Comment on hess-2021-609

Anonymous Referee #1

Referee comment on "Technical note: A procedure to clean, decompose, and aggregate time series" by François Ritter, Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2021-609-RC1, 2021

Overall, I find this submission not suitable for HESS due to the following major issues.

1) The title says it is a technical note, but the structure does not really fit that. According to instructions, submissions of technical notes should be only a few pages, while this submission is substantially longer. Also, in the abstract it is stated that the submission wants to propose a standardized way to pre-process time series, which is not really something that can be done in a technical note. However, to be a full research article several important parts are missing in the submission (see comments further down)

2) The article lacks references and introduction in the field of HESS. The introduction is quite general on time-series and R packages, but does not discuss what is commonly used in hydrology or earth sciences and to which extend there is a need for additional improvement of outlier detection and gap imputation within these areas. The only part that connects to the journal are the case studies, which are relevant, but as the results are not compared with other approaches or articles that have used these series before in a different way conclusions are difficult to make.

3) Generally, in environmental and earth science time series can have a large variety of different structures and the questions to be investigated vary widely. Depending on which statistical analysis is to be made, the filling of gaps or outlier detection can be more or less important. For many approaches, outliers or gaps are not a crucial problems, but can be handled intrinsically. It is, thus, not obvious that a standardized way to preprocess is desirable. Obviously, when several series are within the same academic study they should be handled similarly, but no examples of this being a real problem at present is given. Also, in earth sciences there are few situations where only single time series need to be handled. Either there are several variables observed at the same time point, which can be

used to identify if there is something wrong with the sample altogether one variable specifically, or there are nearby stations available that can be used to identify outliers or fill gaps.

Potentially the submission could be resubmitted as a purely technical note describing the R package and discussing the possible inputs to the function (k, …) and with some examples of different choices on the output. Such a submission should more intuitively describe which effect a change on k has, rather than rely on a simulation study that is difficult to relate to in practice. For example, describing how it might work for a normal and a log-normal distribution often met in earth science.

Specific comments:

There are several important parts missing or unclear:

- A new boxplot rule is suggested and motivated by that using this rule leads to far less false positives, i.e. the type I error is improved. No mention is made on the type II error, which is typically increases, when the type I error decreases. Clearly, this is not easy to study as, in a univariate time series, only outliers above a certain threshold can be detected. In this study this threshold is chosen to be very high, leading probably to situations where few (real) outliers are detected. This is also one of the reasons outlier detection methods flag rather many observations as outlies, giving the user the possibility to doublecheck the correctness of those and keep the ones that seem reasonable. In the recommendations it is stated that the value of k=0.6 will minimize the type I and type II errors, but it is very unclear how this determined and generally it is not possible to minimize type I and type II errors at the same time.
- It is not discussed which definition of outlier is used in this context, and especially it would be important to define outliers in highly skewed distributions and how it would be possible to distinguish them from observations that belong to the distribution.
- In the case studies, outliers are introduced and can be identified with the proposed method, but the outliers are completely unrealistic and could be identified by visual inspection only, no advanced methods are needed.
- The breakdown points of the outlier detection methods are not given.
- As a technical note on an R package the code should be made available, e.g. on GitHub or similar, as this is a major part of the submission.

- It is rather unclear how well the suggested values of 3.8 and 9.4 work in practice as they are the median of values achieved in the simulation. This means probably that these values work considerably worse for some specific distribution. No discussion is made about this.
- It is rather unclear how the value of k are determined. Are simulations in Figure 1a-1c made for several sample sizes and their medians are shown in panel d?
- In many cases in environmental and earth science it is common to work with e.g. log-transformed values (or models that use a log-link) to account for single high values in skewed distributions. This would make it unnecessary to develop outlier detection methods for skewed distribution. Instead, conventional methods can be used on the symmetric transformed values. Has typical handling of skewed distributions in earth science been studied? Is there a need to find outliers in skewed distributions?
- For comparison between outlier detection method 600 distributions were selected to give the same weight to different types of distributions. For determining the value of alpha and k all 9702 distribution are used. It is unclear why.
- It is also not clear how the 9702 distributions are defined and how they are chosen. At one place, a reference to the supplementary is given, but there is no info on distributions in the supplementary.
- At least one of the case studies has a seasonal pattern, which would allow a comparison to STL or STLplus
- It is argumented that STLplus has severe disadvantages compared to the proposed method. For example, it is said that the trend modelled with loess needs to be parametrized. No reference is given and it is unclear what is meant by this, as loess is a non-parametric regression methods and does not need a parametrization.

Comments on structure

- Referencing within the article is not clear. Often Figures are referenced already in the methods description, making the reading difficult. A better separation of method and results would be helpful
- Also sections called context and method are not clearly divided.
- References to the use of outlier detection methods in earth sciences are missing.