Earth Syst. Sci. Data Discuss.,
https://doi.org/10.5194/essd-2019-232-RC2, 2020

Earth System Science Data Discussions

Interactive comment on "GloFAS-ERA5 operational global river discharge reanalysis 1979–present" *by* Shaun Harrigan et al.

Anonymous Referee #2

Received and published: 9 April 2020

The paper entitled "GloFAS-ERA5 operational global river discharge reanalysis 1979-present" presented by Harrigan et al., describes re-analysis driven global river discharge simulations that are updated in near real time and distributed through the Copernicus Climate Change Service Climate Data Store. Overall, the paper is well written and provides the reader with an overview on the methods used for data production, file formats and the performance of the data set.

Given, that this paper is a data-descriptor and neither a model documentation nor a research article there is little to criticize. Nonetheless, some aspects of the paper would benefit from additional information. My main points are summarized below:

1. Terminology: The data product presented is referred to as "reanalysis". Although the runoff data used to drive lisflood stem from a reanalysis, the presented data product

is not an integral part of ERA5. In addition, observational discharge data are only used for calibrating lisflood, but are (to my understanding) not assimilated through a state-updating procedure. Given that the term reanalysis is often associated with state updating, I would find a clarification of the chosen terminology helpful in order to avoid confusion about the nature of the presented data set.

2. Transparency of the data production process: Although the paper does a good job in summarizing the workflow resulting in the presented data set, the amount of information presented is not sufficient to replicate the data. While I acknowledge that a description of ERA5 or Lisflood are beyond the scope of the paper, there are a number of essential technical steps that are not described. Open questions include, but are not limited to, (i) how was ERA5 output disaggregated to the finer resolution, (ii) how was lisflood calibrated (are the data used for validation independent of the data used for calibration), (iii) what does it mean that reservoirs are included (e.g. is management also simulated), etc. I realize that some of these questions are also treated in other publications but for a user of the data set a comprehensive overview with more details would be essential to fully understand the capabilities (and limitations) of the data.

3. The output variable (discharge) is "Volume rate of water flow, including sediments, . . .". While I acknowledge that this is likely the variable of interest for flood forecasting, I would appreciate if the volume (or mass) of pure H2O could also be made available (if this does not differ significantly, then a statement explaining this might be useful).

4. What is the time resolution of the observations used for validation? I assume daily, but this was not stated explicitly.

5. Stations used for evaluation come "predominantly" from the GRDC. This is not transparent at all and hinders reproducibility of the study. I assume that some of the data cannot be re-distributed, but an overview (e.g. supplementary table) on the considered stations including some key properties (geolocation, river and station names, data-provider, catchment area, . . .) foster reproducibility of the results.

6. If there is more than one station per grid-cell only one station is selected. This is OK. However, what is the criterion to select a particular station (random, expert judgment, catchment size,. . .)?

7. I personally would find extended global summaries (in addition to medians and IQRs) of the performance metrics useful (e.g. tables with percentiles, or empirical cumulative distribution functions).

8. The performance assessment focusses predominantly on the skill of the full time series at daily resolution. For some users information focussing on different modes of variability (e.g. seasonal cycle, anomalies of the seasonal cycle, year-to-year fluctuations) would be also of great interest.

9. Accessibility of the data product. I am aware of and support the effort of the Copernicus Climate Change Service but I don't have an account for this at the time being. I am also reluctant to create "random" accounts that I need to keep track of if not really needed. Given the fact that the data are produced by one of the world leading institutions for global weather data (ECMWF) and hare hosted on the Copernicus platform, I assume that the data format will be state of the art.

Interactive comment on Earth Syst. Sci. Data Discuss., https://doi.org/10.5194/essd-2019-232, 2020.