Response to Reviewer #2

We would like to thank Reviewer #2 for the fair and thoughtful comments which will help to improve the presented manuscript. In the following we provide point by point replies to the reviewer's comments. For the sake of clarity, we first repeat the reviewer's comments in blue before including our response.

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.

We would like to thank Reviewer #1 for this encouraging verdict of our approach.

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.

We would like to thank Reviewer #2 for pointing out that the presentation of the quality control and homogeneity testing results could be improved. In the revised version the section concerned with the presentation of these results will be re-structured, aiming at a more in-depth presentation of the available information. To this end we will also consider creating more informative graphics (e.g. global time series of number of available stations, histograms/boxplots of time series lengths, regional summaries of data quality). In addition, we may include a more detailed discussion of how the information contained in the GSIM data can be used for data-selection. Nevertheless, we would like to emphasise that the scope of Earth System Sciences Data is the documentation of data and not their analysis.

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).

Thank you for noting that it was not always clear which information will be available in the data-stet. To mitigate this issue, we will consider including a new section "Description of data-files", in which we provide the necessary details.

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.

Thanks for emphasising the need to revisit the structure and the organisation of the text. We will consider this while revising the article. One topik that we will likely highlight, is the need for improved methods of quality checking of streamflow data. To this end we will revisit the approach used in this paper and point to avenues for future research in this field.

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.

We will likely reformat the content of theses sections into two tables as suggested by Referee #2. Any advise by the editor would be very welcome.

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.

In our evaluation, pointers to studies that make use of the considered indices do provide valuable background information to potential users. Nevertheless, we will consider moving this information into an appendix.

Generally, the manuscript will also require another careful proofread to correct 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))

The manuscript will be revised accordingly.

Selected specific comments

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

Depending on the body of literature considered, both terminologies are used. We will clarify this in the revised manuscript.

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.

The criteria mentioned in the paper are based on recommendations of ECA&D13. We acknowledge the caveat mentioned by Reviewer #2 and we will discuss this accordingly.

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

Thanks for noting. We will revise accordingly.

27 typo: appraise

Thanks for noting.

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.

The figure label will be revised. Note that the aim of this figure is to provide the "full view" on a time series. We will consider additional "zoomed" panel for the individual flags.

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

The figure will be revised accordingly.

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.

Please note that presenting >30000 locations on a world-map will always bee a compromise between providing the full spatial picture and the loss of some detail. If individual points should remain visible they have to have a certain size and will hence overlap in regions with high station density. We do, however acknowledge that the resolution of the figures is not optimal, and we will make sure to provide sufficiently resolved files for final production. In addition, we may opt for providing a set of regional figures in higher resolution in a supplement.