Open Access

Earth System
Science
Data

Discussions

ESSDD

Interactive
comment

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.

Anonymous Referee #2

Received and published: 31 December 2017

General comments

I found this second, and (acknowledged) more difficult-to-do, part useful and publishable in general. But it reads less finished than the first part of the companion papers and my assessment is it still requires a bit more work on text and structure to be as influential in the end as perhaps can be on the community's use of the presented dataset. Again, of course the data quality control approach chosen, selection of indices for the dataset and the homogeneity tests chosen are subjective and pragmatic and not everyone needs to agree with it and it will never be possible to satisfy everyone's needs,

so I don't see a point in criticising individual choices and decisions made here. What I think, however, would help the paper's acceptance and impact in particular regarding the use of the quality assessments given is to bring out the assessment of the result of these better, incl. for example a nicer clearer presentation and assessment of the summary statistics of quality flags, homogeneity tests and indices (Figs 4-7 are a bit uninspiring and very simple - maybe combining maps and box plots somehow or something like that) and more comment/assessment on the patterns on the maps that will allow at least some guidance to the data usage. Partly this is can perhaps be done by just a bit of a better organisation and presentation of the material that is there, but also by a more structured discussion, also linking to metadata from part 1.

It wasn't always clear to me, which information I will find in the dataset and which are only steps of production described here. This could be make clearer overall and in some cases removing repetitions may help (first an overview paragraph, then reading the same thing again later in the individual steps is unnecessary). Section 6 in particular is a bit random in what is covered and highlighted and thus presents not a strong conclusion. I suggest to give this some thought and better organise and bring out the highlights. This is perhaps an editorial decision, but I find sections 3.2.1 to 3.2.15 not a correct nor a very useful text format as they contain neither a list nor paragraphed text and the use of the many subheadings is an unnecessary waste of space. Tables have been invented to reduce repetitive headings/descriptors. So why not a table with the name, abbreviation, unit, resolution and definition followed then by the more descriptive text paragraphs providing additional info. Alternatively, just a series of paragraphs always starting in similarly structured sentences would do as well. The selection of example studies is a bit random. Is it really necessary? This paper is supposed to describe the dataset created and it may be enough to use some of these references exemplary in a summary-motivation for the selection of indices or rather in a discussion on possible use of the dataset information.

Generally, the manuscript will also require another careful proofread to cor-

rect several typos and some inconsistent formatting (italic or quotes for dataset flags/categories/...confusing!), some terminology (examples below), tenses (e.g. what 'was' done to the data - use part tense consistently - and what 'is' provided in the dataset - use present tense consistently) and notation (examples below - not an exhaustive list). In particular: see Journal's Manuscript guidelines for symbols, exponents and units (e.g change sec to s and make format exponents as superscripts (most figures))

Selected specific comments

3.3.1. Isn't 'reference period' the more commonly used terminology (instead of base period)?

3.1.3 Requirements for number of valid data to estimate a 'reliable index'. These are very subjective, which I know is a necessary pragmatic solution. However, it creates a bias to less 'reliable' indices in climates streamflow gauging isn't possible or meaningful part of the year (seasonally dry climates and cold climates). This needs to be discussed.

Harmonize the current mix of Q/C, Qc, qc, and define what is meant by it initially as common definitions vary.

27 typo: appraise

Fig. 1 Typo in legend: 'equal', change axis label sec –> s and proper superscript (also in Fig. 3). I suggest to zoom in more as like this there is actually little to see.

Fig. 3 Since a) daily data won't be provided by the dataset anyway if I understand correctly and b) one cannot see anything in the daily graph anyway, I suggest to remove it from Fig.3

Similar to part 1, but perhaps even more so here, are the global maps. At that size and resolution it's impossible to see anything and not enough credit is given to why these difference may not simply reflect a lack of data but an inherent feature of the variable

covered and which may not be present or measurable (see earlier comment). When I zoom in I see grids rather than station location points, but didn't read anything on gridding the point information. This is not acceptable and needs to be changed or very clearly described.

_____

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