

## ***Interactive comment on “The Global Streamflow Indices and Metadata Archive (GSIM) – Part 2: Quality Control, Time-series Indices and Homogeneity Assessment” by Lukas Gudmundsson et al.***

**W. Grabs (Referee)**

grabsw@gmail.com

Received and published: 18 December 2017

ESSD- 2017- 104

Review

Wolfgang E. Grabs

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

25 Although monitoring data are increasingly posted on national and regional websites,

C1

this does not relate to entire time series of data. In many cases, data are presented in graphical format only to prevent potential misuse

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

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

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

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

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

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

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the RivDis Data of UNESCO more than 20 years ago that still is being used although it contains errors and is long-since superseded

C2

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

#### 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.

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

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

---

Interactive comment on Earth Syst. Sci. Data Discuss., <https://doi.org/10.5194/essd-2017-104>, 2017.