

Response to Reviewer #1 (W. Grabs)

We would like to thank W. Grabs for the open comments, which will clearly help to improve the presented manuscript. In the following we provide point by point replies to the reviewer's suggestions. For the sake of clarity, we first repeat the reviewer's comments in blue before including our response.

10 be consistent with number of station in first paper (35.000 stations)

Thanks for noting, we will ensure consistency in the revised manuscript

25 Although monitoring data are increasingly posted on national and regional websites, this does not relate to entire time series of data. In many cases, data are presented in graphical format only to prevent potential misuse

Thank you for noting that not all regional/national streamflow archive are available for download. Nevertheless, we think that it is noteworthy to emphasise that a substantial number of time series is now publicly accessible. This is documented in Part 1 of this paper series.

p 2, para 5 The data contained in different data bases also stem from different sources and often unknown quality control procedures

Thank you for noting, we will highlight this issue in the revised manuscript

2, 25 typ-o: "In cases where access. . ."

Thank you for noting, the paper will be carefully checked for typos before re-submission.

3, 15 A critical question is whether this data set is a closed data set or whether there is a strategy to keep the data set alive through updates In the latter case, a possible procedure could be outlined how to apply the calculation of the indices to updated time series. Possibly, a tool kit could be proposed to calculate indices on new or updated datasets

As mentioned in our response to the review for Part 1, GSIM will likely become a closed data set. The reason for this is that we do not have the resources for regular updates. However, we would welcome any initiative aiming at keeping the presented collection alive.

4,5 In its latest version (summer 2017), GRDC is no longer providing flags and continues to abstain from flagging data

Thank you for this valuable information. We will incorporate this in the revised manuscript.

4,10 Routine QC of data by GRDC (and most other data centres) is limited to plausibility checks and correction of obvious errors

Thank you for this information. We will mention this in the revised manuscript. We will also put this into context of our daily QC criteria which are also focussing on plausibility checks.

4, 10-15 It should be noted that base or reference periods are an important feature to characterize deviations from base period statistics that is vital for a number of management decisions in water management. The use of climate normal is such an example and efforts are underway to link climate normal to hydrological normals calculated for rivers. Although such normal are not globally valid, they are regionally significant

We assume this comment refers to page 5 (not 4). We are aware that indices that require reference periods are regularly used for both scientific and management purposes. We investigated the use of

such indices in our preliminary assessments, but these indices had some issues and it proved impractical to develop a “one size fits all” solution. For example, large differences in temporal coverage prevented us from finding a reference period that is applicable around the world. These issues are outlined in the manuscript.

4,15 The notion that “time series indices cannot be extended when new data becomes available” is a serious constraint that limits the utility of the data set and the approach chosen. Moreover, a closed data set is aging fast. This needs to be discussed more in the paper. There is also the danger that such a data set becomes “orphaned” and continues to be used over a long period of time even when better data set have become available, similar like the RivDis Data of UNESCO more than 20 years ago that still is being used although it contains errors and is long-since superseded

We agree that there are many challenges related to maintaining the currency of a dataset. These challenges are various, and are both bureaucratic and technical. To this end, please see our response to the comments on Part 1 for a detailed discussion related to issues with orphaned data sets. Regarding the technical challenge of updating the dataset once new data are available, the index-computation for all indices in the manuscript does not depend on previous values and we consider this to be a relatively easy task. These points will be elaborated upon in the revised manuscript.

9,20 typ-o: “. . .to a year for which. . .”

Thanks for noting.

Quality control

The indices should be divided in those that actually express data quality and those that are significant for science applications such as water balance indications etc. A transparent approach should be taken to categorize data as a result of the data quality indices, making however clear that QC procedures cannot replace the responsibility of data owners and providers who in the first place provided data to data centres. The QC indices alone are already a valuable asset of the data set provided, guiding researchers and practitioners in the selection of data sets for their purposes.

The reviewer suggests providing an overall classification into indices into time-series suitable for specific applications based on the quality of the indices. Our approach has been to provide general information on quality rather than pre-determine their possible usage. That is, we (i) check the plausibility of daily values, (ii) provide the number of time-points that were used for computing index values at each time step (e.g. each year) and (iii) employ a set of previously suggested homogeneity tests, and for evaluating these results, we point to previously suggested criteria.

We consider that the “usefulness” of data is always very context dependent (which was also pointed out by referee # 2). For example, a study focussing on climate change detection will have different data requirements than a model validation exercise. Some studies will require highly selective quality criteria even though it results in few sites, while other studies will more readily trade-off quality for gauge density or record length. Others will implement or develop more rigorous quality procedures of their own by which to assess the data rather than be restricted by the basic checks we provide.

We appreciate that the above-mentioned quality control procedure may not be suitable for selecting data for all science applications and we will discuss this in the revised manuscript. We note also, that the scope of Earth System Sciences Data is on presenting data sets and not on developing new methodologies (which would be the case for a more sophisticated classification of data quality).

Discuss minimum length of time series needed for specific applications. For climate studies, i.e. WMO recommends a minimum length of 30 years.

We will mention this in the revised manuscript

The paper should discuss the robustness of the chosen indices on time series with differing qualities. A procedure could be proposed or recommended to check robustness of indices under varying conditions

Please see our response to the previous comments. We have implemented a basic set of quality checks on the indices and in the revised manuscript. We anticipate that there are numerous opportunities for proposing or developing new methods, but these are beyond the scope of the current paper. As a recommendation of the revised manuscript, we will identify the potential for further classification of robustness of indices and the need for methods to optimally classify the quality of time series in large streamflow data bases.