GROW: A GLOBAL TIME SERIES DATASET FOR GROUNDWATER STUDIES WITHIN THE EARTH SYSTEM

Annemarie Bäthge[1]*, Claudia Ruz Vargas[2], Gunnar Lischeid[3,4], Raoul Collenteur[5], Mark Cuthbert[6], Jan Fleckenstein[7], Martina Flörke[8], Inge de Graaf[9], Sebastian Gnann[10], Andreas Hartmann[11], Xander Huggins[12], Nils Moosdorf[13,14], Yoshihide Wada[15], Thorsten Wagener[4], Robert Reinecke[1]

[1]Institute of Geography, Johannes Gutenberg-University, Mainz, Mainz, Germany
[2]International Groundwater Resources Assessment Centre (IGRAC), Delft, The Netherlands
[3]Research Area 4: "Simulation & Data Science", Leibniz Centre for Agricultural Landscape Research, Müncheberg, Germany
[4]Institute of Environmental Science and Geography, University of Potsdam, Germany
[5]Department Water Resources and Drinking Water, Eawag, Dübendorf, Zürich, Switzerland
[6]School of Earth and Environmental Sciences, Cardiff University, UK
[7]Department of Hydrogeology, Helmholtz Center for Environmental Research, Leipzig, Germany
[8]Institute of Engineering Hydrology and Water Resources Management, Ruhr University Bochum, 44801, Bochum, Germany
[9]Earth Systems and Global Change group, Wageningen University and Research, Wageningen, the Netherlands
[10]Chair of Hydrology, University of Freiburg, Freiburg, Germany
[11]Institute of Groundwater Management, Dresden University of Technology
[12]Institute for Resources, Environment, and Sustainability, University of British Columbia, Vancouver, BC, Canada
[13]Leibniz Centre for Tropical Marine Research (ZMT), Bremen, Germany
[14]Kiel University, Kiel, Germany
[15]Biological and Environmental Science and Engineering Division, King Abdullah University of Science and Technology, Thuwal, Saudi Arabia
* Correspondence: a.baethge@uni-mainz.de

Earth System Modeling

JOHANNES GUTENBERG UNIVERSITÄT MAINZ

JG|U

GROW (global integrated GROundWater package) is a global, analysis-ready dataset that combines observed groundwater time series data with potential associated Earth system variables.

MOTIVATION

- Access to groundwater is pivotal for humans and ecosystems that depend on it as freshwater source[1,2,3]
- However, we have a limited systematic understanding of the complex interlinkage between groundwater and its environmental controls within the Earth system (here called Earth system variables)[4,5,6,7]
- Observations play a key role in filling these knowledge gaps[7,8,9], but…
- So far, there is no existing groundwater dataset that:

(1) is **global** (captures the natural variability of the included variables);
(2) contains **time series data** that provide insights into temporal patterns (e.g., seasonality), extreme events, and cause-effect lags[10,11,12];
(2) combines groundwater observations with **associated Earth system variables** that can be used to explain groundwater dynamics[8,13,14];
(4) include **metadata** to improve comparability and the assessment of data uncertainty[15,5,6];
(5) is **standardized** and **freely available** according to the FAIR principles[15,16,17].

→ Therefore, we developed GROW (global integrated GROundWater package), a quality-controlled dataset that accompanies groundwater depth and level time series from around the world with associated Earth system variables.

OUTLOOK

- With GROW, we offer a dataset to the global community that can be used to understand spatiotemporal groundwater dynamics in the context of the Earth system. For example:
  - Cumulative effects of multiple controls on groundwater time series can be studied in space and time (e.g. with machine learning or principal component analysis)
  - Processes conceptualized for specific environmental settings can be transferred to regions with no data but the same environmental conditions
  - The standardized groundwater time series in GROW provide a ready-to-use calibration and validation dataset for global hydrological, land-surface, or climate modeling
- Still, GROW inherits uncertainty from the groundwater observations (e.g. spatial data bias) and Earth system variables (e.g. model uncertainty, interpolation errors, commensurability errors). To reduce uncertainty:
  - GROW can benefit from additional groundwater time series and metadata, especially from regions that are underrepresented in the current dataset
  - GROW is envisioned to evolve into a dynamic community dataset that continually grows and benefits from the addition of new data

THE DATASET CONTAINS

Abstract | References | This presentation participates to OSPP – Outstanding Student & PhD candidate Presentation contest | Sharing is encouraged

GROUNDWATER TIME SERIES

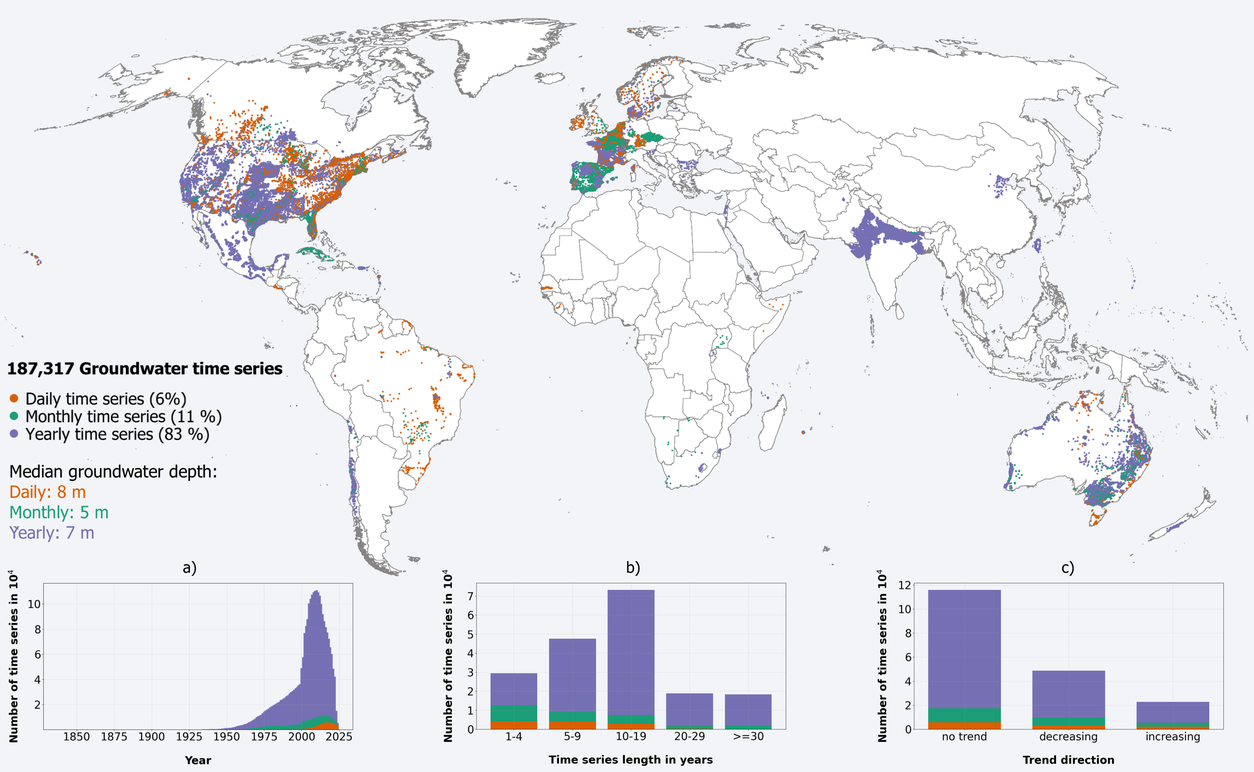
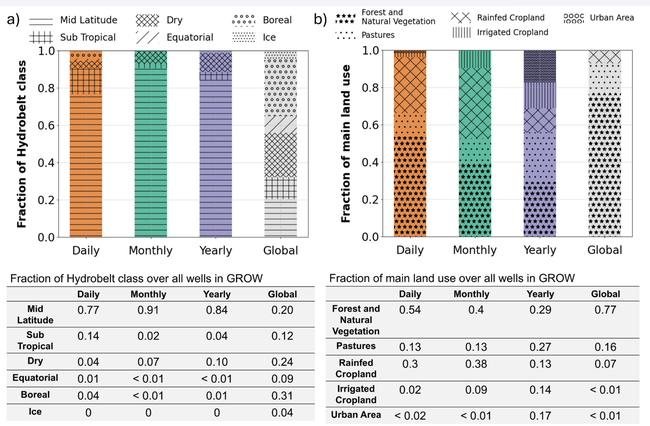


Figure 1- World map with locations of every time series and classified by temporal resolution: The global groundwater data was derived from igrac's Global Well and Monitoring Data and Other Data[18,19]. The color indicates whether a time series has a daily (orange), monthly (green) or yearly (purple) resolution. Additionally, the median groundwater depth per well is given for each temporal resolution. Below, a) the number of time series available per year, b) the number of time series per length class in years and c) the number of time series with no trend, increasing or decreasing trend (Mann Kendall trend direction) are displayed. The trend direction was switched for groundwater table depth time series so that a rising groundwater table (decreasing depth) is flagged as "increasing".

GROUNDWATER ASSOCIATED EARTH SYSTEM VARIABLES

Table 1: List of all 34 Earth system variables that were merged with the groundwater data classified by Earth system component. In contrast to attribute variables, time series are written in italics.

Atmosphere	Geosphere	Hydrosphere	Cyrosphere	Biosphere	Anthroposhere
• *Precipitation*[20,21] • *Potential Evapotranspiration*[22,33] • *Actual evapotranspiration*[23] • *Air temperature*[22] • Köppen-Geiger classification[24] • Hydrobelts[25]	• Ground elevation[26] • Topographic slope[27] • Rock type[28] • Rock permeability[29] • Rock porosity[30] • Aquifer type[28,31] • Saturated hydraulic conductivity for topsoil and subsoil[32] • Soil texture class for topsoil and subsoil[32]	• Distance between perennial streams[33] • Drainage density[34]	• *Snow depth*[22] • Glacier cover[34] • Permafrost cover[34]	• *Interception loss*[23] • *NDVI*[35] • *Leaf Area Index of low vegetation*[22] • *Leaf Area Index of high vegetation*[22] • Groundwater dependent ecosystem[36]	• *Total water withdrawal for domestic use*[37] • *Total water withdrawal for industrial use*[37] • *Land use fractions (Forest and Natural Vegetation, pastures, rainfed cropland, irrigated cropland, urban area)*[38] • Groundwaterscapes[39]

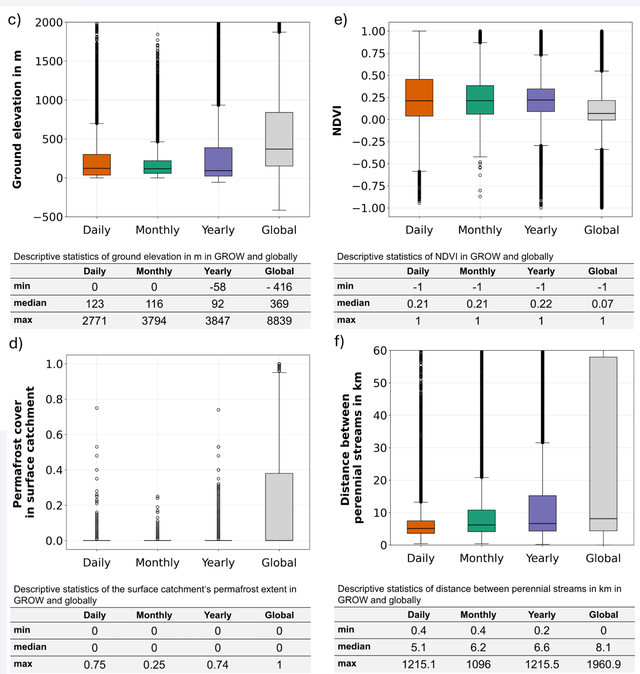


Fraction of Hydrobelt class over all wells in GROW

	Daily	Monthly	Yearly	Global
Mid Latitude	0.77	0.91	0.84	0.20
Sub Tropical	0.14	0.02	0.04	0.12
Dry	0.04	0.07	0.10	0.24
Equatorial	0.01	< 0.01	< 0.01	0.09
Boreal	0.04	< 0.01	0.01	0.31
Ice	0	0	0	0.04

Fraction of main land use over all wells in GROW

	Daily	Monthly	Yearly	Global
Forest and Natural Vegetation	0.54	0.4	0.29	0.77
Pastures	0.13	0.13	0.27	0.16
Rainfed Cropland	0.3	0.38	0.13	0.07
Irrigated Cropland	0.02	0.09	0.14	< 0.01
Urban Area	< 0.02	< 0.01	0.17	< 0.01



Descriptive statistics of ground elevation in m in GROW and globally

	Daily	Monthly	Yearly	Global
min	0	0	-58	- 416
median	123	116	92	369
max	2771	3794	3847	8839

Descriptive statistics of NDVI in GROW and globally

	Daily	Monthly	Yearly	Global
min	-1	-1	-1	-1
median	0.21	0.21	0.22	0.07
max	1	1	1	1

Descriptive statistics of the surface catchment's permafrost extent in GROW and globally

	Daily	Monthly	Yearly	Global
min	0	0	0	0
median	0	0	0	0
max	0.75	0.25	0.74	1

Descriptive statistics of distance between perennial streams in km in GROW and globally

	Daily	Monthly	Yearly	Global
min	0.4	0.4	0.2	0
median	5.1	6.2	6.6	8.1
max	1215.1	1096	1215.5	1960.9

Figure 2 - Distribution of selected Earth system variables in the GROW dataset compared with their global distribution: This overview is showing the distribution of the following Earth system variables in GROW classified by temporal resolution of the time series in comparison to the global range of the respective variable: a) the fraction of Hydrobelt classes as a stacked bar chart and b) the fraction of main land use as stacked bar chart, c) ground elevation (not full value range), d) distance between perennial streams (not full value range), e) permafrost cover in surface catchment and f) NDVI as box plot. Box plots show the median of a distribution as a black line inside the box. The upper and lower edges of the box are determined by the 25th and 75th percentiles (interquartile range). Whiskers indicate the farthest data point within 1.5 times the interquartile range. Outliers outside this range are displayed as dots. Below the plots, a table with either the respective fractions per class and group or the min, median and max value per group is given.