Author Responses to Editor Comments

1) The data on the repository: While the ESSD paper provides the opportunity to describe the data, measurements and field sites in much detail, the information on the repository should be sufficient alone to work with the data without the manuscript and needs some improvement.

a. First of all, the description is simply a copy of the paper abstract. This is confusing because some of the wording does not fit. (“this article”, “this paper”)

This has now been addressed, and does not refer to the dataset as “this paper” or “article.”

b. While the general abstract is suited to give an overview of the intentions behind the measurements, there should be a readme file alongside the data,

(i) Listing the sites and the respective measurements, probably also including the map from the paper.

(ii) A brief description of sampling methods and data processing.

(iii) Describing the folder structure including the naming of the files and also of the columns within the files.

A readme file has been created and added to the data repository. All information requested above is now included in this readme file.

2) The overview table that Referee #2 requested: I also think this is a very valuable asset for the users of the dataset, to have the information which of the measurements is available for which time periods.

The sentences you added for the streamflow data just indicate that there are gaps. For modellers for example, the information when there are meteorological and streamflow data in the same area available so that they can judge the length of consistent time series for the models would be incredibly valuable. A suggestion would be to structure the table additionally in a way so that it is visible which of the datasets belong to the same catchment, without looking up the station names.

We have decided that a figure showing all gaps in the datasets would be more effective than a lengthy table. We have added these figures as a supplement so it can be found as needed by the user. These figures are also referenced in the paper in appropriate locations.

Also, please indicate which is the period you are referring to within this paper and the referenced dataset. If stations are still running you can indicate that with a symbol or something, but the basis for the manuscript is a definite dataset.

The period of operation for all stations is listed in Table 1 and 3. Their operational status has been updated in the table with an asterisk as an indicator and has been reflected upon in the caption for both tables.

The current status of data acquisition is addressed in lines 226-229:

> *The supplemental figures S1 and S2 indicate missing data beyond 2020, however due to travel restrictions over the past two years we have not been able to collect the up-to-date measurements. Our intent is to sustain these hydrological and meteorological measurements into the future once travel to the region is possible again.*

3) Uncertainty/errors of discharge measurements: Although Referee #1 did not insist on the rating curves, a measure of uncertainty and a discussion about this is still needed in the data and manuscript.

Did you use the rating curve for all the measurements in the dataset? Did the rating curves change? Based on the quality of the rating curve and the uncertainty in the water level measurements you can give an uncertainty range for the discharge measurements. If the curves changed and you did not update older data, then the discharge values would have a different basis for the uncertainty range as well. So please add this information.

Uncertainty in discharge measurements is not directly calculated in this paper for a variety of reasons. We have however added some statements in the paper to address the possible uncertainties and have added some discussion of the range of uncertainty in the rating curves. We have also added citations that support our stance of providing R^2 and error percentages for rating curves, while also describing the fluid nature of the stream sites. Our citations also provide evidence that the errors we are seeing in rating curves is typical for hydrological data when using stage-discharge rating curves.

This has now been fully addressed in Lines 417-440:

> *Many factors influence the uncertainty of discharge measurements, especially in smaller, turbid streams (McMillan et al., 2012). First, the Solinst levelogger instruments used have accuracies and resolutions provided by the company as a percentage of FS as indicated above. Second, we estimate a ± 2 mm variation in water stage due to the turbulent nature of the streams in the region. These streams have a variety of bed surfaces, ranging from weedy to rocky, and flat to rolling, which are constantly modified by high flows during each year. Each stage and discharge field measurement are influenced by these small to large changes in the stream bed and the timing and location at which the discharge is measured. Finally, stage-discharge rating curves inherently introduce variable amounts of error depending on high or low flows, and how frequently they are updated. Stage-discharge rating curves at all stations were established and verified by conducting discharge measurements using the velocity area method. Prior to publishing this dataset, one updated rating curve was fitted for all of the data at each site, helping standardize the measurements and uncertainty in the data. All rating curves developed from discharge measurements in the field are assessed by their fit and significance to a quadratic function. All rating curves display R2 values above 0.85, except for Cuchillacocha, which provides an R2 value of 0.71. This slightly lower R2*

value at Cuchillacocha is likely due to the under-representation of high flows in the field measurements, and overall lower flow volumes recorded at this site. Uncertainties (using standard error at 95% uncertainty intervals) were calculated for all rating curves, indicating average error ranging from ± 3–20% for low flows, and ± 4–70% during high flows. These error values are standard when using a stage-discharge rating curve as described in Kiang et al. (2018), and McMillan et al. (2012). Rating curves can be further constrained, and their variation better understood with the addition of more high flow discharge point measurements (McMillan et al., 2010; Coxon et el., 2014). All rating curves are available upon request.

If you used one rating curve for and backcorrected all water level values, it would be also good to put the curves for the different stations as graphs in the appendix and the data of the rating curve measurements in the repository.

The other authors and I have again discussed the request for rating curves to be included in the paper as a supplement. We would greatly prefer to leave the rating curves as available upon request because of the fluid nature of these equations. As indicated in the newly added text, each station has one rating curve, and is "backcorrected" to all water level values in each time series and is fully up to date with all recent data in each time series.