
Anonymous Referee #2

Received and published: 5 October 2019

The authors have done tremendous work in compiling a data set of global river water levels based on radar satellite altimetry from Jason-2 and Envisat. This data set also includes several arctic rivers, which are absent in other altimetry services. The paper is generally well written and easy to read. However, the sections regarding the evaluation of the data sets need some clarifications. I hope that the authors find the comments to be constructive.

General comments Section 3.3-3.4 describing the evaluation of the data set needs to be presented more clearly. These sections contain a lot of numbers in the text, which makes it difficult to read. I recommend putting the summary statistics in tables instead.
I find figure 3 confusing. An improvement could be to have a column of plots for each summary statistics.

Since this paper describes a data product I would expect to have a description of a data file/product. The description could just be put in an appendix. Please also expand section 4. Hence, add some general information about the product, and the map webpage.

On the interactive map, the individual time series are shown with error bars. It is however not clear what these error bars represent. This is important for the user when applying the data. Please make this clear in the paper. Please also add error bars to the time series in figure 2.

Specific comments:

p2,l9-10: What is meant by this sentence "Newer radar altimeter missions like Sentinel-3 are improving the contemporary record with features like automated processing."

p2,l28: "be accurately be measured" -> "be accurately measured"

p5, l27-28: Maybe the authors could add approximately dates for the winter period

p6, l7-8: "Analyses showed that VS-stream gage distance was often not an accurate predictor of height anomaly differences". please, clarify/comment on this statement. Why should the satellite data have a better fit for a station farther away? Why not look at the resemblance among the gauges along the river, this will give you an idea.

p6,l25-28: The description of your qualitative evaluation is a bit vague. And in the result section, you do not present any summary measure. Hence, what are the limits for getting grades A, B, C or D? An alternative measure could be to evaluate the along-track height variation. Please expand your description.

p9, l25: Please, clarify what is meant by this: "We suggest that individual VS data point error be estimated as the STDE of the time series they are a component of."
Figure 1: The red color in the two evaluations is a bit similar. Maybe make the colors more different or make two figures.

Figure 2: Please add error bars and describe what they represent.

Figure 3: Reorganize for clarity. Make the font of the axis and labels larger.