Lessons learnt from checking the quality of openly accessible river flow data worldwide

L. Crochemore, K. Isberg, R. Pimentel, L. Pineda, A. Hasan & B. Arheimer

OPEN ACCESS

Check for updates

Lessons learnt from checking the quality of openly accessible river flow data worldwide

L. Crochemore, K. Isberg, R. Pimentel*, L. Pineda[†], A. Hasan[‡] and B. Arheimer

Hydrological Research Unit, Swedish Meteorological and Hydrological Institute, Norrköping, Sweden

ABSTRACT

Advances in open data science serve large-scale model developments and, subsequently, hydroclimate services. Local river flow observations are key in hydrology but data sharing remains limited due to unclear quality, or to political, economic or infrastructure reasons. This paper provides methods for quality checking openly accessible river-flow time series. Availability, outliers, homogeneity and trends were assessed in 21 586 time series from 13 data providers worldwide. We found a decrease in data availability since the 1980s, scarce open information in southern Asia, the Middle East and North and Central Africa, and significant river-flow trends in Africa, Australia, southwest Europe and Southeast Asia. We distinguish numerical outliers from high-flow peaks, and integrate all investigated quality characteristics in a composite indicator. We stress the need to maintain existing gauging networks, and highlight opportunities in extending existing global databases, understanding drivers for trends and inhomogeneity, and in innovative acquisition methods in data-scarce regions.

1 Introduction

The increasing availability of hydro-meteorological datasets has been a crucial enabler in the race for continental and global hydrological models (Eagleson 1986, Archfield *et al.* 2015, Bierkens 2015). The current emphasis on water and climate services operating at broad scale further accentuates the relevance of such datasets (WMO 2011, Brooks 2013, Hurk *et al.* 2016, Donnelly *et al.* 2018). In hydrological modelling, datasets used throughout model delineation, parametrization, calibration, and validation, are essential and integral parts of the modelling, which will later influence model performance and may limit model applications (e.g. Arheimer *et al.* 2012, 2019, Crochemore *et al.* 2019).

With the emergence of open remote and *in situ* data, the plethora of datasets being available to hydrologists (e.g. Lehner and Döll 2004, Portmann *et al.* 2010, Yamazaki *et al.* 2014) has further enhanced opportunities for large-scale and large-sample hydrological studies (e.g. Pechlivanidis and Arheimer 2015, Siqueira *et al.* 2018, Arheimer *et al.* 2019). In hydrology, river flow is one of the most crucial variables for water resources projects, such as energy production, irrigation planning, water quality improvements or waterway transport. Moreover, river flow observations, provided that they are of reasonable quality, represent the Grail of hydrological modelling, which, for instance, provides the base for flood or drought forecasting.

River flow information has long been recorded through networks of flow gauging stations designed to support water resources planning at the regional or national scales. Observed river flows can be assessed through a range of measurement techniques, the most common ones taking advantage of water levels or velocities and based on simple established relationships between the measured variable and river flow (Sauer 2002, WMO 2008a, WMO 2010). These records are further complemented by regional and sometimes transboundary datasets collected for research purposes. At global scale, we can cite two efforts to collect river flow information from regional and national providers: the global database proposed by the Global Runoff Data Centre (GRDC; www.bafg.de/GRDC/), which is to date the largest database of river flow time series, and the Global Monthly River Discharge dataset (Vörösmarty *et al.* 1998). Both are openly available online.

Whether it is for model setup, calibration, validation (e.g. Donnelly *et al.* 2016, Beck *et al.* 2017) or statistical or extreme analyses (e.g. Blöschl *et al.* 2017, Kuentz *et al.* 2017, Gudmundsson *et al.* 2018b), a minimum record and a quality assurance of available river flow observations are always recommended (e.g. WMO 2008b). However, open datasets rarely come along with quality information. To ensure quality, visual hydrograph inspection is probably the most thorough method, which can be complemented by numerical evaluations. However, individual visual checks become strenuous when performed on large observation samples. Quality metrics can thus help identify records that require in-depth investigations. Many recent studies have made use of datasets of river flow at the global scale (e.g. Arheimer *et al.* 2019, Beck *et al.* 2013, 2017, Gudmundsson *et al.* 2018a, 2018b). These use different ways of ensuring a minimum level of trust in the river flow observations, such as removing anthropogenic influences based on land-use

---

information (e.g. Beck *et al.* 2017), removing large catchments that may be impacted by channels (e.g. Beck *et al.* 2013, 2017), removing time series containing negative values or steps (e.g. Gudmundsson *et al.* 2018a) and ensuring 5–10 years of available data (e.g. Beck *et al.* 2013, 2017).

Here, we propose a minimum of quality checks that should be done after downloading river flow data from open and readily available databases globally. We present the results from applying these methods starting with information from 64 187 stations and 36 652 river flow time series from 13 data providers worldwide. Data availability is here defined as record length and spatial distribution of available data, while quality assurance is addressed by outlier, inhomogeneity and trend detection. These were examined for 21 586 unique time series across the globe. The incentive for this work was the setup of the HYPE model at global scale (Worldwide-HYPE; Arheimer *et al.* 2019), which required the location and time series for a large number of river flow stations to define catchments, calibrate model parameters, and assess model performance. To the authors' knowledge, no database of quality indicators covering simultaneously all 13 datasets compiled in this study is currently available. Although, with similar motives, Do *et al.* (2018) and Gudmundsson *et al.* (2018a) proposed the Global Streamflow Indices and Metadata Archive, which offers a database of river flow indicators based on time series collected worldwide. The present paper aims to highlight data sources with readily available time series for downloading. It gives a temporal and spatial overview of openly available river flow data worldwide, a first quality check, as well as methods to screen potentially doubtful time series before using them in hydrological sciences. Finally, we discuss the opportunities and barriers that emerge from using open-source hydrological data.

2 Data and methods

2.1 Openly accessible data sources

The criterion used in this paper for selecting data sources was that these should be openly available and easily accessible for download (cf. Table A1 in the Appendix), while being open for use. River flow time series were first collected from two global datasets: the Global Runoff Database from the GRDC and the Global River Discharge Data (RIVDIS; Vörösmarty *et al.* 1998). National datasets were then also downloaded including Surface-Water Data from the US Geological Survey (USGS), HYDAT from the Water Survey of Canada (WSC), WISKI from the Swedish Meteorological and Hydrological Institute (SMHI), Hidroweb from the Brazilian National Water Agency (ANA), National data from the Australian Bureau of Meteorology (BOM), and Spanish river flow data from the Ecological Transition Ministry (Spain). Lastly, these global and national datasets were complemented by regional and research-based datasets including R-ArcticNet v. 4.0 from the Pan-Arctic Project Consortium (R-ArcticNet), Russian River data (NCAR-UCAR; Bodo 2000), Chinese river flow data from the China Hydrology Data Project (CHDP; Henck *et al.* 2010, 2011), the European Water Archive from GRDC – EURO-FRIEND-Water (EWA), and the GEWEX Asian Monsoon Experiment (GAME) – Tropics dataset provided by the Royal Irrigation

Department of Thailand. More datasets were found but were less readily available and therefore were not included in this paper (e.g. India-WRIS).

2.2 Procedure for downloading and selecting stations

2.2.1 Download and formatting of river flow time series

For all providers mentioned above, information from 64 187 stations and 36 652 openly accessible time series were downloaded (for links and information, please refer to Table A1 in Appendix and to Do *et al.* 2018). Depending on the providers, the time series and station metadata can exhibit different characteristics, which can impede their use and harmonization into a database. Differences include: (a) language; (b) file formats ranging from webpages, text files, Excel tables, SQL databases; (c) data structures; (d) time steps; (e) missing values identification and flagging; (f) river flow units and coordinate projections; and (g) precision of station coordinates or river flow records. Caution is thus necessary when building a common river flow database and specific treatment is needed for each provider. For the purpose of this paper, we harmonized all downloaded time series to common file format (here, R time series), data structure, missing value identification (NA), and units (m^3/s).

2.2.2 Duplicate identification and association of stations and river flow time series

Duplicate stations were present in our original list of 36 652 time series. Judgement and numerical analysis were combined in order to identify duplicate stations in the 13 river flow datasets. Potential duplicates were first numerically detected based on the proximity between station coordinates. These potential duplicates were then manually discarded or confirmed based on visual inspections and station metadata such as station and river name. In this process, only stations with coordinates that were not significantly different (based on the third decimal point) from one another and stations that had a common station name were eliminated. Stations that were very close or on the same river but that had different station names were not considered as duplicate stations.

When several time series were available from duplicate stations for one location, a "best" time series was identified based on (a) the availability of daily or monthly data, (b) the length of the available time series, and (c) how recent the available data were. The procedure was the following: (1) if daily data were available, monthly data were discarded, (2) if several daily time series were available, an indicator was calculated, giving equal weight to the length (number of days with data normalized by the entire 1961–2019 time period) and to the latest date with data (number of years between the latest date and 1961, normalized by the number of years between 1961 and 2019). The time series with the highest indicator value was selected as "best" time series.

Figure 1 shows the resulting 21 586 time series across the globe after duplicate removal. Global and transboundary datasets underwent the largest reductions (Table 1). This was expected because global and transboundary datasets often gather time series from national datasets, and therefore
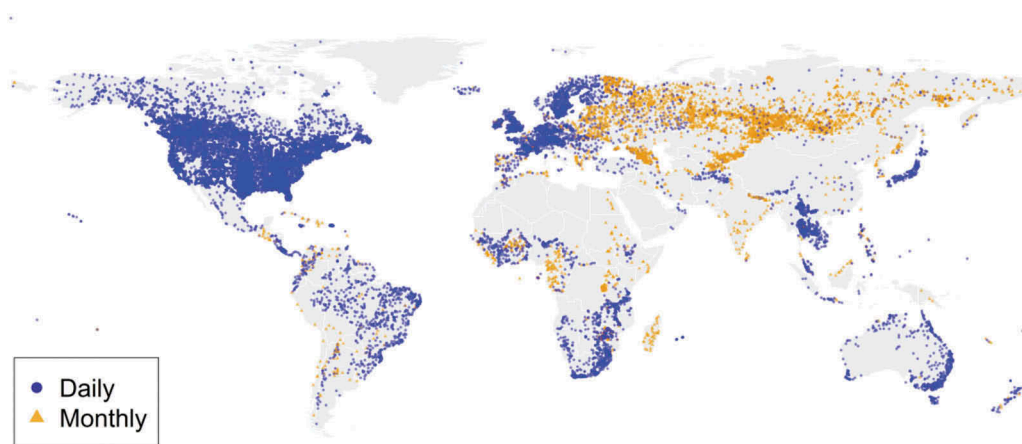
Figure 1. Location of the stations corresponding to the collected time series after duplicate removal. Colors indicate the finer time step found for each location.

Table 1. Summary per provider for the duplicate removal step.

Provider	Dataset type	All openly accessible time series	Time series after removing duplicates	Reduction after duplicate removal
GRDC	Global	12 634	6 875	−45.6%
USGS	National	6 984	6 860	−1.8%
WSC	National	6 326	5 032	−20.4%
R-ArcticNet	Regional	5 827	754	−87.1%
NCAR-UCAR	Regional	2 458	805	−67.2%
RIVDIS	Global	1 018	10	−99.0%
SMHI	National	439	426	−3.0%
ANA	National	283	198	−30.0%
BOM	National	217	212	−2.3%
CHDP	National	165	148	−10.3%
EWA	Continental	147	120	−18.4%
GAME	National	132	125	−5.3%
Spain	National	22	21	−4.5%
Total		36 652	21 586	−41.1%

mostly contain duplicates. Moreover, national datasets are more likely to include the most recent records and thus prevailed in our selection process. This is in accordance with the methodology used by Do *et al.* (2018), which consisted in systematically replacing global and continental datasets with corresponding national information. The following analysis of the quality of river flow data worldwide focuses on these remaining 21 586 time series.

2.3 Methods for quality assurance

A range of quality assurance characteristics covering availability, outliers, homogeneity and trends were computed using 21 586 time series from 13 data providers worldwide. Here, we describe the methods used to compute these characteristics, which Table 2 summarizes.

2.3.1 Availability

Availability aims at assessing available records, both spatially and temporally, and reflects hydrological stations installation, maintenance and dismantling. Similarly to when removing duplicates (Section 2.2.2), availability was characterized by the fraction (%) of the reference time period 1961–2019 covered by observed data. For each time series, we assessed the overall availability, computed as the total fraction of 1961–2019 covered

with data, as well as the longest availability, computed as the longest fraction of 1961–2019 covered with continuous data (i.e. without missing values). The longest availability was compared to the overall availability to assess the continuity of the time series and to highlight fragmented time series. Lastly, availability was calculated for each month of the year and compared to the overall availability to highlight time series in which one month of the year, and therefore parts of the hydrological regime, may not be well represented.

2.3.2 Outliers

Outliers in a time series correspond to measurements that stand out from the rest of the time series. These may be caused by false measurements when acquiring the data. Here, the number of outliers was assessed based on the median and standard deviation of time series. Time series were first standardized by subtracting their respective median river flow. All values greater than five times the standard deviation (5SD) were then flagged as outliers. Outliers were finally presented as fractions (%) of the entire time series. While the median is not sensitive to outliers, the standard deviation is sensitive to the existence of outliers. This approach was chosen based on hydrograph inspection in order to limit detecting extreme flood peaks as outliers in time series with highly variable hydrological regimes.

In addition, we visually inspected outliers and tried to distinguish high-flow peaks, which, by nature, will occur repetitively, from numerical outliers caused by data acquisition or unit errors. Here, outliers were considered as events rather than single days, with one outlier corresponding to consecutive days above the threshold (5SD) and distinct outliers being separated by at least 10 days. Periodicity was defined as the average number of days between two outliers. Deviation was defined as the average outlier magnitude above the threshold, which was then normalized by the long-term flow average.

2.3.3 Homogeneity

Inhomogeneity can be due to natural variability, anthropogenic changes, such as dam constructions and deviations, or to changes in the data acquisition itself, such as changes in rating curves, errors in units and faulty recording devices. In

Table 2. Description of the eight quality assurance characteristics assessed for 21 586 collected river flow time series.

Quality assurance characteristics	Objective	Worst case	Best case
Overall availability for 1961–2019	Length of the observation series, as a fraction of the 1961–2019 period	No data from 1961 to 2019	All data from 1961 to 2019
Longest availability without gaps for 1961–2019	Length of the longest observation series without gaps, as a fraction of the 1961–2019 period	No continuous data from 1961 to 2019	Continuous data from 1961 to 2019
Continuity in available data	Length of the longest observation series without gaps, as a fraction of the overall availability, i.e. ratio between longest availability and overall availability	Only gaps in available data	No gaps in available data
Minimum relative availability for a calendar month	Minimum ratio between month availability and overall availability of the time series. Indicates whether available data are equally spread between months of the year	At least one month has no data	Data are evenly spread among all months of the year
Ratio of outliers	Outliers in the time series as a fraction of the length of available data	All data are considered as outliers	No outliers
Homogeneity of annual flows	Number of statistical tests that detect inhomogeneity in yearly-averaged time series	The four tests report a shift or jump in the time series	No test reports a shift or jump in the time series
Trend in annual flows	Whether a significant trend was detected in yearly-averaged river flows	Significant trend	No significant trend
Trend in one month of the year	Whether at least one calendar month shows a significant trend in river flows	Significant trend in any of the months	No significant trend in any of the months

order to assess homogeneity, we applied the standard normal homogeneity test, the Buishand range test, the Pettitt test and the rank version of the von Neumann ratio test (see Wijngaard *et al.* 2003, and references therein). The null hypothesis of these four tests is that all values from the tested time series follow a similar statistical behavior (e.g. mean, distribution). If the p-value is below a defined threshold, the null hypothesis shall be rejected. Following the methodology proposed by Wijngaard *et al.* (2003), we then analysed the number of tests that rejected the null hypothesis. In this paper, inhomogeneity was assessed at the significance level of 5%. The best (worst) case is no test (all four tests) rejecting time series homogeneity. The tests were applied on yearly-averaged time series provided that time series included at least ten years of data (14 598 stations).

2.3.4 Trend
Trends reflect long-term changes and non-stationarity, which may be caused by natural or anthropogenic changes but also by drifts in measuring devices. While the former is a more likely cause for detected trends, the latter can nonetheless induce artificial signals that are worth detecting, when possible. These can for example limit statistical analyses or modelling exercises such as calibration and validation carried out on independent time periods. The existence of trends in the time series is evaluated with a modified version of the Mann-Kendall non-parametric test (Mann 1945, Kendall 1975). Several papers have highlighted the sensitivity of the Mann-Kendall test to serial correlation in time series, and thus proposed alternatives and pre-treatment of the time series to account for the correlation. Here, we applied the method proposed by Yue *et al.* (2002), which additionally accounts for the relationship between trends and serial correlation and was applied in a number of hydrological studies (e.g. Burn *et al.* 2008, Yeh *et al.* 2015). The test was applied on yearly-averaged and monthly-averaged river flow time series including at least ten years of data (14 598 stations). The existence of a trend was assessed at the significance level of 5%. When a trend was detected by the statistical test, the slope was assessed with the Sen slope estimator (Sen 1968, see

Gudmundsson *et al.* 2018b for a global application) and converted to percent per year (% per year). As suggested by Serinaldi *et al.* (2018), the pre-whitened Mann-Kendall test is applied solely for quality control screening. All statistical tests and results are presented at the global scale, but individual time series investigations are necessary to infer non-stationarity.

2.3.5 Composite indicator
Lastly, an indicator compiling the above-mentioned quality assurance characteristics (Table 2) was produced. All characteristics were normalized between 0 and 1, 1 (0) being the best-case (worst-case) scenario. Then, the composite indicator was produced based on the mean of all available characteristics for the 14 598 time series that have at least ten years of data. The higher the composite indicator (1), the higher the quality of the time series with regards to data availability, outliers, homogeneity and trends.

3 Results

3.1 Overview of time series characteristics
We first provide an overview of the results obtained from applying the eight quality indicators presented in Table 2 to all river flow time series collected (Fig. 2). Data availability appears as the first limitation with more than half of the stations with a reduced overall availability (shorter than 40%), and two thirds of the stations with a longest availability shorter than 40%. Nevertheless, a third of the stations have an overall availability greater than 60%, and two thirds of stations have continuous data (continuity greater than 60%). The existence of monthly trends in river flows can also be limiting with 52% of stations displaying a trend in at least one month of the year. Low values in availability and monthly trends seem equally spread over the globe, despite patches with only high availability characteristics in Europe. The fraction of potential outliers in time series is considerable with outliers being detected in 86% of time series. This is the case
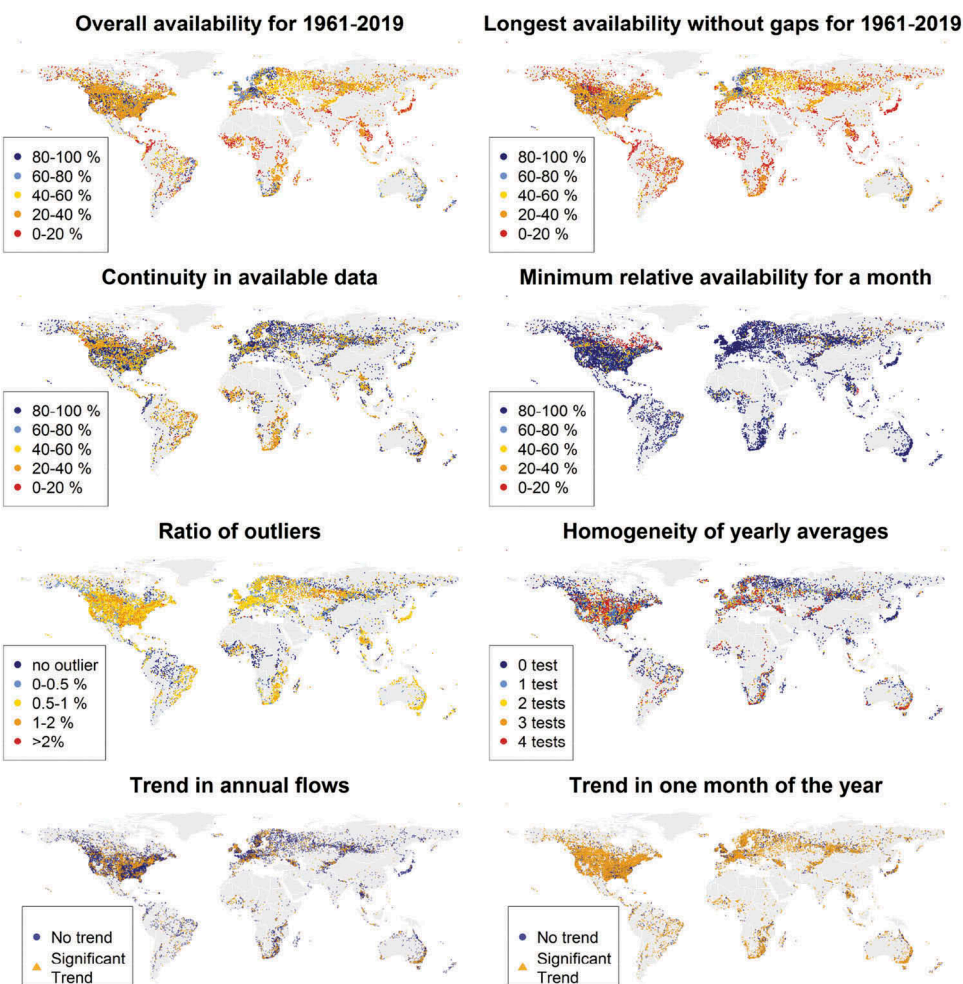
Figure 2. Maps of the eight indicators presented in Table 2: overall availability, longest availability without gaps, continuity, minimum relative availability for a month, ratio of outliers, homogeneity of yearly averages, trend in annual flows and trend in one month of the year. Points indicating low quality are superimposed on points indicating high quality.

in most regions, with the exception of central South America and Central Africa, suggesting that further investigations are required.

The least impactful characteristic is the monthly availability with available data being evenly spread over the calendar months. This suggests that all seasons are well represented in most time series. Trends in yearly-averaged flows are detected much less often than trends in monthly flows with a significant trend being detected in 14% of the time series. Finally, inhomogeneity is detected in 64% of the time series, though all four statistical tests agree on 6% only. One can note that southwestern Canada stands out in terms of continuity, monthly availability, outliers and homogeneity. This is explained by measurements separated by long periods of inactivity in a number of WSC time series.

3.2 Data availability

3.2.1 Spatially: all regions are not equal

We found discrepancies in data availability when investigating the average river flow data availability for the entire 1961–2019 period per country and per continent (Fig. 3). At

the continental level, Oceania (667 stations), for which 74% of records belong to the Australian territory, has the best availability, with almost 60% of its stations with more than 60% of 1961–2019 covered (i.e. more than 35 years). Europe (5 359 stations) and North and Central America (12 262 stations) come next with about 30% of their stations covering at least 35 years. Africa (1 516 stations) is equivalent to North and Central America in terms of its fraction of stations covering at least 10 years (about 60%), even though it has eight times fewer stations. In contrast, Asia has the lowest data availability with 93% of its stations with a data availability lower than 40% (i.e. less than 24 years), and 68% with an availability lower than 20%, which approximately corresponds to the 10-year limit for hydrological analyses.

At country level, central European and Scandinavian countries, as well as South Africa, Namibia and Australia have the longest average temporal availability. An analysis of station density also showed high densities (>50 stations per 100 km^2) in northern and central Europe, Southeast Asia, Central America, the USA and Canada, and southeastern Africa. In agreement with WMO (2008a), small islands have the highest station densities while arid areas have the lowest station densities.
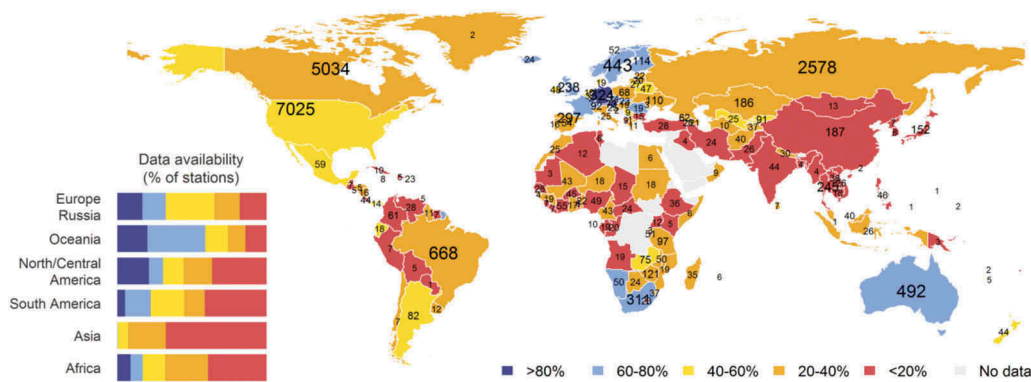
Figure 3. Barplot of data availability in each continent, and corresponding number of stations, and map of data availability and number of stations per country over 1961 to 2019. Colours indicate the average percentage of data availability over 1961–2019 for all stations in the country. Figures indicate the number of stations in each country.

However, the collected datasets do not cover some countries where hydrometric stations are known to exist even though they may not be openly accessible, such as the Democratic Republic of Congo, Iraq, Israel, Lebanon, South Sudan, and some other countries where no hydrometric station is known to the authors, such as Jordan, Libya, Saudi Arabia and Yemen. Overall, gaps in spatial data coverage extend to southern Asia, the Middle East, North and Central Africa and the western coast of South America,

potentially because of aridity, political and financial reasons, or simply due to a lack of open databases.

3.2.2 Over time: towards less river flow information?
We found trends when examining the evolution of data availability from 1961 to 2019 per continent (Fig. 4). The highest fraction of active stations with a good global coverage is achieved between the 1970s and the 1990s. However, apart from North and Central America where the fraction of active
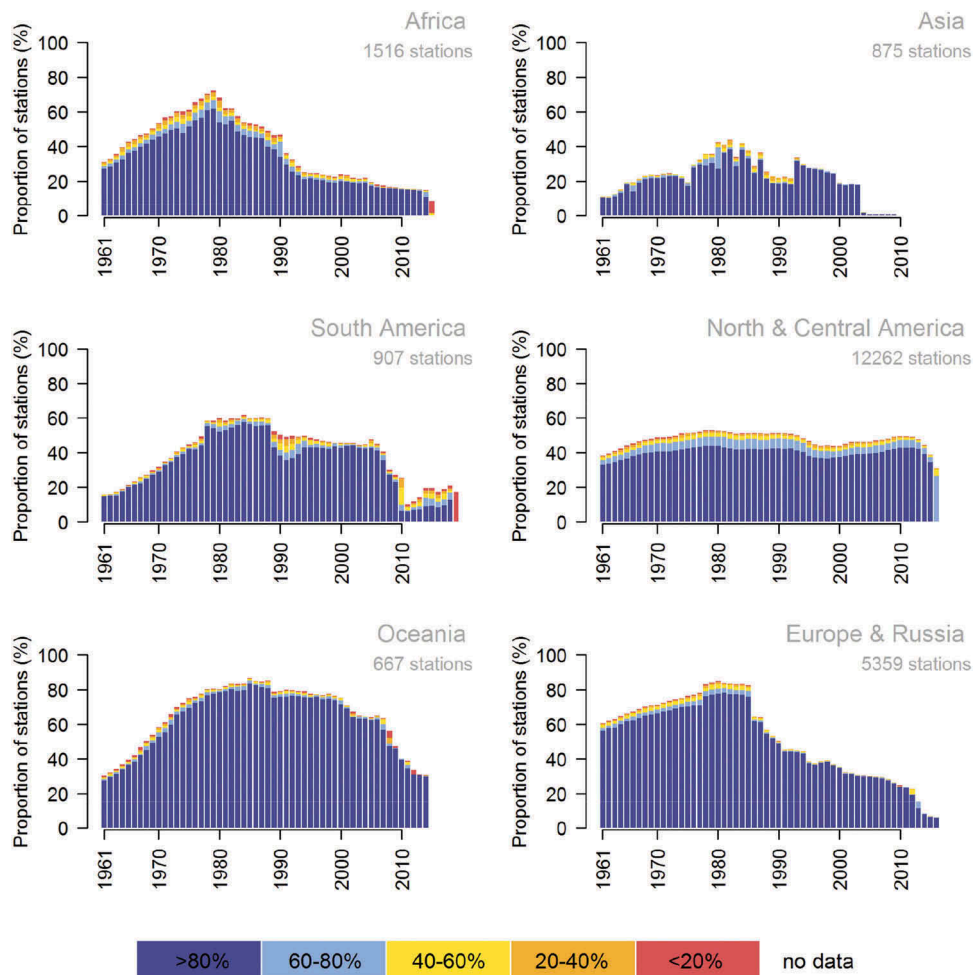

Figure 4. Proportion of river-flow data availability per year and per continent from 1961 to 2019.

stations is remarkably constant with 40%–60% of the stations always being active from 1961 to 2016, all continents display an increasing trend until the end of the 1970s or 1980s, before a clear downward trend until nowadays. This downward trend starts from the beginning of the 1980s in Africa and Asia, and from the mid-1980s in South America, Oceania, Europe and Russia. The most drastic drop is in Europe and Russia where a peak at 80% of the stations being active at the beginning of the 1980s drops to 20–30% by 2010. In Africa as well, the fraction of active stations peaks at 70% at the end of the 70s before dropping to 20% by 2010. In Asia, this fraction of active stations drastically drops from 20% to nearly zero in 2004 and to zero in 2010.

Figure 5 shows that Russia, southern Asia, Southern Europe, Africa and Central America are particularly affected by this decrease in available data. This may be due to dismantled or no longer maintained stations, but also to more recent time series not being shared, which could be explained by a range of political or economic reasons, or by the fact that the latest observations are kept confidential, for example for financial reasons or risk prevention. From this map, we can also observe steadiness or gain in available data in some regions, such as Japan, Sweden (SMHI), some regions of Brazil (ANA), spatially heterogeneous regions in the United States (USGS) and Canada (WSC), Indonesia or Namibia.

3.3 Outliers: distinguishing high-flow peaks from numerical outliers

The outlier detection method proposed in this paper detected that 86% of the stations contained at least one outlier in the time series. The method used is a very common but simplistic one, as confirmed by this percentage. High-flow peaks are detected as outliers in rivers with highly variable regimes. The results from Fig. 2 thus overestimate the number of stations flagged for outliers.

The second method, based on visual inspections, showed that time series with more than 100 outliers (305 series) or with an outlier deviation smaller than 1 (1 051 series) were in almost all cases time series with highly-varying flows and therefore with low recurring peak flows being wrongly detected as outliers (Fig. 6). Conversely, if the outlier deviation was

greater than 200 (106 series), the time series were likely to be suspect. Nevertheless, peak flows detected as outliers can hide outliers, and high deviations can be caused by intermittent rivers, so no rule of thumb can replace visual inspections and local knowledge.

Periodicity of outliers was hardly an indicator because it is sensitive to the length of the time series. Moreover, outliers can be periodic if they are caused by material failure or errors in the rating curve linked to specific flow magnitudes. Nevertheless, we can note that high-flow peaks falsely detected as outliers occurred with a periodicity between 100 and 500 days (from three high-flow peaks a year to one high-flow peak every two years). Also, short periodicity (less than 50 days; 130 series) was always found in short time series which are more sensitive to outliers and often correspond to one high-flow peak. Altogether, this still leaves 80% of the stations with identified outliers.

Figure 6 summarizes this visual inspection and localizes suspected outliers and high-peak flows following these criteria. Overall, high numbers of outliers likely reveal peak flows screened as outliers. Conversely, high outlier deviations likely reveal numerical errors. The regional patterns further highlight the need for local expertise on hydrology and measurement techniques, since some localized high deviation patterns may be due to intermittent rivers in dry regions.

3.4 Homogeneity: a robust detection requires consensus

Out of the 14 598 stations screened for inhomogeneity, 7 756 stations (i.e. 36% of all stations and 53% of the stations with at least 10 years of data) were not identified as inhomogeneous by any of the four statistical tests applied in this study. Conversely, a total of 6 842 stations (i.e. 31.7% of all stations and 47% of the stations with at least 10 years of data) were detected as potentially inhomogeneous by at least one of the statistical tests (Fig. 7). The four tests investigated in this study agreed only on 20% of these potentially inhomogeneous time series; the remaining 80% being detected inhomogeneous by one (47%), two (21%) or three (12%) tests.

Figure 7 clearly shows that inhomogeneity detection becomes more robust when the number of tests agreeing increases. The tests alone detect between 0.7% (Pettitt test) and 14% (modified
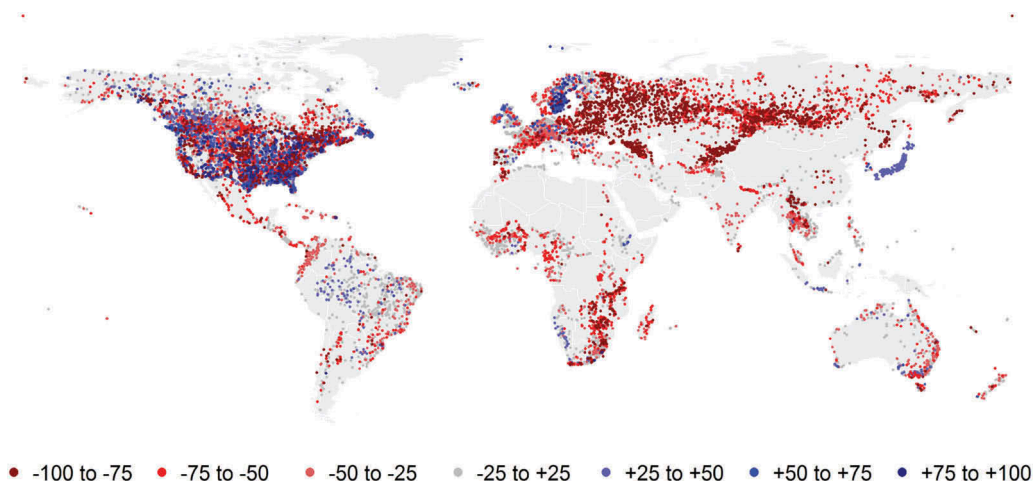

● -100 to -75 ● -75 to -50 ● -50 to -25 ● -25 to +25 ● +25 to +50 ● +50 to +75 ● +75 to +100

Figure 5. Percent change in data availability between the periods 1961–1990 and 1991–2019.
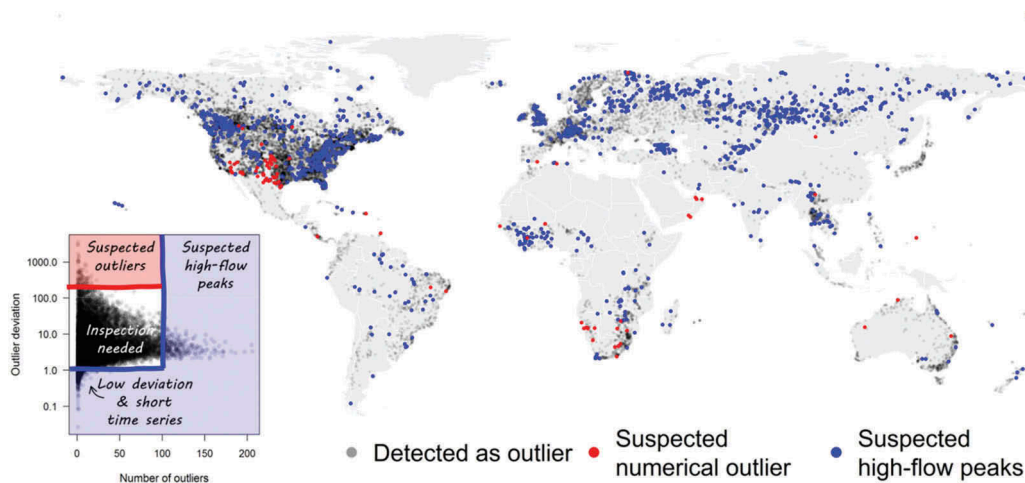
Figure 6. Map of suspected outliers and suspected high-flow peaks, and graphical distinction between suspected outliers and suspected high-flow peaks based on outlier deviation and number of outliers.
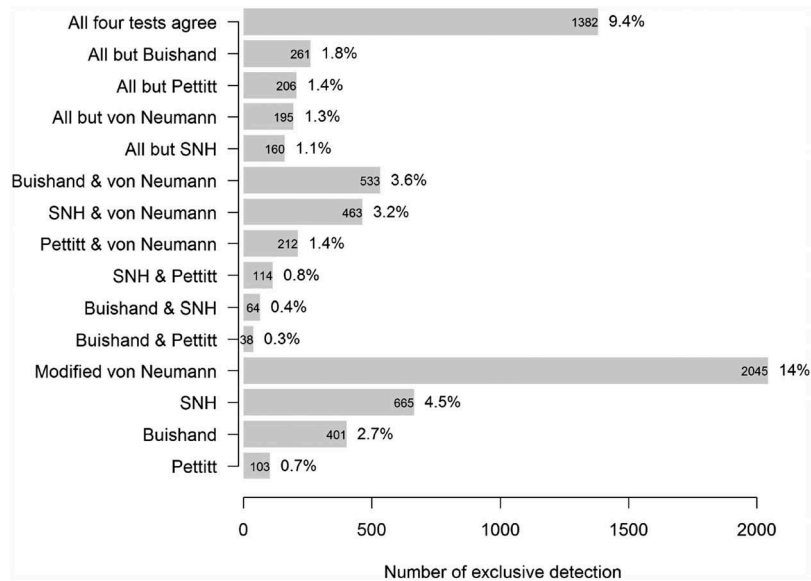

Figure 7. Barplot of the number of stations detected exclusively by either one of the statistical homogeneity tests, by a combination of two tests, by all tests but one, or by all tests. Numbers at the end of each bar indicate the number of stations in each category, as well as the corresponding fraction relative to all 14 598 stations with at least 10 years of records.

von Neumann ratio test) of stations as potentially inhomogeneous, whereas an agreement of three tests would result between 1.1% and 1.8% of detected stations. The results from the exclusive detection by one of the tests are consistent with remarks from Wijngaard *et al.* (2003) pointing out that the Pettitt test is less sensitive to outliers. No clear spatial patterns in the stations detected by one or part of the tests could be observed, which highlights the need for further individual time series checks based for instance on catchment characteristics, water management, or human influence information before any hydrological impact study.

3.5 Trends: towards a change in river flow distribution

Over the collected datasets, a significant trend in yearly-averaged streamflow was found in 8% of the time series, and no significant trend in 60%. The remaining 32% did not have long enough records. This preliminary screening for trends was furthered by analysing the sign and amplitude of the trends. Approximately half of the time series with a significant trend have a slope greater than 1% per year. Figure 8 presents the location and the sign of the trend for these time series with a significant trend and a slope greater than 1%.

Spatially consistent trends can be seen for instance in West Africa, southern Europe and western North America, which exhibit negative trends. Other areas are characterized by opposite phenomena occurring in different regions, as in Australia, where the southeast part displays negative trends, while the northwestern coast displays increasing trends. In Southeast Asia, the northern region is characterized by negative trends, while Malaysia is characterized by positive trends. In Western Europe, the northern part
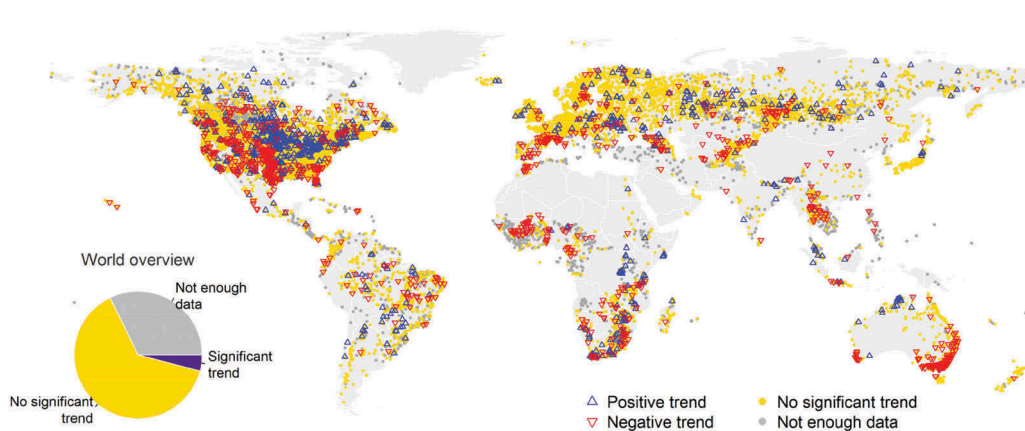
Figure 8. Results from the trend analysis conducted for time series over the globe. Significant trends are limited to time series where the slope of the trend is greater than 1% per year.

exhibits some positive trends while the south clearly displays negative trends. Finally, in North America, clear but nonetheless overlaying patterns appear between the southwest and northeast.

However, caution is necessary because a slope of 1% per year still remains moderate. An investigation of regions with slopes greater than 1% shows that the patterns for North America, most of Europe, South America and Russia tend to disappear, while patterns for Africa, Australia, southwest Europe and southeast Asia clearly remain.

3.6 Composite indicator for quality assurance and hydrograph examples

We finally examine the results obtained with the composite indicator based on all quality assurance characteristics proposed in this paper. Figure 9 shows the composite indicator for the 14 598 stations that have at least 10 years of available data. The maps show that a large majority of stations have time series with a composite indicator greater than 0.6 (76%) or greater than 0.8 (21%). The remaining 23% have a composite indicator below 0.4, with only 2% of the time series having a composite indicator below 0.2. Regions with particularly high indicator values (i.e. of high quality
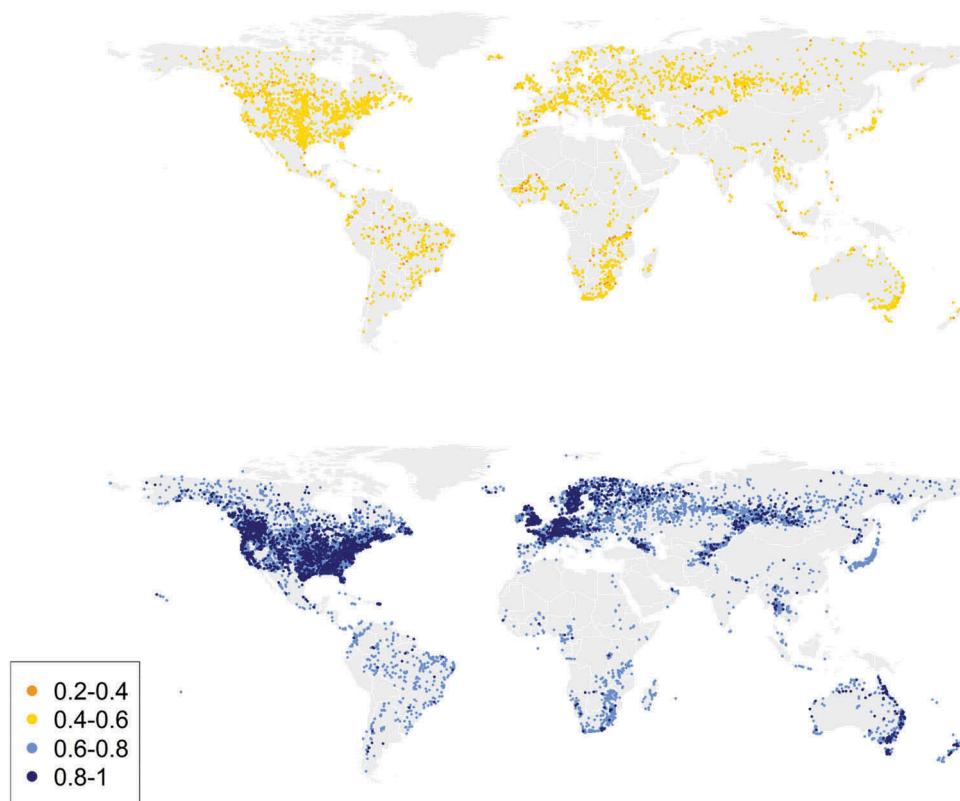

Figure 9. Composite indicator compiling data availability, outliers, homogeneity and trends per station for stations with a composite indicator smaller than 0.6 (upper map) and stations with a composite indicator greater than 0.6 (lower map). Composite indicator values closer to 1 (0) indicate better-case (worse-case) scenarios.

with regards to data availability, outliers, homogeneity and trends) include Central and Northern Europe, the USA and Canada, the eastern coast of Australia, the eastern region of the Black Sea and some parts of Central Asia. The maps indicate that all regions in the world have both high-quality and low-quality time series based on the composite indicator, which suggests that there is access to high-quality time series in all regions covered by worldwide datasets, but quality checks are crucial everywhere to detect potentially problematic time series.

Finally, Fig. 10 gives four examples of hydrographs for stations with long time series. In these four stations coming either from the GRDC or the WSC datasets, the corresponding homogeneity, outlier and trend characteristics, and composite indicator are displayed. The first hydrograph presents the example of a time series for the Syr Darya River in Kazakhstan detected by all four tests as potentially inhomogeneous and that exhibits a significant trend. This time series is given a composite indicator value of 0.48, the lowest of the four hydrographs. The Credit River close to Toronto, Canada, whose time series is detected as inhomogeneous by one test and that supposedly contains a large number of outliers receives a reasonable value

of 0.67. The time series for the Colorado River in the USA clearly shows inhomogeneity, which also leads to a large number of detected outliers. Finally, the time series for the Muricizal River in Brazil receives the highest composite indicator with no significant trend, inhomogeneity detected by one test, and few but clear outliers.

4 General outlook on opportunities and barriers with open data

The large quantity of openly accessible river flow time series nowadays enables large-scale and observation-based hydrologic studies, which gives new opportunities for research and can accelerate knowledge in hydrological sciences. We observed for instance good river flow data availability worldwide between the 1970s and the 1990s, suggesting great opportunities for research on global hydrology during this time period. Furthermore, based on the preliminary quality check presented in this paper, many times series display adequate quality criteria for use in hydrological modelling. This is very promising.
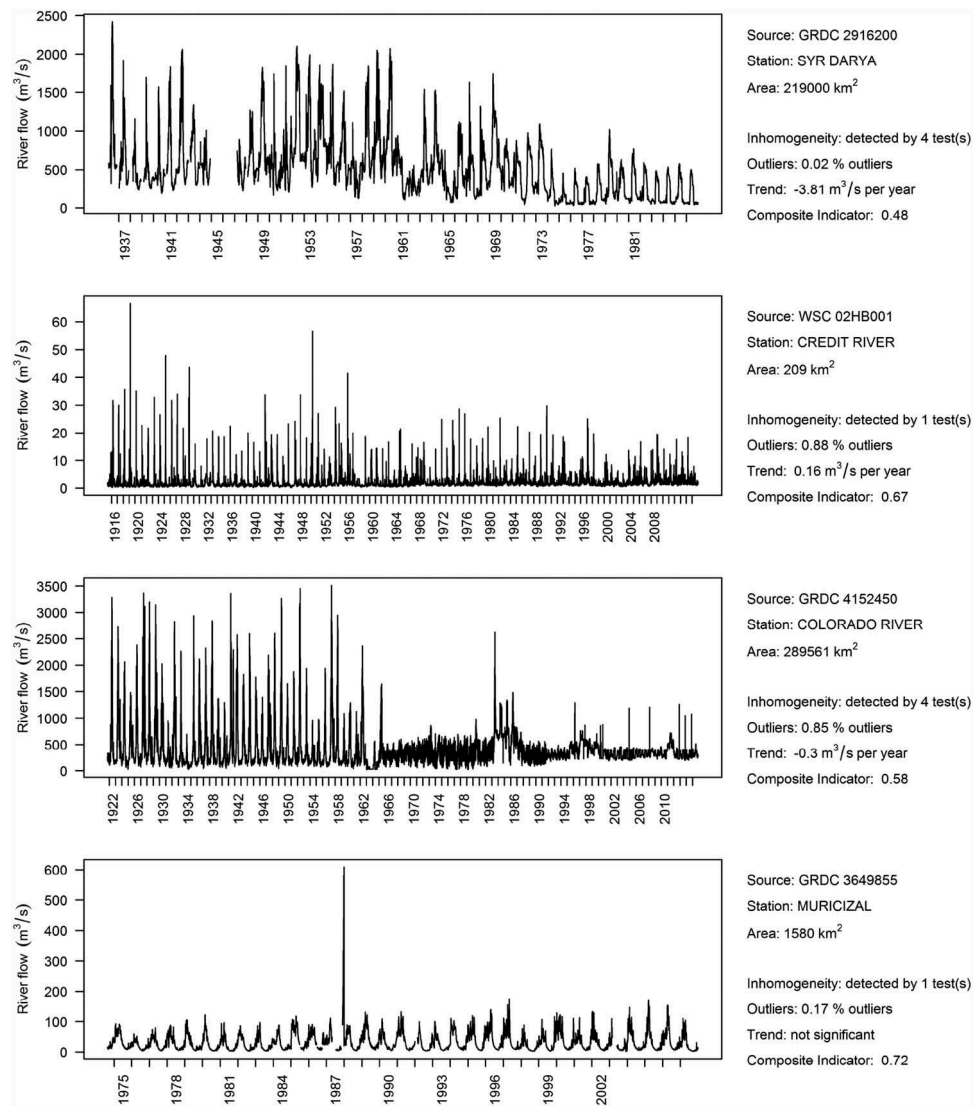

Figure 10. Examples of hydrographs, along with corresponding homogeneity, outliers and trend characteristics and composite indicator.

However, negative trends in data availability have become severe in most continents since the 1980s. If this downward trend, also noted by Bierkens (2015) and explained for New Zealand by Pearson (1998), is due to stations being dismantled or no longer maintained, or to networks being rethought for economic reasons, there is a risk for river flow data availability to further decrease in the coming years. In some regions, this decrease may be due to data being restricted because they are too recent and therefore economically or politically critical. In this latter case, we could either expect an increase in available data as time passes, or increasing restrictions on water-related data as political and economic climates aggravate.

There is also a lack of spatial coverage in some regions, including Southern Asia, the Middle East, or Northern and Central Africa. We can only hypothesize on the reasons for this lack of coverage. Political conflicts, climate features, or simply confidentiality leading to access restrictions can be cited as potential impediments to openly accessible river flow data. Unless data are available through local authorities, it is impossible to carry out modelling in these regions based on observed river flow data, or to include these regions in large-scale studies. Nevertheless, the increasing trend toward open databases with numerous sources for station information and river flow time series may compensate for these spatial and temporal gaps, in regions where monitoring networks exist but are not easily accessible. Moreover, spatial gaps offer research and business opportunities for innovating remote data acquisition techniques.

Data quality can also be an impediment. Here for instance, we detected potential trends in at least one calendar month in most stations. Trends are associated with non-stationarity which can limit some of the scientific investigations carried out in hydrological studies. Nevertheless, yearly and monthly trends, identified here in Africa, Australia, Southwest Europe and Southeast Asia (in accordance with the findings from Gudmundsson *et al.* 2018b) offer great opportunities for research on global environmental changes. Identification of the drivers causing these changes can bring key elements for decision-making in global water resources management, provided that analyses are based on sound statistical methodologies. Similarly, inhomogeneity can be the focus of hydrological studies rather than an impediment, depending on the geographical area and the research objectives. Therefore, all stations detected as inhomogeneous by at least one statistical test (half of the stations with sufficient data) also provide a set of stations of interest to study anthropogenic changes impacts.

The frequent detection of outliers in this paper suggests, among others, that for local as well as large-scale hydrological studies, an in-depth quality assurance beyond the quality characteristics used in this paper is required. Visual inspection is one of the necessary steps, which can quickly become tedious for large river flow data samples. This is why this paper proposes a composite indicator that can give preliminary hints on where in-depth checks are required before use in hydrological studies. Given the dual nature of some of the quality characteristics used in this paper, i.e. trends and inhomogeneity which can be both barriers and opportunities, the composite indicator could be tailored in specific areas to exclude or weigh some of its components that are not bottlenecks in specific hydrological impact studies. Lastly, even though visual inspection is

time consuming, one could think of automatizing its main mechanisms (e.g. through machine learning techniques) as done previously for hydrograph comparison and evaluation (e.g. Ehret and Zehe 2011, Ewen 2011).

Finally, compiling a global database requires a laborious work to collect and harmonize data across data providers. The next obvious step of selecting pertinent information for global hydrologic studies or modelling can subsequently be cumbersome. The statistical tests used in this paper should not be used as stand-alone quality checks. Hydrological time series are known to be non-independent and auto-correlated, which limits the application of most statistical tests and therefore requires case-by-case investigations of trend and inhomogeneity in each time series. These points relate to the challenges of big data, which highlights the large quantity of information openly available, and brings us back to the opportunities identified from the amount of collected river flow time series. The scattered databases also leave room for open global databases gathering the different information from open databases worldwide. Two of the databases used in this study, namely GRDC and RIVDIS, are based on such ideas, but our findings as well as those from Do *et al.* (2018) show that these could be further extended. Furthermore, increasing the diversity of measurements and variables being openly accessible could allow for more in-depth quality checks. For instance, the availability of water stages together with river flows could help shed light on potential causes for inhomogeneity and outliers.

5 Conclusions

Open data are important for accelerating science but may be difficult to access in some regions and of uncertain quality. This study shows different ways to quickly screen the quality of open data, and results imply that most regions worldwide have access to some high-quality time series of river flow. Availability in open hydrological data continuously evolves as long as measurement stations are maintained. Therefore, the present study provides a picture in time that will change under political, economic and climatic constraints, or thanks to scientific advances and innovating technologies.

The openly available data worldwide offer good opportunities for research on global hydrology and environmental changes. However, severe downward trends in data availability demonstrate the potential for new data acquisition techniques to maintain the current river flow data coverage, and to extend this spatial coverage to data-scarce regions. The results also indicate that it is now timely to implement new methods and facilities in global data management to better harvest from the open data providers at national level.

More specifically, we found that:

- the access to open and readily available river flow data is not equal across the globe, Asia having the lowest availability from river flow monitoring stations (followed by Africa and South America); and
- all continents display a decreasing trend in data availability, starting around the 1980s for most regions.

From the screening of quality characteristics, we found that the 14 598 flow stations with more than 10 years of continuous data:

- are homogenous in 53% of the stations;
- have outliers in 80% of the stations that could not be explained in a straightforward manner by low recurring high flows; and
- show significant trends ($p < 0.05$) with more than 1% slope in yearly-averaged streamflow in 4% of the time series, while 60% of time series show no significant trends in river flow.

Acknowledgements

Disclosure statement

No potential conflict of interest was reported by the authors.

Open data

The quality characteristics presented in this paper are openly available in the dataset "Quality check of river flow data worldwide" (http://doi.org/10.5281/zenodo.2611858). The open R package "stats" was used to assess the longest part of the time series without missing values. The open R package "trend" was used to apply the four homogeneity tests. The open R package "modifiedmk" was used to apply the modified Mann-Kendall statistical test.

References

Archfield, S.A., et al., 2015. Accelerating advances in continental domain hydrologic modeling. Water Resources Research, 51 (12), 10078–10091. doi:10.1002/2015WR017498

Arheimer, B., et al., 2012. Water and nutrient simulations using the HYPE model for Sweden vs. the Baltic Sea basin – influence of input-data quality and scale. Hydrology Research, 43 (4), 315–329. doi:10.2166/nh.2012.010

Arheimer, B., et al., 2019. Global catchment modelling using World-Wide HYPE (WWH), open data and stepwise parameter estimation. Hydrology and Earth System Sciences Discussions, 2019, 1–34. doi:10.5194/hess-2019-111

Beck, H.E., et al., 2013. Global patterns in base flow index and recession based on streamflow observations from 3394 catchments. Water Resources Research, 49 (12), 7843–7863. doi:10.1002/2013WR013918

Beck, H.E., et al., 2017. Global evaluation of runoff from 10 state-of-the-art hydrological models. Hydrology and Earth System Sciences, 21 (6), 2881–2903. doi:10.5194/hess-21-2881-2017

Bierkens, M.F.P., 2015. Global hydrology 2015: state, trends, and directions. Water Resources Research, 51 (7), 4923–4947. doi:10.1002/2015WR017173

Blöschl, G., et al., 2017. Changing climate shifts timing of European floods. Science, 357 (6351), 588. doi:10.1126/science.aan2506

Bodo, B., 2000. Russian river flow data by Bodo. Boulder, CO: Research Data Archive at the National Center for Atmospheric Research, Computational and Information Systems Laboratory. Available from http://rda.ucar.edu/datasets/ds553.1/

Brooks, M.S., 2013. Accelerating innovation in climate services: the 3 E's for climate service providers. Bulletin of the American Meteorological Society, 94 (6), 807–819. doi:10.1175/BAMS-D-12-00087.1

Burn, D.H., Fan, L., and Bell, G., 2008. Identification and quantification of streamflow trends on the Canadian Prairies. Hydrological Sciences Journal, 53 (3), 538–549. doi:10.1623/hysj.53.3.538

Crochemore, L., Ramos, M.-H., and Pechlivanidis, I.G., 2019. Can continental models convey useful seasonal information at the catchment scale? Water Resources Research. Submitted.

Do, H.X., et al., 2018. The Global Streamflow Indices and Metadata Archive (GSIM) – Part 1: the production of a daily streamflow archive and metadata. Earth System Science Data, 10 (2), 765–785. doi:10.5194/essd-10-765-2018

Donnelly, C., Andersson, J.C.M., and Arheimer, B., 2016. Using flow signatures and catchment similarities to evaluate the E-HYPE multi-basin model across Europe. Hydrological Sciences Journal, 61 (2), 255–273. doi:10.1080/02626667.2015.1027710

Donnelly, C., Ernst, K., and Arheimer, B., 2018. A comparison of hydrological climate services at different scales by users and scientists. Climate Services, 11, 24–35. doi:10.1016/j.cliser.2018.06.002

Eagleson, P.S., 1986. The emergence of global-scale hydrology. Water Resources Research, 22 (9S), 6S–14S. doi:10.1029/WR022i09Sp0006S

Ehret, U. and Zehe, E., 2011. Series distance – an intuitive metric to quantify hydrograph similarity in terms of occurrence, amplitude and timing of hydrological events. Hydrology and Earth System Sciences, 15 (3), 877–896. doi:10.5194/hess-15-877-2011

Ewen, J., 2011. Hydrograph matching method for measuring model performance. Journal of Hydrology, 408, 178–187. doi:10.1016/j.jhydrol.2011.07.038

Gudmundsson, L., et al., 2018a. The Global Streamflow Indices and Metadata Archive (GSIM) – Part 2: quality control, time-series indices and homogeneity assessment. Earth System Science Data, 10 (2), 787–804. doi:10.5194/essd-10-787-2018

Gudmundsson, L., et al., 2018b. Observed trends in global indicators of mean and extreme streamflow. Geophysical Research Letters, 46 (2), 756–766. doi:10.1029/2018GL079725

Henck, A.C., et al., 2011. Spatial controls on erosion in the three rivers region, southeastern Tibet and southwestern China. Earth and Planetary Science Letters, 303 (1–2), 71–83. doi:10.1016/j.epsl.2010.12.038

Henck, A.C., et al., 2010. Monsoon control of effective discharge, Yunnan and Tibet. Geology, 38 (11), 975–978. doi:10.1130/G31444.1

Hurk, B.J.J.M., et al., 2016. Improving predictions and management of hydrological extremes through climate services: www.imprex.eu. Climate Services, 1, 6–11. doi:10.1016/j.cliser.2016.01.001

Kendall, M., 1975. Multivariate analysis. London: Charles Griffin.

Kuentz, A., et al., 2017. Understanding hydrologic variability across Europe through catchment classification. Hydrology and Earth System Sciences, 21 (6), 2863–2879. doi:10.5194/hess-21-2863-2017

Lehner, B. and Döll, P., 2004. Development and validation of a global database of lakes, reservoirs and wetlands. Journal of Hydrology, 296 (1), 1–22. doi:10.1016/j.jhydrol.2004.03.028

Mann, H.B., 1945. Nonparametric tests against trend. Econometrica, 13 (3), 245–259. doi:10.2307/1907187

Pearson, C.P., 1998. Changes to New Zealand's national hydrometric network in the 1990s. Journal of Hydrology (New Zealand), 37 (1), 1–17.

Pechlivanidis, I.G. and Arheimer, B., 2015. Large-scale hydrological modelling by using modified PUB recommendations: the India-HYPE case. Hydrology and Earth System Sciences, 19 (11), 4559–4579. doi:10.5194/hess-19-4559-2015

Portmann, F.T., Siebert, S., and Döll, P., 2010. MIRCA2000—Global monthly irrigated and rainfed crop areas around the year 2000: a new high-resolution data set for agricultural and hydrological modeling. Global Biogeochemical Cycles, 24 (1). doi:10.1029/2008GB003435

Sauer, V.B., 2002. Standards for the analysis and processing of surface-water data and information using electronic methods. Washington, DC: US Geological Survey, Water-Resources Investigations Report No. 01–4044, 2–4.

Sen, P.K., 1968. Estimates of the regression coefficient based on Kendall's Tau. Journal of the American Statistical Association, 63 (324), 1379–1389. doi:10.1080/01621459.1968.10480934

Serinaldi, F., Kilsby, C.G., and Lombardo, F., 2018. Untenable nonstationarity: an assessment of the fitness for purpose of trend tests in

hydrology. *Advances in Water Resources*, 111, 132–155. doi:10.1016/j. advwatres.2017.10.015

Siqueira, V.A., *et al.*, 2018. Toward continental hydrologic–hydrodynamic modeling in South America. *Hydrology and Earth System Sciences*, 22 (9), 4815–4842. doi:10.5194/hess-22-4815-2018

Vörösmarty, C.J., Fekete, B.M., and Tucker, B.A., 1998. *Global river discharge, 1807–1991, V[ersion]*. 1.1 (RivDIS). doi:10.3334/ornldaac/199

Wijngaard, J.B., Klein Tank, A.M.G., and Können, G.P., 2003. Homogeneity of 20th century European daily temperature and precipitation series. *International Journal of Climatology*, 23 (6), 679–692. doi:10.1002/joc.906

WMO, 2008b. *Manual on low-flow estimation and prediction*. Geneva, Switzerland: World Meteorological Organization, WMO Report No. 1029. Operational Hydrology No. 50.

WMO, 2010. *Manual on stream gauging*. Geneva, Switzerland: World Meteorological Organization, WMO No. 1044.

WMO, 2011. *Climate knowledge for action: a global framework for climate services*. Geneva, Switzerland: World Meteorological Organization, WMO Report No. 1065.

WMO (World Meteorological Organization), 2008a. *The guide to hydrological practices. Volume I hydrology - from measurement to hydrological information*. Geneva, Switzerland: World Meteorological Organization, WMO Report No. 168, 296.

Yamazaki, D., *et al.*, 2014. Development of the global width database for large rivers. *Water Resources Research*, 50 (4), 3467–3480. doi:10.1002/2013WR014664

Yeh, C.-F., *et al.*, 2015. Spatial and temporal streamflow trends in Northern Taiwan. *Water*, 7 (2), 634–651. doi:10.3390/w7020634

Yue, S., *et al.*, 2002. The influence of autocorrelation on the ability to detect trend in hydrological series. *Hydrological Processes*, 16 (9), 1807–1829. doi:10.1002/hyp.1095

Appendix A

Table A1. List of the collected datasets with corresponding references and links.

Dataset	References	Link
Global Runoff Database from the Global Runoff Data Center (GRDC),	-	https://www.bafg.de/GRDC/EN/Home/homepage_node.html
Surface-Water Data from the USGeological Survey (USGS)	-	https://waterdata.usgs.gov/nwis/sw
HYDAT from the Water Survey of Canada (WSC)	-	https://www.canada.ca/en/environment-climate-change/services/water-overview/quantity/monitoring/survey/data-products-services/national-archive-hydat.html
R-ArcticNet v. 4.0 from the Pan-Arctic Project Consortium (R-ArcticNet)	-	http://www.r-arcticnet.sr.unh.edu/v4.0/index.html
Russian River data (NCAR-UCAR)	Bodo (2000)	https://rda.ucar.edu/datasets/ds553.2/
Global River Discharge Data (RIVDIS)	Vörösmarty et al. (1998)	https://daac.ornl.gov/RIVDIS/guides/rivdis_guide.html
WISKI from the Swedish Meteorological and Hydrological Institute (SMHI)	-	https://vattenwebb.smhi.se/station/
Hidroweb from the Brazilian National Water Agency (ANA)	-	http://hidroweb.ana.gov.br
National data from the Australian Bureau of Meteorology (BOM)	-	http://www.bom.gov.au/water/hrs/
Chinese river flow data from the China Hydrology Data Project (CHDP)	Henck et al. (2010, 2011)	https://depts.washington.edu/shuiwen/index.html
European Water Archive from GRDC – EURO-FRIEND-Water (EWA)	-	https://www.bafg.de/GRDC/EN/04_spcldtbss/42_EWA/ewa.html
GEWEX Asian Monsoon Experiment (GAME) – Tropics dataset provided by the Royal Irrigation Department of Thailand	-	http://hydro.iis.u-tokyo.ac.jp/GAME-T/GAIN-T/routine/rid-river/index.html
Spanish river flow data from the Ecological Transition Ministry (Spain)	-	https://www.mapama.gob.es/es/ministerio/funciones-estructura/organizacion-organismos/organismos-publicos/confederaciones-hidrograficas/default.aspx