub link, so that a user can also always access the most recent version of the code. Since your original code is hosted on GitHub, the easiest way to produce a persistent archive of a precise version is to use GitHub's Zenodo integration. For more details, see: https://guides.github.com/activities/citable-code/.

Data only available on request

Some of the data used in the experiments is only available on request from the authors. Because authors may not always have access to this data (for example they may move institution, or data may be deleted), this is not usually an acceptable practice. This data should also be archived in a public, persistent archive such as Zenodo. It is only in the case where the authors legally cannot publish this data (for example because they only have it under licence and the licence does not permit republication) that it would be acceptable to supply data only on request. In these cases, the reason why the data cannot be published must be given in the manuscript, and the data must be identified with sufficient detail that it will always be possible for the precise data used to be identified.