Review for R journal-2022-1871

RCTS- Robust Clustering of TimeSeries with Interactive Fixed Effects

Thank you for submitting "hydrotoolbox (v 1.0.1), a Package for Hydrometeorological Data Management" to R Journal. The authors of this paper present `hydrotoolbox`, an R package that provides a general framework to efficiently store and manipulate various types of hydrometeorological series. The paper discusses an important topic because accurate and efficient handling of vast and heterogeneous amounts of hydrometeorological data is a critical preprocessing step to increasing data confidence and creating coherent datasets as required to realize the full potential of the available information. However, there are a few places in the paper where clarification is required. With some changes, revisions (to both the paper and the package implementation), and additional explanation, this paper could still be significant.

1. Because there are numerous R packages for managing hydrological data, I recommend that the authors summarize the main contribution and novelty (the unique features) of the techniques/functionalities available in this R package implementation in a few sentences in the introduction. Currently, the exposition appears to be lacking.

2. In order to have a proper understanding of the classes and subclasses, it is better to provide some examples in the package help pages of `hydromet-class`, `hydromet_station-class`, and `hydromet_compact-class`.

3. Page 3: *"To make the package more user-friendly, all the methods for manipulating the objects follow the syntax hm_accion"*

I think this should be corrected as **hm_action**

4. It's hard to follow the description regarding Table 1. *"Hydrometeorological variables supported in the hydromet_station class (8 out of 37)"*. How can I get access to all the 37 variables? Is this number 37 related to a specific station? The help page of `getSlots` only provides a few examples with dummy data.

5. Page 3 : *"The second method, `hm_mutate()`, allows the use of any external (or internal) package function to manipulate the data."*

This statement is vaguely written. I would suggest adding more details or at least an example of how external package functions can be used to manipulate the data.

6. According to the Abstract of the package, the main objective of the package is to provide a "**general** framework to work with **any kind** of hydrometeorological series". But the first example starts with a very specific case using mnemos data, which seems to be specific to a regional bureau's software. To highlight the generalizability of the proposed framework, I encourage the authors to consider prominently presenting documentation covering how to use this package for "any kind of" general hydrometeorological dataset, rather than limiting the presentation to a specific case study using the "mnemos" dataset.

7. What is `bureau` in the `hm_build` function? Can we expect this input for any dataset obtained from other databases other than the SNIH of Argentina? If the function is specifically built for the datasets obtained from the SNIH of Argentina website, it is better to define the scope of the package/function both in the paper and the package help page. To help the users, it's better to add more details in the help page related to the `bureau` input rather than introduce it as a "string value".

8. Packge implemention : `hm_build`. I think the function has the ability to work with both regular and irregular time series by setting the input of the `by` variable to "none". I suggest the authors discuss these features in the manuscript as well.

9. Page 5: *"To view the slots with data..."* `hm-show` has been used. However, the `hm-show` function only displays the **'head' or 'tail'** of a specific slot. There is another function `hm_get` to extract a full series from a given slot. It's better to introduce both functions in the example to add clarity.

10. In the example, under `hm_build`, the order of the slot names is given as `"c("qd", "evap", "tair", "tmax", "tmin", "wspd")`. However, `hm_show` displays them in a different order. How do they decide the default ordering of the slots?

11. Page 6: The output generated by `guido %>% hm_report(slot_name = "qd")`

Under `missing_data`, the fifth row contains `NA`s under the two columns `first` and `last`. I suggest the authors provide some additional details and interpretations of these outputs.

11. In the `hm_build` function, if the `out_name` is dropped, then the variable names are labeled as `Medición`. It's better to set the defaults for `out_name` input to `slot_names` instead of setting all the measurements to `Medición`.

12. Page 6: *"The time series shows oscillations during the low water flow periods that can be interpreted as noise."* All figures and tables must be referenced in the text by their number. Figures 3, 4, and 5 are not cited in the text.

13. Section *"Post processing of the HBV.IANIGLA hydrological model"*: It is preferable to provide the code outputs to add clarity to the discussion. If Figure 5 is provided as a part of this section, refer to this output in the discussion guide under this section.

14. Page 9: *"As an example, the hydrotoolbox vignettes show how to combine it with two packages for downloading data from Canada (see vignette(topic ="tidyhydat_can", package = "hydrotoolbox") and vignette(topic = "tidyhydat_can", package = "hydrotoolbox")) (LaZerte and Albers, 2018; Albers, 2017)."*

The same vignette was mentioned twice. The second topic, "tidyhydat_can," I believe, should be replaced by "weathercan_can".

15. Page 10. *d. The package allows the user to easily define new methods and objects to further expand its capabilities.* I suggest adding more details here on how this is achieved.

16. Page 10: *"provide basic data for each basin, being streamflow records the cornerstone. The metadata should include the name, unique identifier, river and geographical coordinates of each stream-gauge, catchment area and elevation info. hydrotoolbox provides all the objects with this metadata."*

According to the description given under `hm_get`, metadata can also be extracted from this function. I suggest adding an example to extract metadata from a selected station both in the package help page and the manuscript.