



OPEN

DATA DESCRIPTOR

A groundwater well database for Brazil (GWDBrazil)

José Gescilam S. M. Uchôa¹, Paulo Tarso S. Oliveira^{1,2}✉, André S. Ballarin^{1,2},
Didier Gastmans³, Jamil A. A. Anache¹, Bridget R. Scanlon⁴, Clyvihk R. Camacho⁵,
Valmor J. Freddo Filho⁵ & Edson C. Wendland¹

Sufficient spatiotemporal *in-situ* groundwater-level measurements are essential for sustainable water management. Despite their importance, lack of harmonized, quality-controlled datasets has hindered large-scale groundwater studies in Brazil. In collaboration with the Geological Survey of Brazil, we present the Groundwater Well Database for Brazil (GWDBrazil), which consolidates and standardizes information from over 351,000 wells, with records dating from 1900 to 2024, including about 450 wells with continuous daily monitoring from 2010 to 2024. Cross-verification steps were applied to ensure data accuracy. GWDBrazil is available in both tabular format and vector points, comprising information such as location, well depth, and well purpose. The dataset also provides data to support integrated surface water and groundwater management, such the distance from each well to the nearest river and aquifer information. This dataset is intended to serve as a valuable resource for researchers, decision-makers, and stakeholders, providing essential information to support comprehensive water management strategies in Brazil.

Background & Summary

Groundwater is critical for meeting human and ecosystem water needs, being the largest available freshwater resource^{1,2}. It provides drinking water to billions of people³, supplies nearly half of global irrigation demand^{4,5}, and sustains ecosystems, including rivers, lakes, wetlands, and other terrestrial habitats^{6,7}. The importance of groundwater is expected to increase due to rising water demand and climate change impacts^{8–10}. This is particularly true for the Southern Hemisphere given the evolving climate impact on surface water sources and the ongoing expansion of the agricultural footprint^{11–13}. For instance, Brazil faces declines in groundwater levels and streamflow leakage due to the highly uneven distribution of water resources and intensified water uses^{14,15}. However, limitations in the spatiotemporal coverage of groundwater monitoring networks continue to constrain our understanding of the spatiotemporal dynamics of groundwater systems^{16,17}, which is essential for effective water resource management^{18,19}.

The potential applications of a groundwater dataset are wide-ranging, including studies on groundwater recharge^{20,21}, seawater intrusion^{22,23}, land subsidence^{24,25}, groundwater dependent ecosystems^{26,27}, vegetation drought resilience²⁸, streamflow depletion^{15,29} and water resource planning^{30,31}. Hence, a user-friendly and reliable dataset may serve as a valuable resource for researchers studying groundwater management, improving understanding and modeling^{32,33}. Besides, it enables researchers to integrate groundwater data into their studies^{34,35}, driving advancements in the understanding of surface water and groundwater resources interactions^{36,37}. Furthermore, such a database may help policymakers develop evidence-based policies and optimize resource allocation and conservation^{38,39}.

Here, we present the Groundwater Well Database for Brazil (GWDBrazil) compiled and standardized from Geological Survey of Brazil projects. This initiative seeks to contribute to a more holistic view of groundwater management and access, in line with previous open groundwater datasets^{40–42}. This paper addresses the current limitation of groundwater dataset in Brazil, which typically require manual well-by-well handling and lack a harmonized format^{43,44}. These limitations can be time-consuming and challenging to work with, particularly for researchers outside of the groundwater field, for policymakers, and for local communities. Our harmonized

¹Department of Hydraulics and Sanitation, São Carlos School of Engineering, University of São Paulo, São Carlos, SP, Brazil. ²Faculty of Engineering, Architecture and Urbanism, and Geography, Federal University of Mato Grosso do Sul, Campo Grande, MS, Brazil. ³Environmental Studies Center, São Paulo State University, Av. 24A, 1515, Bela Vista, Rio Claro, SP, Brazil. ⁴Bureau of Economic Geology, Jackson School of Geosciences, University of Texas at Austin, Austin, TX, USA. ⁵Geological Survey of Brazil, Rio de Janeiro, Brazil. ✉e-mail: paolutarsoms@gmail.com

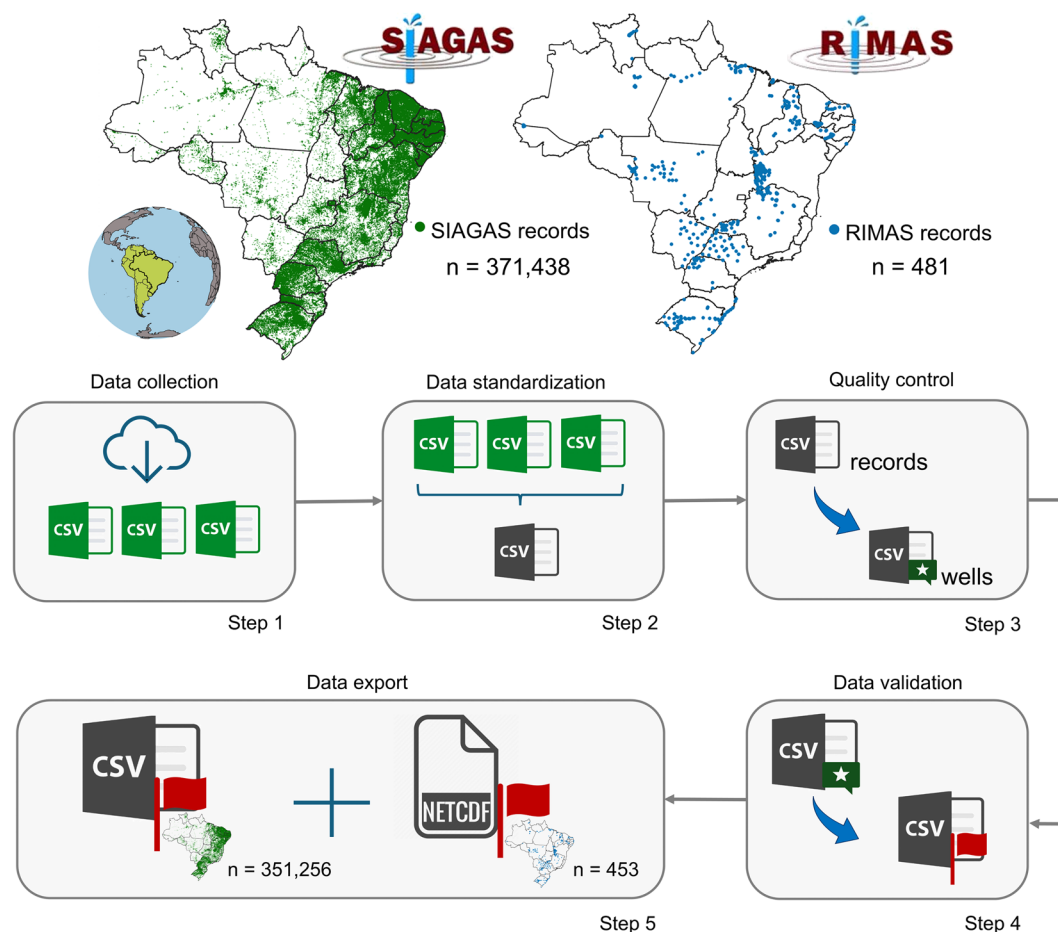


Fig. 1 Overview of the process to develop the groundwater well data. Data were collected from two projects of Geological Survey of Brazil: Groundwater Information System (SIAGAS)⁴³ and Integrated Groundwater Monitoring Network Project (RIMAS)⁴⁴. The data were harmonized (step 2) and then quality control was applied (step 3) to ensure that the records are accurate groundwater wells. When possible, groundwater well data were cross-verified and marked with a quality flag (step 4). In the end, the data were made available in tabular (csv) and geospatial formats (shapefile, and netCDF for monitoring wells). All of the steps were reviewed by members of the Geological Survey of Brazil. The inset map (top left) highlights Brazil within South America in green.

dataset, validated by the Geological Survey of Brazil, underwent quality assurance and quality control procedures to ensure accuracy, adhering to the principles of transparency and data integrity. The dataset is available in both tabular (csv) and geospatial (shapefile) formats for 351,256 wells, including about 450 continuously monitored wells (2010–2024, also in netCDF format). Ultimately, we hope that this accessible database fosters collaboration across the fields of groundwater, surface water hydrology, and water policy and management^{33,45}.

Methods

The development of our groundwater well database includes five initial steps: data collection, data standardization, quality control, data validation, and data export. These steps are summarized in Fig. 1 and explained in detail in this section. After finalizing the initial groundwater well database, we also added additional variables from remote sensing products and other databases that may be useful for researchers and managers, particularly for comprehensive studies of surface water and groundwater resources.

Data sources. The groundwater well database presented here was built in collaboration with the Geological Survey of Brazil (SGB – in Portuguese: *Serviço Geológico do Brasil*) through the Groundwater Information System (SIAGAS – in Portuguese: *Sistema de Informações de Águas Subterrâneas*)⁴³ and the Integrated Groundwater Monitoring Network Project (RIMAS – in Portuguese: *Rede Integrada de Monitoramento das Águas Subterrâneas*)⁴⁴. The Geological Survey of Brazil granted us explicit permission to publish their groundwater well records and had multiple opportunities to review this database.

SIAGAS is a repository primarily consisting of water wells, with data typically provided by several different drilling companies or state agencies, which may result in some inconsistencies and a non-harmonized database⁴³. The Geological Survey of Brazil is responsible for storing the data, but the accuracy of the information remains the responsibility of the data providers. This database contains five data categories: (i) general data,

Type	Header in the study	Definition	NA values	Data Type
General data	ID_well	Unique well ID	0.0%	Integer
	Latitude	Well latitude	0.0%	Float
	Longitude	Well longitude	0.0%	Float
	ID_city_IBGE	IBGE City ID	0.0%	Integer
	City	Brazilian city where the well is located	0.0%	String
	State	Brazilian state where the well is located	0.0%	String
	Types_of_wells	Type of well	5.2%	String
	Well_status	Well status	40.8%	String
	Well_water_use	Primary use of the groundwater	43.4%	String
	Surface_elevation	Surface elevation of the well [m.a.s.l]	77.9%	Float
	Year_reported	Year in which it was entered into SIAGAS system	4.1%	String
	Last_updated_reported	Last updated reported into SIAGAS system	82.4%	String
Drilling data	Drilling_reported	Drilling date	27.1%	String
	Well_depth	Well depth [m]	21.1%	Float
Pumping data	Pumping_test_date	Pumping test date	56.4%	String
	Static_water_level	Static water level [m]	57.2%	Float
	Dynamic_water_level	Dynamic water level [m]	65.7%	Float
	Well_capacity	Well capacity [m ³ /h]	61.7%	Float
Aquifer data	Aquifer_broad	General aquifer, or aquifer systems, that the well overlays	69.2%	String
	Aquifer_confinement	Aquifer confinement	75.8%	String
Lithological data	Lithological_data	Lithological records (Yes, NA)	65.0%*	String
Water quality data	Water_quality_data	Water quality data records (Yes, NA)	66.4%*	String

Table 1. Overview of the variables extracted from the Groundwater Information System (SIAGAS) project. The table includes the type, definition, and data type of the variables, following the original classification from the Geological Survey of Brazil. NA values (missing data) calculated in this study. *For lithological data and water quality data, NA values were assigned if the data was not provided.

such as well location and purpose; (ii) drilling data, such as well depth; (iii) geological data, such as the pumped aquifer; (iv) pumping test data; and (v) chemical analysis data.

RIMAS is a continuous groundwater level monitoring project focused on Brazil's main sedimentary aquifers⁴⁴. The Geological Survey of Brazil is responsible for both data generation and quality control. Although groundwater quality monitoring is not the primary focus of this project, some wells may include hydrogeochemical analyses, upon request by local authorities. In addition to the information already present in SIAGAS, RIMAS includes two additional data categories: (vi) groundwater level monitoring; and (vii) hydrogeochemical monitoring.

Both databases are updated daily and have been used in several scientific studies^{14,15,46}. However, since these databases require manual well-by-well handling and do not present a fully harmonized dataset, their use can be challenging for users outside of the scientific field. Even for researchers, a complete, updated, quality controlled, and harmonized dataset, in a centrally located archive, can save precious time, avoid duplicating the efforts of other researchers in cleaning the data.

Data collection and standardization. The SIAGAS data collection was completed in March 2024, containing 371,438 records at that time (see Fig. 1). The primary SIAGAS raw data was divided into six raw spreadsheets (in csv format), which can be found in the Zenodo repository⁴⁷. We directly communicated with the members of the Geological Survey of Brazil, who provided insights into the data collection methods, definitions, and notations used. With their support, we aggregated the data into a single spreadsheet using a unique identifier for each well. Table 1 presents the variables extracted from SIAGAS. To enhance the accessibility of the database by the international community, we adopted the Latin-ASCII format, and all dates were transformed into the American format (month/day/year). Empty variables, as well as inconsistent values and dates, were assigned the value "NA". Additional information related to the original SIAGAS header (in Portuguese), and the specific raw spreadsheet each variable is derived from can be found in the supplementary material (see Table S1).

In the general data category, we selected variables such as well location, type (e.g., water well, monitoring well), status (e.g., active or abandoned), and primary water use (e.g., domestic, agriculture or industry). In collaboration with the Geological Survey of Brazil, we standardized this information by merging synonyms and translating terms into those commonly used in international literature. Additionally, we ensured that the well locations are presented in the SIRGAS 2000 datum using decimal degrees. Further details on this standardization process can be found in the supplementary material (see Table S2). Moreover, in the drilling data category, we selected two variables: the drilling date and well depth. Information such as the well screened interval was not included in this database, as it was not explicitly available in the downloaded data. However, this information may be available for some wells on the SIAGAS website⁴³.

In the pumping data category, we collected key information such as the date of the pumping test, the static water level, and the well capacity (see Table 1). Many records lack pumping test data, possibly due to state-owned

Header in the study	Definition	NA values	Data Type
First_year	Year of the first available data	4.0%	String
Final_year	Year of the final available data	4.0%	String
Number_data	Number of available data (daily)	4.0%	Integer
Hydrochemical_monitoring	Hydrochemical monitoring records (Yes, No)	0.0%	String

Table 2. Overview of the variables extracted from the Integrated Groundwater Monitoring Network Project (RIMAS).

companies not providing this information, or because these records were added through field campaigns where general well information was identified, but detailed pumping test data could not be collected. Similarly, most records do not contain data about the aquifer. For most of the records with aquifer data (~96%), only a single overlying aquifer is indicated. In cases where multiple overlying aquifers are present, we present them as “multiple aquifers” in variable ‘Aquifer_broad’ in Table 1. For records, we classified the aquifer as confined if at least one of the overlying aquifers is confined. If none of the overlying aquifers is confined but at least one is semi-confined, the record was classified as semi-confined. Otherwise, we classified the aquifer as unconfined. Information about wells with multiple overlying aquifers can be found in the supplementary material (see Table S3).

Due to the challenges in consolidating and validating information on lithology and water quality, we only indicate whether a record contains such data. A record was marked as containing water quality or lithological data if it includes consistent information. For example, a record is said to contain water quality information if it clearly specifies the sampling date and includes at least one physical or chemical parameter (e.g., pH, electric conductivity, water color, temperature, water turbidity, total suspended solids, settleable solids). Recent local efforts have been made to standardize lithological data in SIAGAS⁴⁸, however, this process requires local aquifer knowledge and validation of lithology data is beyond the scope of this study. We recommend that users who require detailed water quality or lithological data access the raw spreadsheet in the Zenodo repository⁴⁷ or refer to SIAGAS website⁴³.

The RIMAS data collection was completed in July 2024, including 481 records at that time (see Fig. 1). We identified about nine potentially erroneous or corrupt records that, despite being included in the RIMAS record list, contained no data in any category. This may be due to differences in the data download period or the possibility that some information was unavailable at the time. We verified the compatibility of information between the RIMAS and SIAGAS projects. Where there were discrepancies, we prioritized the RIMAS data, as it is maintained under the responsibility of the Geological Survey of Brazil.

In addition to the variables already included in Table 1, we provide the initial and final year of the data timeseries and the amount of daily information each series contains in RIMAS (see Table 2). We identified 19 records that lacked a groundwater data timeseries among the 472 RIMAS records; however, 10 of these 19 records included hydrogeochemical data. Most of these records without a groundwater data timeseries correspond to wells drilled in 2023, which may explain why their groundwater data timeseries are not yet available. Given the variability in hydrogeochemical parameters analyzed for each well, we only indicate whether hydrogeochemical data is present in each record. The original data downloaded from RIMAS can be found in the Zenodo repository⁴⁷.

All steps were validated by members of the Geological Survey of Brazil. Additionally, we visually checked a random sample of 100 records to ensure that the information in Tables 1, 2 is consistent with the data in the downloaded raw spreadsheets and on the SIAGAS and RIMAS websites.

Quality control. This step ensures that the records in the database refer to water or monitoring wells in Brazil, not inferred from surface water, such as springs or wetlands. As a comprehensive definition, we consider a groundwater well a man-made excavation used for extracting or monitoring water in an aquifer⁴¹. To achieve this, we applied four steps commonly used to create well databases^{29,31}: (i) removing records that do not correspond to well construction, (ii) removing duplicate records, (iii) removing records outside the Brazilian territory, and (iv) removing records with unclear construction dates (see Fig. 2). Following the principles of transparency and data integrity, all records removed during this process are included in the supplementary material.

To remove records that do not correspond to well construction (i), we applied a pre-filter during the previous standardization step in the ‘Types_of_wells’ variable (see Table 1). These records were typically springs and lakes (see Table S2), resulting in the removal of 2,696 records. Additionally, 19,388 records had no information (NA values) in this variable, among these, 6,959 records lacked information on the drilling date, pumping test date, or aquifers data, which could confirm that they were wells. Therefore, we excluded these records from our database, resulting in the removal of 9,655 records (2.60% of the total, see Figure S1a). The records removed in this step can be found in the supplementary material (Table S4).

We then checked the remaining 361,783 records for duplicates (ii). If two records were less than 10 meters apart, we compared their information. This threshold accounts for locations errors (e.g., GPS errors), which are typically within a few meters. We identified 14,837 pairs of records less than 10 meters apart. If the records had the same information in the drilling and pumping data categories, we considered them duplicates. We found 129 pairs of duplicate records, none of which were duplicated more than twice. For 9,625 pairs of records, we could not fully compare drilling and pumping data due to missing values. Among these, 8,507 pairs were in the same location and were treated as duplicates as a precaution. This resulted in 2,933 groups of duplicates,

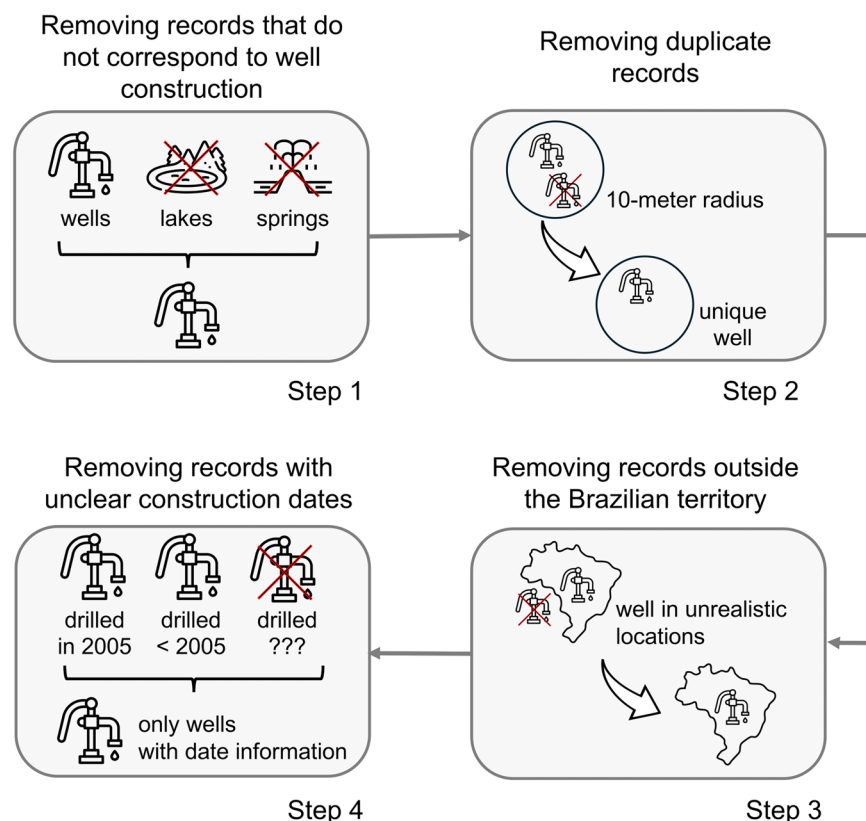


Fig. 2 Overview of the quality control step. The data was structured to exclude any records that did not correspond to water or monitoring wells (Step 1). Then, all wells located within 10 meters of each other were checked to eliminate potential duplicates (Step 2). Next, all wells outside Brazilian territory were removed (Step 3), and wells with missing drilling dates or the date they were entered into the database were also excluded. After these steps, 351,256 records were identified as likely unique wells, including approximately 450 monitoring wells.

meaning a record could be duplicated multiple times. This process led to exclusion of 4,522 records. For the remaining 1,118 pairs of records less than 10 meters apart, but not in the same location, we visually identified 64 pairs as duplicates based on other matching information, such as well depth. The remaining 1,054 pairs were not excluded because there was insufficient evidence to classify them as duplicates. In some cases, wells may have been placed in the same location due to the lack of detailed positional data (e.g., wells located at the center of an industrial park). In the end, preserving the record that was first added into the SIAGAS database, we removed 4,711 duplicate records (1.27% of the total, see Figure S1b). The records removed in this step can be found in the supplementary material (Table S5).

Next, we checked if the remaining 357,072 records were outside Brazilian territory (iii). We found only two records in unrealistic locations (see Figure S1c), likely due to location errors. We then checked whether the remaining 357,070 records included any information that could indicate when they were drilled (iv). A total of 276,145 records contained either the drilling date or the pumping test date. However, the SIAGAS project also consists of wells that were entered into the system through field work. Considering this, we found 351,256 records with information about when they were entered into the database, indicating that these wells were drilled before that date. Since information from these wells can be relevant to policymakers and some scientific applications, we excluded only 5,814 records (1.56% of the total, see Figure S1d) that lacked any date information (see Table S6).

In the end, the quality control steps indicated that 351,256 records (94.57% of the total) are likely unique wells, including 453 monitoring wells. Recent estimates indicate the presence of over 2.5 million tubular wells in Brazil, with most (~88%) operating without official authorization^{49,50} (i.e., without a license or registration for pumping). Among these millions of wells, we believe that the Geological Survey of Brazil projects provide enough information to identify 351,256 wells. This suggests that a substantial contribution of groundwater to water supply in Brazil is unaccounted for in official statistics. In fact, among the wells in this database with a stated purpose (~58%), nearly half are dedicated to domestic use or public water supplies (Fig. 3b), even though agricultural wells likely contribute to high-volume groundwater extraction (Fig. 3c). This unaccounted for water supply may lead to misinterpretations about the country's groundwater usage and dependence, despite reports of localized aquifer overexploitation, induced contamination, and reduced baseflow (groundwater discharge) to local rivers^{51–53}.

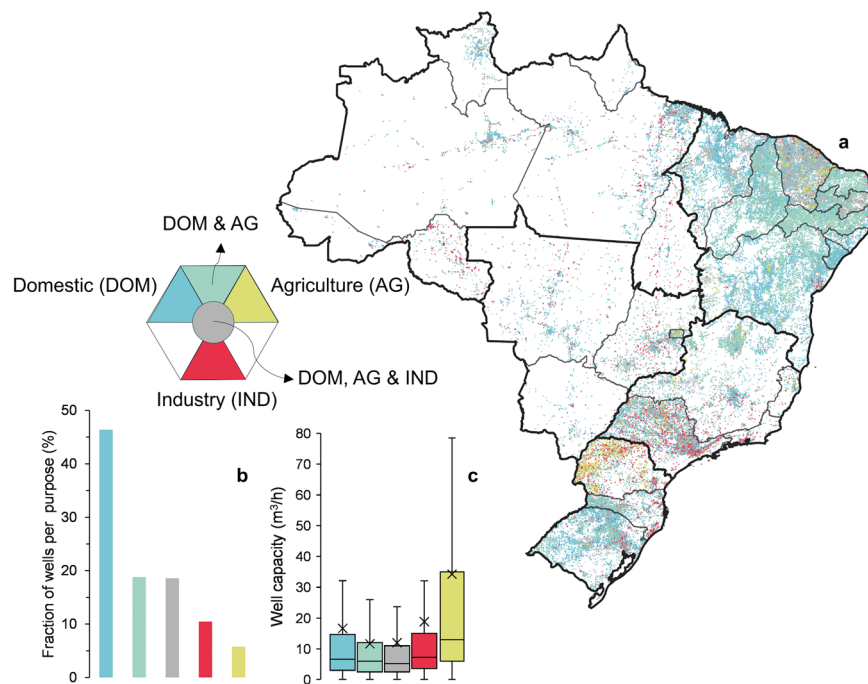


Fig. 3 Overview of the water purpose of the groundwater wells in the database. (a) Each point on the map represents the recorded location of a well. Blue dots represent wells constructed for domestic or public water supply (DOM), green dots for agriculture/ livestock (AG) and red dots for industry (IND). Mixed colors represent wells with more than one use (e.g., DOM & AG; DOM, AG & IND). Wells without an indication of purpose are not shown. The fraction of wells per purpose in Brazil is shown, with the majority primarily dedicated to domestic or public water supply. (b) Well capacity by purpose, indicates that wells for agriculture have the highest average capacity. (c) Box plots indicate 25th and 75th percentiles, bars represent 10th and 90th percentiles, solid horizontal lines indicate median, and x marks mean values. The dark black lines delineate the geographic regions of Brazil, the lighter black lines represent the Brazilian states and the federal district.

Data validation. Data validation of the 351,256 wells included cross-verification checks for 14 of the variables in Table 1: 'Latitude'; 'Longitude'; 'ID_city_IBGE'; 'City'; 'State'; 'Surface_elevation'; 'Year_reported'; 'Last_updated_reported'; 'Drilling_reported'; 'Well_depth'; 'Pumping_test_date'; 'Static_water_level'; 'Dynamic_water_level' and 'Well_capacity'. We used a flagging system to indicate data quality: -1 for inconsistent data, and 0 for raw data (i.e., data where no inconsistencies were identified, or data quality could not be assessed). This system safeguards the integrity of the database, and allows users to apply the appropriate level of rigor depending on their application.

We verified the well's city and state information by cross-referencing with data from the Brazilian Institute of Geography and Statistics⁵⁴, using the well's latitude and longitude. We found discrepancies in the city information for 905 wells (see Figure S2a) and flagged them as inconsistent. Figure 4a shows that cities with the highest number of wells per unit area are concentrated in Northeast Brazil, particularly in metropolitan regions (see Figure S3), consistent with previous local studies^{55,56}. This distribution might be explained by the region's vulnerability to recurring drought events⁵⁷. Similarly, 20 wells showed discrepancies in their state data (see Figure S2b) and were also flagged as inconsistent. Most of these wells with inconsistent city and state data are located near the borders of two cities or states, which may explain their inconsistencies. Outside the Northeast, the highest concentrations of wells by state are found in São Paulo (see Figure S4), a state with a long history of groundwater use^{49,52}.

Next, we verified the latitude and longitude variables by using a 100-meter radius buffer around each well and calculating the percentage of land use and land cover based on the MapBiomas Collection 8 project for 2022^{58,59}. We selected a 100-meter radius to provide a representative value based on the 30 × 30-meter resolution of the land use and land cover dataset. We flagged the location of well as inconsistent if more than 50% of the area was classified as 'water' or 'not observed'. A total of 1,661 wells were classified as having inconsistent locations, with most situated in water bodies or very close to the Brazilian coast (see Figure S5). Some regions in Brazil, such as the Pantanal ecoregion, can be flooded for months, and wells near rivers may be submerged for extended periods⁶⁰. However, as a conservative approach, we still classified these wells' location (latitude and longitude) as inconsistent.

Given the importance of surface elevation for various scientific applications, we compared the database values with the ANADEM product, a digital terrain model for South America based on the Copernicus DEM GLO-30, which includes vegetation bias correction and has a resolution of 1 arc-second⁶¹ (~30 m). If the database value fell outside the ANADEM tolerance range, we flagged them as inconsistent. As a conservative measure, we set the tolerance range as the model bias (1.50 m) plus two times the root mean squared deviation⁶¹ (6.99 m).

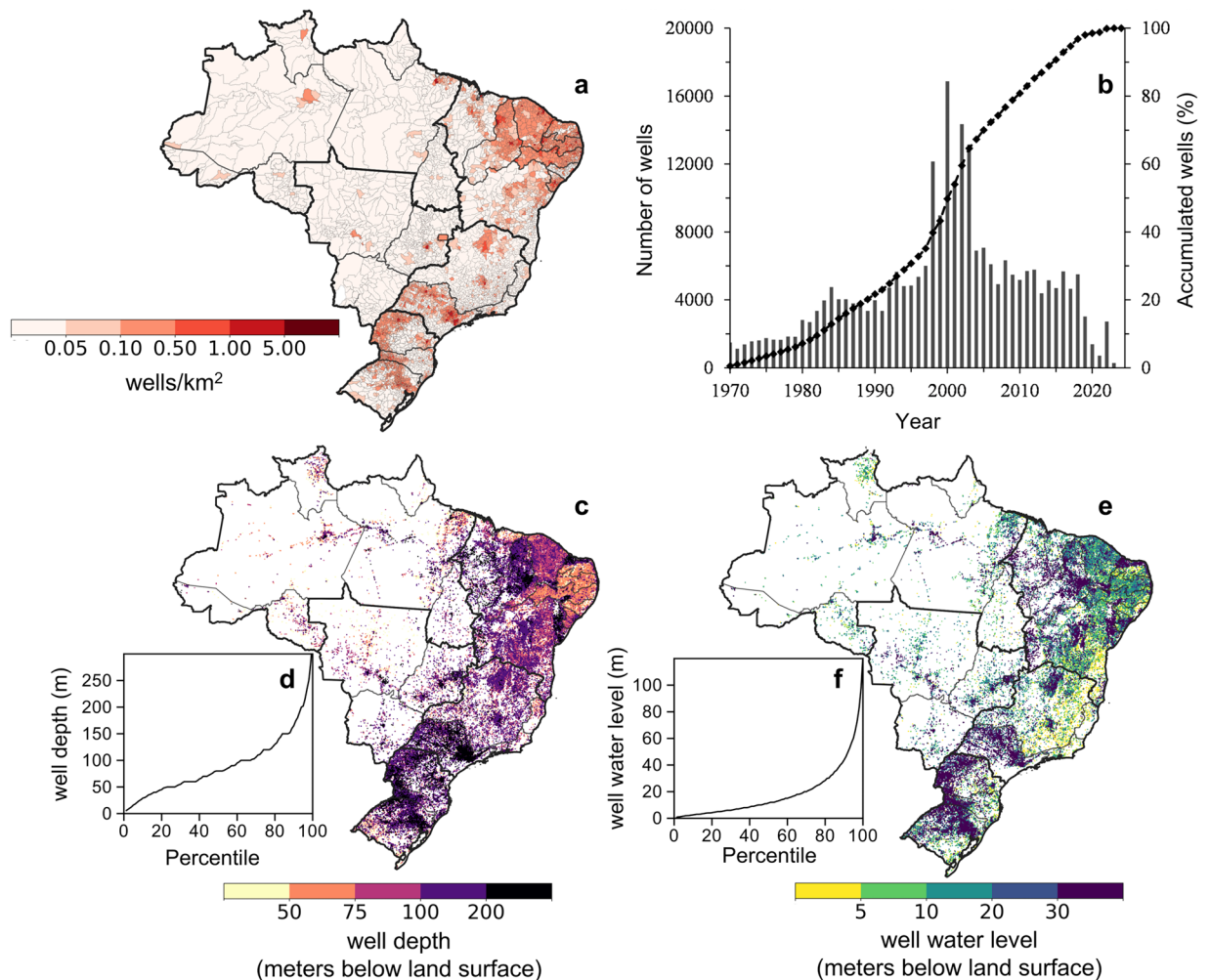


Fig. 4 Heat map showing the well density across Brazilian cities (a), number of wells constructed between 1970 and 2024 (b), well depth distribution (c) and its percentile graph (d), and well water level distribution across Brazil (e) and its percentile graph (f). In (a) and (b), wells with inconsistent city data and drilling date inconsistencies were not included, respectively. In (c) and (e), wells with inconsistent well depths and water levels are not shown, respectively. Only 99% of the data is shown in (d) and (f) for better visualization.

We found 26,379 wells that fell outside the established tolerance range (see Figure S6a). We selected ANADEM due to its high resolution and greater accuracy in Brazil compared to other digital elevation models⁶¹. We also performed a relaxed comparison using the Multi-Error-Removed Improved-Terrain (MERIT) model⁶², which has a resolution of 3 arc-seconds (~90 m), and obtained similar results (see Figure S6b). For wells with surface elevation values within the ANADEM tolerance range, we obtained a median of 475 meters above sea level (see Figure S7).

In addition, we checked the date variables consistency by comparing them with the date the files were downloaded. The year in which the well was registered into SIAGAS system was considered inconsistent if it was later than the downloaded date. We did not find any wells with this inconsistency. Similarly, the last year that the well was updated into SIAGAS system was considered inconsistent if it was earlier than the 'year_reported' or later than the downloaded date, but again, we did not find inconsistencies. The well drilling date was considered inconsistent if it was later than either the 'year_reported' or the downloaded date. We found 1,305 wells with this inconsistency (see Figure S8a). The pumping test date was considered inconsistent if it was later than the 'year_reported' or the downloaded date, or earlier than the well drilled date. We found 10,077 wells with this inconsistency (see Figure S8b), most of which had a pumping test date only a few days earlier than the well drilled date. This may be due to some data consistency issues, but as a conservative approach, we still classified them as inconsistent. Figure 4b shows the drilling dates of wells with consistent and non-empty dates from 1970 to 2024, indicating a strong increase in the number of wells in Brazil around the year 2000. The Geological Survey of Brazil began its activities in the 1970s, so data from periods before this date might not be accurately documented. However, since we lacked the means to verify the consistency of these earlier data, they were classified as raw data.

Moreover, we checked the well depth variable and identified 288 wells with potentially unrealistic depths (see Figure S9a), such as 9,999 meters, which were flagged as inconsistent. Figure 4c shows the well depth data across

Brazil, excluding inconsistent values, and indicates that more than 50% of the wells have depths of less than 80 meters (see Fig. 4d). Similarly, we checked the well water level variable and found 3,433 wells with inconsistent well water level (see Figure S9b), such as values exceeding the well depth. Figure 3e shows the well water level data across Brazil, excluding inconsistent values, and indicates that more than 50% of the wells have a water level shallower than 12 meters (see Fig. 4f), in line with previous studies^{15,46}.

Likewise, we checked the dynamic water level variable and found 121 wells (see Figure S9c) with inconsistent values, including dynamic water level during the pumping test shallower than the static water level or deeper than the well depth. The median dynamic water level, excluding inconsistent values, is around 40 meters (see Figure S10a). Moreover, we checked the well capacity variable and, we flagged 520 wells with zero values or unrealistically high capacities exceeding 1,000 m³/h — substantially higher than values typically reported for major Brazilian aquifers^{63,64}. Furthermore, we identified 3,681 wells with well capacity data but with missing or inconsistent static or dynamic water level. We marked all these 4,182 wells as having inconsistent well capacity, noting that some wells had more than one inconsistency (see Figure S9d). The median well capacity, excluding inconsistent values, is 6 m³/h, a value commonly used by the Brazilian National Water and Sanitation Agency to calculate groundwater availability⁶⁵.

To summarize, a total of 309,564 wells had no inconsistencies, which account for over 88% of the total dataset (see Table S7). Additionally, we could not cross-reference the qualitative variables, as the database is compiled from information provided by various well drillers, government agencies, and field records. However, regarding the type of well, the majority (~96%) can be classified as drilled wells (see Figure S11). For the status of the wells, most (~70%) are classified as active (see Figure S12). In terms of water use, as indicated in Fig. 3, almost half of the wells with a specified water use (~46%) are dedicated to domestic or public water supply. Concerning aquifer data, about one-third of the wells have some information about the aquifer (~31%; see Figure S13a). Among the wells with information on the aquifer confinement status, the majority is located in unconfined aquifers (~71%; see Figure S13b). Regarding lithological data, about one-third of the wells (~36%) include some lithological information (see Figure S14a). Similarly, about one-third of the wells (~34%) contain information on water quality (see Figure S14b).

Among the RIMAS continuous monitoring wells, 453 wells have at least one recorded daily observation, with a median monitoring duration of 3,074 days (Fig. 5a, Table S8). While numerous methods exist for detecting errors and outliers, most require local hydrogeological expertise and manual inspection. Consequently, large-scale studies often bypass this step or rely on subjective assessments, which may introduce biases in trend analyses and groundwater model calibration. To minimize this bias, we employed the HydroSight toolbox⁶⁶, as it does not require adjustments for monitoring frequency, in conjunction with the best available hydrogeological information to systematically flag data.

Each daily observation was assigned one of two flags: an error flag or an outlier flag. The error flag was set to -1 if the data met any of the following criteria: (i) duplicate values, (ii) absolute daily water level changes exceeding 10 meters, (iii) constant head values persisting for more than 30 days, or (iv) physically implausible readings (e.g., values exceeding well depth or negative values). Less than 0.5% of the dataset exhibited such errors (see Fig. 5). The outlier flag was set to -1 for data identified as outliers by the HydroSight toolbox double exponential smoothing time-series model. Less than 1% of the dataset was flagged as a potential outlier (see Fig. 5). The threshold for identifying outliers — expressed as the number of noise standard deviations (η) — ranged from 3 to 6, depending on hydrogeological conditions. Data without detected errors or outliers — or cases where data quality could not be assessed — were assigned a flag of 0.

While the HydroSight toolbox provides an automated and reproducible approach, it has limitations. It may misclassify outliers, with false negatives occurring during periods of low variance and false positives during periods of high variance. A detailed discussion of these limitations is available in the supplementary material (see Figure S15). We recommend that users exclude data flagged as errors and critically evaluate data flagged as outliers based on their specific applications.

Data export. In addition to the traditional variables available in SIAGAS and RIMAS, we incorporated 10 additional key variables from remote sensing products and other databases into this dataset (see Table 3). These variables may be particularly valuable for researchers and water resource managers, especially for integrated studies of surface water and groundwater interactions. When data from a specific product was unavailable at a well's location, an 'NA' value was assigned.

In the climate category, we included long-term mean annual precipitation (Figure S16a), long-term mean annual potential evapotranspiration (Figure S16b), and the ratio of long-term mean annual precipitation to potential evapotranspiration, also known as the aridity index, from 1981 to 2024. These data were obtained from the Brazilian Daily Weather Gridded Data⁶⁷, interpolated from ground observations at a 0.1° × 0.1° (~10 m) grid resolution. A concentration of wells is observed in Brazil's semi-arid region (aridity index < 0.5; see Fig. 6a).

The topography category includes ground elevation data from the ANADEM product, which has a resolution of 1 arc-second⁶¹ (~30 m), and from the MERIT model, which has a resolution of 3 arc-seconds⁶² (~90 m) (see Figure S17). A concentration of wells is found near sea level (ground elevation < 100 m; see Fig. 6b).

In the land use and land cover (LULC) category, we identified the predominant land use and land cover within a 100-meter radius of each well using data from the MapBiomas Collection 8 project for 2022^{58,59}, which has a 30 × 30-meter resolution. Most wells are in areas classified as agricultural (see Fig. 6c), including pasture, croplands, forest plantations, and mixed-use mosaics (see Figure S5 for details on the major land use and land cover classes used in MapBiomas Collection 8).

For the aquifer category, we incorporated permeability and porosity data from the GLObalHYdrogeology MaPS (GLHYMPS)^{68,69}, which provides global lithologies for approximately 1.8 million polygons with an

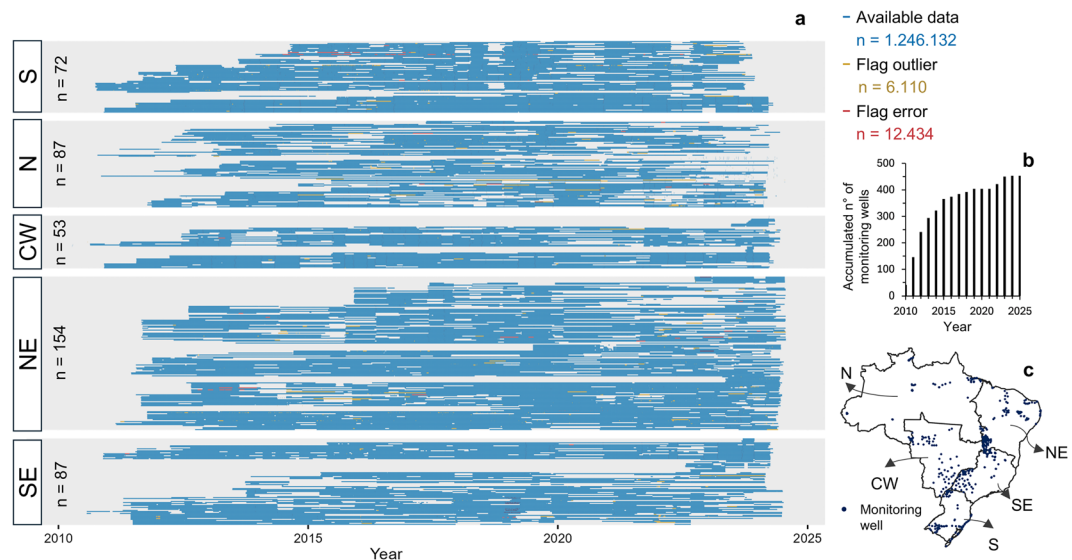


Fig. 5 Heat map showing groundwater level monitoring in RIMAS project across Brazil with flagged potential outliers or errors (a). Evolution of the number of monitoring wells installed by the RIMAS project (b). Brazilian region (S: South, N: North, CW: Central-West, NE: Northeast, SE: Southeast) with monitoring wells (c). The number of wells per region is indicated next to each region in subplot (a). Below the color legend, the total number of raw daily groundwater level observations, potential outlier observations, and observations with errors are indicated.

Category	Header in the study	Definition	Data Type	Source
Climate	Mean_pr_an	Mean annual precipitation (mm)	Float	67
	Mean_eto_an	Mean annual potential evapotranspiration (mm)	Float	67
	Aridity_index	Aridity index (-)	Float	67
Topography	Sur_ele_ANADEM	Surface Elevation from ANADEM (m.a.s.l)	Float	61
	Sur_ele_MERIT	Surface Elevation from MERIT (m.a.s.l)	Float	62
LULC	Main_landuse	Main land use around the well	String	58
Aquifer	Aq_permeability	Log aquifer permeability (m ²)	Float	69
	Aq_porosity	Aquifer porosity [%]	Float	68
	Overlaying_aq	Aquifer outcrop zone overlaid by the well	String	70
Surface waters	Dis_stream	Distance to the closest stream 2nd order or higher (m)	Float	71

Table 3. Overview of the variables added to this database for integrated studies of surface water and groundwater interactions, categorized into five main groups: Climate, Topography, Land Use and Land Cover (LULC), Aquifer, and Surface Waters.

average size of 100 km² (see Figure S18). The distribution of wells does not indicate specific concentration at a particular permeability (see Fig. 6d). Additionally, we included the aquifer outcrop zone overlaid by each well, based on data from the Brazilian National Water and Sanitation Agency⁷⁰ (see Figure S19).

Finally, in the surface waters category, we introduced the distance from each well to the nearest second-order or greater stream using the HydroATLAS database⁷¹. This information is useful for various applications related to integrated surface water and groundwater studies^{72,73}. A concentration of wells is observed within 1 km of the nearest river (see Fig. 6e).

Data Records

The Groundwater Well Database for Brazil (GWDBrazil) includes SIAGAS well data (available in CSV and Shapefile formats) and RIMAS well data (also available in NetCDF format). All products generated in this study are available in the Zenodo repository⁴⁷.

- SIAGAS_data.csv – SIAGAS dataset (see Table 1) without the flagging system.
- SIAGAS_data_flagged – SIAGAS dataset (see Table 1) with the flagging system.
- RIMAS_ID – RIMAS dataset with the flagging system (see ‘Section: Quality Control’).
- Additional_data.csv – Includes 10 additional key variables from remote sensing products and other databases (see Table 3).

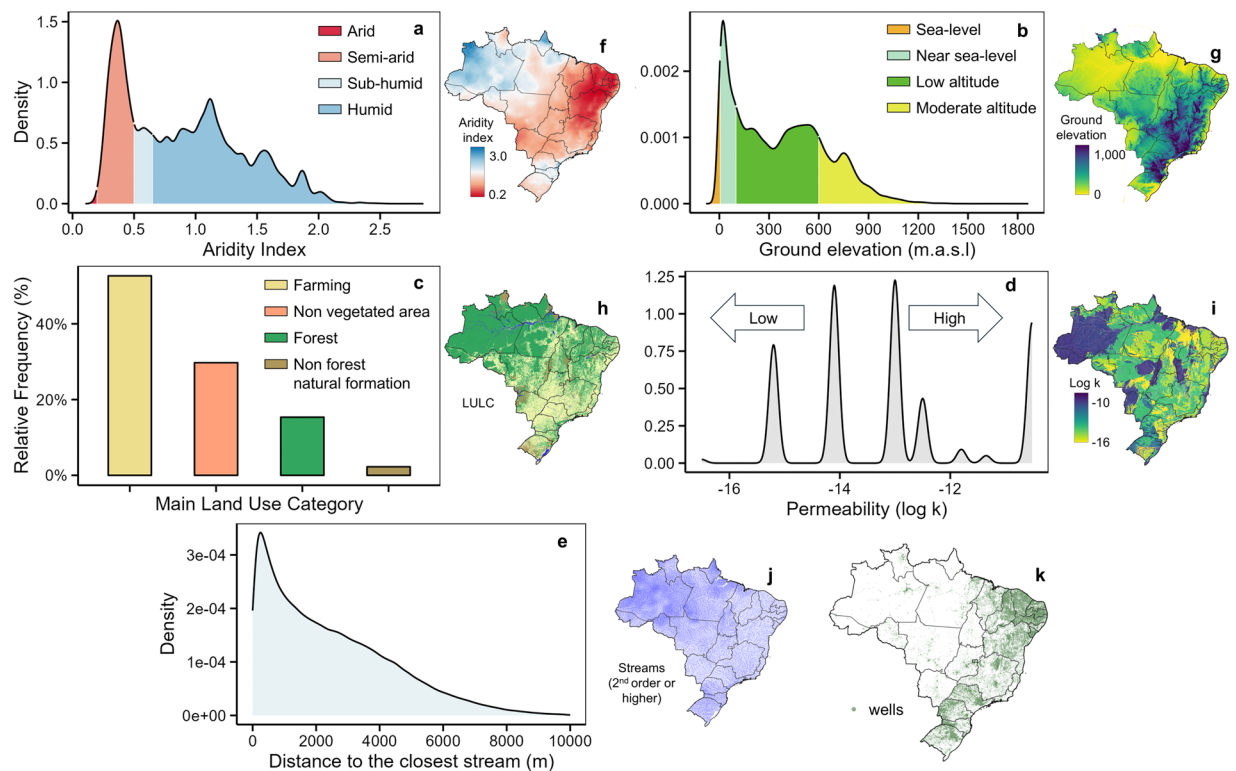


Fig. 6 Probability density distribution of the long-term mean annual aridity index⁶⁷ (a) and of ground elevation per well⁶¹ (b). Predominant land use and land cover (LULC) around each well⁵⁸ (c). Probability density distribution of permeability⁶⁸ (d) and of the distance between each well and the closest stream per well⁷¹ (e). Spatial distribution of the long-term mean annual aridity index (f), and of ground elevation (g), of main LULC types (h), of permeability (i), of streams (j), and wells included in this database across Brazil (k). Aridity index classes follow the classification from the Food and Agriculture Organization of the United Nations⁸². Elevation classes are based on the Brazilian Institute of Geography and Statistics⁸³. The LULC classification is from the MapBiomas Collection 8 project⁵⁸. Permeability is represented on a logarithmic scale for better visualization.

Note that in the SIAGAS well data, we provide as many variables as possible (see Table 1), even when some contain a high proportion of missing values (NAs). The percentage of NA values for each variable in the Zenodo repository version of the database, i.e., after excluding all inconsistent data identified during our validation process, is presented in the supplementary material (see Table S7). To further support users and facilitate data exploration, we also present the number of wells and the percentage of NA values per variable for each Brazilian state and the Federal District, and by the decade in which the well was entered into the SIAGAS system (see Tables S9–S35). This allows users to quickly assess data availability for their region and intended study period. For a complete description of all available spreadsheets, refer to the supplementary material (see ‘S1. Description of Supplementary Files’ and ‘S2. Supplementary Tables’).

Technical Validation

During the construction of this well database, a series of steps were taken to ensure data harmonization (see ‘Section: Data collection and standardization’), followed by additional steps to verify that the records in our database are indeed wells (see ‘Section: Quality control’). For quantitative variables, cross-checks were performed, and a flag system was implemented to indicate potential inconsistencies (see ‘Section: Data validation’). Additionally, for daily monitoring data, a flagging system was introduced to identify potential errors and outliers (see ‘Section: Data validation’).

In this section, we qualitatively compare the classification of public water supply by city, as provided by the Brazilian National Water and Sanitation Agency⁵⁶, with the wells classified in this database as domestic or public water supply wells (see Fig. 7). We found that cities classified as having exclusively surface water public supply have domestic wells. However, these cities exhibit the lowest median number of domestic wells per area compared to cities with mixed water sources or those that rely exclusively on groundwater (see Fig. 6d–f).

Next, we analyze trends in monitoring wells with long-term records free of flagged errors or outliers, comparing them with regional and local studies. As an example of this flagged monitoring data, we selected 77 wells from the 453 monitoring wells in Brazil that have over seven consecutive years of data with gaps of less than 15 consecutive days, excluding flagged data. We applied the nonparametric Theil–Sen slope estimator to assess long-term trends in these wells^{74,75}. The statistical significance of trends was evaluated using the Mann–Kendall test at a 0.05 significance level⁷⁶.

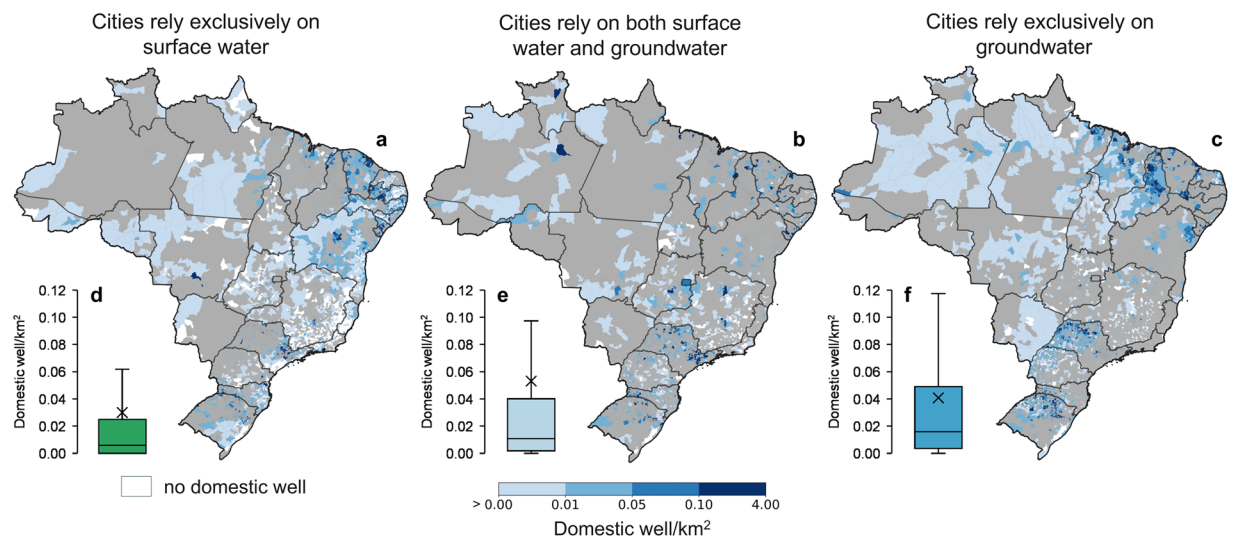


Fig. 7 Domestic wells per area in cities that rely exclusively on surface water (a), both surface water and groundwater (b), and exclusively on groundwater (c), based on the classification from the Brazilian National Water and Sanitation Agency⁵⁶. Boxplots of domestic wells per area unit for cities in each category (d–f) with boxes indicating 25th and 75th percentiles and bars representing 10th and 90th percentiles, solid horizontal lines denoting median, and x marks indicating mean value.

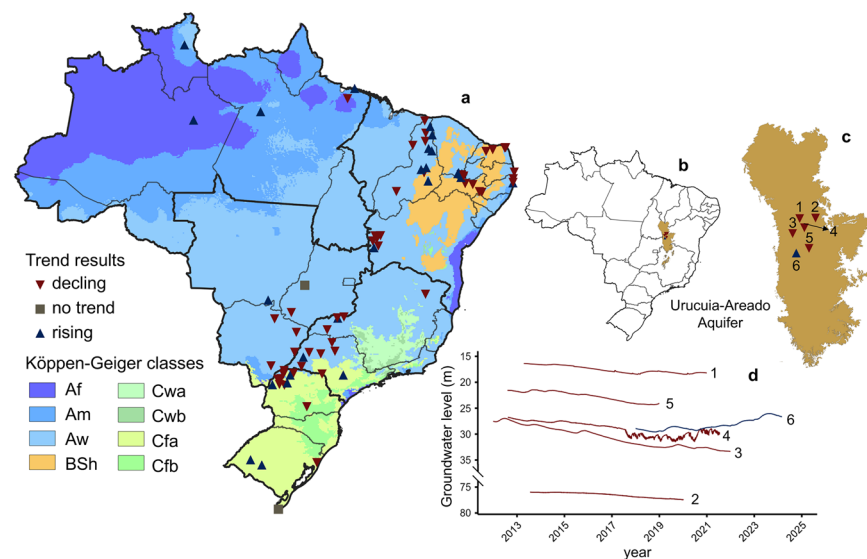


Fig. 8 Trend analysis results for RIMAS monitoring wells with more than seven consecutive years of daily observation across Brazil and their distribution within Brazilian Köppen-Geiger climate classifications⁸⁴ (Af: Tropical rainforest; Am: Tropical monsoon; Aw: Tropical savanna; BSh: Arid steppe, hot; Cwa: Temperate, dry winter, hot summer; Cwb: Temperate, dry winter, warm summer; Cfa: Temperate, no dry season, hot summer; Cfb: Temperate, no dry season, warm summer); (a). Location of the Urucua-Areado Aquifer in Brazil⁷⁰ (b). (c) Detailed view of six RIMAS monitoring wells (1: id2900020674; 2: id2900020672; 3: id2900021798; 4: id2900024872; 5: id2900024877; 6: id2900024879) within the Urucua-Areado Aquifer (c). Groundwater levels are displayed as depth below ground level (d). Note that the groundwater level axes do not start at 0 m, and a break is introduced between 35 m and 75 m to enhance data visualization.

The results indicate that 49 wells exhibit a declining trend, 26 show a rising trend, and two display no significant trend (see Fig. 8a). Wells with long-term declining trends are primarily concentrated in specific regions of the country. As an example, we analyzed the Urucua-Areado Aquifer, which includes six monitoring wells. Of these, five exhibit significant declining trends (see Fig. 8d), aligning with previous studies that indicate a drop in the water table within this aquifer^{14,77}.

Uncertainty and future needs. This study is the first to publicly release a nationwide, standardized dataset describing both water and monitoring wells in Brazil. Because SIAGAS data primarily consist of water wells, inconsistencies in drilling reports may exist and might not have been fully identified during the quality control and validation steps. However, given the high cost of maintaining a long-term monitoring network^{78,79}, many large-scale studies integrate water well data with monitoring well data to enhance spatial coverage^{15,29,46}. Therefore, we believe that this integration is valuable for advancing sustainable water management in Brazil.

Additionally, many well records are not routinely updated, meaning that unreported modifications to wells may not be reflected in the data. Despite this, our database includes both the date each well was entered into SIAGAS and the most recent update date (see Table 1) to provide users with relevant temporal information.

The Geological Survey of Brazil continuously receives data on new and existing wells, meaning our database represents the most up-to-date information available at the time of collection. However, further investments are needed to expand and maintain groundwater monitoring efforts, and strategies should be implemented to regularize undocumented wells, thereby increasing the amount of available data in future updates^{52,79}.

The results presented in Figs. 4, 6 highlight biases and gaps in Brazilian groundwater data, which could inform strategic placement of monitoring wells in priority areas to build a more representative and robust observational network for integrated surface and groundwater management³⁵. Furthermore, this database serves as a foundation for integrating other well datasets, including those from universities, private companies, and Brazilian states, which already contribute data to the Geological Survey of Brazil. Such integration would facilitate the development of a comprehensive and easily accessible national well database, benefiting both the scientific community and decision-makers.

Code availability

Most statistical analyses were conducted using the R language⁸⁰, with additional manual analyses performed (i.e., certain processes required visual inspections), supported by the Geological Survey of Brazil. Geospatial analyses were conducted using the QGIS 3.28.3 software⁸¹. The main R scripts used are available in the Zenodo repository⁴⁷. Analyses presented here do not depend on specific code. The approach can be reproduced following the procedures described in the work.

Received: 30 March 2025; Accepted: 18 August 2025;

Published online: 29 September 2025

References

- Gleeson, T., Wada, Y., Bierkens, M. F. & van Beek, L. P. Water balance of global aquifers revealed by groundwater footprint. *Nature* **488**, 197–200 (2012).
- Bierkens, M. F. & Wada, Y. Non-renewable groundwater use and groundwater depletion: a review. *Environ. Res. Lett.* **14**, 063002 (2019).
- United Nations. The United Nations World Water Development Report 2022: groundwater: making the invisible visible (UNESCO, Paris, 2022) (2022).
- Siebert, S. *et al.* Groundwater use for irrigation – a global inventory. *Hydrology and Earth System Sciences* **14**(10), 1863–1880, <https://doi.org/10.5194/hess-14-1863-2010> (2010).
- Döll, P. *et al.* Impact of water withdrawals from groundwater and surface water on Continental Water Storage Variations. *Journal of Geodynamics*, 59–60, 143–156, <https://doi.org/10.1016/j.jog.2011.05.001> (2012).
- Alley, W. M., Healy, R. W., LaBaugh, J. W. & Reilly, T. E. Flow and storage in groundwater systems. *Science* **296**, 1985–1990 (2002).
- Fan, Y., Miguez-Macho, G., Jobbágy, E. G., Jackson, R. B. & Otero-Casal, C. Hydrologic regulation of plant rooting depth. *Proceedings of the National Academy of Sciences* **114**(40), 10572–10577, <https://doi.org/10.1073/pnas.1712381114> (2017).
- Taylor, R. G. *et al.* Ground water and climate change. *Nature Climate Change* **3**(4), 322–329, <https://doi.org/10.1038/nclimate1744> (2012).
- Famiglietti, J. S. The Global Groundwater Crisis. *Nature Climate Change* **4**(11), 945–948, <https://doi.org/10.1038/nclimate2425> (2014).
- Lutz, A. F. *et al.* South Asian agriculture increasingly dependent on meltwater and groundwater. *Nature Climate Change* **12**(6), 566–573, <https://doi.org/10.1038/s41558-022-01355-z> (2022).
- Potapov, P. *et al.* Global maps of cropland extent and change show accelerated cropland expansion in the twenty-first century. *Nature Food* **3**(1), 19–28, <https://doi.org/10.1038/s43016-021-00429-z> (2021).
- Ballarin, A. S. *et al.* Brazilian water security threatened by climate change and human behavior. *Water Resources Research*, **59**(7), <https://doi.org/10.1029/2023wr034914> (2023).
- Uchôa, J. G. *et al.* Tropical Cities Research Boundaries: A bibliometric analysis to bridge the gaps through multi-dimensional and cross-disciplinary features. *Frontiers in Sustainable Cities*, **5**, <https://doi.org/10.3389/frsc.2023.1154667> (2023).
- Camacho, C. R., Getirana, A., Rotunno Filho, O. C., & Mourão, M. A. Large-scale groundwater monitoring in Brazil assisted with satellite-based Artificial Intelligence Techniques. *Water Resources Research*, **59**(9), <https://doi.org/10.1029/2022wr033588> (2023).
- Uchôa, J. G. *et al.* Widespread potential for streamflow leakage across Brazil. *Nature Communications*, **15**(1), <https://doi.org/10.1038/s41467-024-54370-3> (2024).
- IPCC - Intergovernmental Panel on Climate Change. Climate Change 2022: Impacts, Adaptation and Vulnerability: Contribution of Working Group II to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change (Cambridge University Press) (2022).
- Reinecke, R. *et al.* Uncertainty in model estimates of global groundwater depth. *Environmental Research Letters* **19**(11), 114066, <https://doi.org/10.1088/1748-9326/ad8587> (2024).
- Ballarin, A. S. *et al.* The impact of an open water balance assumption on understanding the factors controlling the long-term streamflow components. *Water Resources Research*, **58**(10), <https://doi.org/10.1029/2022wr032413> (2022).
- Schwaback, D. *et al.* Are Brazilian Catchments Gaining or Losing Water? *The Effective Area of Tropical Catchments*. <https://doi.org/10.22541/au.162384891.15325943> (2021).
- Cuthbert, M. O. *et al.* Observed controls on resilience of groundwater to climate variability in sub-Saharan africa. *Nature* **572**(7768), 230–234, <https://doi.org/10.1038/s41586-019-1441-7> (2019).
- Moeck, C. *et al.* A global-scale dataset of direct natural groundwater recharge rates: A review of variables, processes and relationships. *Science of The Total Environment* **717**, 137042, <https://doi.org/10.1016/j.scitotenv.2020.137042> (2020).

22. Werner, A. D. *et al.* Seawater intrusion processes, investigation and management: Recent advances and future challenges. *Advances in Water Resources* **51**, 3–26, <https://doi.org/10.1016/j.advwatres.2012.03.004> (2013).
23. Jasechko, S., Perrone, D., Seybold, H., Fan, Y., & Kirchner, J. W. Groundwater level observations in 250,000 coastal us wells reveal scope of potential seawater intrusion. *Nature Communications*, **11**(1), <https://doi.org/10.1038/s41467-020-17038-2> (2020).
24. Galloway, D. L. & Burbey, T. J. Review: Regional land subsidence accompanying groundwater extraction. *Hydrogeology Journal* **19**(8), 1459–1486, <https://doi.org/10.1007/s10040-011-0775-5> (2011).
25. Sharifi, A., Khodaei, B., Ahrari, A., Hashemi, H. & Torabi Haghighi, A. Can river flow prevent land subsidence in urban areas? *Science of The Total Environment* **917**, 170557, <https://doi.org/10.1016/j.scitotenv.2024.170557> (2024).
26. Klöve, B. *et al.* Groundwater dependent ecosystems. part I: Hydroecological status and trends. *Environmental Science & Policy* **14**(7), 770–781, <https://doi.org/10.1016/j.envsci.2011.04.002> (2011).
27. Eamus, D., Zolfaghar, S., Villalobos-Vega, R., Cleverly, J. & Huete, A. Groundwater-dependent ecosystems: Recent insights from satellite and field-based studies. *Hydrology and Earth System Sciences* **19**(10), 4229–4256, <https://doi.org/10.5194/hess-19-4229-2015> (2015).
28. Costa, F. R., Schietti, J., Stark, S. C. & Smith, M. N. The other side of tropical forest drought: Do shallow water table regions of Amazonia act as large-scale hydrological refugia from drought? *New Phytologist* **237**(3), 714–733, <https://doi.org/10.1111/nph.17914> (2022).
29. Jasechko, S., Seybold, H., Perrone, D., Fan, Y. & Kirchner, J. W. Widespread potential loss of streamflow into underlying aquifers across the USA. *Nature* **591**(7850), 391–395, <https://doi.org/10.1038/s41586-021-03311-x> (2021).
30. Scanlon, B. R. *et al.* Global Water Resources and the role of groundwater in a resilient water future. *Nature Reviews Earth & Environment* **4**(2), 87–101, <https://doi.org/10.1038/s43017-022-00378-6> (2023).
31. Jasechko, S. *et al.* Rapid groundwater decline and some cases of recovery in aquifers globally. *Nature* **625**(7996), 715–721, <https://doi.org/10.1038/s41586-023-06879-8> (2024).
32. Bierkens, M. F. Global Hydrology 2015: State, trends, and directions. *Water Resources Research* **51**(7), 4923–4947, <https://doi.org/10.1002/2015wr017173> (2015).
33. Condon, L. E. *et al.* Global groundwater modeling and monitoring: Opportunities and challenges. *Water Resources Research*, **57**(12), <https://doi.org/10.1029/2020wr029500> (2021).
34. Fan, Y. Groundwater in the Earth's critical zone: Relevance to large-scale patterns and processes. *Water Resources Research* **51**(5), 3052–3069, <https://doi.org/10.1002/2015wr017037> (2015).
35. Gleeson, T. *et al.* GMD perspective: The quest to improve the evaluation of groundwater representation in continental- to global-scale models. *Geoscientific Model Development* **14**(12), 7545–7571, <https://doi.org/10.5194/gmd-14-7545-2021> (2021).
36. Winter, T. C., Harvey, J. W., Franke, O. L. & Alley, W. M. Ground Water and Surface Water: A Single Resource US Geological Survey Circular 1139, <https://doi.org/10.3133/cir1139> (USGS) (1998).
37. Bierkens, M. F., Sutanudjaja, E. H. & Wanders, N. Large-scale sensitivities of groundwater and surface water to groundwater withdrawal. *Hydrology and Earth System Sciences* **25**(11), 5859–5878, <https://doi.org/10.5194/hess-25-5859-2021> (2021).
38. Konikow, L. F. & Leake, S. A. Depletion and capture: Revisiting “the source of water derived from wells”. *Groundwater* **52**(S1), 100–111, <https://doi.org/10.1111/gwat.12204> (2014).
39. Perrone, D. & Jasechko, S. Deeper well drilling an unsustainable stopgap to groundwater depletion. *Nature Sustainability* **2**(8), 773–782, <https://doi.org/10.1038/s41893-019-0325-z> (2019).
40. IGRAC - International Groundwater Resources Assessment Center. Global Groundwater Monitoring Network (GGMN). Retrieved from <https://ggis.un-igrac.org/view/ggmn/> Last access: 08/05/2024 (2021).
41. Lin, C.-Y., Miller, A., Waqar, M. & Marston, L. T. A database of groundwater wells in the United States. *Scientific Data*, **11**(1), <https://doi.org/10.1038/s41597-024-03186-3> (2024).
42. Venegas-Quinones, H. L. *et al.* Development of groundwater levels dataset for Chile since 1970. *Scientific Data*, **11**(1), <https://doi.org/10.1038/s41597-023-02895-5> (2024).
43. SGB - Geological Survey of Brazil [Serviço Geológico do Brasil]. *Sistema de Informações de Águas Subterrâneas*. Retrieved from <https://siagasweb.sgb.gov.br/layout/apresentacao.php>. Last access: 07/30/2024 (2024a).
44. SGB - Geological Survey of Brazil [Serviço Geológico do Brasil]. *Projeto Rede Integrada de Monitoramento das Águas Subterrâneas*. Retrieved from <https://rimasweb.sgb.gov.br/layout/apresentacao.php>. Last access: 07/30/2024 (2024b).
45. Orduña Alegria, M. E. *et al.* Unlocking aquifer sustainability through irrigator-driven groundwater conservation. *Nature Sustainability*, <https://doi.org/10.1038/s41893-024-01437-0> (2024).
46. Fan, Y., Li, H. & Miguez-Macho, G. Global patterns of groundwater table depth. *Science* **339**(6122), 940–943, <https://doi.org/10.1126/science.1229881> (2013).
47. Uchôa, J. G. *et al.* A Groundwater Wells Database for Brazil (GWDBrazil) (Version 2) [Data set]. *Zenodo*. <https://doi.org/10.5281/zenodo.16755455> (2025).
48. Quaggio, C. S., Gastmans, D., Kirchheim, R. & Batista, L. V. Variações na Composição das águas subterrâneas do Sistema Aquífero Serra geral em território brasileiro e sua relação com anomalias hidrogeoquímicas. *Águas Subterrâneas* **32**(3), 283–294, <https://doi.org/10.14295/ras.v32i3.29085> (2018).
49. Hirata, R. *et al.* O sistema Aquífero Guarani e a crise hídrica NAS regiões de Campinas e São Paulo (SP). *Revista USP*, (106), 59–70, <https://doi.org/10.11606/issn.2316-9036.v0i106p59-70> (2015).
50. Conicelli, B. *et al.* Groundwater governance: The Illegality of exploitation and ways to minimize the problem. *Anais Da Academia Brasileira de Ciências*, **93**(1), <https://doi.org/10.1590/0001-376520210200623> (2021).
51. Hirata, R., & Suhogusoff, A. V. How much do we know about the groundwater quality and its impact on Brazilian society today? *Acta Limnologica Brasiliensia*, **31**, <https://doi.org/10.1590/s2179-975x4419> (2019).
52. Hirata, R. & Foster, S. The Guarani aquifer system – from regional reserves to local use. *Quarterly Journal of Engineering Geology and Hydrogeology*, **54**(1), <https://doi.org/10.1144/qjegh2020-091> (2020).
53. Lucas, M. *et al.* Significant Baseflow Reduction in the Sao Francisco River Basin. *Is It Time to Call for Attention on Groundwater Withdrawals?* <https://doi.org/10.22541/au.159225069.93952235> (2020).
54. IBGE - Brazilian Institute of Geography and Statistics. *Malha Municipal*. Retrieved from <https://www.ibge.gov.br/geociencias/organizacao-do-territorio/malhas-territoriais/15774-malhas.html> Last access: 09/17/2024 (2022).
55. Petelet-Giraud, E. *et al.* Multi-layered Water Resources, management, and uses under the impacts of global changes in a Southern Coastal Metropolis: When will it be already too late? crossed analysis in Recife, NE Brazil. *Science of The Total Environment* **618**, 645–657, <https://doi.org/10.1016/j.scitotenv.2017.07.228> (2018).
56. ANA - National Water and Sanitation Agency. *Atlas Águas* (2021): segurança hídrica do abastecimento urbano. <https://metadados.snirh.gov.br/geonetwork/srv/api/records/d77a2d01-0578-4c71-a57e-87f5c565aacf> (2021b).
57. Pereira, B. S. *et al.* Hydrological heritage: A historical exploration of human–water dynamics in northeast brazil. *Hydrological Sciences Journal*, <https://doi.org/10.1080/02626667.2024.2446272> (2025).
58. MapBiomas. *Coleção 8 da Série Anual de Mapas de Cobertura e Uso da Terra do Brasil*. Retrieved from <https://brasil.mapbiomas.org/colecoes-mapbiomas/> (2023).
59. Souza, C. M. *et al.* Reconstructing three decades of land use and land cover changes in Brazilian biomes with Landsat Archive and Earth engine. *Remote Sensing* **12**(17), 2735, <https://doi.org/10.3390/rs12172735> (2020).

60. Boin, M.N., Martins, P.C.S., da Silva, C.A., Salgado, A.A.R. Pantanal: The Brazilian Wetlands. In: Salgado, A., Santos, L., Paisani, J. (eds) *The Physical Geography of Brazil. Geography of the Physical Environment*. Springer, Cham. https://doi.org/10.1007/978-3-030-04333-9_5 (2019).
61. Laipelt, L. *et al.* ANADEM: A Digital Terrain Model for South America. <https://doi.org/10.3390/rs16132321> (2024).
62. Yamazaki, D. *et al.* A high-accuracy map of global terrain elevations. *Geophysical Research Letters* **44**(11), 5844–5853, <https://doi.org/10.1002/2017gl072874> (2017).
63. ANA - National Water and Sanitation Agency. *Estudos Hidrogeológicos e de Vulnerabilidade do Sistema Aquífero Urucuia e Proposição de Modelo de Gestão Integrada e Compartilhada*. <https://metadados.snirh.gov.br/geonetwork/srv/api/records/b26a6c8e-affa-4766-8cb7-ccdaaadb3453> (2017).
64. Schuch, C. S., Galvão, P., de Melo, M. C. & Pereira, S. Overexploitation assessment in an urban karst aquifer: The case of Sete Lagoas (MG), Brazil. *Environmental Research* **236**, 116820, <https://doi.org/10.1016/j.envres.2023.116820> (2023).
65. ANA - National Water and Sanitation Agency. Atlas irrigação: uso da água na agricultura irrigada. *National Water and Sanitation Agency*, 2ed. <https://portal1.snirh.gov.br/ana/apps/storymaps/stories/a874e62f27544c6a986da1702a911c6b> (2021a).
66. Peterson, T. J., Western, A. W. & Cheng, X. The good, the bad and the outliers: Automated detection of errors and outliers from groundwater hydrographs. *Hydrogeology Journal* **26**(2), 371–380, <https://doi.org/10.1007/s10040-017-1660-7> (2017).
67. Xavier, A. C., Scanlon, B. R., King, C. W. & Alves, A. I. New improved Brazilian daily weather gridded data (1961–2020). *Int. J. Climatol.* **42**, 8390–8404 (2022).
68. Gleeson, T., Moosdorf, N., Hartmann, J. & van Beek, L. P. A glimpse beneath Earth's surface: Global Hydrogeology maps (GLHYMPS) of permeability and porosity. *Geophysical Research Letters* **41**(11), 3891–3898, <https://doi.org/10.1002/2014gl059856> (2014).
69. Huscroft, J., Gleeson, T., Hartmann, J. & Börker, J. Compiling and mapping global permeability of the unconsolidated and Consolidated Earth: Global Hydrogeology Maps 2.0 (GLHYMPS 2.0). *Geophysical Research Letters* **45**(4), 1897–1904, <https://doi.org/10.1002/2017gl075860> (2018).
70. ANA - National Water and Sanitation Agency. *Sistemas Aquíferos*. <https://metadados.snirh.gov.br/geonetwork/srv/api/records/3ec60e4f-85ea-4ba7-a90c-734b57594f90> (2016).
71. Linke, S. *et al.* Global Hydro-environmental sub-basin and river reach characteristics at high spatial resolution. *Scientific Data*, **6**(1), <https://doi.org/10.1038/s41597-019-0300-6> (2019).
72. Zipper, S. C. *et al.* Rapid and accurate estimates of streamflow depletion caused by groundwater pumping using analytical depletion functions. *Water Resources Research* **55**(7), 5807–5829, <https://doi.org/10.1029/2018wr024403> (2019).
73. Ma, Y., Koch, J. & Maxwell, R. M. Using random forests to explore the feasibility of groundwater knowledge transfer between the contiguous US and Denmark. *Environmental Research Communications* **6**(12), 121005, <https://doi.org/10.1088/2515-7620/ad9b08> (2024).
74. Theil, H. A rank-invariant method of linear and polynomial regression analysis. *Proc. R. Neth. Acad. Sci.* **53**, 386–392 (1950).
75. Sen, P. K. Estimates of the regression coefficient based on Kendall's Tau. *J. Am. Stat. Assoc.* **63**, 1379–1389 (1968).
76. Mann, H. B. Nonparametric tests against trend. *Econometrica* **13**, 245 (1945).
77. Rodrigues, A. F. *et al.* Has unsustainable groundwater use induced low flow regimes in the Urucuia Aquifer System? an urgent call for Integrated Water Management. *Journal of Environmental Management* **370**, 122979, <https://doi.org/10.1016/j.jenvman.2024.122979> (2024).
78. Famiglietti, J. S. *et al.* Satellites measure recent rates of groundwater depletion in California's Central Valley. *Geophysical Research Letters*, **38**(3), <https://doi.org/10.1029/2010gl046442> (2011).
79. Foster, S., Hirata, R. & Custodio, E. Waterwells: How can we make legality more attractive? *Hydrogeology Journal* **29**(4), 1365–1368, <https://doi.org/10.1007/s10040-021-02319-x> (2021).
80. R Core Team. A language and environment for statistical computing [Software]. R Foundation for Statistical Computing. Retrieved from <https://www.R-project.org/> (2024).
81. QGIS.org QGIS Geographic Information System. *Open Source Geospatial Foundation Project*. Retrieved from <http://qgis.org> (2022).
82. FAO - Food and agriculture organization of the United Nations. Trees, forests and land use in drylands: the first global assessment. FAO Forestry Paper No. 184, <https://www.fao.org/dryland-assessment/en/> (2019).
83. IBGE - Brazilian Institute of Geography and Statistics. Manual técnico da vegetação brasileira. Retrieved from <https://loja.ibge.gov.br/manual-tecnico-da-vegetacao-o-brasileira.html> Last access: 02/18/2025 (2012).
84. Beck, H. E. *et al.* Present and future Köppen-Geiger climate classification maps at 1-km resolution. *Scientific Data*, **5**(1), <https://doi.org/10.1038/sdata.2018.214> (2018).

Acknowledgements

We thank the Geological Survey of Brazil (SGB) for their scientific support, especially Dr. Francisco Marcuzzo and Dr. Franco Buffon. This study was supported by grants from the Ministry of Science, Technology, and Innovation – MCTI and the National Council for Scientific and Technological Development – CNPq (Grants 406647/2022-4 and 446044/2023-7) and Fundação de Apoio ao Ensino, Ciência e Tecnologia do Estado de Mato Grosso do Sul - Fundect (Grant 71/032.795/2022). The authors acknowledge the Graduate Program in Hydraulic Engineering and Sanitation (PPGSHS) at the University of São Paulo (USP) for their scientific support. J.G.S.M.U acknowledges the São Paulo Research Foundation—FAPESP (Grant 2020/15434-0, 2023/13160-8 and 2023/18011-0) for supported part of this study. A.S.B acknowledges the FAPESP (2020/08140-0 and 2022/06017-1) for supported part of this study. P.T.S.O acknowledges the CNPq (Grant 309048/2023-0) and the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brazil - CAPES (Capes Print). We thank Professor Scott Jasechko for his support and suggestions throughout the development of this database. We thank Professor Ying Fan for her valuable input during the database construction and for kindly reviewing an earlier version of this manuscript.

Author contributions

J.G.S.M.U., P.T.S.O. and E.C.W. conceived the study. J.G.S.M.U. download and pre- and post- processed the data. J.G.S.M.U. carried out the analyses. J.G.S.M.U. wrote the draft manuscript. P.T.S.O. and E.C.W. managed funding, and supervised the project. All authors, J.G.S.M.U., P.T.S.O., A.S.B., D.G., J.A.A.A., B.R.S., C.R.C., V.J.F.F., E.C.W., discussed the results and edited the paper at multiple stages.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41597-025-05843-7>.

Correspondence and requests for materials should be addressed to P.T.S.O.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2025