

We would like to thank the reviewer for the time and effort they have put into reviewing our manuscript, as well as for the helpful comments they have provided. Reviewer comments are shown below in italics, and our response in plain type. We have numbered and labeled the reviewer's comments, "General Comment 1", etc.

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 comment 1

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.

Response:

We agree that a table will be useful for aiding readers in understanding our validation process and results. We have added table 1 and referenced it in the beginning of section 3.3. We carefully reviewed the text and found that all of the sentences are important to the paper, so have kept this section as is. We think having the Table to refer to will aid in reading these sections as well.

General comment 2

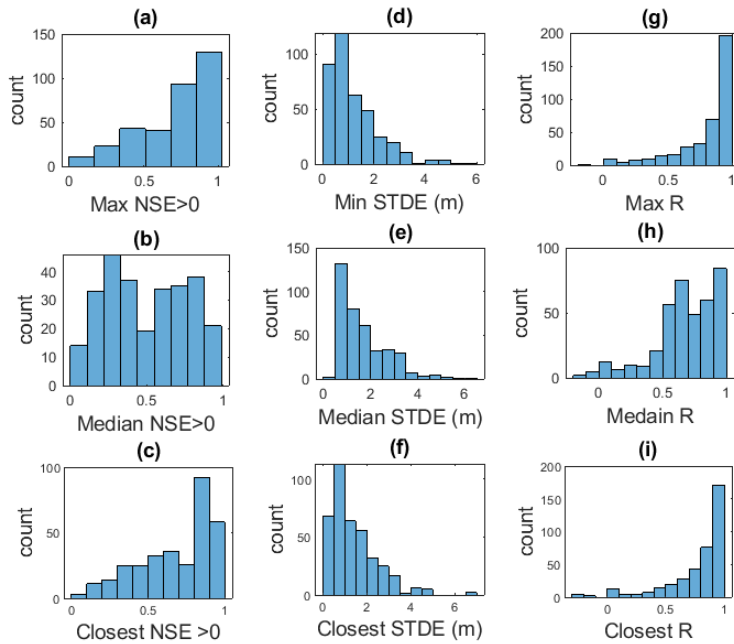
I find figure 3 confusing. An improvement could be to have a column of plots for each summary statistics

Response:

Based on the reviewer's recommendation, we have updated figure 3 so that each fit statistic is in its own column. In order to make the spacing of the figure less awkward, we removed NSE plots that included values <0 . We have updated the text in section 3.3 to reflect the new figure labels. We have also updated the figure caption to describe the amount of data missing in figures 3a-c. It reads as follows:

Figure 3 Virtual Station fit statistics computed with all available evaluation gages located in the same river and closest comparison. Please note that NSE values are plotted here only when greater than 0 to enable readers to more easily see the majority of the data. 12.2%, 28.8% and 17.2% of the total data are not shown in panel a, b, and c respectively. Panel (a) histogram of the max NSE >0 at each VS in the dataset, Panel (b) histogram of the median NSE >0 at each VS in the dataset, Panel (c) histogram of closest NSE >0 , Panel (d) A histogram of the minimum STDE in the dataset, Panel (e) A histogram of the median STDE all the VSs in the dataset, Panel

(f) histogram of closest STDE, Panel (g) histogram of the max R at each VS in the dataset, Panel (h) histogram of the median R at each VS in the dataset, Panel (i) histogram of closest R.



General comment 3

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.

Response:

We added an appendix with a file description and updated section 4 to read as follows:

GRRATS (DOI 10.5067/PSGRA-SA2V1) is available at https://podaac.jpl.nasa.gov/dataset/PRESWOT_HYDRO_GRRATS_L2_VIRTUAL_STATION_HEIGHTS_V1 for non-commercial use only (Durand et al., 2016). Data are provided in NETCDF format. For a file content description please see Appendix A. An interactive map of the data is located at <http://research.bpcrc.osu.edu/grrats/>. This tool is intended for exploration only, and may not reflect the most up-to-date version of the data. As with figure 2, error bars represent the range of the values that were averaged to generate each data point (does not include filtered data points).

General comment 4

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.

Response:

We have added error bars to figure 2 and described them in the figure 2 caption, the text of section 3.2 and the text of section 4 with the following description:

Error bars represent the range of the values that were averaged to generate each data point (does not include filtered data points). Data necessary to compute error bars are a part of the data product.

We updated the figure 4 caption with the following to explain the lack of error bars:

GRRATS error bars are not shown to improve readability.

Specific comment 1

p2,19-10: What is meant by this sentence " Newer radar altimeter missions like Sentinel-3 are improving the contemporary record with features like automated processing. "

Response:

Updated to read as follows:

"Newer radar altimeter missions like Sentinel-3 are improving the contemporary record with features like automated processing, alleviating the need for retracking and other post processing to generate useful measurements."

Specific comment 2

p2,128: "be accurately be measured" -> "be accurately measured"

Response:

Changed as requested.

Specific comment 3

p5, 127-28: Maybe the authors could add approximately dates for the winter period

Response:

Updated to read as follows:

“If ice breakup data were not available, we applied broad date limits regionally, using observations from the Pavelsky and Smith (2004) study of Arctic river ice breakup timing. Breakup dates range from late September to early June.”

Specific comment 4

p6, 17-8: "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.

Response:

We updated this section to read as follows:

“Analyses showed that VS-stream gage distance was often not an accurate predictor of height anomaly differences. This is likely due to the hydraulics (width, nearby dams, confluences) of a more distant gage being more similar to the location of the VS than the most proximal gage.”

We believe that a detailed study of river-by-river comparison of in situ gage observations is out-of-scope for this manuscript.

Specific comment 5

p6,125-28: 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 alongtrack height variation. Please expand your description.

Response:

We added the following to this section:

“Letter grades are take in to consideration all of these criteria, but in general, VSs with an A rating would have 1 or fewer obvious outliers per year, no more than 2 cycles filtered out per year, and will fit nicely above VS downriver and below VS upriver. A D rating might be applied to a VS with 3 or more outliers per year, 5 or more cycles missing per year, and might fall below VS downriver from it, and above VS upriver from it.”

Additionally, we added a table with the letter grade results and referenced it in the results section (table2).

P7, 16-7 “We gave the remaining 27 rivers qualitative letter grades based on number of missing data points, obvious outliers, and agreement with nearby stations. These grades are included with the data for end users (Table 2). The majority of rivers evaluated this way fall into the B or C category (~61%), with only ~15% getting an A rating”

Specific comment 6

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."

Response:

Updated to the following for clarity:

"While we do not provide error estimates at the individual data point level, we suggest that individual VS data point error be treated as the STDE of the time series they are a component of."

Specific comment 7

Figure 1: The red color in the two evaluations is a bit similar. Maybe make the colors more different or make two figures.

Response:

This is a good point: the colors are similar. However, note that each river is evaluated entirely qualitatively or quantitatively. Also note that the qualitative (cool) and quantitative (warm) color maps are easily distinguished. Thus, the similarity of the colors does not affect readers' ability to interpret the figure. To prevent confusion, we have added the following to the figure caption:

"Each river is evaluated using only one of these methods."

Specific comment 8

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

Response:

Error bars are added and described in the figure caption, as well as the main text.

Specific comment 9

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

Response:

Figure reorganized based on the reviewer's recommendation. Font size increased.