

Non-piped groundwater sources supply over 2 billion people in low- and middle-income countries

Rebekah Hinton

hinton@iiasa.ac.at

International Institute for Applied Systems Analysis (IIASA)

Dor Fridman

IIASA

Florian Sorger-Domenigg

IIASA <https://orcid.org/0000-0002-5403-8472>

Bárbara Willaarts

IIASA

Taher Kahil

International Institute for Applied Systems Analysis <https://orcid.org/0000-0002-7812-5271>

Physical Sciences - Article

Keywords:

Posted Date: August 6th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-10430797/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: There is **NO** Competing Interest.

Non-piped groundwater sources supply over 2 billion people in low- and middle-income countries

Abstract:

The source through which people obtain their drinking water, whether through a household tap, shallow well or a tanker truck, underpins water security, including safety^{1,2}, cost³, and climate resilience⁴. Yet source-explicit drinking water use remains poorly understood and absent from global assessments^{5,6}; masked by broad classifications, such as the improved/unimproved dichotomy, that obscure profound differences between source types^{7,8}. Using household survey data from 53 countries to train and validate random forest models, we generate the first high-resolution (5 km), source-explicit estimates of drinking water source use across low- and middle-income countries (LMICs). Here we show that over 2 billion people (31% of the LMIC population) use non-piped groundwater, including boreholes, dug wells, and springs, as their primary drinking water source, rivalling piped supplies in scale and exceeding them in sub-Saharan Africa and South Asia. We further identify that bottled and delivered water markets each serve over 450 million people. These non-piped sources are almost entirely absent from the global databases⁹ and vulnerability indices used to assess water security, leaving the drinking water use of nearly half the LMIC population effectively invisible in global monitoring, with direct consequences for hydrological modelling^{10,11}, climate vulnerability assessment¹², and water governance¹³.

Introduction

Access to safe and reliable drinking water underpins health, economic opportunity, and human wellbeing^{14–16}. How households access that water, whether from a piped-water tap, a community borehole, a tanker-truck or a water bottle, shapes almost every dimension of water security, including quality^{1,2}, reliability¹⁷, accessibility¹⁸, cost³, governance¹³, and the quantity^{19,20} of drinking water, ultimately influencing the hydrological systems on which water security depends²¹. Furthermore, such distinctions result in divergent vulnerabilities to climatic stresses⁴, making source a critical factor in drinking water vulnerability and environmental sustainability assessments^{22,23}.

Global monitoring, as embedded in the Millennium and Sustainable Development Goals, has largely focused on a broad classification of whether a drinking water source is considered 'improved' or 'unimproved'²⁴. Whilst beneficial for tracking progress, this classification aggregates sources with profoundly distinct characteristics: piped water supplies, boreholes, protected wells and springs, delivered water, and bottled water are all grouped as 'improved', whilst unprotected wells, springs, and surface water are 'unimproved'⁷. More recently focus has been placed on safely managed drinking water sources, defined as use of improved sources located on premises, available when needed, and free from contamination, yet while it offers more insight into the nature of drinking water access, the classification still clusters sources within the improved/unimproved dichotomy. This obscures critical distinctions between source types and their implications for water security⁸. This is particularly relevant in low- and middle-income countries (LMICs) which are home to approximately 85% of the world's population and

have been the primary focus of global efforts to expand drinking water access over the past three decades⁷.

Spatial resolution is another limitation in understanding water source use. Water access varies substantially within countries, yet monitoring data are often aggregated to national or regional scales⁷. Recent advances in high-resolution mapping have substantially improved understanding, including subnational estimates of improved water access across LMICs by the Deshpande *et al.*, (2020)⁶ and, more recently, high-resolution estimates of safely managed drinking water use by Greenwood *et al.* (2024)⁵. While these studies reveal marked subnational heterogeneity in access, they remain limited to the broad categories of improved or unimproved sources^{5,6} and, although some distinguish piped sources^{6,7}, they offer limited insight into the specific types of drinking water sources people rely on. Existing source-explicit maps are largely restricted to individual cities or countries^{25,26}. No study has generated high-resolution, source-explicit estimates of drinking water access for LMICs, limiting the ability to identify source-specific vulnerabilities and constraining hydrological understanding, particularly regarding groundwater dependence²¹.

Here we address this gap by mapping household drinking water source use in LMICs. Using household survey data from 53 countries to train source-specific random forest models on climatic, hydrological, topographic, and socioeconomic predictor variables, we generate high-resolution (5 km) spatial estimates of twelve distinct drinking water sources across LMICs and identifying source heterogeneity in how people access drinking water. The findings reveal a domestic water supply landscape far more diverse than captured under current monitoring frameworks, with direct consequences for how water resources are modelled and managed, and how drinking water governance must be designed to reflect the diversity of drinking water access.

Results:

Source-explicit drinking water access across LMICs

Across LMICs, drinking water access is characterised by high heterogeneity in water source use with non-piped groundwater sources in particular emerging as a major, yet often overlooked, pillar of supply comparable in scale to piped infrastructure (Figure 1). Piped water is the single most used type of water access, accessed by half of the population (50.7% [42.4-58.8%], 3.37 billion people [2.82-3.91 billion]), consistent with previous estimates⁶. Yet over a quarter of piped water users (27.3% [18.8-37.3%]) do not access drinking water on premises but rather rely on water from public standpipes or neighbours, with implications for collection burden, effective water availability, and water demand estimates. This is particularly notable in Sub-Saharan Africa where half (50.2% [25.7-73.3%]) of piped water supply is accessed off premises.

Non-piped groundwater sources (boreholes, dug wells, or springs) accessed directly by households serve almost a third of the population (31.2% [27.2-35.2%], 2.08 billion people [1.81-2.34 billion]), rivalling on-premises piped water use (36.8% [29.4-44.2%]) as the dominant mode of drinking water access. Boreholes are the primary infrastructure through which this is accessed, supplying drinking water for 19.4% [15.9-23.0%] of the population. Other direct groundwater sources, including protected and unprotected dug wells and springs, serve smaller

fractions of the population individually (ranging from 1.5% to 5.2%) but collectively contribute substantially to water access (11.9% [9.81-13.6%]). Non-piped groundwater dependence is highest where piped infrastructure is least developed: in sub-Saharan Africa and South Asia it constitutes the primary mode of drinking water access for 51.1% [40.7-61.5%] and 45.6% [35.8-55.4%] of the population, respectively, in both cases exceeding piped supply.

Commercial, often high-cost and largely unregulated, sources of delivered (or tanker) and bottled (or packaged) water are also identified as significant sources, each used by over 450 million people (6.78% [4.81-8.75%] and 6.83% [5.81-7.85%] of the population, respectively). Reliance on commercial supply is strongly regionally concentrated. Delivered water accounts for 16.5% [10.2-22.8%] of primary drinking water supply in the water scarce areas of the Middle East and North Africa. Bottled and packaged water exceeds 10% of primary supply in Latin America and the Caribbean (13.6% [10.3-17.0%]) and East Asia and Pacific (12.8% [10.1-15.6%]). Surface water and rainwater contribute modestly across most regions, though surface water use in sub-Saharan Africa (3.82%, [1.66-5.98%]) represents almost 50 million people (49.2 million [21.4-77.0 million]) retaining dependence on the most contamination-vulnerable source category. Source breakdowns by region are found in Extended Data Table 3.

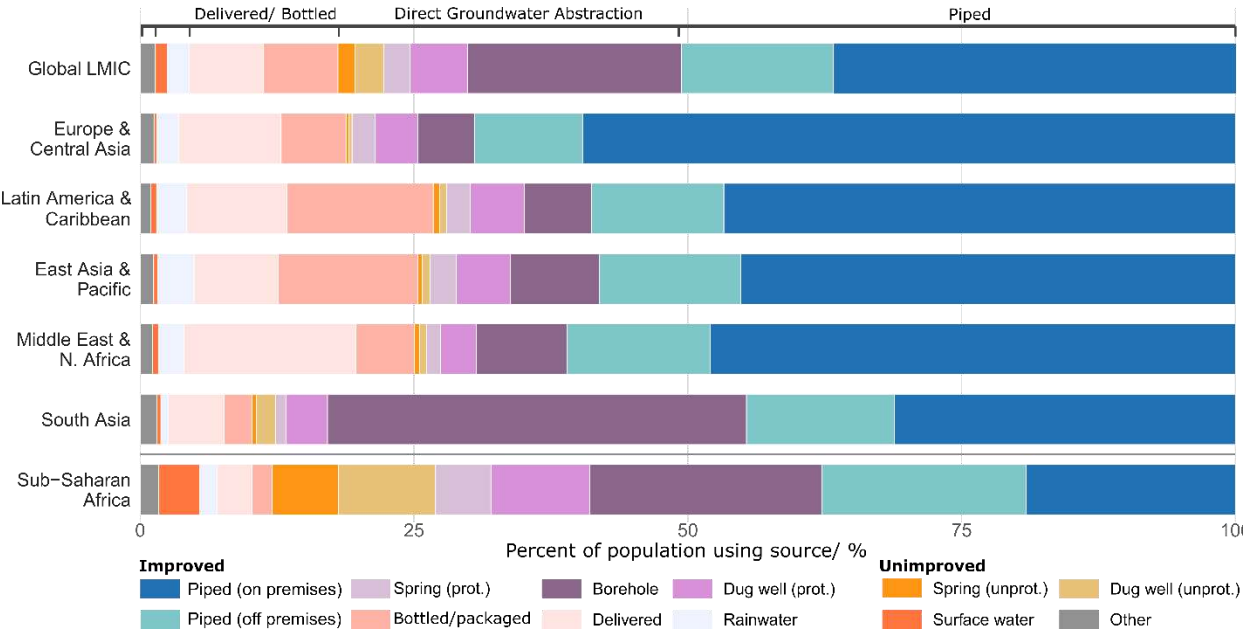


Figure 1: Drinking water source composition by region across low and middle-income countries. The estimated proportion of the population using each drinking water source type as their primary source, by World Bank region and for LMICs overall.

At subregional and subnational scale, Figure 2 reveals source-specific hotspots. Borehole reliance is particularly high in the Sahel, Southeastern Africa, and Northeast India while spring use is most notable in the Congo Basin reflecting both hydrogeological opportunity and the absence of alternative improved sources. Hotspots of high delivered water dependence are evident across the Middle East and North Africa, whilst bottled water use is particularly elevated in Thailand. Areas of high off-premises piped water supply are also evident in specific regions, notably Lesotho.

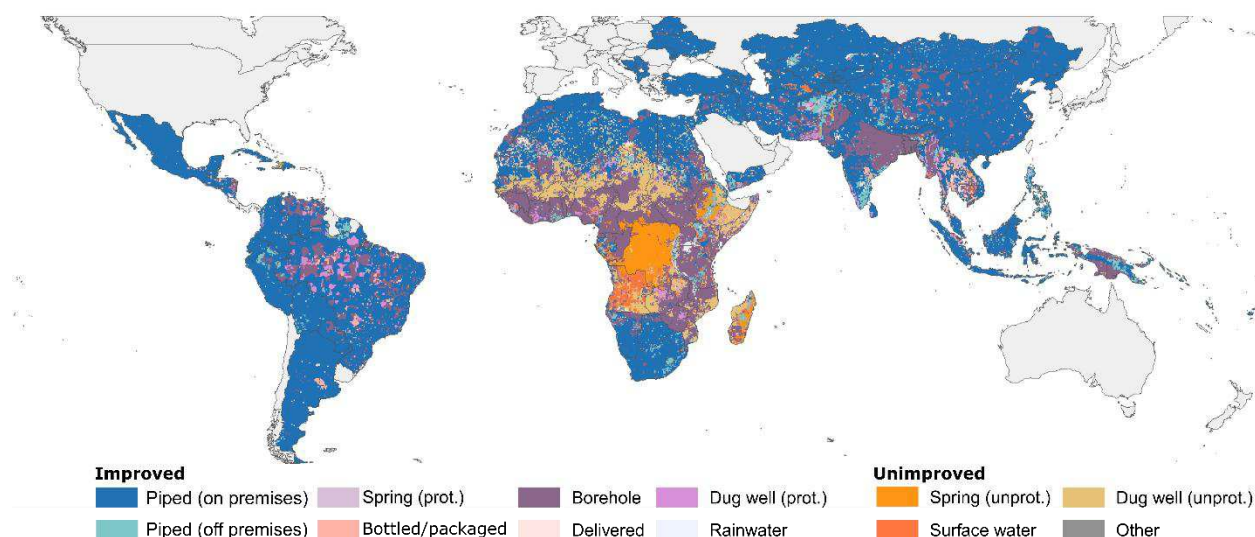


Figure 2: Dominant drinking water source type across low and middle-income countries at 5 km resolution. Each grid cell shows the dominant drinking water source, the source used by the largest proportion of the population, derived from source-specific random forest models.

Model performance

Model performance varies by source type. The highest model explanatory power is observed for borehole/tubewell ($R^2 = 0.674$, $MAE=0.101$), bottled/packageged water ($R^2 = 0.658$, $MAE=0.0372$), piped on premises ($R^2 = 0.566$, $MAE= 0.184$), delivered water ($R^2 = 0.563$, $MAE=0.0666$), and protected dug well ($R^2 = 0.522$, $MAE= 0.0291$) with moderate performance for piped off premises, protected wells, rainwater, and springs ($R^2 = 0.488-0.245$). Lower R^2 values are observed for marginal sources, unprotected dug wells ($R^2 = 0.198$, $MAE= 0.0190$) and surface water ($R^2 = 0.136$, $MAE= 0.0085$) though these collectively represent 3.69% of total source use and absolute prediction errors remain small across all categories. Mean absolute errors remain small in absolute terms across categories, ranging from 0.0059 to 0.184, indicating limited magnitude of prediction error even where R^2 is lower. Performance is comparable to previously reported estimates for aggregated improved and unimproved access at this scale⁵. Per source model performance is found in Extended Data Table 4.

Validation against national estimates from the WHO/UNICEF Joint Monitoring Programme (JMP) confirms broad model agreement, Extended Data Figure 3. We estimate 281 million people use unimproved sources (4.2%) and 71 million use surface water (1.1%) in LMICs, comparable to, though underestimating, JMP global estimates of 302 million (unimproved) and 106 million (surface water)²⁷. Across 122 countries for which national JMP data were available, improved water coverage is predicted with $MAE=0.061$ and $R^2=0.56$, consistent with literature performance⁵. Regional errors (Extended Data Table 5) are highest in sub-Saharan Africa ($MAE=0.112$), where coverage variance is greatest, and lowest in Europe and Central Asia ($MAE=0.022$), where near-universal access compresses the variance. Piped water supply explains a similar amount of variation ($R^2= 0.58$) but higher error, largely due to underprediction

at high coverage (MAE= 0.242) whilst surface water use is less well explained by the model ($R^2= 0.34$) but has lower error, due to lower usage (MAE= 0.0167), Extended Data Table 6.

Overall prediction uncertainty is low, with a median coefficient of variation (CoV) of 0.168 across dominant source predictions and 5.3% of grid cells classified as high uncertainty ($\text{CoV} \geq 0.5$). Prediction uncertainty varies by source. Piped water on premises has a median coefficient of variation (CoV) of 0.209, with 4.1% of grid cells with high uncertainty ($\text{CoV} \geq 0.5$). Strongly hydro(geo)logically constrained sources also have lower uncertainty, with fewer areas of high uncertainty ($\text{CoV} \geq 0.5$) for springs (0.7% unprotected and 5.6% protected) and surface water (4.3%), as well as delivered water (3.8%) which is often restricted to highly water scarce areas. Uncertainty is higher for boreholes and protected dug wells (both with 10.7% of cells with high uncertainty, largely in areas where there is uncertainty over whether the source is a borehole or protected dug well), bottled water (11.1%), rainwater (13.2%), and unprotected dug wells (18.6%) also all exceed a CoV of 0.5. Uncertainty maps with CoV and IQR are provided in Extended Data Figure 2.

Variable importance

Socioeconomic development is the strongest and most consistent predictor of drinking water source use across all source models (Figure 3). The Subnational Human Development Index ranked in the top ten predictors for all eleven directly trained source models whilst nighttime light intensity, infant mortality rate, and poverty indices did so for ten of eleven. Climatic factors including mean temperature, mean precipitation, and precipitation variability have a strong influence, notably mean precipitation has high importance for the use of delivered (tanker) water, associated with more arid areas, whilst spring presence is associated with high precipitation. Local hydrogeological characteristics, such as water table depth and permeability were relevant, especially for certain groundwater sources, such as borehole and spring use but generally less dominant than broader climatic or human settlement factors. Figure 3 summarises mean Shapley values across models with a rank weighted score. Source specific summaries of variable importance are found in Extended Information Figure 4.

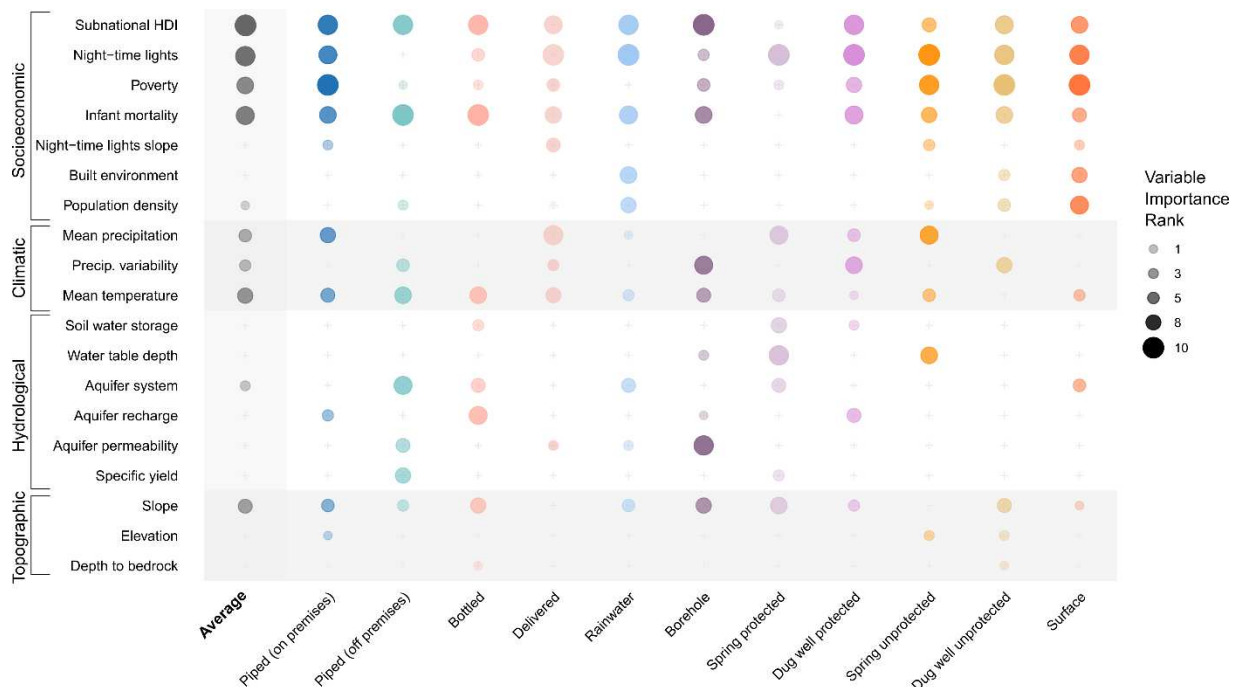


Figure 3: Variable importance across source-specific random forest models. Dot size and shading indicate the rank-weighted importance of each variable within each source-specific model, scored from 10 (most important variable in that model) to 1 (tenth most important); variables ranking below the top ten are marked with a cross. The average column shows the mean rank-weighted score across all eleven directly trained source models. Variables are grouped by thematic category (socioeconomic, climatic, hydrological, topographic). Only variables which are in the top 10 most important variables in at least one model are shown.

Discussion:

Global monitoring of drinking water access in LMICs has largely ignored the source through which people actually access water, masking profound distinctions within broad classifications such as improved, unimproved, and safely managed. Through source-explicit mapping, we reveal a highly diverse water supply landscape, particularly shedding light on the scale of non-piped groundwater, bottled, and delivered water, which operate largely outside of regulatory frameworks.

Non-piped groundwater sources are identified as a central pillar of drinking water supply across LMICs, yet one that remains systematically underrepresented in the monitoring and modelling frameworks on which water security assessment depends. Used by over 2 billion people in LMICs (31% of the population), a figure comparable to on-premises piped supply (37%), they form the primary source of drinking water in sub-Saharan Africa and South Asia^{28–30}. This oversight reflects in part the governance structure of such sources; non-piped groundwater almost entirely self- or community-supplied, operating largely outside the utility frameworks that generate the data on which national monitoring and hydrological models are built^{31–34}. Boreholes and tubewells are the dominant infrastructure for such supply, serving 1.3 billion people. Their use has grown substantially³⁵ often as a way for households to shift away from unimproved

sources and supplement unreliable piped supply³⁶. Whilst being promoted as safer³⁷ and more resilient to drought than shallower dug wells³⁵, challenges of microbial³⁸ and geogenic^{37,39} contamination, as well as groundwater declines⁴⁰ remain a concern while high drilling costs⁴¹ and infrastructure failure⁴² limit availability. Integrating non-piped groundwater into hydrological models and water security assessments is essential for accurately representing demand, abstraction, and vulnerability across most of the LMIC population.

Over half of the population of LMICs (50.7%) is identified as using piped drinking water which is consistent with previous analysis⁶. Yet even among the population accessing piped sources, source-level heterogeneity is significant; over a quarter of piped water users access water off-premises, from public standpipes or neighbours, rather than their own dwelling/plot. The importance of on-premises access has been widely documented, motivating a focus on ensuring on-premises water under the safely managed drinking water framework. On-plot connections reduce collection time and physical burden, increase effective water availability, and alter storage practices, which in turn influences contamination risks. Source also influences behaviour^{19,20}; when water collection time is reduced to under five minutes, compared to collection times of 5-30 minutes, household water use more than doubles, and consumption increases up to fivefold when access is provided through multiple taps within the dwelling²⁰. This has implications not only for health and hygiene behavior²⁰, but also for water demand estimates, infrastructure investment, and cost-recovery. Considering piped water access without specifying location may be misleading⁶, especially in areas with high off-plot access such as sub-Saharan Africa, where nearly half of all piped users rely on off-premises piped sources, piped coverage figures substantially overstate the population benefiting from the reduced collection burden and increased effective water quantity that on-plot connection enables.

Delivered and bottled water collectively serve over 900 million people across LMICs, identifying them not as marginal sources but a primary drinking water supply system. These sources have increased significantly in recent years^{7,43} partly driven by perceived health risks associated with alternative sources, including intermittent piped systems⁴³. However, challenges surrounding limited regulatory oversight⁴⁴, and highly variable pricing structures⁴⁵⁻⁴⁷ exist. Expenditure on tanker water has been estimated to account for up to 16% of average household income⁴⁸ presenting a severe economic burden on households that have typically no other alternatives. Quality is highly variable and difficult to assure⁴⁸, with challenges of storage conditions and handling practices⁴⁵ presenting a concern for delivered/tanker water and inconsistent quality control, and emerging concerns related to microplastics threatening bottled water safety⁴⁹. The environmental implications of high delivered/tanker and bottled/package water are also notable. Many tankers rely on illegal water abstraction⁴⁷, representing a major challenge for water management, especially in water scarce regions, whilst plastic waste generation presents a significant concern associated with high bottled water dependence⁴⁹. The highly concentrated dependence of such commercial supplies, notably delivered/tanker water in the Middle East and Northern Africa (16.5% of the population) and bottled/package water dependence in Latin America and the Caribbean (13.6%) and East Asia and the Pacific (12.8%), suggest that these sources are not marginal coping strategies but embedded primary supply systems for large shares of population. Given their rapid growth, substantial user base, and projected future

reliance⁴⁷, increased attention must be paid to the role and risks of packaged and delivered water sources in water security.

The source-level diversity documented here has direct consequences for how water resources are measured, modelled, and governed. Notably, the more than 2 billion people drawing directly from non-piped groundwater, and the nearly 900 million accessing delivered or bottled water, are almost entirely absent from the water resources databases and models used to assess scarcity, groundwater sustainability, and climate impacts. Domestic water withdrawal is often used interchangeably with municipal abstraction, the volume extracted by public distribution networks⁵⁰, in global databases including AQUASTAT, which underpins SDG indicator 6.4.2 on water stress⁹. As such, non-networked services are systematically omitted from global monitoring and reporting⁵¹. AQUASTAT formally aims to include self-abstraction but currently holds no data for this component⁹; the result is that the water demand of roughly half the LMIC population is either omitted or opaquely estimated. This omission is particularly consequential in sub-Saharan Africa, where diffuse domestic groundwater abstraction represents the dominant withdrawal sector in many areas, exceeding industrial and agricultural demand, yet is often omitted in national or regional abstraction accounting⁵¹. The level of non-piped water sources, especially groundwater, raises challenges for global understanding of domestic water use which are often unable to account for direct abstraction¹⁰; notably the frequently cited estimate that groundwater supplies 50% of global domestic water^{30,52–56}, fails to explicitly account for the extent of direct groundwater abstraction³⁰, as identified here, likely underestimating groundwater dependence. Alongside conceptually limiting our understanding of the hydrological system, these gaps propagate directly into global hydrological models, many of which derive domestic demand estimates from global databases such as AQUASTAT^{10,11}, systematically misrepresenting both the quantity and source of domestic abstraction.

Despite source being a central component in exposure of drinking water to many stressors, many vulnerability metrics also fail to capture the impact of source type on the exposure to stressors including water scarcity, extreme events such as droughts and floods⁴, and exhibit distinct contamination pathways⁵⁷ and risks⁵⁸, especially under climate change⁴. Widely used vulnerability frameworks, including the Water Poverty Index⁵⁹ and the UNICEF Extreme Water Vulnerability Index¹², assess water-related vulnerability without distinguishing source type, and therefore cannot represent these dynamics. The widespread reliance on decentralised sources identified here also demands governance frameworks that extend beyond piped utility regulation, recognising that for the majority of the LMIC population, drinking water security depends on systems beyond traditional regulatory and monitoring infrastructure^{13,30,31,33,34}. The spatially explicit, source-resolved data produced here provide a foundation for integrating drinking water source diversity into hydrological models, vulnerability assessments, governance approaches, and the SDG monitoring frameworks that must account for it if they are to accurately reflect water security for the populations most at risk.

Methods:

The full methodological pipeline is summarised in Extended Data Figure 1a and is explained in detail below.

Household water source data:

Household survey data were collated from Demographic and Health Surveys⁶⁰ (DHS) and Multiple Indicator Cluster Surveys⁶¹ (MICS) from 2015-2024, Extended Data Figure 1b. Only the latest survey was retained for each country, to ensure that certain countries were not overrepresented. Only surveys with associated GPS coordinates were retained. Data were available for 53 countries with a total of 2.21 million household surveys. All spatial analyses and predictive rasters were generated at 5 km resolution (0.045°), which is consistent with the maximum systematic standard displacement of DHS and MICS surveys (up to 5km in rural areas and 2km in urban areas)⁶².

Original DHS water source categories were harmonised and minimally aggregated into functionally and hydrologically meaningful classes based on Bain et al. (2014)¹⁶. These groups preserve key distinctions relevant to access and water security whilst aggregating to ensure sufficient data and minimise national differences in surveying and classification. Water classes are summarised in Extended Data Table 1.

For each DHS cluster, the proportion of the population using each water source group was estimated using DHS household sampling weights. Weighted household counts were multiplied by the number of de jure household members to derive population-representative estimates of water source use. DHS cluster-level proportions of water source use were rasterised to 5km resolution, consistent with the spatial scale of covariates and the DHS displacement radius⁶.

Random forest:

To estimate the proportion of the population using each water source type, we employed a random forest approach. Random forest (RF) models provide a powerful tool for analysis of complex data⁶³, and are particularly effective for predicting spatially distributed outcomes^{64,65}, including water access⁵. An RF consists of an ensemble of decision trees, with each tree using a different subset of predictor variables to predict the response^{63,66}. RF models are especially advantageous in settings with non-linear relationships and collinearity among covariates⁶⁷.

Variable selection.

A range of global climatic (3), hydrological/ hydrogeological (8), topographic (4), and socioeconomic (7) variables were included to capture factors likely to influence household water source use⁵. Full details of variables and data sources are provided in the Extended Data Table 2.

Several variables were constructed to represent proximity to surface water sources. Distance to freshwater was constructed using data from Hydrolakes⁶⁸ and Hydrorivers⁶⁹ datasets. River data were derived from HydroRIVERS, with only rivers exceeding a mean discharge of 1 m³/s retained in order to represent relatively reliable surface water sources⁶⁹. As households could be displaced by up to 5km to actual distance between households and freshwater sources could not be calculated. Rather, to reflect proximity to surface water whilst respecting the displacement of household locations, a binary indicator was generated to represent the presence of a surface water at 5 km resolution of each location. If a household location falls within this raster cell, it indicates the presence of surface water within 5km of the household, considered a plausible collection distance for surface water use. In addition, stream density was

also included as predictor variable to further capture surface water availability. This variable was calculated from the HydroRIVERS dataset⁶⁹ and reflects the density of river channels within a given area, providing a proxy for broader hydrological accessibility to surface water resources.

To account for rapid socioeconomic and demographic changes during the study period, key time-sensitive covariates, population and nighttime lights, were matched to the year of each household survey. Population data, from WorldPop⁷⁰, and VIIRS nighttime lights⁷¹, were extracted for the year corresponding to the year in which the survey was conducted. For surveys spanning multiple years, the year of completion was used. Prediction models were then trained using the most recent data for population (2026) and nighttime lights (2024).

Random Forest training

Random cross-validation can result in many training and testing datasets that are close together, resulting in more similar attributes and overestimating performance due to spatial autocorrelation. To prevent overestimation, and increase the independence between testing and training data, we implemented spatial cross-validation⁷², using spatial blocks of 100 km, which were generated and assigned to one of 5 folds randomly.

Following spatial cross-validation, the dataset was partitioned into independent training and testing sets using a 70:30 split of all DHS clusters with valid covariate information. Random forest models were trained separately for each drinking water source category. The response variable for each model was the proportion of the population within a survey cluster that use a given source as their primary drinking water source. In total, 11 source categories were directly modelled (all source types excluding other which was the residual). All models were trained using the same set of predictor variables and fitted using the ranger package⁷³ in R with 500 trees.

Model performance was evaluated using population-weighted metrics calculated on the independent (30%) test dataset calculating the coefficient of determination (R^2) and mean absolute error (MAE).

Hierarchical normalisation

A hierarchical normalisation procedure was applied reflecting both the relative reliability of model predictions and the underlying JMP access hierarchy as outlined in Deshpande et al., 2020⁶; piped, improved, unimproved and surface. In cases where the sum of all sources was less than 1, the residual was assigned to the class 'other'. Where the sum of all sources exceeded 1, the hierarchical normalisation process was implemented; estimates using the piped water models which typically represents the highest level of service and had highest model reliability (pre-normalisation OOB R^2 = 0.540) were first anchored. Non-piped improved source predictions (pre-normalisation OOB R^2 =0.537) were then rescaled proportionally together, preserving the relative proportions of each individual model and constraining the total to the remaining residual space following piped if the cumulative sum exceeded unity. This logic was applied sequentially to the unimproved (pre-normalisation OOB R^2 =0.300) and surface water (pre-normalisation OOB R^2 = 0.16) categories, where each subsequent model was normalised based on the residual space from higher-tier indicators.

Validation with JMP data

In addition to evaluating model performance using independent test data, model outputs were validated against national estimates from the WHO/UNICEF Joint Monitoring Programme (JMP) 2025 dataset²⁷. Predicted grid-cell values were aggregated to national-level population-weighted shares for each water source category and compared with corresponding JMP country estimates. Because JMP reports data use broader service categories, model predictions were grouped to match available JMP classifications of piped water, surface water, unimproved water (excluding surface water), and improved water. Performance metrics (MAE and R²) were calculated both globally and regionally as well as for individual source classes.

Uncertainty

Spatial uncertainty in water source predictions was quantified using a geographically stratified bootstrap procedure. For each water source model 30 bootstrap iterations were performed (the number was determined by stability analysis), each using 100 decision trees. For each iteration, a training sample was drawn with replacement, stratified by continental region to avoid over-representation of data-dense regions. Each model was applied to generate a full spatial prediction surface. For each pixel, the interquartile range (IQR) and the coefficient of variation (CoV) were calculated. A dominant source plot of uncertainty was produced with the uncertainty of the dominant source used in each pixel.

Variable importance

Variable importance was assessed using SHapley Additive exPlanations (SHAP), to calculate the contributions of variables to model predictions by computing the average contribution of that variable for a given prediction⁷⁴. For each of the eleven directly modelled water source models, SHAP values were estimated using the fastshap package⁷⁵ with 30 simulations per observation on a random sample of 1,000 training observations (for computational efficiency). To summarise importance across sources, a rank-weighted score was calculated by assigning points to each predictor based on the variable importance for a given model (most important 1 = 10 points, 10th importance = 1 point, below top 10 variables = 0). Summary of average variable importance was calculated as the average rank weighted variable importance for each variable and ranked by importance.

Limitations

A few limitations should be acknowledged. First, water source categories remain broad. In particular, the piped category does not distinguish between underlying supply types (e.g., surface water, groundwater, or desalination), limiting the ability to assess source-specific hydrological vulnerability⁷⁶. Some level of aggregation is also conducted, such as the aggregation of piped water to neighbours and public standpipes in the same category. In addition, certain sources (e.g., water kiosks) are not explicitly modelled due to inconsistent categorisation across datasets. Second, estimates are based on reported primary drinking water source and do not capture multi-source use or seasonal switching, which are common coping strategies^{30,77}. Nor does it reflect drinking water sources in public facilities, such as schools or hospitals⁷⁸. As a result, the analysis may underrepresent the role of certain sources in water access. Third, governance and management arrangement, such as distinctions between community-managed systems and self-supply^{33,34}, are not directly observed. The type of

regulation has significant implications for reliability, regulation, and long-term sustainability and further studies should aim to represent these distinctions.

Acknowledgements

The authors are grateful to the households and communities who participated in the Demographic and Health Surveys (DHS) and Multiple Indicator Cluster Surveys (MICS). We acknowledge the DHS Program, funded by the United States Agency for International Development (USAID) and implemented by ICF, for the collection and dissemination of DHS survey data. We acknowledge UNICEF for the design and implementation of the MICS programme and for making survey data publicly available.

Funding

Removed for anonymity

Author contributions

Removed for anonymity

Competing interests

The authors declare no conflicts of interest.

Data availability

Household survey data was obtained from DHS⁶⁰ and MICS⁶¹ surveys through their websites. Covariate data sources are found in Extended Data Table 2. A synthetic small (1000 household) sample database for code testing has been generated. The synthetic household survey data, source-explicit spatial estimates, associated uncertainty metrics, and summary data have been deposited to the Open Science Framework (doi: 10.17605/OSF.IO/JDV7A), accessible to reviewers [here](#).

Code availability

All code has been deposited to the Open Science Framework and available [here](#).

References

1. Bain, R. *et al.* Fecal Contamination of Drinking-Water in Low- and Middle-Income Countries: A Systematic Review and Meta-Analysis. *PLoS Med.* **11**, e1001644 (2014).
2. Foster, T. *et al.* Does the source of water for piped supplies affect child health? Evidence from rural Vanuatu. *Journal of Water, Sanitation and Hygiene for Development* **9**, 591–595 (2019).
3. Dakyaga, F., Schramm, S., Lupala, J. M. & Magembe-Mushi, D. L. Charging the non-networked: Water pricing governance of the heterogeneous infrastructures beyond the utility network in Dar es Salaam. *Environ. Plan. E Nat. Space* **7**, 1868–1892 (2024).
4. Howard, G., Calow, R., Macdonald, A. & Bartram, J. Climate Change and Water and Sanitation: Likely Impacts and Emerging Trends for Action. *Annu. Rev. Environ. Resour.* **41**, 253–276 (2016).
5. Greenwood, E. E. *et al.* Mapping safe drinking water use in low- and middle-income countries. *Science* (1979). **385**, 784–790 (2024).

- 430 6. Deshpande, A. *et al.* Mapping geographical inequalities in access to drinking water and sanitation facilities
431 in low-income and middle-income countries, 2000–17. *Lancet Glob. Health* **8**, e1162–e1185 (2020).
- 432 7. World Health Organization (WHO) and the United Nations Children’s Fund (UNICEF). *Progress on Household*
433 *Drinking Water, Sanitation and Hygiene. Special Focus on Inequalities.* (2025).
- 434 8. Shaheed, A., Orgill, J., Montgomery, M. A., Jeuland, M. A. & Brown, J. Why “improved” water sources are
435 not always safe. *Bull. World Health Organ.* **92**, 283–289 (2014).
- 436 9. Food and Agriculture Organization of the United Nations (FAO). AQUASTAT.
437 <https://data.apps.fao.org/aquastat/> (2026).
- 438 10. Nazari, S., Reinecke, R. & Moosdorf, N. Global estimates of groundwater withdrawal trends and
439 uncertainties. *Environmental Research Letters* **20**, 094043 (2025).
- 440 11. Rameshwaran, P. *et al.* Use of Abstraction and Discharge Data to Improve the Performance of a National-
441 Scale Hydrological Model. *Water Resour. Res.* **58**, (2022).
- 442 12. UNICEF. *The UNICEF Extreme Water Vulnerability Index (EWVI) – Methodology Paper.* (2021).
- 443 13. Foster, S., Hirata, R., Eichholz, M. & Alam, M.-F. Urban Self-Supply from Groundwater—An Analysis of
444 Management Aspects and Policy Needs. *Water (Basel)*. **14**, 575 (2022).
- 445 14. Prüss-Ustün, A. *et al.* Burden of disease from inadequate water, sanitation and hygiene for selected adverse
446 health outcomes: An updated analysis with a focus on low- and middle-income countries. *Int. J. Hyg.*
447 *Environ. Health* **222**, 765–777 (2019).
- 448 15. Wolf, J. *et al.* Burden of disease attributable to unsafe drinking water, sanitation, and hygiene in domestic
449 settings: a global analysis for selected adverse health outcomes. *The Lancet* **401**, 2060–2071 (2023).
- 450 16. Bain, R. *et al.* Global assessment of exposure to faecal contamination through drinking water based on a
451 systematic review. *Tropical Medicine & International Health* **19**, 917–927 (2014).
- 452 17. Murray, A. L. *et al.* Rural Water Point Functionality Estimates and Associations: Evidence From Nine
453 Countries in Sub-Saharan Africa and South Asia. *Water Resour. Res.* **60**, (2024).
- 454 18. Staddon, C. & Brewis, A. Household Water Containers: Mitigating risks for improved Modular, Adaptive, and
455 Decentralized (MAD) water systems. *Water Secur.* **21**, 100163 (2024).
- 456 19. Sequeira, A. R. *et al.* Assessing the short-term outcomes of a piped water supply intervention in peri-urban
457 Mozambique. *Journal of Water, Sanitation and Hygiene for Development* **9**, 348–355 (2019).
- 458 20. Howard, G. *et al.* Domestic Water Quantity, Service Level and Health Second Edition. ISBN: 97892-4-
459 001524-1 (2003).
- 460 21. Huggins, X., Gleeson, T., Villholth, K. G., Rocha, J. C. & Famiglietti, J. S. Groundwaterscapes: A Global
461 Classification and Mapping of Groundwater’s Large-Scale Socioeconomic, Ecological, and Earth System
462 Functions. *Water Resour. Res.* **60**, (2024).
- 463 22. Leão, S., Roux, P., Loiseau, E., Junqua, G. & Rosenbaum, R. K. Operationalisation and application of water
464 supply mix (WSmix) at worldwide scale: how does WSmix influence the environmental profile of water
465 supply for different users? *Int. J. Life Cycle Assess.* **24**, 2255–2267 (2019).

- 466 23. Fridman, D., Biran, N. & Kissinger, M. Beyond blue: An extended framework of blue water footprint
467 accounting. *Science of The Total Environment* **777**, 146010 (2021).
- 468 24. Onda, K., LoBuglio, J. & Bartram, J. Global Access to Safe Water: Accounting for Water Quality and the
469 Resulting Impact on MDG Progress. *Int. J. Environ. Res. Public Health* **9**, 880–894 (2012).
- 470 25. Tetteh, J. D. *et al.* Spatial heterogeneity in drinking water sources in the Greater Accra Metropolitan Area
471 (GAMA), Ghana. *Popul. Environ.* **44**, 46–76 (2022).
- 472 26. Yu, W. *et al.* Mapping access to domestic water supplies from incomplete data in developing countries: An
473 illustrative assessment for Kenya. *PLoS One* **14**, e0216923 (2019).
- 474 27. WHO/UNICEF Joint Monitoring Programme for Water Supply, S. and H. (JMP). JMP Global Database. (2025).
- 475 28. Lapworth, D. J. *et al.* Drinking water quality from rural handpump-boreholes in Africa. *Environmental*
476 *Research Letters* **15**, 064020 (2020).
- 477 29. Gaye, C. B. & Tindimugaya, C. Review: Challenges and opportunities for sustainable groundwater
478 management in Africa. *Hydrogeology Journal* vol. 27 1099–1110 Preprint at
479 <https://doi.org/10.1007/s10040-018-1892-1> (2019).
- 480 30. Grönwall, J. & Danert, K. Regarding Groundwater and Drinking Water Access through A Human Rights Lens:
481 Self-Supply as A Norm. *Water (Basel)*. **12**, 419 (2020).
- 482 31. Aina, I. V., Thiam, D. R. & Dinar, A. Substitution of piped water and self-supplied groundwater: The case of
483 residential water in South Africa. *Util. Policy* **80**, 101480 (2023).
- 484 32. Galada, H. C. *et al.* Assessing preferences regarding centralized and decentralized water infrastructure in
485 post-earthquake Leogane, Haiti. *Earth Perspectives* **1**, 5 (2014).
- 486 33. Foster, T. *et al.* Self-supplied drinking water in low- and middle-income countries in the Asia-Pacific. *NPJ*
487 *Clean Water* **4**, 37 (2021).
- 488 34. Fischer, A. *et al.* Risky responsibilities for rural drinking water institutions: The case of unregulated self-
489 supply in Bangladesh. *Global Environmental Change* **65**, 102152 (2020).
- 490 35. Healy, A. *et al.* Perceptions of Trends in the Development of Private Boreholes for Household Water
491 Consumption: Findings from a Survey of Water Professionals in Africa. RIGSS Working Paper, Cardiff
492 University, UK. (2018) 476 477
- 493 36. Danert, K. & Healy, A. Monitoring Groundwater Use as a Domestic Water Source by Urban Households:
494 Analysis of Data from Lagos State, Nigeria and Sub-Saharan Africa with Implications for Policy and Practice.
495 *Water (Basel)*. **13**, 568 (2021).
- 496 37. Moyo, N. A. G. An analysis of the chemical and microbiological quality of ground water from boreholes and
497 shallow wells in Zimbabwe. *Physics and Chemistry of the Earth, Parts A/B/C* **66**, 27–32 (2013).
- 498 38. Hinton, R. G. K. *et al.* Mixed method analysis of anthropogenic groundwater contamination of drinking
499 water sources in Malawi. *Science of The Total Environment* **957**, 177418 (2024).
- 500 39. Podgorski, J. & Berg, M. Global threat of arsenic in groundwater. *Science (1979)*. **368**, 845–850 (2020).

- 501 40. Jasechko, S. & Perrone, D. Global groundwater wells at risk of running dry. *Science* (1979). **372**, 418–421
502 (2021).
- 503 41. MacDonald, A., Davies, J., Calow, R. & Chilton, J. Prelims - Developing Groundwater. in *Developing*
504 *Groundwater* i–xx (Practical Action Publishing, 2005). doi:10.3362/9781780441290.000.
- 505 42. Foster, T., Furey, S., Banks, B. & Willetts, J. Functionality of handpump water supplies: a review of data from
506 sub-Saharan Africa and the Asia-Pacific region. *Int. J. Water Resour. Dev.* **36**, 855–869 (2020).
- 507 43. Family, L. *et al.* Reasons why low-income people in urban areas do not drink tap water. *The Journal of the*
508 *American Dental Association* **150**, 503–513 (2019).
- 509 44. Wheeler, S. A. Informal and formal markets in meeting water needs. *Nat. Sustain.* **6**, 1291–1292 (2023).
- 510 45. Zozmann, H., Morgan, A., Klassert, C., Klauer, B. & Gawel, E. Can Tanker Water Services Contribute to
511 Sustainable Access to Water? A Systematic Review of Case Studies in Urban Areas. *Sustainability* **14**, 11029
512 (2022).
- 513 46. Alba, R., Bruns, A., Bartels, L. & Kooy, M. Water Brokers: Exploring Urban Water Governance through the
514 Practices of Tanker Water Supply in Accra. *Water (Basel)*. **11**, 1919 (2019).
- 515 47. Klassert, C. *et al.* Unexpected growth of an illegal water market. *Nat. Sustain.* **6**, 1406–1417 (2023).
- 516 48. Constantine, K., Massoud, M., Alameddine, I. & El-Fadel, M. The role of the water tankers market in water
517 stressed semi-arid urban areas: Implications on water quality and economic burden. *J. Environ. Manage.*
518 **188**, 85–94 (2017).
- 519 49. Abraham, A., Cheema, S., Chaabna, K., Lowenfels, A. B. & Mamtani, R. Rethinking bottled water in public
520 health discourse. *BMJ Glob. Health* **9**, e015226 (2024).
- 521 50. World Bank Group. Databank: Metadata Glossary. Annual freshwater withdrawals, domestic (% of total
522 freshwater withdrawal). [https://databank.worldbank.org/metadataglossary/world-development-](https://databank.worldbank.org/metadataglossary/world-development-indicators/series/ER.H2O.FWDM.ZS)
523 [indicators/series/ER.H2O.FWDM.ZS](https://databank.worldbank.org/metadataglossary/world-development-indicators/series/ER.H2O.FWDM.ZS) (2026).
- 524 51. BGS. FAO AQUASTAT Information. https://earthwise.bgs.ac.uk/index.php/FAO_AQUASTAT_Information
525 (2026).
- 526 52. Madani, K. Global Water Bankruptcy: Living Beyond Our Hydrological Means in the Post-Crisis Era. United
527 Nations University Institute for Water, Environment and Health (UNU-INWEH), Richmond Hill, Ontario,
528 Canada. (2026).
- 529 53. Smith, M., Cross, K., Paden, M. & Laban, P. *Spring : Managing Groundwater Sustainably*. International
530 Union for Conservation of Nature and Natural Resources (2016).
- 531 54. Margat, Jean. Groundwater around the World: A Geographic Synopsis. CRC Press, Taylor & Francis Group.
532 (2013).
- 533 55. Zektser, I. & Everett, L. Groundwater resources of the world and their use. United Nations Educational,
534 Scientific and Cultural Organization, 7, place de Fontenoy, 75352 Paris 07 SP, (2004).
- 535 56. Wada, Y. *et al.* Global monthly water stress: 2. Water demand and severity of water stress. *Water Resour.*
536 *Res.* **47**, (2011).

537 57. Graham, J. P. & Polizzotto, M. L. Pit Latrines and Their Impacts on Groundwater Quality: A Systematic
538 Review. *Environ. Health Perspect.* **121**, 521–530 (2013).

539 58. Genter, F., Willetts, J. & Foster, T. Faecal contamination of groundwater self-supply in low- and middle
540 income countries: Systematic review and meta-analysis. *Water Res.* **201**, 117350 (2021).

541 59. Sullivan, C. Calculating a Water Poverty Index. *World Dev.* **30**, 1195–1210 (2002).

542 60. The DHS Programme. Demographic and Health Surveys (DHS). <https://dhsprogram.com/>. (2026)

543 61. United Nations Children’s Fund (UNICEF). Multiple Indicator Cluster Surveys (MICS) Surveys.
544 <https://mics.unicef.org/surveys>. (2026)

545 62. Gashi, N., Burdziej, J. & Beshanski-Pedersen B. *Geographic Displacement Procedures and Dissemination of*
546 *Geocoded Data in the Multiple Indicator Cluster Surveys (MICS) Programme. MICS Methodological Papers,*
547 *No. 12 (Version 2).* (2026).

548 63. Breiman, L. Random Forests. *Mach. Learn.* **45**, 5–32 (2001).

549 64. Hengl, T. *et al.* SoilGrids250m: Global gridded soil information based on machine learning. *PLoS One* **12**,
550 e0169748 (2017).

551 65. Couronné, R., Probst, P. & Boulesteix, A.-L. Random forest versus logistic regression: a large-scale
552 benchmark experiment. *BMC Bioinformatics* **19**, 270 (2018).

553 66. Rokach, L. & Maimon, O. Data mining with decision trees. Theory and applications. in 1–11. *World Scientific*
554 *Pub Co Inc*, (2008).

555 67. Louppe, G. Understanding Random Forests: From Theory to Practice. (2014).

556 68. Messenger, M. L., Lehner, B., Grill, G., Nedeva, I. & Schmitt, O. Estimating the volume and age of water stored
557 in global lakes using a geo-statistical approach. *Nat. Commun.* **7**, 13603 (2016).

558 69. Lehner, B. & Grill, G. Global river hydrography and network routing: baseline data and new approaches to
559 study the world’s large river systems. *Hydrol. Process.* **27**, 2171–2186 (2013).

560 70. Tatem, A. J. WorldPop, open data for spatial demography. *Sci. Data* **4**, 170004 (2017).

561 71. Elvidge, C. D., Baugh, K., Zhizhin, M., Hsu, F. C. & Ghosh, T. VIIRS night-time lights. *Int. J. Remote Sens.* **38**,
562 5860–5879 (2017).

563 72. Sun, K., Hu, Y., Lakhanpal, G. & Zhou, R. Z. Spatial Cross-Validation for GeoAI. in *Handbook of Geospatial*
564 *Artificial Intelligence* 201–214. CRC Press, Boca Raton, (2023).

565 73. Wright, M. N. ranger: A Fast Implementation of Random Forests. *CRAN: Contributed Packages* Preprint at
566 <https://doi.org/10.32614/CRAN.package.ranger> (2015).

567 74. Lundberg, S. & Lee, S.-I. A Unified Approach to Interpreting Model Predictions. *ArXiv*
568 <https://doi.org/10.48550/arXiv.1705.07874> (2017).

569 75. Greenwell, B. fastshap: Fast Approximate Shapley Values. <https://doi.org/10.32614/CRAN.package.fastshap>
570 (2024).

571 76. Lester, S. & Rhiney, K. Going beyond basic access to improved water sources: Towards deriving a water
572 accessibility index. *Habitat Int.* **73**, 129–140 (2018).

573 77. Elliott, M. *et al.* Addressing how multiple household water sources and uses build water resilience and
574 support sustainable development. *NPJ Clean Water* **2**, 6 (2019).

575 78. Hope, R. Four billion people lack safe water. *Science* (1979). **385**, 708–709 (2024).

576

577

578 **Extended Data**

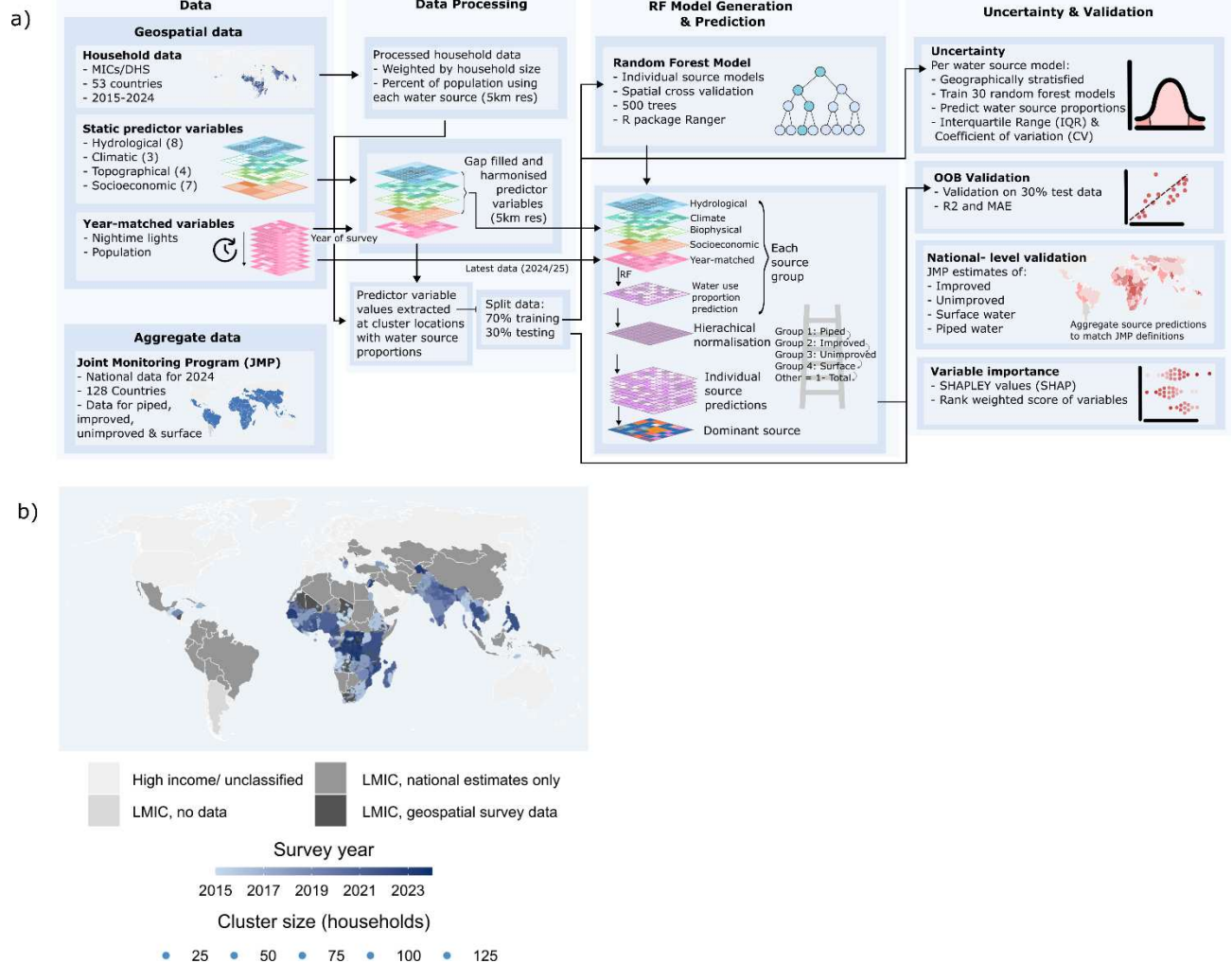
579 Extended Data Table 1: Water classes used to aggregate water sources, as in Bain et al.,
 580 2014¹⁶, within this analysis and their associated vulnerabilities. Improved water sources
 581 highlighted in blue, water sources classed as unimproved are marked with an asterisk*.

Water source	Description and core vulnerabilities
Piped water on premises	Piped water, often large scale from utility companies, especially in urban areas, but can be small scale local schemes piped to dwelling or yard. Reliability is highly context dependent; many areas face challenges of intermittent supply ^{7,79,80} , reducing reliability and linked to microbial contamination ^{80,81}
Piped water off premises	Water piped either to a public tap/ standpipe or to a neighbour. Can be from utility companies or small-scale schemes. May face challenges of intermittent supply ^{7,79,80} , contamination ^{80–82} , and challenges of access, especially if located on someone else's property.
Tube well/borehole	A narrow tube, usually mechanically, drilled into the ground enabling (deep) groundwater abstraction (typically 20-100m deep) ⁴¹ . They can be expensive to construct/repair and require maintenance which often results in high levels of breakdown. Whilst generally considered safer than shallow wells ⁴¹ , they can be at risk of geogenic (dependent on geology) and microbial contamination ⁸³ , especially from sanitation in densely populated areas ^{38,57} . Whilst generally allowing access to more reliable groundwater ⁴¹ , they can still experience seasonal availability fluctuations ⁸⁴ .
Protected spring	Protected springs are areas of natural groundwater flow from the ground engineered to prevent contamination and increase access, typically with a collection tank and tap, often built away from the eye of the spring to minimise contamination risk ⁴¹ . They are particularly vulnerable to contamination and seasonal water scarcity ^{4,41} .
Protected dug well	Protected wells are hand-dug or shallow wells (typically <20m) which are constructed to minimise contamination through protective structures (lining, apron, and cover, typically concrete) ⁴¹ . Hand-dug wells require minimal equipment for construction making them a relatively accessible source of water. Their design, enabling water accumulation when not in use, makes them suitable even in poorly permeable aquifers ⁸⁴ . Whilst protective infrastructure minimises surface contamination risk there is still a risk of subsurface contamination, especially from sanitation infrastructure ^{57,85} . Greater risk to drought and water scarcity than boreholes due to relying on shallower groundwater ^{41,84}
Rainwater harvesting	Water collected, and stored, from rainfall, often from roof run-off. Typically managed at a household, or individual level. They can often lack adequate protection ³³ , creating challenges of contamination, especially from storage ^{86,87} . They are also especially vulnerable to seasonal water shortage, making them particularly vulnerable to increased rainfall variability under climate change ^{4,33} .
Delivered water (Tanker truck, cart with small tank)	Tanker trucks or carts used to deliver drinking water, they can deliver water directly to users or to a central location which users travel to collect it. Usually operated by private water vendors ⁴⁵ , that are subject to limited regulation ⁴⁴ . They are often used in water scarce contexts or where other safe water supply systems are unavailable ⁴⁵ . There have been reports of poor water-quality ⁴⁵ , often related to poor water-handling practices. They often have high costs ⁷⁹ , which can vary seasonally ⁴⁵ .

Bottled water	Often used as a drinking water source in tandem with other water sources, used for other domestic water use. Bottled/package water is typically purchased from private vendors, resulting in inconsistent regulation ⁴⁹ , though can be supplied by government/ NGOs, especially during disasters. It has a very high volumetric cost ⁷⁹ . Whilst often considered safer than alternative sources, concerns have been raised over contamination, particularly plastic contaminants ⁴⁹ . Concerns have also been raised regarding the high environmental pollution ⁴⁹ from plastic packaging.
Unprotected spring*	Areas of natural groundwater flow (see protected spring) but without protection to minimise contamination or collect the water. They face challenges of high contamination vulnerability ^{1,41} and high susceptibility to water scarcity and drought ⁴¹ (higher than protected sources due to the absence of a collection tank).
Unprotected dug well*	Shallow dug wells (see protected dug well) with no protection (lining or cover) ⁴¹ resulting in a high risk of (microbial) contamination, including from both sanitation infrastructure and surface run-off ^{57,83} . Also vulnerable to seasonal water availability ⁴¹ .
Surface water (lake, river, stream)*	Direct water abstraction from surface water sources including lakes, rivers and streams. It has high risk of contamination (especially faecal) as well as other water quality concerns including high turbidity. Surface water is also particularly vulnerable to seasonal fluctuations, making it often unavailable during periods of drought or water scarcity. Both water quantity and quality of surface water is also particularly vulnerable to climate change ⁸⁸ .
Other*	Includes non-specified, ambiguous, or country specific sources.

582

583



Extended Data Figure 1a: Modelling pipeline for source-explicit estimation of drinking water access. Household survey data from MICs/DHS clusters and 22 geospatial predictor variables (hydrological, climatic, topographical, socioeconomic) were processed to 5 km resolution. Source-specific random forest models (500 trees, spatial cross-validation) were trained for each drinking water source type, and predictions hierarchically normalised to ensure proportions sum to one. Uncertainty was quantified through a geographically stratified bootstrap (30 models per source). Model performance was assessed through hold-out validation and comparison against national JMP estimates; variable importance was evaluated using SHAP values. Figure 1b) Geographic distribution and characteristics of survey data used for model training and validation. Survey clusters from DHS and MICS surveys (2015–2024) are shown for the 53 LMICs with geospatial data available, coloured by survey year and scaled by cluster size (number of households). Countries are classified by data availability: LMICs with geospatial survey data (dark grey), LMICs with national estimates only (medium grey), LMICs with no data (light grey), and high-income or unclassified countries (white).

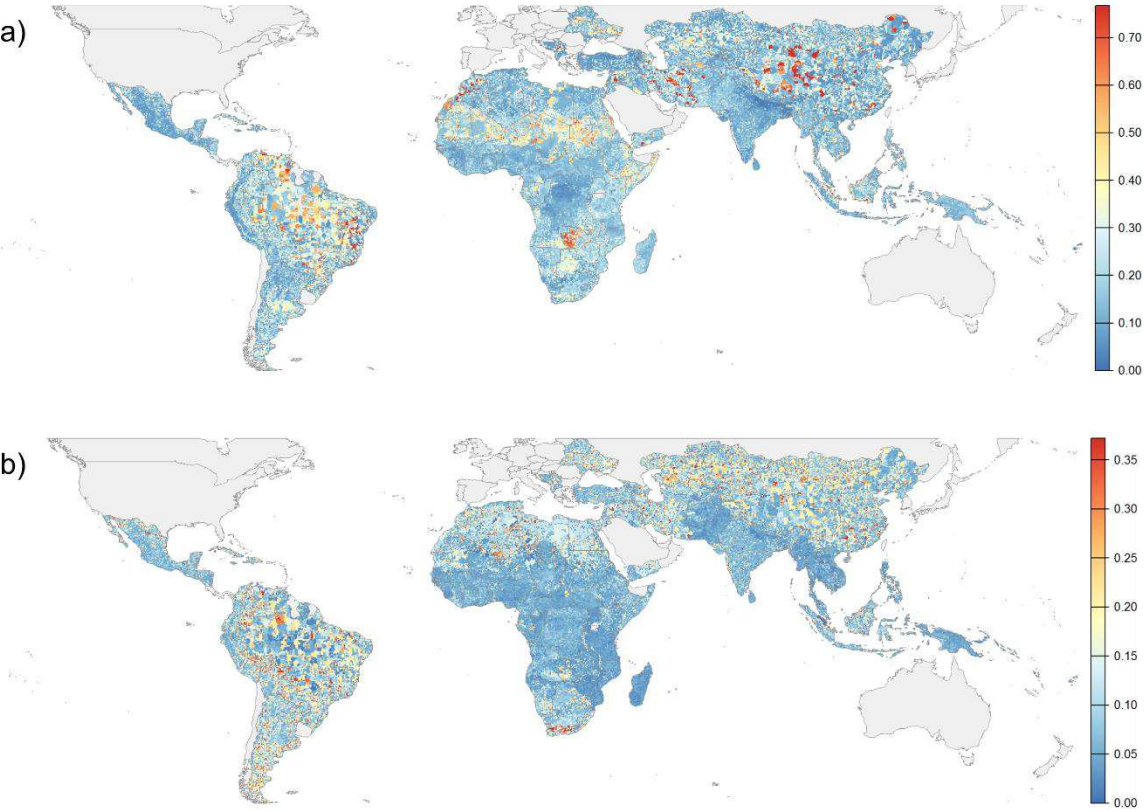
600 Extended Data Table 2: Model variable inputs and sources in group of variables with resolution
601 and year.

Variable group	Data	Source	Resolution	Year/s
Climate	Average Precipitation	MSWX Global Long-term average precipitation ⁸⁹ https://journals.ametsoc.org/view/journals/bams/103/3/BAMS-D-21-0145.1.xml	5 min	2000-2023
	Precipitation variability	Calculated as average annual standard deviation between months for precipitation averaged for 2000-2023 ⁸⁹	5 min	2000-2023
	Average temperature	MSWX Global Long-term average temperature ⁸⁹ https://journals.ametsoc.org/view/journals/bams/103/3/BAMS-D-21-0145.1.xml	5 min	2000-2023
Hydrology/hydrogeology	Soil water storage capacity	Fischer, G., F. Nachtergaele, S. Prieler, H.T. van Velthuizen, L. Verelst, D. Wiberg, 2008. <i>Global Agro-ecological Zones Assessment for Agriculture (GAEZ 2008)</i> . IIASA, Laxenburg, Austria and FAO, Rome, Italy. ⁹⁰	30 arc seconds	2008
	Distance to inland freshwater	Created from Hydorrivers (version 1.0) ⁶⁹ and Hydrolakes ⁶⁸ database.	5km	2019
	Stream density	Created from Hydorrivers (version 1.0) ⁶⁹ and Hydrolakes ⁶⁸ database.	5km	2019
	Depth to water table	Fan, Y., Li, H., & Miguez-Macho, G. (2013). Global Patterns of Groundwater Table Depth. <i>Science</i> , 339(6122), 940-943. doi:10.1126/science.1229881. ⁹¹	30 arc seconds	2010
	Aquifer system group	RICHTS, A., STRUCKMEIER, W. & ZAEPKE, M. (2011): WHYMAP and the Groundwater Resources of the World 1:25,000,000. In: Jones J. (Eds.): <i>Sustaining Groundwater Resources</i> . International Year of Planet Earth; Springer. doi: 10.1007/978-90-481-3426-7_10	0.5 degrees	2011
	Aquifer recharge class	RICHTS, A., STRUCKMEIER, W. & ZAEPKE, M. (2011): WHYMAP and the Groundwater Resources of the World 1:25,000,000. In: Jones J. (Eds.): <i>Sustaining Groundwater Resources</i> . International Year of Planet Earth; Springer. doi: 10.1007/978-90-481-3426-7_10 ⁹²	0.5 degrees	2011
	Permeability	Huscroft, J., Gleeson, T., Hartmann, J., & Börker, J. (2018). Compiling and mapping global permeability of the unconsolidated and consolidated Earth: GLobal HYdrogeology MaPS 2.0 (GLHYMPS 2.0). <i>Geophysical Research Letters</i> , 45, 1897–1904. https://doi.org/10.1002/2017GL075860 ⁹³	5 min	2014
	Specific yield	Huscroft, J., Gleeson, T., Hartmann, J., & Börker, J. (2018). Compiling and mapping global permeability of the unconsolidated and consolidated Earth: GLobal HYdrogeology MaPS 2.0 (GLHYMPS 2.0). <i>Geophysical Research Letters</i> , 45, 1897–1904. https://doi.org/10.1002/2017GL075860 ⁹³	5 min	2014
Socioeconomic	Global Gridded Relative Deprivation Index (GRDI)	Center For International Earth Science Information Network-CIESIN-Columbia University. (2022). <i>Global Gridded Relative Deprivation Index (GRDI), Version 1</i> (Version 1.00) [Data set]. Palisades, NY: NASA Socioeconomic Data and Applications Center (SEDAC). https://doi.org/10.7927/3XXE-AP97 Date Accessed: 2025-11-25 ⁹⁴	1 km	2010-2020
	Infant mortality rate (IMR)	Center For International Earth Science Information Network-CIESIN-Columbia University. (2022). <i>Global Gridded Relative Deprivation Index (GRDI), Version 1</i> (Version 1.00) [Data set]. Palisades, NY: NASA Socioeconomic Data and Applications Center (SEDAC).	1 km	2010-2020

		https://doi.org/10.7927/3XXE-AP97 Date Accessed: 2025-11-25 ⁹⁴ *IMR component of GRDI, originally derived from IHME Global Burden of Disease estimates		
	Nighttime lights (VNL) trend	Center For International Earth Science Information Network-CIESIN-Columbia University. (2022). <i>Global Gridded Relative Deprivation Index (GRDI), Version 1</i> (Version 1.00) [Data set]. Palisades, NY: NASA Socioeconomic Data and Applications Center (SEDAC). https://doi.org/10.7927/3XXE-AP97 Date Accessed: 2025-11-25 ⁹⁴ * NTL trend component of GRDI, derived from VIIRS annual composites	1 km	2010-2020
	Subnational Human Development Index (SHDI)	Smits, J. & Permanyer, I. (2019). The Subnational Human Development Database. <i>Scientific Data</i> , 6, 190038. https://doi.org/10.1038/sdata.2019.38 . ⁹⁵ Available via GRDI (CIESIN, 2022). Accessed via: Center For International Earth Science Information Network-CIESIN-Columbia University. (2022). <i>Global Gridded Relative Deprivation Index (GRDI), Version 1</i> (Version 1.00) [Data set]. Palisades, NY: NASA Socioeconomic Data and Applications Center (SEDAC). https://doi.org/10.7927/3XXE-AP97 Date Accessed: 2025-11-25 ⁹⁴	1 km	2010-2020
	Built Environment	Pesaresi M., Politis P. (2023). GHS-BUILT-S R2023A – GHS built-up surface grid, derived from Sentinel2 composite and Landsat, multitemporal (1975-2030). European Commission, Joint Research Centre (JRC). https://doi.org/10.2905/9F06F36F-4B11-47EC-ABB0-4F8B7B1D72EA . ⁹⁶ Available via GRDI (CIESIN, 2022). Accessed via: Center For International Earth Science Information Network-CIESIN-Columbia University. (2022). <i>Global Gridded Relative Deprivation Index (GRDI), Version 1</i> (Version 1.00) [Data set]. Palisades, NY: NASA Socioeconomic Data and Applications Center (SEDAC). https://doi.org/10.7927/3XXE-AP97 Date Accessed: 2025-11-25 ⁹⁴	1 km	2010-2020
	Population	Bondarenko M., Priyatikanto R., Tejedor-Garavito N., Zhang W., McKeen T., Cunningham A., Woods T., Hilton J., Cihan D., Nosatiuk B., Brinkhoff T., Tatem A., Sorichetta A.. The spatial distribution of population in 2015-2030 at a resolution of 30 arc (approximately 1km at the equator) R2025A version v1. Global Demographic Data Project - WorldPop - School of Geography and Environmental Science, University of Southampton. DOI:10.5258/SOTON/WP00845 ⁹⁷	30 arc seconds	2015-2026
	Nighttime lights	Elvidge, C.D.; Zhizhin, M.; Ghosh, T.; Hsu, F.-C.; Taneja, J. Annual Time Series of Global VIIRS Nighttime Lights Derived from Monthly Averages: 2012 to 2019. <i>Remote Sens.</i> 2021 , <i>13</i> , 922. https://doi.org/10.3390/rs13050922	500m	2015-2026
Topographic	Landcover	Friedl, M., & Sulla-Menashe, D. (2022). MODIS/Terra+Aqua Land Cover Type Yearly L3 Global 500m SIN Grid V061 ⁹⁸	500m	2025
	Elevation	Jarvis, A., H. I. Reuter, A. Nelson and E. Guevara (2008). Hole-filled SRTM for the globe Version 4, available from the CGIAR-CSI SRTM 90m Database (http://srtm.csi.cgiar.org) ⁹⁹ Lehner, B., K. Verdin and A. Jarvis (2008). "New global hydrography derived from spaceborne elevation data." <i>Eos</i> 89(10): 93-94. ¹⁰⁰	5 min	2008

	Slope	Jarvis, A., H. I. Reuter, A. Nelson and E. Guevara (2008). Hole-filled SRTM for the globe Version 4, available from the CGIAR-CSI SRTM 90m Database (http://srtm.csi.cgiar.org) ⁹⁹ Lehner, B., K. Verdin and A. Jarvis (2008). "New global hydrography derived from spaceborne elevation data." Eos 89(10): 93-94. ¹⁰⁰	5 min	2008
	Depth to bedrock	Shangguan, W., Hengl, T., de Jesus, J.M., Yuan, H. and Dai, Y. (2017). Mapping the global depth to bedrock for land surface modeling. Journal of Advances in Modeling Earth Systems, 9(1) ¹⁰¹	250m	2017

Uncertainty



Extended Data Figure 2: Uncertainty plots for dominant water source a) Coefficient of Variation (CoV) for dominant water source and b) interquartile range (IQR) for dominant water source.

607 **Water source:**

608 Extended Data Table 3: Percentage of the population (%) and total population using each water
609 source by region (/million users). Brackets denote 95% standard deviation intervals. The 'Other'
610 category reflects the residual population; its uncertainty bounds indicate the range of potential
611 residual users.

Region		Piped on premises	Bottled	Spring protected	Piped off premises	Delivered	Dug protected	Dug unprotected	Spring unprotected	Bore hole	Surface	Rainwater	Other
Global LMIC	Population /%	36.8 % (29.4 %, 44.2 %)	6.8 % (5.8 %, 7.9 %)	2.4 % (1.6 %, 3.2 %)	13.8 % (10.2 %, 17.5 %)	6.8 % (4.8 %, 8.8 %)	5.2 % (3.8 %, 6.6 %)	2.6 % (1.8 %, 3.4 %)	1.5 % (0.9 %, 2.2 %)	19.4 % (15.9 %, 23.0 %)	1.1 % (0.6 %, 1.5 %)	2.0 % (1.4 %, 2.7 %)	1.4 % (0.0 %, 2.9 %)
	Population/ million users	2447 (1957, 2936)	454 (387, 522)	159 (110, 209)	920 (678, 1162)	451 (320, 582)	348 (256, 440)	173 (118, 228)	103 (62, 143)	1292 (1058, 1526)	70 (42, 98)	136 (93, 179)	92 (0, 193)
East Asia & Pacific	Population /%	45.2 % (26.7 %, 63.7 %)	12.8 % (10.1 %, 15.6 %)	2.4 % (1.0 %, 3.9 %)	12.8 % (6.6 %, 19.1 %)	7.6 % (2.9 %, 12.3 %)	5.0 % (2.2 %, 7.8 %)	0.70 % (0.25 %, 1.15 %)	0.39 % (0.19 %, 0.59 %)	8.1 % (3.5 %, 12.7 %)	0.37 % (0.21 %, 0.53 %)	3.4 % (1.5 %, 5.3 %)	1.2 % (0.0 %, 3.7 %)
	Population/ million users	943 (556, 1329)	268 (210, 326)	51 (20, 81)	268 (137, 399)	159 (61, 257)	104 (45, 162)	15 (5, 24)	8.14 (3.99, 12.3)	169 (72, 265)	7.77 (4.39, 11.2)	71 (31, 111)	26 (0, 76)
Europe & Central Asia	Population /%	62.4 % (47.2 %, 77.5 %)	5.3 % (3.6 %, 7.0 %)	2.2 % (1.1 %, 3.2 %)	8.4 % (3.3 %, 13.6 %)	9.7 % (4.6 %, 14.8 %)	3.7 % (1.4 %, 6.0 %)	0.24 % (0.12 %, 0.36 %)	0.24 % (0.15 %, 0.33 %)	4.4 % (1.8 %, 6.9 %)	0.23 % (0.16 %, 0.30 %)	2.0 % (1.2 %, 2.8 %)	1.3 % (0.0 %, 3.6 %)
	Population/ million users	155 (117, 192)	13 (9, 17)	5.35 (2.71, 7.99)	21 (8, 34)	24 (11, 37)	9.2 (3.55, 14.9)	0.597 (0.291, 0.902)	0.587 (0.368, 0.807)	11 (5, 17)	0.572 (0.388, 0.755)	4.86 (2.86, 6.86)	3.22 (0, 9.02)
Latin America & Caribbean	Population /%	47.0 % (29.7 %, 64.2 %)	13.6 % (10.3 %, 17.0 %)	2.2 % (1.1 %, 3.3 %)	11.7 % (5.2 %, 18.1 %)	9.2 % (4.3 %, 14.2 %)	4.9 % (1.9 %, 8.0 %)	0.58 % (0.12 %, 1.05 %)	0.62 % (0.31 %, 0.92 %)	5.9 % (2.2 %, 9.6 %)	0.53 % (0.34 %, 0.72 %)	2.8 % (1.4 %, 4.1 %)	0.99 % (0.0 %, 3.00 %)
	Population/ million users	286 (181, 391)	83 (62, 103)	13 (7, 20)	71 (32, 110)	56 (26, 86)	30 (11, 49)	3.56 (0.742, 6.38)	3.76 (1.9, 5.61)	36 (13, 59)	3.21 (2.04, 4.37)	17 (9, 25)	6.03 (0, 18.3)
Middle East & North Africa	Population /%	48.5 % (30.0 %, 67.0 %)	4.9 % (2.3 %, 7.6 %)	1.3 % (0.7 %, 1.9 %)	12.6 % (5.4 %, 19.7 %)	16.5 % (10.2 %, 22.8 %)	3.1 % (1.2 %, 4.9 %)	0.60 % (0.20 %, 1.00 %)	0.46 % (0.24 %, 0.69 %)	8.1 % (3.3 %, 12.9 %)	0.56 % (0.32 %, 0.81 %)	2.3 % (0.8 %, 3.8 %)	1.1 % (0.0 %, 3.9 %)

	Popula tion/ million users	215 (133, 297)	22 (10, 34)	5.68 (3.05, 8.31)	56 (24, 87)	73 (45, 101)	14 (5, 22)	2.67 (0.906, 4.44)	2.04 (1.04, 3.04)	36 (15, 57)	2.5 (1.4, 3.59)	10 (4, 17)	5 (0, 17.2)
South Asia	Popula tion /%	31.0 % (18.5 %, 43.4 %)	2.4 % (1.1 %, 3.7 %)	0.94 % (0.49 %, 1.39 %)	13.4 % (5.5% , 21.3 %)	5.1% (1.5% , 8.6%)	3.8% (2.0% , 5.6%)	1.8% (0.6%, 2.9%)	0.42% (0.21%, 0.62%)	38.7 % (29.2 %, 48.3 %)	0.34 % (0.1 4%, 0.55 %)	0.63% (0.23 %, 1.03%)	1.6 % (0.0 %, 5.4 %)
	Popula tion/ million users	610 (364, 855)	47 (22, 73)	18 (10, 27)	264 (109, 419)	100 (30, 170)	74 (39, 109)	35 (12, 57)	8.2 (4.21, 12.2)	763 (574, 952)	6.79 (2.6 8, 10.9)	12 (5, 20)	31 (0, 106)
Sub- Sahar an Africa	Popula tion /%	18.5 % (10.2 %, 26.7 %)	1.6 % (1.3 %, 2.0 %)	5.1% (2.2% , 8.1%)	18.7 % (9.3% , 28.1 %)	3.0% (1.0% , 5.1%)	9.1% (4.6% , 13.7 %)	9.1% (5.3%, 12.9%)	6.2% (3.1%, 9.3%)	21.6 % (14.2 %, 28.9 %)	3.8 % (1.7 %, 6.0 %)	1.6% (0.8%, 2.4%)	1.7 % (0.0 %, 4.9 %)
	Popula tion/ million users	238 (132, 344)	21 (17, 26)	66 (28, 104)	240 (119, 361)	39 (13, 65)	117 (59, 176)	117 (68, 166)	80 (39, 120)	277 (183, 372)	49 (21, 77)	21 (11, 31)	22 (0, 63)

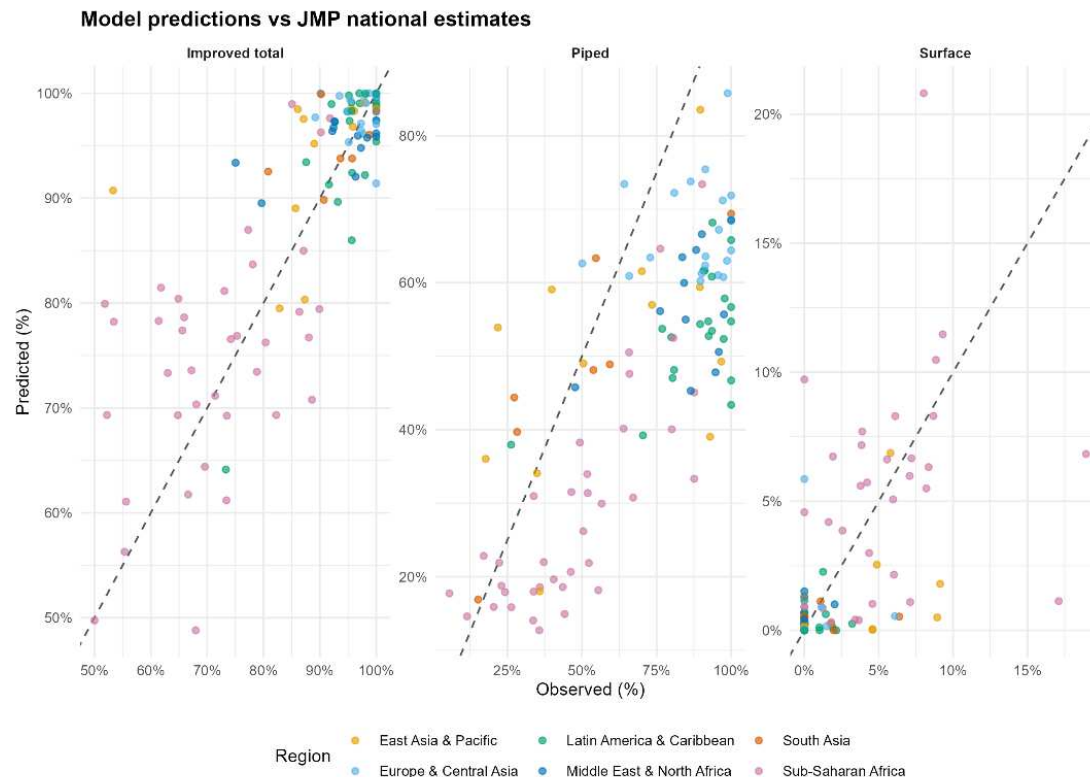
612

613 **Model performance:**

614 Extended Data Table 4: Model performance on independent test set, post normalisation

Water Source	R²	MAE
Piped on premises	0.566	0.1843
Piped off premises	0.488	0.1201
Borehole/ Tubewell	0.674	0.1009
Protected dug well	0.522	0.0291
Protected spring	0.473	0.0107
Rainwater	0.333	0.0059
Bottled/Packaged water	0.658	0.0372
Delivered water	0.563	0.0666
Unprotected dug well	0.198	0.0190
Unprotected spring	0.245	0.0147
Surface water	0.136	0.0085

615



Extended Data Figure 3: Validation with nation JMP monitoring data. Observed (JMP monitoring data) on the x axis with predicted percent of the population using each source on the y axis. Each point represents an individual country, coloured by their region.

Extended Data Table 5: Model performance by comparison of predicted percent of the population using improved water sources at the national level to JMP national data for improved water source, summarised by region.

Region	Number of countries	R ²	MAE
East Asia and Pacific	19	0.089	0.0680
Europe and Central Asia	17	0.001	0.0217
Latin America and Caribbean	20	0.712	0.0409
Middle East and North Africa	13	0.137	0.0315
South Asia	7	0.038	0.0478
Sub-Saharan Africa	46	0.311	0.112
Total	122	0.564	0.0612

625 Extended Data Table 6: Model performance for predicted percent of population using each
626 source class by comparison to JMP national data by source type

Region	MAE	RMSE	R^2
Piped water	0.242	0.278	0.576
Improved water	0.0612	0.0880	0.564
Surface water	0.0167	0.0315	0.336

627



628

629

630

631

632

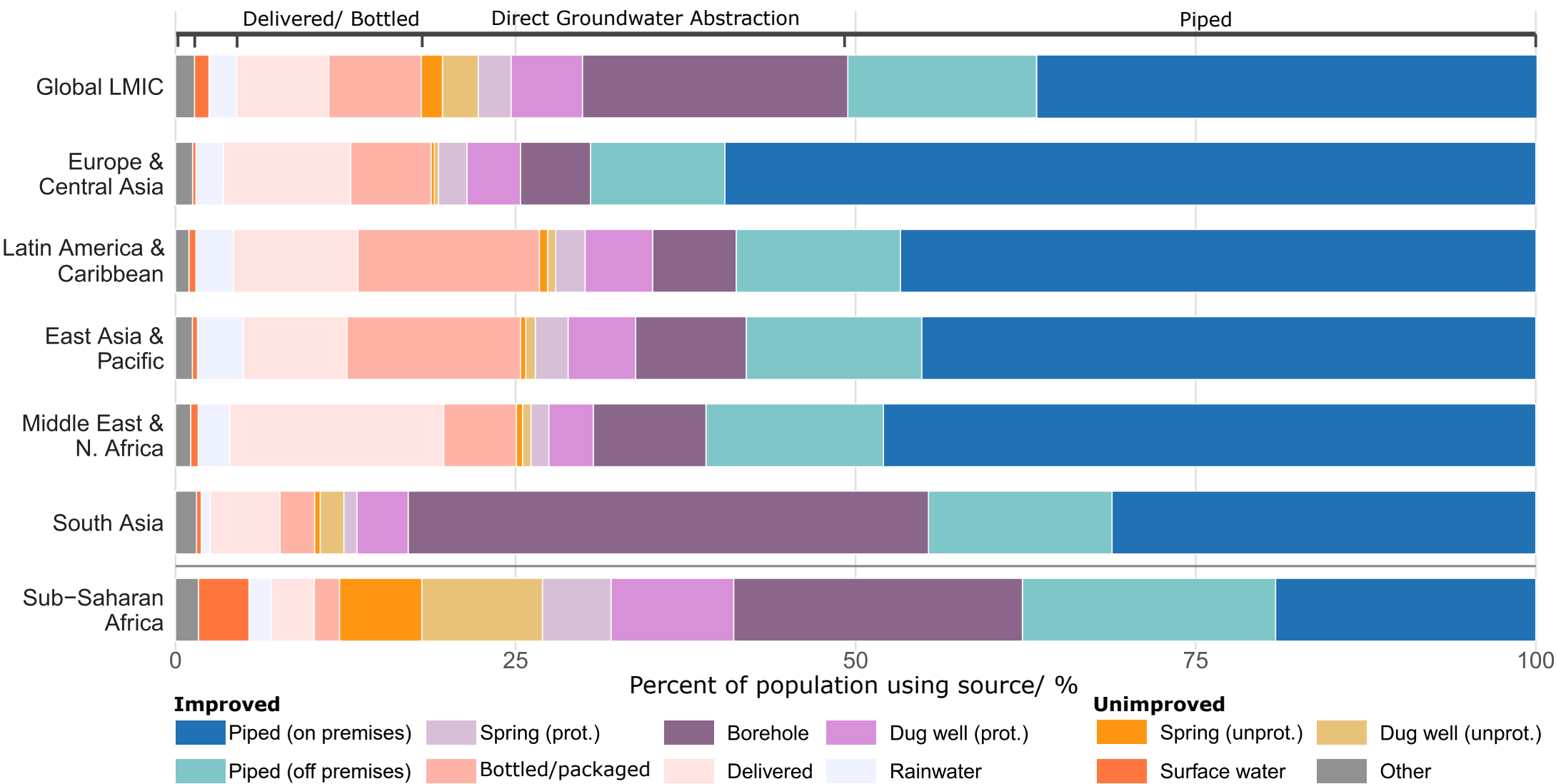
Extended data figure 4: Variable importance (SHAPley values) for individual water source models. Each point represents a training observation (n=1,000). Results are aggregated across 30 bootstrap iterations (nsim=30). SHAP values indicate the contribution of each predictor to model output. Variables are ordered by mean absolute SHAP value.

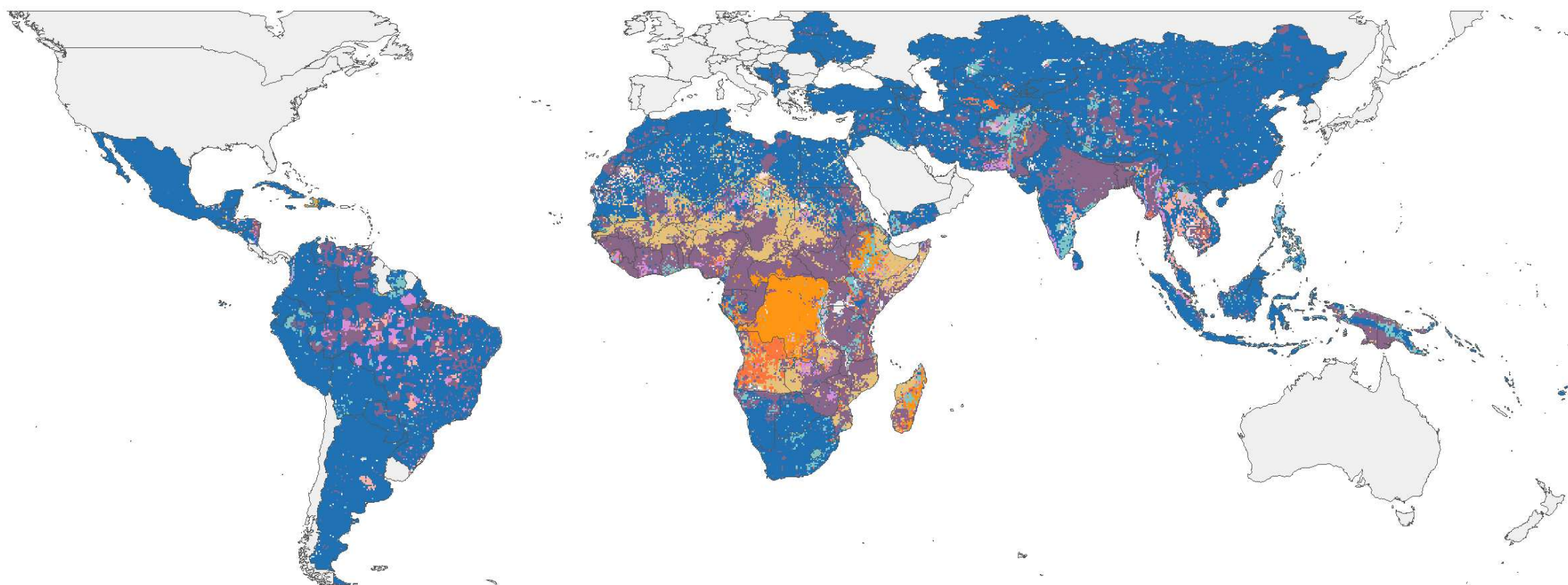
633 References Extended Data

- 634 79. Beard, V. A. & Mitlin, D. Water access in global South cities: The challenges of intermittency and
635 affordability. *World Dev.* **147**, 105625 (2021).
- 636 80. Kumpel, E. & Nelson, K. L. Intermittent Water Supply: Prevalence, Practice, and Microbial Water Quality.
637 *Environ. Sci. Technol.* **50**, 542–553 (2016).
- 638 81. Erickson, J. J., Smith, C. D., Goodridge, A. & Nelson, K. L. Water quality effects of intermittent water supply
639 in Arraiján, Panama. *Water Res.* **114**, 338–350 (2017).
- 640 82. Rosario-Ortiz, F., Rose, J., Speight, V., Gunten, U. von & Schnoor, J. How do you like your tap water? *Science*
641 (1979). **351**, 912–914 (2016).
- 642 83. Abanyie, S. K., Apea, O. B., Abagale, S. A., Amuah, E. E. Y. & Sunkari, E. D. Sources and factors influencing
643 groundwater quality and associated health implications: A review. *Emerg. Contam.* **9**, 100207 (2023).
- 644 84. MacAllister, D. J., MacDonald, A. M., Kebede, S., Godfrey, S. & Calow, R. Comparative performance of rural
645 water supplies during drought. *Nat. Commun.* **11**, 1099 (2020).
- 646 85. Mbae, M. *et al.* Onsite sanitation systems and contamination of groundwater: A systematic review of the
647 evidence for risk using the source-pathway-receptor model. *PLOS Water* **3**, e0000167 (2024).
- 648 86. de Kwaadsteniet, M., Dobrowsky, P. H., van Deventer, A., Khan, W. & Cloete, T. E. Domestic Rainwater
649 Harvesting: Microbial and Chemical Water Quality and Point-of-Use Treatment Systems. *Water Air Soil*
650 *Pollut.* **224**, 1629 (2013).
- 651 87. Nnaji, C. C., Nnaji, I. V. & Ekwule, R. O. Storage-induced deterioration of domestic water quality. *Journal of*
652 *Water, Sanitation and Hygiene for Development* **9**, 329–337 (2019).
- 653 88. Delpla, I., Jung, A.-V., Baures, E., Clement, M. & Thomas, O. Impacts of climate change on surface water
654 quality in relation to drinking water production. *Environ. Int.* **35**, 1225–1233 (2009).
- 655 89. Beck, H. E. *et al.* MSWX: Global 3-Hourly 0.1° Bias-Corrected Meteorological Data Including Near-Real-Time
656 Updates and Forecast Ensembles. *Bull. Am. Meteorol. Soc.* **103**, E710–E732 (2022).
- 657 90. Fischer, G., Shah, M., Van Velthuisen, H. & Nachtergaele, F. *Agro-Ecological Zones Assessment*. (2006).
- 658 91. Fan, Y., Li, H. & Miguez-Macho, G. Global Patterns of Groundwater Table Depth. *Science* (1979). **339**, 940–
659 943 (2013).
- 660 92. Richts, A., Struckmeier, W. F. & Zaepke, M. WHYMAP and the Groundwater Resources Map of the World
661 1:25,000,000. in *Sustaining Groundwater Resources* 159–173. Springer Netherlands, Dordrecht, (2011).
- 662 93. Huscroft, J., Gleeson, T., Hartmann, J. & Börker, J. Compiling and Mapping Global Permeability of the
663 Unconsolidated and Consolidated Earth: GLobal HYdrogeology MaPS 2.0 (GLHYMPS 2.0). *Geophys. Res. Lett.*
664 **45**, 1897–1904 (2018).
- 665 94. Center For International Earth Science Information Network-CIESIN-Columbia University. Global Gridded
666 Relative Deprivation Index (GRDI), Version 1 (Version 1.00) .(2022).
- 667 95. Smits, J. & Permanyer, I. The Subnational Human Development Database. *Sci. Data* **6**, 190038 (2019).

- 668 96. Pesaresi M & Politis P. GHS-BUILT-S R2023A – GHS built-up surface grid, derived from Sentinel2 composite
669 and Landsat, multitemporal (1975-2030). (2026)
- 670 97. Bondarenko, M. *et al.* The spatial distribution of population in 2015-2030 at a resolution of 30 arc
671 (approximately 1km at the equator) R2025A version v1. (2025).
- 672 98. Friedl, M. & Sulla-Menashe, D. MODIS/Terra+Aqua Land Cover Type Yearly L3 Global 500m SIN Grid V061.
673 (2022).
- 674 99. Jarvis, A., Reuter, H. I., Nelson, A. & Guevara, E. Hole-filled SRTM for the globe Version 4, available from the
675 CGIAR-CSI SRTM 90m . <http://srtm.csi.cgiar.org/> (2008).
- 676 100. Lehner, B., Verdin, K. & Jarvis, A. New Global Hydrography Derived From Spaceborne Elevation Data. *Eos,*
677 *Transactions American Geophysical Union* **89**, 93–94 (2008).
- 678 101. Shangguan, W., Hengl, T., Mendes de Jesus, J., Yuan, H. & Dai, Y. Mapping the global depth to bedrock for
679 land surface modeling. *J. Adv. Model. Earth Syst.* **9**, 65–88 (2017).

680





Improved

■ Piped (on premises)	■ Spring (prot.)	■ Borehole	■ Dug well (prot.)
■ Piped (off premises)	■ Bottled/packaged	■ Delivered	■ Rainwater

Unimproved

■ Spring (unprot.)	■ Dug well (unprot.)
■ Surface water	■ Other

