

Comment on essd-2021-215

Anonymous Referee #2

Referee comment on "High-temporal-resolution hydrometeorological data collected in the tropical Cordillera Blanca, Peru (2004–2020)" by Emilio I. Mateo et al., Earth Syst. Sci. Data Discuss., <https://doi.org/10.5194/essd-2021-215-RC2>, 2021

In this paper, the author's present a hydrometeorological data set with high temporal resolution in the tropical Andes of Peru. This region is considered data scarce and hydrometeorological data has -if available at all- usually a daily time resolution. They authors describe the study area and it's history and development with respect to hydrological and meteorological data and research. The presented data set is put into this context and the difficulties with collecting data in such remote regions are addressed. Therefore, the data set presented by the authors is unique and potentially very useful for hydrological and meteorological studies in this region. The manuscript is well written and the data and their collection are described adequately. However, I have some minor to major comments regarding the manuscript and the data itself that should be addressed before final publication.

General comments:

- The data set presented by the authors is very heterogeneous in terms of instrumentation, temporal resolution and time series length. This is of course due to the (as described by the authors) adverse field conditions in this region. Nevertheless, I suggest that the authors add a section (or at least some information) about the data processing/quality control (QC), e.g. checks for physical limits, outliers, inhomogeneity (caused by changes in instrumentation among others). Information in the provided data point to some form of correction or QC applied, e.g. "bad or missing data" at Quilcayhuanca, "RH" and "RH corrected", etc. which undermines the necessity for a section where (pre-)processing and/or QC are explained.

- Other data in the region are mentioned in the manuscript, e.g. the gauging station La Balsa (also in Fig. 1), temperature and precipitation data from SENAMHI or the meteorological data from the UNASAM. Since these are not included in the data set presented in this article it would be desirable to provide information where these records can be obtained from.

- Since the covered time period and temporal resolution of these data differ, I suggest that the authors add an overview (figure and/or table) of data availability for each site/sub catchment, data gaps and/or percentage of data missing, temporal resolution, etc.

- An overview of the measured parameters in form of a table (similar to table 2) where these parameters are described with used abbreviations, descriptions, units and type of sampling (Average, Sum, Min, Max, Sample,) would be helpful.

- An explanation and discussion about errors and the problems gathering data in this remote area is given, especially with respect to the unshielded Lascar Temperature Loggers. However, this particular section (4.1) appears a bit misplaced under "Data Availability". I suggest that you move this to section 3 where the data are explained (or even to a new section on QC as suggested above)

- Could you comment on and/or discuss changes and validity in the rating curves over time as well as a quantification of uncertainty of these rating curves. Furthermore, a figure showing one or all rating curve(s) would be interesting.

- I assume that the authors have a long experience in field work and data recording and analysis in this region. Therefore, it would be helpful if you could give an estimation/assessment for which kind of research questions the data are suitable (e.g. could these data be used for hydrological modeling?) and what kind of problems might be expected when working with these data in terms of data quality, reliability, etc.

Comments regarding the data set:

- I noticed a diurnal cycle in the discharge data, is this related to water/air temperature changes? Or does this have other reasons?

- For the sake of clarity, I would suggest that the authors use a consistent naming for the time information and time zone in the data. This is very heterogeneous right now (e.g. "Date Time GMT-05:00", "timestamp", "time", etc.).

- Try to avoid special characters in the data headers and use plain ASCII.

- The location of the stations (Lat/Lon, UTM or similar) should also be provided.

Specific manuscript comments:

Figure 1: It's not coherent to display the gauge at La Balsa with the same symbol as the other gauging stations since this data is provided here.

- I. 211 Isn't it hornfels?

- I. 229 To my knowledge, at least some of the precipitation data from SENAMHI is/was recorded twice a day (7:00 and 19:00). Can you confirm this?

- I. 229 ff. Do you know if this network is still operational? And if so, is there a reason why this is not included in your data set?

- I. 236 ff. More examples for recent meteorological measurements in the region are the meteorological stations from <http://agroclim-huaraz.info> and the Micro Rain Radar at the UNASAM in Huaraz (<https://doi.org/10.3390/rs11242985>)

- Table 1 What does "Lascar Error Adjustment" mean or refer to? This is not explained in the manuscript.

- I. 266 ff. and I. 283 ff. Could you provide a comment or discussion as to how these changes in instrumentation influenced the homogeneity of these data (c.f. general comments above)?

- Figure 4: Showing all Lascar Sensors in Figure 4 is not very meaningful, since the instruments cannot be seen on the relatively small pictures. One example of such a Lascar sensor site (e.g. attached to a weather station) would be sufficient from my point of view. Or even leave this figure out completely.

- Figure 6: Units (m^3/s) missing on the graphs. For the sake of comparability of the different gauge catchments you could also consider plotting the specific discharges instead.