www.nature.com/scientificdata

SCIENTIFIC DATA

OPEN

DATA DESCRIPTOR

Global karst springs hydrograph dataset for research and management of the world's fastest-flowing groundwater

Tunde Olarinoye *et al.*#

Karst aquifers provide drinking water for 10% of the world's population, support agriculture, groundwater-dependent activities, and ecosystems. These aquifers are characterised by complex groundwater-flow systems, hence, they are extremely vulnerable and protecting them requires an in-depth understanding of the systems. Poor data accessibility has limited advances in karst research and realistic representation of karst processes in large-scale hydrological studies. In this study, we present World Karst Spring hydrograph (WoKaS) database, a community-wide effort to improve data accessibility. WoKaS is the first global karst springs discharge database with over 400 spring observations collected from articles, hydrological databases and researchers. The dataset's coverage compares to the global distribution of carbonate rocks with some bias towards the latitudes of more developed countries. WoKaS database will ensure easy access to a large-sample of good quality datasets suitable for a wide range of applications: comparative studies, trend analysis and model evaluation. This database will largely contribute to research advancement in karst hydrology, supports karst groundwater management, and promotes international and interdisciplinary collaborations.

Background & Summary

Karst aquifers are essential sources of drinking water to about 10% of the world's population[1]. In many regions across the globe, karst groundwater is also an indispensable resource for ecosystems, agriculture and, economic activities, as well as for tourism and recreation[2,3]. For example, in Europe, 21.6% of the land surface is underlain by carbonate rock[4] which contributes up to 50% of supplied drinking water in some countries[5–7]. However, groundwater flow in karst aquifers is characterised by a complex interplay of fast-flowing conduit and slow-flowing matrix systems[8,9]. Hence, the storage capacity of karst aquifers is variable and systems are extremely vulnerable to climatic pressures, human impacts and contamination[10]. In order to ensure adequate protection of karst water sources, in-depth hydrogeological knowledge is necessary.

Large-scale modelling and comparative water resource research have shown the great value of large datasets in hydrology[11]. Numerous studies have applied these large datasets for several purposes such as model evaluation, global parameter estimations, impact studies, statistical and comparative analyses. For instance, large-scale hydrological models such as WaterGAP[12] used discharge data from the Global Runoff Data Centre (https://www.bafg.de/GRDC) for parameter estimation. Likewise, streamflow data from the Model Parameter Estimation Experiment (MOPEX)[13] and the Global Runoff Data Centre (GDRC) were combined to derive global base flow indexes and recession constants[14]. Streamflow observations of near-natural catchments obtained from UNESCO's European Water Archive (EWA) were used to investigate the streamflow trends across Europe and differentiated the impacts from climatic variability and anthropogenic drivers[15]. The same dataset was also applied to assess the sensitivity of streamflow to storage changes in Europe[16].

Even though the existing large hydrological databases (e.g. MOPEX, GRDC) have brought great advances to the understanding of hydrological processes and their simulation, these databases do not explicitly consider

#A full list of authors and their affiliations appears at the end of the paper.

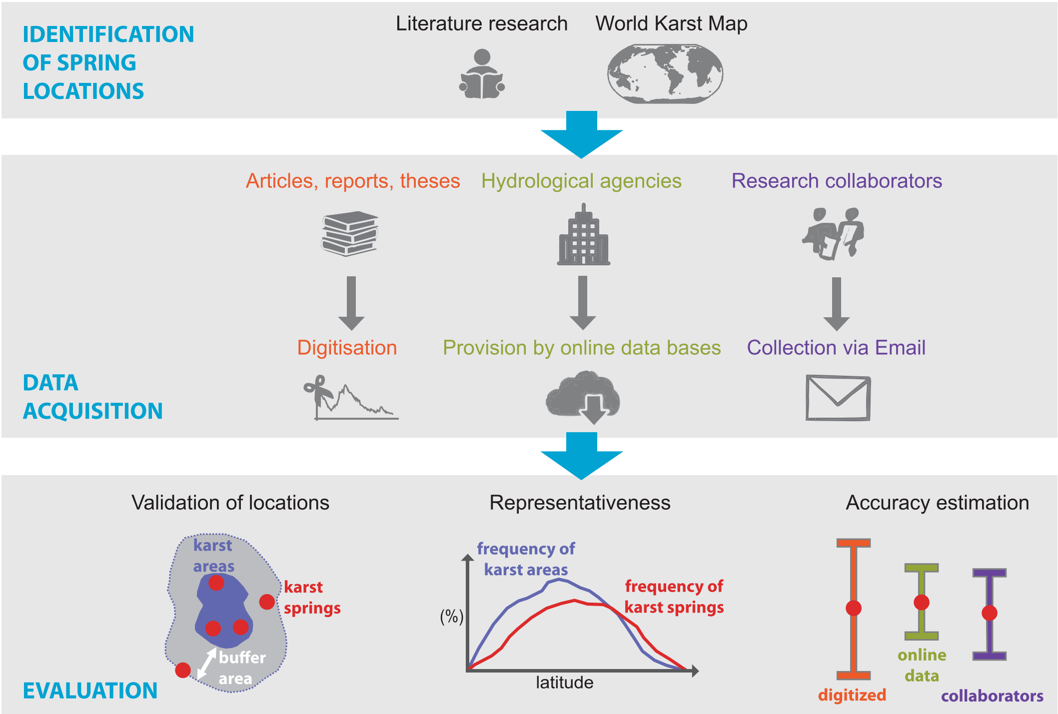

Fig. 1 Workflow of the karst spring discharge observation database development.

karst areas as karst spring discharges are under-represented. Access to data has been identified as a major impediment in quantifying karstification, modelling flow dynamics and transport processes of karst groundwater[2,17]. Studies involving large-scale parameter estimation or comparative studies in karst hydrology are still fairly rare and unrealistic representations of hydrological processes in karst regions can still be found in many large-scale hydrological models[17,18]. The need to advance research in karst hydrology especially on larger spatial scales, combined with the importance and peculiarities of karst aquifers, therefore requires a consolidated, global database for karst systems. Recent advances in providing large-scale information on karst aquifers include the development of World Karst Aquifer Map (WOKAM)[4,19], which is the first to accurately map karst regions worldwide, or the SNO KARST database[20] that provides long-term observations of hydro meteorological and geochemical variables of several karst observatories across France, including karst spring discharge. The SNO KARST also offers a new tools for characterizing and modelling flow in karst aquifers[21] or assess their vulnerability[22].

With the WoKaS database, we provide the first known effort to create a global database of karst spring discharge observations. It is the result of an intense and global effort to make a large number of karst spring hydrographs accessible for karst researchers and the wider hydrological communities. Data from individual researchers and research networks like the Karst Commission of the International Association of Hydrogeologists (IAH) was combined with karst spring hydrographs from national databases and digitized spring discharge data from scientific publications. Access to WoKaS database will motivate large scale and comparative karst hydrology studies, help to improve representation of karstic processes in large-scale models, improve management of karst groundwater, and will promote international and interdisciplinary collaborations. We encourage future users of the datasets to contact researcher or agency that provided the datasets to start a fruitful research collaboration.

Methods

The development of the WoKaS database followed three steps: (i) identification of karst spring locations across the globe; (ii) sourcing for discharge observations of the identified springs; and (iii) evaluation of collected datasets, which included technical validation and quality assessment. The workflow of these steps is illustrated in Fig. 1.

Identifying karst spring locations. Firstly, we assembled the list of karst springs in countries with carbonate outcrops identified from the World Karst Aquifer Map[4,19]. For each country with carbonate outcrops, we performed an extensive literature search with a set of keywords consisting of: (1) country's name; (2) karst; (3) spring; and (4) hydrology. From all the identified materials (articles, conference proceedings, reports, theses, news bulletins, books), we extracted karst spring names, location coordinates, elevation as well as land cover, catchment area (km^2), defined as the topographic boundary within which the spring is located, recharge area (km^2), defined as the area contributing to the recharge of the aquifer drained by the spring, and factors influencing discharge if such information were available. Several spring locations were also collected from WOKAM, which provides a list of relevant karst springs for each country, and from reviewing national databases.

Country	Database/agency name	Database access	Automatic download
France	Ministère de l'Écologie, du Développement Durable et de l'Énergie (BANQUE HYDRO)	http://www.hydro.eaufrance.fr/indexd.php	Yes
France	SNO KARST, OSU OREME	https://data.oreme.org/observation/snokarst	No
Austria	Bundesministerium für Nachhaltigkeit und Tourismus (eHYD)	https://ehyd.gv.at/	Yes
Germany	Bayerisches Landesamt für Umwelt	https://www.gkd.bayern.de/de/grundwasser/quellschuettung//tabellen	Yes
Germany	Landesanstalt für Umwelt Baden-Württemberg (LUBW)	http://udo.lubw.baden-wuerttemberg.de/public/pages/selector/index.xhtml	Yes
Ireland	Environmental Protection Agency (EPA HydroNet)	http://www.epa.ie/hydronet/#Groundwater	Yes
Slovenia	Agencija Republike Slovenije za okolje (ARSO)	http://vode.arso.gov.si/hidarhiv/pov_arhiv_tab.php	Yes
US	U.S. Geological Survey, National Water Information System (USGS)	https://waterdata.usgs.gov/nwis/uv/?referred_module=gw	Yes
UK	Nation River Flow Archive (NRFA)	https://nrfa.ceh.ac.uk/data/search	Yes
Croatia	Croatian Meteorological and Hydrological Service (DHMZ)	http://hidro.dhz.hr/	No

Table 1. Hydrological databases where datasets were downloaded. If automatic download is "Yes" the corresponding database is included in the automatic download routine; see Hydrological agencies subsection. All databases were last accessed in September 2019.

Data acquisition. Time series of karst spring discharge observations were collected from three sources: (1) published data including scientific articles, reports and theses; (2) project partners and collaborators; and (3) public databases of national hydrological services. For each source, the method for data extraction, collection and gathering were different.

Published articles, reports and theses. A web search routine protocol was developed to look-up all publications in karst and non-karst hydrology containing karst spring hydrographs. Firstly, karst spring hydrographs of identified locations (see previous subsection) were searched in published journal articles. The keyword "hydrograph" was added to the set of keywords used in location identification (country's name, karst, spring and hydrology). Occasionally, the country's name was substituted with the spring's name for a more specific web search. The search was further extended to published reports from NGOs, government agencies and PhD theses. The web search protocols for karst spring hydrographs and location identification were similar, hence, the two processes were usually run concurrently.

To extract the spring discharge observations from the published articles, theses and reports, we used *WebPlotDigitizer* (https://github.com/ankitrohatgi/WebPlotDigitizer). WebPlotDigitizer is an open source, web-based, semi-automatic digitization tool developed with HTML5 that works on most common web browsers. The hydrographs were cut out from the original publications, saved as image or pdf files and imported to WebPlotDigitizer. The raw discharge values for the total duration of the observation period were then extracted. Python codes for daily time step interpolation were used to post-process the extracted raw values.

Spring discharge observation time step are not usually stated in publications. Therefore, when the temporal resolution of the observation was unknown, the interpolation time step used was irregular and dependent on the resolution of the extracted figure: plot quality, number of plotted variables, and length or duration of the hydrograph. For instance, hydrographs that covered longer time periods only show seasonal and annual events, hence, a discharge variability could only be captured on monthly time steps. Whereas, if the observation period was shorter, individual events could be identified and discharge values could be extracted on a daily temporal resolution.

Research partners and collaborators. Additional data were acquired through the karst research community. Calls for data contribution were made at conferences, through social media platforms (Twitter and Facebook) and emails soliciting data support for the database to various research commissions, institutes, working groups and researchers with relevant datasets.

Hydrological agencies. A large number of karst spring discharge observations were obtained from national hydrological services that provide online access to their datasets. In total, we collected discharge datasets from ten national databases mainly in Europe and the United States of America (Table 1). Most of this data is in the public domain or published under the creative commons (CC-BY) license and could be directly combined with the data obtained from other sources (see above). Data from databases (Banque Hydro, eHYD, LUBW and NRFA) that do not provide their discharge data under the open data or CC-BY license are made available only as annual averages in the data repository[23]. To access this data at daily resolution, we provide an automatic download routine written with R, which extracts the karst spring discharge time series from the databases' webpages directly. In addition, the download procedure updates the spring discharge time series of all databases in case new observations were added after publication of the WoKaS database (see following subsection).

Data Records

The WoKaS database includes over 400 karst spring discharge observations from more than 30 countries across the globe covering a wide range of hydrologic and climatic diversity. The datasets which are freely available for download[23] are accessible in comma-separated values (CSV) file format. Time series datasets cover time spans ranging from a couple of months to a maximum of 120 years (Fig. 2a). Over 60% of the dataset is made by discharge time series observations of up to 20 years and within this subset, over 90% of the time series have discharge

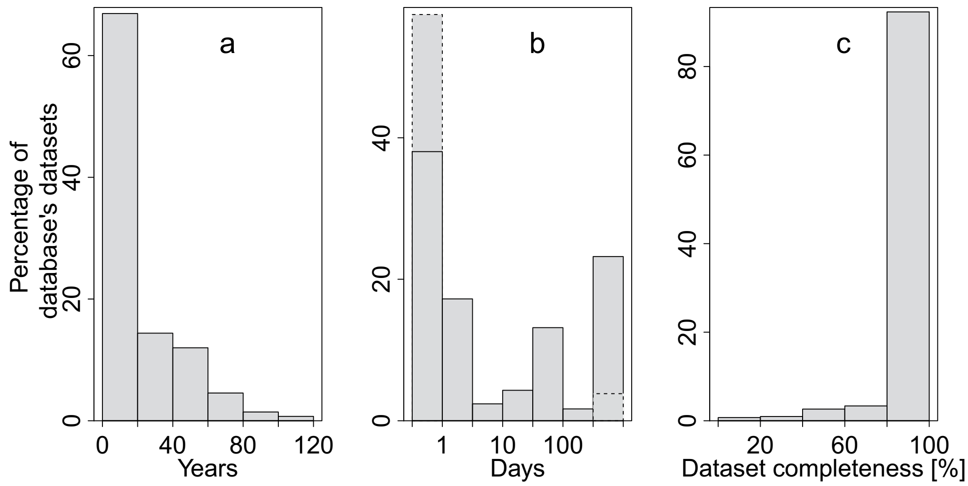

Fig. 2 Properties of the collected datasets in the WoKaS database. (**a**) Time span of spring discharge observations, (**b**) temporal resolution of spring discharge observations with "Days" axis plotted on a Log scale (dashed-line bars indicate the shift in the percentage of datasets with $<=1$ day and a year temporal resolution if the time series from databases that do not hold open data or CC-BY license are replaced with higher resolution time series datasets obtainable through the automatic download routine; see subsection Hydrological Agencies), (**c**) completeness of discharge datasets.

Class	Description	Criteria: Measurement type known	Individual event	Seasonality	Individual recession
A	Very High	✓	✓	✓	✓
B	High	✗	✓	✓	✓
C1	Medium (digitized $<=5$ years)	✗	✓	✓	Recognisable
C2	Low (digitized >5 years)	✗	✗	✓	Recognisable
C3	Very Low (unclear, poor plot)	✗	✗	✓	✗

Table 2. Datasets quality description. The symbol "✓" indicates that the corresponding requirement is fulfilled and "✗" indicates that the requirement is not fulfilled.

observations greater than a year, more than 65% cover an observation periods greater than 5 years and above 35% have more than 10 years of discharge observations. More than 30% have time series measurements of >20 years (Fig. 2a). Considering all collected datasets with those from databases without CC-BY license, which are available as annual averages[23] (see hydrological agencies subsection), ca. 40% of the datasets have temporal resolution of ≤ 1 day and above 20% have a year resolution (Fig. 2b). If the datasets from databases without CC-BY license are substituted by higher resolution time series data which are accessible through the download routine (see hydrological agencies subsection), the percentage of datasets with ≤ 1 day temporal resolution increased to almost 60%. Subsequently, datasets with a year temporal resolution are reduced by 20% (Fig. 2b). Dataset completeness describes the percentage of total discharge observations of a dataset without missing values. More than 90% of the datasets in the WoKaS database are gap-free for the obtained hydrograph duration (Fig. 2c).

Accuracy and quality of datasets. The accuracy and quality of the datasets were defined based on four criteria (Table 2): (1) if the discharge observation measurement is known, (2) recognition of individual events on the spring's hydrograph, (3) recognition of seasonal events on the spring's hydrograph and (4) identification of recession events on the hydrograph. These criteria (mostly based on the data source - see subsection data acquisition) were used to assign five quality classes, from A (very high) to C3 (very low). Generally, datasets from hydrological databases, research partners and collaborators fall within the quality class A or B. Since digitized datasets were extracted from hydrograph plots, inaccuracies were inherited from quality and observation length of the hydrograph plot that was to be digitized. For example, when discharge observation covered several years, only seasonal variability was visually identifiable and individual discharge events could not be recognised. Meanwhile, both seasonal variability and individual events are visually recognisable for discharge observations extending over fewer years (<5years). Consequently, the digitized datasets are sub-divided to class C1 (individual events identifiable, recession periods recognizable), C2 (individual events identifiable, recession periods not clearly recognizable) and C3 (individual events not identifiable, recession periods not clearly recognizable).

Based on the defined quality classes, a high percentage of WoKaS datasets are of good quality (Fig. 3), approximately 62% and 20% of the datasets are of class A and B respectively (Fig. 3). Class A datasets are predominantly found in the northern hemisphere between the latitudes of 30° and 60°, in the same region in which we have

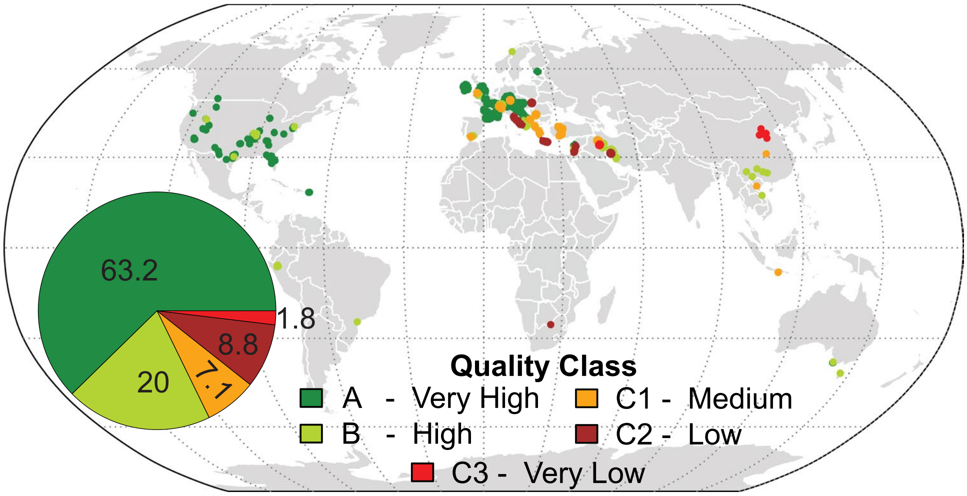

Fig. 3 Distribution of WoKaS datasets based on assigned quality classes. The coloured points on the map are WoKaS locations, attributed colour codes correspond to the quality class. Numbers shown on pie chart in the map are percentage distribution of WokaS datasets based on defined quality classes.

80% of the WoKaS datasets (see subsection spatial representativeness of datasets). Similarly, class B datasets are distributed within these latitudes in Europe and Asia, and also in Australia. The class C datasets are found in the Middle East, Asia and Southern Africa, these are the regions where spring discharge datasets have been digitised from publications due to the scarcity or unavailability of direct spring discharge observations.

Technical Validation

The collected datasets were evaluated by: (1) determining the accuracy of the identified spring locations with respect to carbonate rock outcrop of WOKAM in order to exclude non-karstic springs and (2) determining the spatial representativeness of the database for karst areas by comparing the distribution of the identified locations over all latitudes with the distribution of carbonate rock outcrops over the world's land surface.

Accuracy of identified locations. A fundamental prerequisite for a spring to be considered karstic is that, it must be fed by a karst aquifer. The recharge area of karst aquifer can lie exclusively (autogenic recharge) or partially (allogenic recharge) within carbonate rock areas[9]. In some cases, recharge of karst aquifers is partly due to groundwater flow coming from adjoining aquifers, such as the alluvial ones. Also, karst spring outlets exist within the carbonate rock formation but seldom, a karst aquifer-fed spring may have its discharge outlet in a non-carbonate formation. All collected spring locations were compared with the carbonate rock areas indicated by WOKAM[19] to ensure that they fall within the carbonate rock outcrop. Taking into account karst aquifer-fed springs outside the carbonate rock outcrop, we allowed for a buffer zone around the carbonate rock areas. We used this simplified strategy because detailed, site-specific field information was not obtainable for the large set of collected karst spring locations. We defined the buffer width by the maximum distance of spring locations provided by WOKAM from the WOKAM carbonate rock areas (17.2 km). We consider this buffer distance reasonable as the WOKAM developers could rely on local experts that could confirm the karstic characteristics of all the included spring locations. Using this procedure, over 90% of the identified spring locations fall on the carbonate rock outcrops and approximately 5% are within the buffer area.

Spatial representativeness of datasets. Likewise karst landscape areas, karst springs are not evenly distributed globally. Consequently, it is important to ensure that the WoKaS database is representative of karst's landscape distribution. Therefore, we compared the frequency of karst areas over all latitudes with the frequency of spring locations over the same latitudes. Using 30° grid steps, we found that the distribution of karst areas resembles the distribution of WoKaS spring locations (Fig. 4) but with a considerable bias towards the wealthier and data-rich regions of Europe and North America. At those latitudes (30°N–60°N), we found approximately 50% of the total global karst area and 80% of the WoKaS datasets. More (financial) resources and attention have been directed towards hydrological studies and monitoring in these regions, which is a common and well-known problem of the global representativeness of experimental hydrology[25]. We expect that future experimental works and research collaborations will allow for compensating this imbalances. In some regions, notably the Middle East and China, information relating to hydrological monitoring are considered confidential and only few authorized people can have access to them. We hope that more open data policies will improve access to this data in the future to increase the benefits of scientific exchange for both the research communities and societies.

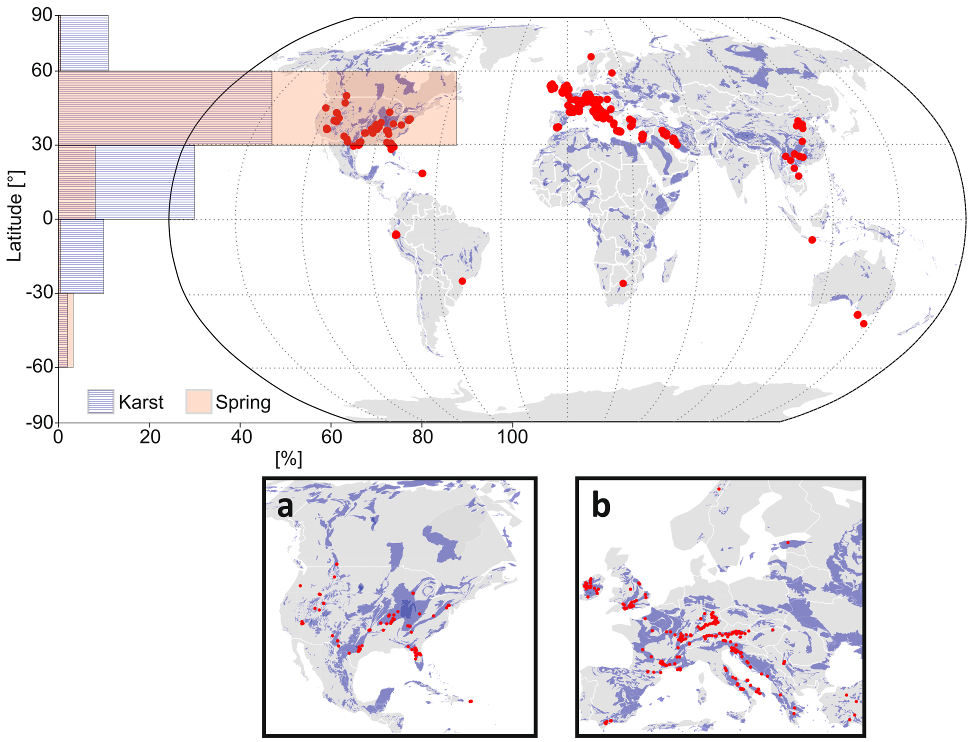

Fig. 4 Global coverage of karst spring discharge observation datasets. Red points on the globe represent WoKaS spring locations and blue areas are the carbonate rock outcrops from WOKAM. The frequencies of WoKaS spring and carbonate rock area distributions across the latitudes are respectively represented by the transparent red and blue bars on the horizontal histogram. Maps insert below are zoom plots of North America (**a**) and Europe (**b**).

Usage Notes

Data repository organisation. WoKaS is a compilation of volumetric karst spring discharge observation datasets from different sources (see subsection data acquisition). The data repository[23] holds a single packaged zip archive named "WoKaS_Data_Record", which contains a "read_me" text file that provides guide information for users and four subfolders: WoKaS_Hydrograph_Metafile, WoKaS_Hydrograph_Datasets, Auto_Download_Routine and WoKaS_Summary_Plot. The contents of the files and subfolders contained in the zip archive are described below:

(A) WoKaS_Hydrograph_Metafile contains a xlsx file "WoKaS_Metafile" which provides a summary of all WoKaS datasets attributes. The file includes information on the dataset's country's name, assigned WoKaS identifier, spring's name, local gauge station identifier which is the assigned identifier in the database of origin of dataset, spring's coordinates, spring's discharge observation length (years), quality flags (as described in subsection accuracy and quality of datasets), dataset's source name and the source type which indicates online or offline accessibility of the datasets (see subsection data acquisition and Table 1). Discharge observation measured at gauge stations located farther downstream of the spring's outlet are usually influenced by superficial flow. When information about the influence of superficial flow is available, it is provided in the "Additional information" section of the metafile. A more comprehensive explanation of the used terms and content of the metafile is provided in the "read_me" document file included in the zip file archive.

(B) WoKaS_Hydrograph_Datasets includes over 400 CSV files of karst spring discharge measurements in cubic metres per second (m^3/s). Headers providing meta-information such as the source of the dataset, spring's name, local gauge station identifier, location coordinates in WGS 84 as well as measurement time format are appended to the csv files. The discharge observations provided in the data repository[23] can be static or dynamic. Datasets obtained from research partners, collaborators and publications are static because they are not updated periodically. Conversely, datasets from hydrological databases (see subsection hydrological agencies) are dynamic and periodically updated through the individual agencies. For users who want the updated datasets, they can be obtained directly from the source online database via an automatic download routine (see below). As described in the Methods sections, for some sources (Banque Hydro, eHYD, LUBW and NRFA), complete datasets at higher temporal resolution are only obtainable through the automatic download routine (see information in "Additional information" column of the metafile).

(C) Auto_Download_Routine includes the R script files for downloading the dynamic datasets from the hydrological agencies online databases. The R codes allow the user to access and download the most recent version of the discharge datasets in their original temporal resolution from the online databases of the hydrological agencies. The downloaded datasets from the different online databases are standardised, having same format as described in "B" above. The newly downloaded version of the dynamic datasets are saved in WoKaS_Dynamic_Datasets folder, which is automatically created while the download routine code is run. In case of changes within the hydrological databases online access link system, the R codes might stop working. However, the R code will be frequently maintained and an updated version will be available on GitHub (https://github.com/KarstHub/WoKaS).

(D) WoKaS_Summary_Plot subfolder contains a pdf file also named "WoKaS_Summary_Plot which includes the hydrograph plots of all the spring discharge datasets. The name of each plot is the name of the corresponding dataset contained in the WoKaS_Hydrograph_Datasets subfolder.

Datasets naming convention. The naming convention used for the datasets is a combination of the International Organisation for Standardisation Alpha-2 (ISO 2) country's code, and a four-digit serial number followed by the spring's name. The ISO 2 code and four-digit serial number are separated by a hyphen "-" and an "@" sign between the serial number and the spring's name. For example, a dataset with the name "FR-0050@ Cent-Fonts" means:

'FR' = ISO-2 country's code for France
'0050' = WoKaS database assigned serial number
'Cent-Fonts' = Name of the spring.

Recommended usage for datasets. Based on the assigned quality classes for the datasets, we provide recommendation on the usage and application of the datasets. The "very high" and "high" quality datasets (Class A and B) are appropriate for all hydrological analyses including statistical and comparative analyses, model evaluation and calibration, impact studies and process understanding. The C1 datasets are suitable for discharge's trend analysis, event-based process understanding and water balance estimation. It should be noted that human impacts such as groundwater pumping for drinking and irrigation could affect spring discharge and trends can't be solely attributed to climatic and landscape changes. In the comment section of the metafile, information about human impacts are only included when available. C2 and C3 quality datasets are most suitable for analysis that does not require much accuracy, such as computing annual averages or monthly spring discharge variations.

The focus of the WoKaS database is to provide easy access to spring discharge dataset, the present structure of the database does not distinguish among different aquifer recharge processes that fed the karst springs (see subsection accuracy of identified locations). Where autogenic recharge prevails, precipitation infiltrates directly into the aquifer through the carbonate rock surface. Whereas, allogenic recharge is due to inflows from non-carbonate units infiltrating into the aquifer through swallow holes, sinking streams, etc[9]. For users interested in distinguishing the recharge processes, recharge processes can be revealed through comparing carbonate rock outcrops and topographic catchment areas. A comprehensive water balance of the spring catchment area may reveal if the aquifer recharge is entirely feed from the carbonate area or if adjacent non-carbonate areas contribute water, as well. Furthermore, allogenic recharge is often associated with sinking streams and disappearing rivers, information on stream density and discontinuities of river networks[26] can provide evidence of allogenic recharge.

Outlook. Presently, access to WoKaS datasets is possible through the figshare repository[23]. In future, we hope that the database can be integrated into a web application platform for visualisation, further data uploads, and easier download.

Code availability

The R code to download datasets directly from the hydrological databases and to combine them with the spring discharge time series obtained from the other sources (see above) is available at https://github.com/KarstHub/WoKaS. The code is provided in R programming language version 3.5.0, and commented following a recommended programming comment guidelines[24]. Comprehensive instructions on how to run the code and system requirements are provided by a "README" file included in the GitHub repository.

Received: 9 September 2019; Accepted: 3 December 2019;
Published online: 20 February 2020

References

1. Stevanović, Z. Global distribution and use of water from karst aquifers. *Geol. Soc. London, Spec. Publ.* **466**, 217–236 (2018).
2. Martos-Rosillo, S. *et al*. Review on groundwater recharge in carbonate aquifers from SW Mediterranean (Betic Cordillera, S Spain). *Environ. Earth Sci.* **74**, 7571–7581 (2015).
3. Bakalowicz, M. Karst groundwater: A challenge for new resources. *Hydrogeol. J.* **13**, 148–160 (2005).
4. Chen, Z. *et al*. The World Karst Aquifer Mapping project: concept, mapping procedure and map of Europe. *Hydrogeol. J.* **25**, 771–785 (2017).
5. Andreo, B. *et al*. Methodology for groundwater recharge assessment in carbonate aquifers: Application to pilot sites in southern Spain. *Hydrogeol. J.* **16**, 911–925 (2008).
6. Hartmann, A., Goldscheider, N., Wagener, T., Lange, J. & Weiler, M. Karst water resources in a changing world: Review of hydrological modeling approaches. *Rev. Geophys.* **52**, 218–242 (2014).
7. Ravbar, N. & Šebela, S. The effectiveness of protection policies and legislative framework with special regard to karst landscapes: Insights from Slovenia. *Environ. Sci. Policy* **51**, 106–116 (2015).
8. Kiraly, L. Karstification and groundwater flow. *Speleogenes. Evol. karst aquifers* **1**, 26 (2002).
9. Goldscheider, N. & Drew, D. *Methods in Karst Hydrogeology. International Contributions to Hydrogeology 26, International Association of Hydrogeology*. (Taylor & Francis, London, 2007).
10. Ford, D. & Williams, P. *Karst Hydrogeology and Geomorphology*. (John Wiley and Sons, Ltd, 2007).
11. Gupta, H. V. *et al*. Large-sample hydrology: A need to balance depth with breadth. *Hydrol. Earth Syst. Sci.* **18**, 463–477 (2014).

12. Döll, P., Kaspar, F. & Lehner, B. A global hydrological model for deriving water availability indicators: model tuning and validation. *J. Hydrol.* **270**, 105–134 (2003).
13. Duan, Q. *et al*. Model Parameter Estimation Experiment (MOPEX): An overview of science strategy and major results from the second and third workshops. *J. Hydrol.* **320**, 3–17 (2006).
14. Beck, H. E. *et al*. Global patterns in base flow index and recession based on streamflow observations from 3394 catchments. *Water Resour. Res.* **49**, 7843–7863 (2013).
15. Stahl, K. *et al*. Streamflow trends in Europe: Evidence from a dataset of near-natural catchments. *Hydrol. Earth Syst. Sci.* **14**, 2367–2382 (2010).
16. Berghuijs, W. R., Hartmann, A. & Woods, R. A. Streamflow sensitivity to water storage changes across Europe. *Geophys. Res. Lett.* **43**, 1980–1987 (2016).
17. Hartmann, A. *et al*. A large-scale simulation model to assess karstic groundwater recharge over Europe and the Mediterranean. *Geosci. Model Dev* **8**, 1729–1746 (2015).
18. Hartmann, A., Gleeson, T., Wada, Y. & Wagener, T. Enhanced groundwater recharge rates and altered recharge sensitivity to climate variability through subsurface heterogeneity. *Proc. Natl. Acad. Sci. USA* **114**, 2842–2847 (2017).
19. Chen, Z. *et al*. World Karst Aquifer Map (WHYMAP WOKAM) BGR, IAH, KIT, UNESCO, https://doi.org/10.25928/b2.21_sfkq-r406 (2017).
20. Jourde, H. *et al*. SNO KARST: A french network of observatories for the multidisciplinary study of critical zone processes in karst watersheds and aquifers. *Vadose Zo. J.* **17**, 180094 (2018).
21. Mazzilli, N. *et al*. KarstMod: A modelling platform for rainfall - discharge analysis and modelling dedicated to karst systems. *Environ. Model. Softw.* **122**, 103927 (2019).
22. Ollivier, C. *et al*. A QGIS Plugin Based on the PaPRIKa Method for Karst Aquifer Vulnerability Mapping. *Groundwater* **57**, 201–204 (2019).
23. Olarinoye, T. *et al*. Global karst springs hydrograph dataset for research and management of the world's fastest-flowing groundwater. *figshare*, https://doi.org/10.6084/m9.figshare.9638939.v2 (2019).
24. Johnson, R. K. *The elements of MATLAB Style*. (Cambridge University Press, 2010).
25. Burt, T. P. & McDonnell, J. J. Whither field hydrology? the need for discovery science and outrageous hydrological hypotheses. *Water Resour. Res.* **51**, 5919–5928 (2015).
26. Lehner, B., Verdin, K. & Jarvis, A. New global hydrography derived from spaceborne elevation data. *Eos, Trans. Am. Geophys. Union* **89**, 93–94 (2008).

Acknowledgements
Support to A.H. and T.O. was provided by the Emmy Noether Programme of the German Research Foundation (DFG; grant no. HA 8113/1-1; project "Global Assessment of Water Stress in Karst Regions in a Changing World"). V.M. was supported by the Innovation Fund of Freiburg University and RiSC of the Ministry for Science, Research and Art of Baden-Wuerttemberg. The authors appreciate Simon Brenner, Laura Vecera, Robin Schwemmle, Mirjam Scheller and Justine Berg for helping with hydrographs digitisation. The authors also appreciate the contributions Laurent Danneville of Parc Naturel Régional des Grands Causses (PRNGC), France; William Santini and Pascal Fraisy who are part of the field and data acquisition team in Peru, which was funded by LMI IRD Paleotraces, SO HYBAM and FONDECYT/CONCYTEC (grant contract 226–2015-FONDECYT). GC was supported by NERC MaRIUS: Managing the Risks, Impacts and Uncertainties of droughts and water Scarcity, grant number NE/L010399/1. DB and GC refer to the UNMIX project, which is supported by the Deutsche Forschungsgemeinschaft (DFG) through the TUM International Graduate School for Science and Engineering (IGSSE), GSC 81. Gabriele Chiogna acknowledges the support of the Stiftungsfonds für Umweltökonomie und Nachhaltigkeit GmbH (SUN). The Spanish team were supported by CGL2015-665858 project of Spanish Ministry of Science and Research Group RNM-308 of Junta de Andalucía. Part of the French data were monitored within the framework of the KARST observatory network (www.sokarst.org) initiative from the INSU/CNRS, which aims to strengthen knowledge-sharing and promote cross-disciplinary research on karst systems. The data collection was kindly supported by the Karst Commission of the International Association of Hydrogeologists. The article processing charge was funded by the German Research Foundation (DFG) and University of Freiburg in the funding programme Open Access Publishing.

Author contributions
T.O. managed the collection and processing of datasets for WoKaS database since 2018. V.M. conducted the collection of the datasets between 2015 and 2017. T.O. wrote the first draft of the manuscript, with the help of A.H. and T.G. T.O. wrote the automatic download routine codes and S.S. tested the compatibility of the code for various computer operating systems. T.O. developed the database structure and data processing methods supervised by A.H. All other co-authors contributed data to the WoKaS database and reviewed the original manuscript before final revision.

Competing interests
The authors declare no competing interests.

Additional information
Correspondence and requests for materials should be addressed to T.O.

Tunde Olarinoye[1 ✉], Tom Gleeson[2], Vera Marx[1], Stefan Seeger[3], Rouhollah Adinehvand[4], Vincenzo Allocca[5], Bartolome Andreo[6], James Apaéstegui[7,8], Christophe Apolit[9], Bruno Arfib[10], Augusto Auler[11], Vincent Bailly-Comte[12], Juan Antonio Barberá[6], Christelle Batiot-Guilhe[13], Timothy Bechtel[14], Stephane Binet[15], Daniel Bittner[16], Matej Blatnik[17], Terry Bolger[18], Pascal Brunet[13], Jean-Baptiste Charlier[19], Zhao Chen[20], Gabriele Chiogna[16,21], Gemma Coxon[22], Pantaleone De Vita[5], Joanna Doummar[23], Jannis Epting[24], Perrine Fleury[12], Matthieu Fournier[25], Nico Goldscheider[20], John Gunn[26], Fang Guo[27], Jean Loup Guyot[28], Nicholas Howden[29], Peter Huggenberger[24], Brian Hunt[30], Pierre-Yves Jeannin[31], Guanghui Jiang[27], Greg Jones[32], Herve Jourde[13], Ivo Karmann[33], Oliver Koit[34], Jannes Kordilla[35], David Labat[36], Bernard Ladouche[19], Isabella Serena Liso[37], Zaihua Liu[27], Jean-Christophe Maréchal[12], Nicolas Massei[25], Naomi Mazzilli[38], Matías Mudarra[6], Mario Parise[37], Junbing Pu[27], Nataša Ravbar[17], Liz Hidalgo Sanchez[39], Antonio Santo[40], Martin Sauter[35], Jean-Luc Seidel[13], Vianney Sivelle[36], Rannveig Øvrevik Skoglund[41], Zoran Stevanovic[42], Cameron Wood[32], Stephen Worthington[43] & Andreas Hartmann[1,29]

[1]Chair of Hydrological Modeling and Water Resources, University of Freiburg, 79098, Freiburg, Germany. [2]Department of Civil Engineering, University of Victoria, Victoria, Canada. [3]Chair of Hydrology, University of Freiburg, 79098, Freiburg, Germany. [4]Department of Earth Sciences, Shiraz University, Shiraz, Iran. [5]Dipartimento di Scienze della Terra, dell'Ambiente e delle Risorse, University of Naples Federico II, Napoli, Italy. [6]Department of Geology and Centre of Hydrogeology of the University of Málaga, Málaga, Spain. [7]Instituto Geofísico del Perú, Lima, Peru. [8]Universidad Nacional Agraria La Molina, Maestria en Recursos Hídricos, Lima, Perú. [9]Parc Naturel Régional des Grands Causses (PRNGC), Saint-Léons, France. [10]Aix Marseille Univ, CNRS, IRD, INRA, Coll France, CEREGE, Aix-en-Provence, France. [11]Instituto do Carste/Carste Ciência e Meio Ambiente, Belo Horizonte, Brazil. [12]BRGM, Univ. Montpellier, Montpellier, France. [13]HydroSciences Montpellier (HSM), Univ. Montpellier, CNRS, IRD, Montpellier, France. [14]Earth and Environment, Franklin and Marshall College, Lancaster, Pennsylvania, USA. [15]ISTO, Université d'Orléans, CNRS, BRGM, OSUC, Orléans, France. [16]Faculty of Civil, Geo and Environmental Engineering, Technical University of Munich, Arcisstr. 21, 80333, Munich, Germany. [17]ZRC SAZU Karst Research Institute, Postojna, Slovenia. [18]Cave and Karst Specialist, Vientiane, Laos. [19]BRGM, Univ. Montpellier, Montpellier, France. [20]Institute of Applied Geosciences, Karlsruhe Institute of Technology (KIT), Kaiserstr. 12, 76131, Karlsruhe, Germany. [21]Institute for Geography, University of Innsbruck, Innrain 52, Innsbruck, Austria. [22]School of Geographical Sciences, University of Bristol, Bristol, BS8 1SS, UK. [23]Department of Geology, American University of Beirut, Beirut, Lebanon. [24]Applied and Environmental Geology, Department of Environmental Sciences, University of Basel, Bernoullistr. 32, 4056, Basel, Switzerland. [25]Normandie Univ, UNIROUEN, UNICAEN, CNRS, M2C, 76000, Rouen, France. [26]School of Geography, Earth and Environmental Science, University of Birmingham, Birmingham, UK. [27]Institute of Karst Geology, Chinese Academy of Geological Sciences, Guilin, China. [28]GET Laboratory, Toulouse University/CNRS/IRD, Toulouse, France. [29]Department of Civil Engineering, University of Bristol, Bristol, UK. [30]Barton Springs/Edward Aquifer Conservation District, Austin, Texas, USA. [31]Institut Suisse de Spéléologie et de Karstologie, ISSKA, CH-2301, La Chaux-de-Fonds, Switzerland. [32]Department for Environment and Water, Government of South Australia, Adelaide, Australia. [33]University of São Paulo, São Paulo, Brazil. [34]Institute of Ecology, School of Natural Sciences and Health, Tallinn University, Uus-Sadama 5, 10120, Tallinn, Estonia. [35]Department of Applied Geology, Georg-August-University Göttingen, Goldschmidstr. 3, 37077, Göttingen, Germany. [36]Géosciences Environnement Toulouse (GET) - CNRS – UPS – IRD – CNES, 14 avenue Edouard Belin, 31400, Toulouse, France. [37]Department of Earth and Environmental Sciences, University Aldo Moro, Bari, Italy. [38]INRAE, Avignon Université, EMMAH, F-84000, Avignon, France. [39]LOCEAN Laboratory, Sorbonne-Université/CNRS/IRD/ MNHN, Paris, France. [40]University Federico II, Naples, Italy. [41]Department of Geography, University of Bergen, Bergen, Norway. [42]Faculty for Mining and Geology, University of Belgrade, Belgrade, Serbia. [43]Worthington Groundwater, Ontario, Canada. ✉e-mail: tunde.olarinoye@hydmod.unifreiburg