Comment on gmd-2021-336

Paul C. Astagneau (Referee)

Referee comment on "RavenR v2.1.4: an open-source R package to support flexible hydrologic modelling" by Robert Chlumsky et al., Geosci. Model Dev. Discuss., https://doi.org/10.5194/gmd-2021-336-RC1, 2021

This manuscript presents a new R package which aims at helping modellers in their use of the Raven hydrologic framework. Most of the package features consist in functions for data wrangling to feed Raven and functions for simulation analyses. Rationales behind the implementation of RavenR are presented. Examples of the RavenR functionalities are introduced using a formerly built perceptual model of the Liard river basin.

Several authors have advocated for the use of flexible structures for systematic testing of multiple working hypotheses in hydrological modelling. The use of such structures inherently results in higher complexity for modellers hence a challenge for reproducibility of methods and results. I think that any attempt at improving the use of these flexible structures is therefore relevant to the community of hydrological modellers. Furthermore, an extensive documentation is introduced to use the RavenR package, lots of interesting functionalities ranging from data preparation to simulation analysis are implemented and feedbacks between users and developers are encouraged to maintain and improve the package.

However, to be able to thoroughly evaluate the added value of using RavenR, I would have needed some experience with the Raven hydrologic framework. As it is not objectively possible in the time required to write a review, the following comments can only be seen as a way to improve the readability of the paper for non-Raven users and broaden the possible reach to the hydrological community.

General comments:

- Two similar flexible hydrological frameworks need to be cited in this work (either in the introduction or in Sect. 2): DECIPHeR (Coxon et al., 2019) and SuperflexPy (Dal Molin

et al., 2020). A short description of the main differences between Raven/RavenR and these frameworks might further demonstrate the added value of using RavenR.

- To improve understanding by new users of Raven (or even new hydrological modellers), I suggest adding a short description of the main choices that were made in the Raven hydrological framework and RavenR in terms of programming languages. The Raven hydrologic framework is coded in a compiled programming language, probably for computational speed and flexibility purposes. To improve its usability, the RavenR package was created. However, some hydrological models are coded in a compiled programming language and interfaced by R using packages (e.g. hydromad; Andrews and Guillaume, 2018). Why is the Raven workflow (in terms of programming languages) more suited for flexible modelling?

- Section 3 is probably the most important section of this paper if we want to use the RavenR package and the Raven hydrologic framework. The steps of the hydrological workflow are presented in Table 1 and the related R code and model files are provided to understand the functionalities of RavenR. However, I found some parts of this section a bit difficult to understand, especially since in the provided R script, the model run command line appears before input file processing.
  - The authors state line 195 that step 4 and 5 will not be presented but it is not clear why. They are important steps of the hydrological workflow especially when performing uncertainty analyses. An explanation of why this is not relevant given the objectives of the paper is needed.
  - Although it is probably relevant to introduce the notion of locked or protected HRUs in Sect. 3.2.4, hydrological modellers with less experience with Raven might need a simpler use case of model discretization first. If the authors want to keep this section as it is, I suggest adding a simpler example in the future vignettes of the package.
  - Sect 3.3 may be too long and its purpose not very clear since the evaluation of what the authors call "model realism" does not lead to questioning the hypotheses behind the Liard basin model. I think this section should be limited to a presentation of the possible analyses of model simulation enabled by RavenR. Possible cuts: l 376 to l 381; l 383 to "Overall" l 386; from "A similar check" l 396 to l 402; from "The model" l 407 to "bias in estimation" l 408; from "The hydrograph" l 430 to "peak" l 433; from "The plot" l 446 to l 448; from "The results" l 452 to l 453; from "The plot shows" l 460 to "measurements" l 464.
  - Overall, I think that the R script provided to understand Sect. 3 could become a vignette but for a very simple use case that would include parameter estimation procedures and questioning of modelling hypotheses. Building a simple model from data preparation to output analysis using a catchment from the Camel dataset (Addor et al., 2017) would allow very different modellers to use the Raven hydrologic framework.

Minor comments:

- I think that lines 60 to 70 could be moved just after line 44 for better links between the paragraphs of the introduction.
- Please add the references of Python, R and C++.
- Line 128, "3) running raven" should be moved before "2) reading output files".
- Line 349/350: please remove "providing…for the right reasons? (Kirchner; Euser et al.,

2013)", as it is not the place to provide insights into a scientific question that was not presented in the introduction.

- Please define "model realism" and "reality checks" in Sect. 3.3.1, as they are vague concepts, especially when no other data than streamflow are available for model validation.
- Line 365: I do not think that the term "observed baseflow" can be used to refer to the results of baseflow separation techniques that rely only on streamflow time series.
- Lines 414 to 418 should not appear in Sect 3.3.
- Line 449: "Figure C" should be "Figure B".
- Line 550: "Figure D" should be "Figure C".

Technical comments:

I noticed a few typos. As I am a non-native English speaker, the following comments might not be relevant.

- L1: "advances…have enhanced" instead of "has enhanced".
- References such as "(e.g. GR4J (Perrin et al., 2003))" should appear as "(e.g. GR4J; Perrin et al., 2003)". The latex command for this is: \citep[e.g. GR4J;][]{citationkey}.
- Line 312: "The development…requires" instead of "require".

Comments specific to the R package documentation:

- From my understanding, the pipe operator is not mandatory to run the Raven package and is only used here for better readability. However, some R users are not familiar with the dplyr syntax. Although this is mentioned in the title of Figure 2 of the article, I would recommend adding this information in the package documentation (if not done already, I might have missed it).
- For some functions (e.g. rvn_annual_peak), the units of the related arguments are mentioned in the detail section. It is always easier for users to find the required unit beside the related argument. I would suggest doing so in future versions of the package.
- I noticed that for some functions, time series must be provided at a daily time step. I thought that the Raven hydrologic framework could run at multiple time steps. Again, I might have missed the explanation at some point. If not, I would suggest adding a warning somewhere to use the time step required by RavenR/Raven.

References

Addor, N., Newman, A. J., Mizukami, N., and Clark, M. P.: The CAMELS data set:

catchment attributes and meteorology for large-sample studies, Hydrol. Earth Syst. Sci., 21, 5293–5313, https://doi.org/10.5194/hess-21-5293-2017, 2017.

Andrews, F. T. and Guillaume, J. H.: hydromad: Hydrological Model Assessment and Development, available at: http://hydromad.catchment.org/ R package version 0.9-26, 2018

Coxon, G., Freer, J., Lane, R., Dunne, T., Knoben, W. J. M., Howden, N. J. K., Quinn, N., Wagener, T., and Woods, R.: DECIPHeR v1: Dynamic fluxEs and ConnectIvity for Predictions of HydRology, Geosci. Model Dev., 12, 2285–2306, https://doi.org/10.5194/gmd-12-2285-2019, 2019.

Dal Molin, M., Fenicia, F., and Kavetski, D.: SuperflexPy: the flexible language of hydrological modelling, https://superflexpy.readthedocs.io/en/latest/index.html version 1.3.1, 2021