Earth System

Open Access

Science

Discussions

Data

ESSDD

Interactive
comment

Interactive comment on "The Global Streamflow Indices and Metadata Archive (GSIM) – Part 1: The production of daily streamflow archive and metadata" *by* Hong Xuan Do et al.

Hong Xuan Do et al.

hong.do@adelaide.edu.au

Received and published: 12 January 2018

We thank the reviewer for taking their time to provide us constructive comments, which have been included below as *italic text*, followed by our response as normal, indented text.

Specific comments

20 Insert references on global runoff estimation using global discharge data sets

> Thank you for your recommendation, we have updated the manuscript to
> include the important application of discharge datasets with three additional

references [1],[2],[3].

[1] Fekete, B., Vörösmarty, C., and Grabs, W.: Global Composite Runoff Fields on Observed River Discharge and Simulated Water Balances/Water System Analysis Group. University of New Hampshire, and Global Runoff Data Centre. Koblenz, Federal Institute of Hydrology (BfG), Koblenz, Germany, Federal Institute of Hydrology (BfG), 2002a. 2002a.

[2] Fekete, B. M., Vörösmarty, C. J., and Grabs, W.: High-resolution fields of global runoff combining observed river discharge and simulated water balances, Global Biogeochemical Cycles, 16, 15-11-15-10, 2002b.

[3] Vörösmarty, C. J., Moore, B., Grace, A. L., Gildea, M. P., Melillo, J. M., Peterson, B. J., Rastetter, E. B., and Steudler, P. A.: Continental scale models of water balance and fluvial transport: an application to South America, Global biogeochemical cycles, 3, 241-265, 1989.

35 GRDC operating under the auspices of the UN - World Meteorological Organization (WMO)

We have included this information into the revision.

40 Update statistics from current GRDC catalogue

The revision now has updated statistics using the latest information from the GRDC (December 05 2017, which is available at ftp://ftp.bafg.de/pub/ REFERATE/GRDC/website/grdc_summary_statistics.pdf).

95 ARCTICNET is a now static database that is mirrored in the GRDC. In GRDC, stations of ARCTICNET are updated based on data deliveries to GRDC. The ARC-TICHYCOS river discharge network is hosted and operated by GRDC with currently over 500 stations that are online available.

We thank the reviewer for additional information about the status of ARC-TICNET, which we have included in the revision to provide a better overview about this data source.

Also we can see that the "GRDC" abbreviation used to represent the database we obtained from the Global Runoff Data Centre might create some confusion. We have carefully checked the Global Runoff Data Centre documentation and determined that "GRDB" is the more precise abbreviation to represent the database of 6,313 time series we obtained from the Global Runoff Data Centre (this abbreviation (GRDB) stands for "Global Runoff Data Base", as described at http://www.bafg.de/GRDC/EN/01_GRDC/13_dtbse/database_node.html). As a result, we have revised the manuscript (including tables, figures) to avoid confusion for users regarding to this database (i.e. GRDB now represents 6,313 stations of the Global Runoff Data Base that we obtained from the Global Runoff Data Centre). With the same intent, we reserve GRDC to refer only to the institution of "the Global Runoff Data Centre". The metadata file (GSIM_metadata.csv) and readme file of the database have also been adjusted to reflect this update.

Based on the most recent download of the GRDB, we suspect that, at present, the ARCTICNET data portal has not been fully integrated into GRDB since there are numerous stations available from ARCTICNET that do not appear in the GRDB (as shown in Figure 1). For this reason we treat ARCTICNET as an independent data source for GSIM but have commented in the manuscript that its future status is likely to be as a part of the GRDB.

115 Most data of the European Water Archive (EWA) hosted by GRDC are available under the GRDC data policy and are no longer restricted to the FRIEND data policy. Based on data deliveries, the EWA is updated.

[Printer-friendly version]

[Discussion paper]

Thank you for this information about the EWA data-policy and we have included it in the revision. Considering data availability, to our knowledge the EWA has not been fully integrated into GRDB. Figure 2 also demonstrates that there are numerous stations available from EWA that do not appear in the GRDB (e.g. in Spain, Italy, and Norway), which is consistent to the result of the de-duplication process in GSIM production (only 781 cases of duplication were detected). Thus we treat EWA and GRDB as two independent data sources for GSIM but have noted the status of EWA as a database hosted by GRDC in the manuscript.

140 ARCTICNET is superseded by the arctic river basin database and ARCTICHY-COS databases that are part of the GRDC database. However, the ARCTICHYCOS database is operated as a project and data are open.

Please see above two responses on this topic. We cannot establish from the latest GRDB dataset that all ARCTICNET gauges have been assimilated (Figure 1).

230 It needs to be noted that data deliveries from national official data suppliers also contain errors. GRDC is performing plausibility checks on these data sets to detect and correct errors and provide feed-back to data suppliers.

We have revised the manuscript to note that the Global Runoff Data Centre also performs quality control procedures to detect and correct errors in supplied data.

Question: Have there been some checks to detect consistency of data sets supplied by national suppliers with data sets from the same stations contained in the GRDC? It is always necessary to check for the latest available versions of databases!

In the initial stage of GSIM, we also made some checks to compare the temporal coverage of GRDB database (obtained in September 2016) and corresponding national suppliers. Generally, timeseries obtained from national suppliers representing the latest version of national streamflow databases as they were downloaded from national data portals, which have been updated regularly by national water agencies.

Table 1 provides a comparison of the first/last year of data entry between time series obtained from the Global Runoff Data Centre and national suppliers in corresponding countries. As mentioned in our manuscript, the number of data in national databases are much higher than the number of stations contained in the GRDB. In addition, many national databases have better coverage in time except for the average length of daily data in Australia, Brazil, Canada and the US. We anticipate this is likely due to strict selection criteria of these national data suppliers when transmitting their data to the Global Runoff Data Centre (i.e. only selecting data with longer periods coverage). Resulting from this information, we have decided to use national databases in preference to GRDB where available and have also included the caveat in the manuscript to ensure GSIM users are fully aware of this procedure.

280 As standardization issues are a prominent issue it's important to describe if the development of the metadata catalogue has followed standards set (and endorsed by WMO as a standard setting organization), by the Open Geospatial Consortium (OGC), using WATER ML-2
320 Is GSIM Metadata compliant with OGC standards?

We thank the reviewer for their comment on WATER ML-2 as endorsed by WMO. During development, the structure of GSIM metadata catalogue was mainly inspired by the Global Runoff Data Centre's data products (e.g.

GRDC catalogue), which have also been used as the standard for the development of a previous data product of our co-authors [4]. The WATER ML-2 is a comprehensive document regarding a standardised structure for the programming and documentation of online databases, as typically developed by national agencies. Providing this level of functionality is well beyond the capacity of our team, so that GSIM was not developed following this standard. Nonetheless, we consider that there are several opportunities for aligning our terminology with that specified by WATER ML-2 and we have endeavoured to match this aspect of WATER ML-2 as best as possible given our constraints.

[4] Gudmundsson, L. and Seneviratne, S. I.: Observational gridded runoff estimates for Europe (E-RUN version 1.0), Earth Syst. Sci. Data Discuss., 2016, 1-27, 2016.

435 ff Provide information on the status of time series. In the case of archived time-series: describe whether there are update mechanisms in place or whether some of the data are closed data sets. There is the danger to generate orphaned data sets with incomplete information (metadata) on the version and last date of such data sets. This has often created confusion as researchers worked with outdated data sets (such as the UNESCO RivDis that still is used although it is outdated since over 20 years. It is used as the data is open without restrictions but the data holdings contain errors and/or have long since been replaced or updated including error correction.

We thank the reviewer for the advice regarding to the status of original databases, which we also agreed is extremely useful to GSIM users. We have updated section 5.1 and Table 1 in the revision to mention this important information. We also added a cautionary sentence to ensure GSIM users are fully aware of possible errors of "static" stations. To avoid the danger of creating such an orphaned dataset, we have also followed the

GRDB metadata structure and included the first and last year of data entry corresponding to each station in the metadata (as described in section 4.1, step 3).

510 Explain in a more transparent manner that GSIM will provide the metadata archive and not the actual time series as a result of different data policies from database operators including national services.

We have included a section at the beginning of section 5 to emphasise that original time series cannot be made available, and thus the metadata catalogue has been developed to address this limitation. We also add some clarification in the Introduction section to ensure that data availability is discussed in a transparent manner.

Discuss in more detail existing update mechanisms of databases and an indication which data sets are closed historic archives and which are living databases that are continuously updated.

Regarding to the update mechanisms of sourced databases, we have provided more information in section 5 (as discussed above). We also revised Table 1 to clearly indicate which sources are closed historic databases and which sources are still being updated by data providers.

Corrections in tables:

Table 1 ARCTICNET is part of GRDC, in addition, GRDC hosts the ARCTIC-HYCOS database; ARCTICNET is a closed historic database.

We have updated Table 1 to clarify this information. We also clarified that (1) CHDP and GAME databases are also closed historic databases, (2)

C7

EWA has been frozen since October 2014, and (3) the other databases are being updated by the data authority.

Table 2 Example for Spain: EWA has 239 stations, GRDC 0 BUT: EWA is hosted under GRDC and data are available under GRDC data policy. These are no longer separated data bases!

As discussed in previous comments, the terminology GRDC is now no longer used to indicate data source for GSIM, and thus this terminology has been replaced in Table 2 as "GRDB". However, we still keep these databases in two separate column to indicate that EWA was not fully integrated into GRDB (see the fifth comments).

Interactive comment on Earth Syst. Sci. Data Discuss., https://doi.org/10.5194/essd-2017-103, 2017.

Figure 1. Available stations in Russia. Red dots represent ARCTICNET database (139 stations). Blue dots represents GRDB (the Global Runoff Data Base) stations with daily record greater than 10 years (102 stations). GRDB stations were plotted on top of ARCTICNET stations.
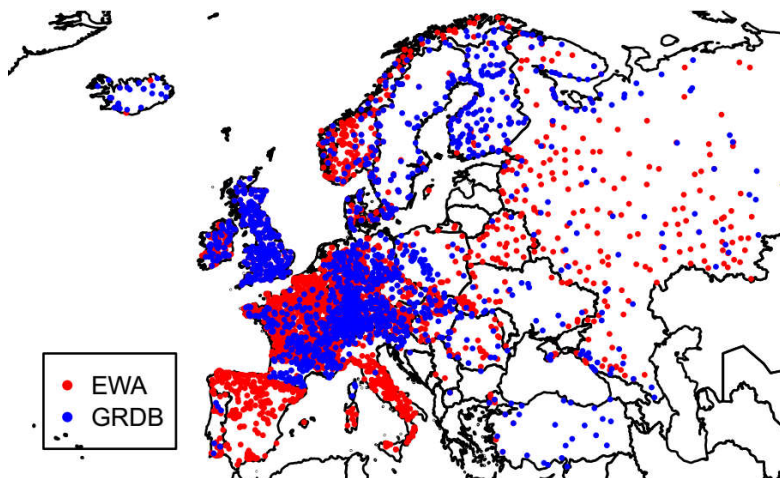
Fig. 1.

Figure 2. Availability of GRDB and EWA databases. Red dots represent EWA database (3,731 stations). Blue dots represents GRDB stations with daily record greater than 10 years (3,104 stations). GRDB stations were plotted on top of EWA stations.

Fig. 2.

Table 1. Comparison of data availability between GRDB database and national databases.

Seq	Country	National database	Data availability (GRDB)				Data availability (national database)			
			No. of station	Earliest	Latest	Average length	No. of station	Earliest	Latest	Average length
1	Australia	bom	358	1886	2012	47.24	2,941	1886	2016	31.25
2	Brazil	ana	439	1910	2010	36.48	3,313	1901	2016	29.29
3	Canada	hydat	1,029	1860	2014	45.84	6,325	1860	2015	26.97
4	India	wris	0	NA	NA	NA	318	1964	2015	30.03
5	Japan	mlit	151	1978	2003	12.42	1,029	1938	2014	22.78
6	Spain	afd	87	1977	1984	7.93	1,197	1912	2011	37.39
7	US	usgs	981	1873	2015	77.88	9,404	1880	2016	53.77

Fig. 3.