Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Global mapping of areas with a double burden of poor water quality and unsafe drinking water access

Matti Kummu

matti.kummu@aalto.fi

Aalto University https://orcid.org/0000-0001-5096-0163

Marko Kallio

Aalto University https://orcid.org/0000-0002-6917-7790

Veronica Baldasso

Aalto University

Saeed Karimzadeh

Aalto University

Veera Saarenheimo

Aalto University

Article

Keywords:

Posted Date: October 16th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10930794/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.

Global mapping of areas with a double burden of poor water quality and unsafe drinking water access

Marko Kallio [1] * #, Veronica Baldasso [1], Saeed Karimzadeh [1], Veera Saarenheimo [1], and Matti Kummu [1]*#

[1] Aalto University, Department of Built Environment, Water and Environmental Engineering Research Group

# contributed equally to the work

*Author to whom any correspondence should be addressed.

E-mail: marko.k.kallio@aalto.fi, matti.kummu@aalto.fi

Abstract

Approximately 500 million people globally still rely on surface water or unimproved sources for drinking water, even as surface water quality deteriorates in many regions, often the same regions where safe access is lacking. Yet this co-exposure is not quantified globally. We constructed the global harmonised subnational drinking water service estimates (G-HSDW), the first fully global subnational dataset of all four lower WHO/UNICEF Joint Monitoring Programme service rungs (basic, limited, unimproved, surface water). G-HSDW covers 3,040 admin 1 units across 240 countries over 1990–2023. We combined it with gridded surface water quality estimates (salinity, organic pollution, faecal coliforms) over 1990–2019. In 2019, only 1 in 10 people relying on unimproved or surface water sources lived in areas with low water quality risk, while 52% of people using unimproved and 60% using surface water sources faced high to severe risk. We identify dark spots where action is most urgent and bright spots where progress provides models for effective intervention.

1. Introduction

Access to safe drinking water is fundamental to human health and development (World Health Organization 2022). Yet half a billion people globally still rely on surface water or unprotected wells and springs for their drinking water (WHO/UNICEF 2023), making them highly vulnerable to contamination of the water bodies they depend on. At the same time, surface waters across many regions are experiencing deteriorating quality, from microbial pollution, salinisation, and organic loading (van Vliet *et al* 2023, Jones *et al* 2024, Desbureaux *et al* 2022), increasing pressure on those least equipped to treat or substitute their water source. Whether improving safe access and deteriorating quality are converging or diverging at policy-relevant scales remains unknown.

The WHO/UNICEF Joint Monitoring Programme (JMP) classifies drinking water services on a five-rung ladder from surface water and unimproved sources through limited, basic, and safely managed services (WHO/UNICEF 2017). While "safely managed" includes water quality criteria (i.e. free from priority chemicals and faecal contamination), the lower rungs are defined by source type and accessibility, leaving the quality of water consumed on these rungs largely unaddressed within the framework. Building on this framework, recent research has integrated source water quality into assessments of water security and exposure risk. At a global scale, microbial contamination has been shown to substantially reshape estimates of water stress and the spatial distribution of high-risk regions, particularly where surface water degradation constrains otherwise available resources. Complementary studies have mapped long-term variability in freshwater quality and its spatial

heterogeneity (Desbureaux *et al* 2022), while health burden assessments show substantial disease impacts from contaminated drinking water, including among populations classified as having improved or safely managed access (Wolf *et al* 2023, Chambers *et al* 2026). Microbial testing has also been integrated into nationally representative surveys and operationalised Sustainable Development Goals (SDG) monitoring by linking *E. coli* measurements with JMP service ladders (Bain *et al* 2014, 2021, Santos *et al* 2023, Irianti *et al* 2024). Geospatial modelling, in turn, has enabled increasingly high-resolution mapping of drinking water access at subnational scales (Ambel *et al* 2023, Greenwood *et al* 2024). However, the only global-scale effort to date (Greenwood *et al* 2024) focuses exclusively on the top JMP rung, safely managed services, leaving populations on the lower rungs unmapped at this resolution.

Despite these advances, three important limitations constrain our ability to assess where unsafe drinking water access and poor source water quality co-occur. First, the evidence base is dominated by microbial indicators, with limited integration of broader physicochemical parameters such as salinity or organic contamination into access frameworks (van Vliet *et al* 2021, Jones *et al* 2024, Hofstra *et al* 2026). Second, most studies rely on cross-sectional or survey-based snapshots, leaving the spatiotemporal co-variation of source water quality and drinking water access largely uncharacterised, including the direction and acceleration of long-term trends. Third, and most importantly, no harmonised, spatially explicit dataset jointly integrates source water quality and drinking water access metrics at scales relevant for policy prioritisation. Where subnational drinking water access has been mapped globally, it has been restricted in scope. The Local Burden of Disease (LBD) WaSH Collaborators (in Deshpande *et al* 2020) mapped source-type access (piped, improved, unimproved, surface water) for 2000–2017, but only across LMICs (low and middle income countries), omitting all high-income countries, most of Europe and the Middle East, parts of South America, and many small island states. Greenwood *et al* (2024) extended subnational mapping to the JMP "safely managed" rung, but only for a single year (2020) and only across 135 LMICs. Thus, no global subnational dataset has yet covered all countries and territories, aligned with the full JMP service-level definitions or extended subnational coverage across the past three decades. Consequently, there is no global estimate of how many people rely on contaminated surface water or unimproved sources, nor of how this exposure is changing over time.

We address this question by first constructing the global harmonised subnational drinking water service estimates (G-HSDW), the first fully global subnational dataset of population shares across the four lower JMP service rungs, covering 3,040 admin 1 units across 240 countries and territories over 1990–2023 (Methods). Second, we couple G-HSDW with gridded surface water quality estimates for total dissolved solids, biochemical oxygen demand, and faecal coliforms (Jones *et al* 2023, DynQual model), jointly capturing salinity, organic pollution, and microbial contamination over 1990–2019. We acknowledge that gridded surface water quality is a spatial proxy for water sources on lower JMP rungs, particularly for unimproved groundwater, but it is the only globally consistent, policy-scale representation currently available.

Using this framework, we characterise the status of co-exposure in 1990 and 2019, and long-term trends in water access and water quality over the intervening period, including whether trends are accelerating or slowing. This allows us to identify dark spots, where reliance on unimproved or surface water sources coincides with poor water quality, and bright spots, where both water quality and drinking water access have improved over the past three decades. Our results offer a spatially

explicit evidence base for prioritising investments in both drinking water service provision and surface water quality protection.

2. Results

2.1 Unimproved drinking water sources

Our G-HSDW dataset (see Methods for how this dataset was constructed) reveals previously unquantified patterns and trends for the different rungs of the JMP drinking water ladder over 1990–2019 (Figure 1). Although the G-HSDW dataset extends to 2023, we focus on 1990–2019 to match the temporal coverage of the DynQual water quality model (1980–2019; Jones *et al* 2022a). Progress has been particularly rapid in Asia, where only a few subnational areas still have populations relying on surface water (Figures 1g–h). The situation has also improved in Africa and Latin America, however large areas remain where more than 4% of the population relies on surface water (river, pond, or lake). For unimproved access to water (unprotected well or spring), coverage remains widespread across sub-Saharan Africa and parts of South and Southeast Asia (Figures 1d–e).

The share of population relying on surface water has decreased across most administrative units, while it has increased in units located in sub-Saharan Africa (Fig. 1i). The share relying on unimproved sources has also decreased, particularly rapidly in Asia as well as West and East Africa (Figure 1f). Rates have increased in Madagascar and the Democratic Republic of Congo, among some other smaller areas. However, in some cases this may represent a positive trend if populations are transitioning from surface water to unimproved sources. Similarly, for limited access (improved sources but >30 min travel time), increasing rates can largely be explained by populations moving up from the two lowest ladder rungs and can thus be interpreted as a positive development.

The total population across these three ladder rungs decreased from 1,282 million in 1990 to 774 million in 2019 (the last year for which we have water quality data) (Supplementary Figure S1). The most drastic decline occurred for surface water, from 305 million to 131 million, while the population relying on unimproved sources dropped from 761 million to 387 million. In contrast, the population with limited access increased from 216 million to 256 million — but as noted above, this largely reflects upward transitions from surface water or unimproved sources and therefore represents an improvement in drinking water service levels.

Our estimates derived from subnational data agree reasonably well with JMP national-level estimates (WHO/UNICEF 2023): 319 million for surface water, 395 million for unimproved, and 270 million for limited access, validating our data on a global scale. We also validated the dataset at national and subnational level against reported values (See Supplementary Figure S2 for comparison between national and subnational estimates, and S3 for validation). We found that our harmonisation with national-level data (see Methods) worked well, resulting in an R^2 of 0.99 for national-level validation across all three service levels (n = 5,157), with RMSE varying from 0.3 percentage-points (pp) (limited) to 0.8 pp (unimproved). The R^2 for subnational validation against reported values not directly used for the dataset varied from 0.69 (limited) to 0.87 (unimproved) (n = 1,027), while the range for RMSE was 5.0 pp – 6.7 pp (see Methods).

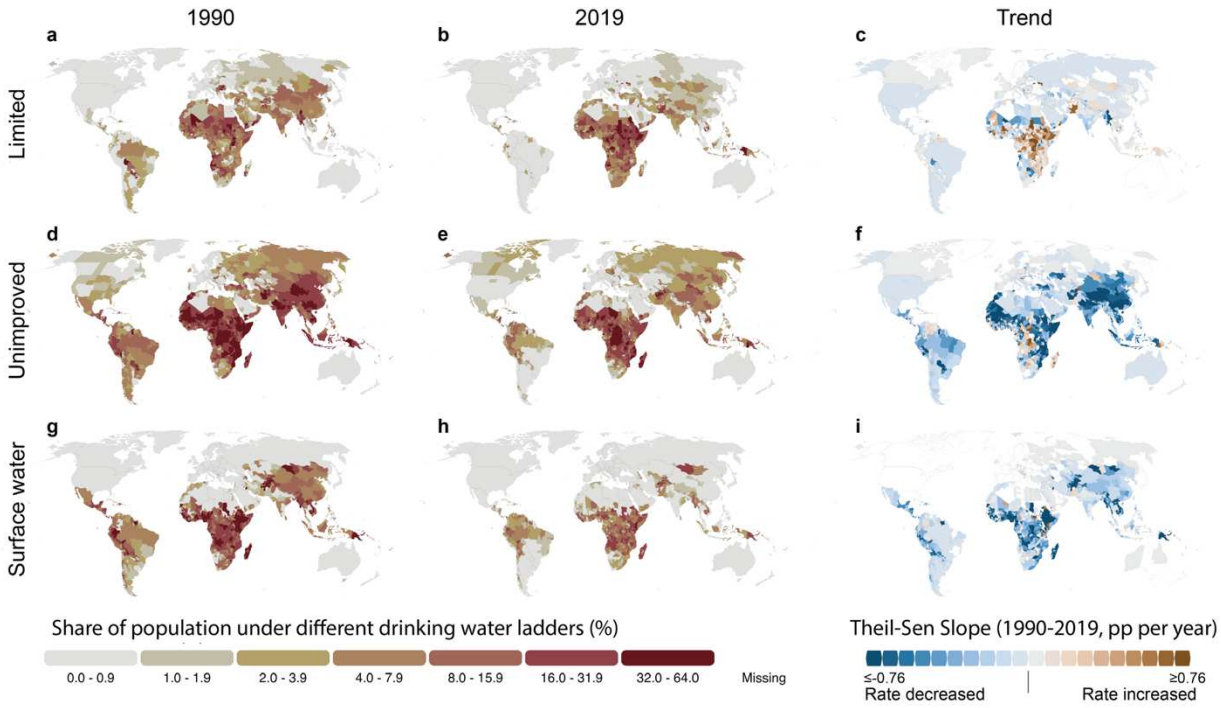


Figure 1. Share of population under different drinking water ladders. Limited access to drinking water (a-c), unimproved access (d-f), and surface water access (g-i). The status is shown for 1990 (a, d, g) and 2019 (b, e, h), the common span of the years of access to drinking water data (1990-2023) and water quality data (1980-2019). The trend (c, f, i) was calculated using Theil-Sen slope, using $p > 0.05$ as a threshold for significant trend.

2.2 Global surface water quality status and trends

We assessed the state and the trend in water quality, using organic contamination, pathogen contamination and salinity as quality parameters, over the study period 1990-2019, to establish a long-term change in risk. Based on our classification we find that, over the whole study period, risk arising from organic contamination, pathogen contamination and salinity is highly uneven across the world (Figure 2). The risk has remained high across 1990-2019 in large parts of the global south, including south, southeast and east Asia, large parts of Africa, and the Americas (see Figure 2). The risk is significantly lower in northern Eurasia and Europe, and these regions are some of the only ones where the trend shows reduction in risk, particularly for organic and pathogen contamination. In addition, the regions most exposed to water quality risk generally have high population, intensive agriculture and large urban populations. Conversely, the areas with least risk are large continuous areas with little population and/or ample water resources. Risk from salinity is overall lower across the world compared to that from organic and pathogen contamination, with hotspots mainly in areas of intensive agriculture and large population in South and East Asia, parts or Europe and North America, and smaller areas across South America and Africa. However, the overall trend indicates that salinity risk is globally increasing, with a few regional exceptions. Overall, the maps show overlapping pressure from the three water quality parameters across a belt that spans North America, Europe, and Central, South and East Asia.

The spatial patterns of water risk from the three water quality risk indicators show considerable similarity Figure 2. To test these associations, we computed Pearson's ρ and Chatterjee's ξ_n correlation coefficients (see Methods) for each admin area. On average, Pearson's ρ is high between organic and pathogen contamination (0.83 with a standard deviation of 0.20) and organic contamination and salinity (0.74, sd = 0.29). However, pathogen contamination and salinity risk

levels are only moderately correlated with one another (0.54, sd = 0.36). Associations estimated using Chatterjee's ξ_n are significantly lower at 0.51, 0.42 and 0.28, respectively, suggesting that although relatively strong linear relationships exist, each risk indicator is influenced by different environmental and anthropogenic processes. The full correlation table is available in Supplementary Table S1.

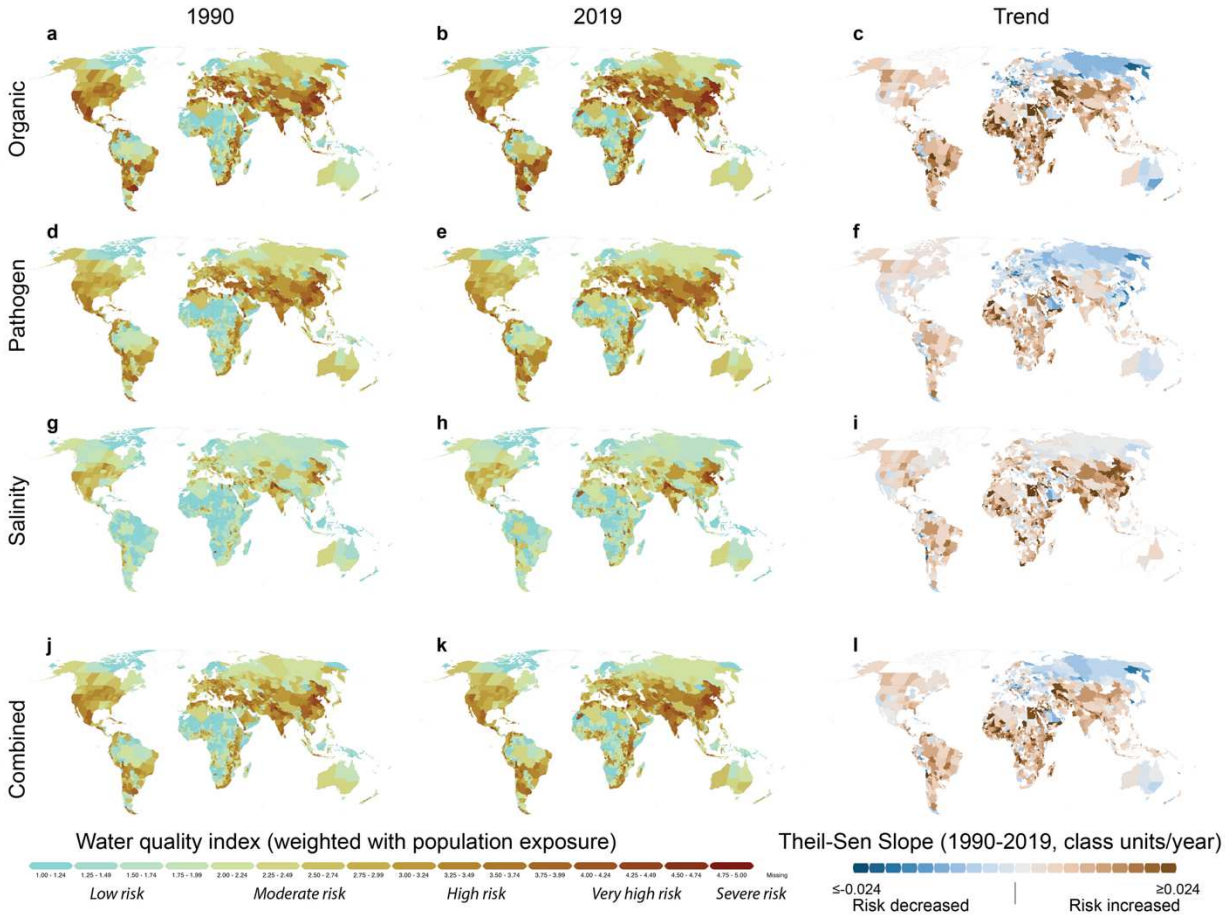


Figure 2. Water quality status and trends. Organic contamination risk (a-c), pathogen contamination risk (d-f), salinity risk (g-i), and combined risk indicator (j-l). The status is shown for 1990 (a, d, g) and 2019 (b, e, h), the common span of the years of access to drinking water data (1990-2023) and water quality data (1980-2019). The trend (c, f, i) was calculated using Theil-Sen slope, using $p < 0.05$ as a threshold for significant trend. See Table 1 for water quality thresholds, and Methods for how the combined risk indicator was calculated.

Since contaminant concentrations can exhibit substantial monthly variation due to variability in contaminant loading, dilution by varying streamflow, and in-stream degradation processes (Jones *et al* 2023), we assess the seasonality in water quality by computing, for each contaminant and continent, population-weighted mean risk classes among people relying on unimproved or surface water sources (Figure 3 and Supplementary Figure S4). Organic contamination, pathogen contamination and salinity risks all show distinct seasonality, and when considered, the associations between water quality parameters are substantially smaller than at annual level (see Supplementary Table S1). The seasonality of risks is considerably different between water quality parameters and continents. The minimum combined WQ risk occurs during early northern hemisphere winter in Africa and during northern hemisphere summer in Americas, during spring in Asia and Europe, and in the autumn in Oceania. In Europe, there is also a shift in seasonality: water quality risks overall have improved during spring months, but have deteriorated during the rest of the year. Globally, water

quality classification varies significantly within a year, with the global mean annual population weighted standard deviation in the class being 0.41 for organic contamination, 0.30 for pathogens, 0.27 for salinity, and 0.32 for combined risk. This seasonality particularly concerns population dependent on unimproved sources or surface water where water quality risks hovers between high and very high/severe (Supplementary Figure S5). Among those experiencing high risk conditions, mean global exposure was 423 ± 48 million in 1990, with a seasonal swing of 158 million between the annual trough (January) and peak (September), declining to 184 ± 13 million by 2019 (~55% reduction) with a narrowed seasonal swing of 41 million. Including very high and severe risk classes, combined exposure declined from 450 ± 51 million in 1990 to 202 ± 13 million in 2019.

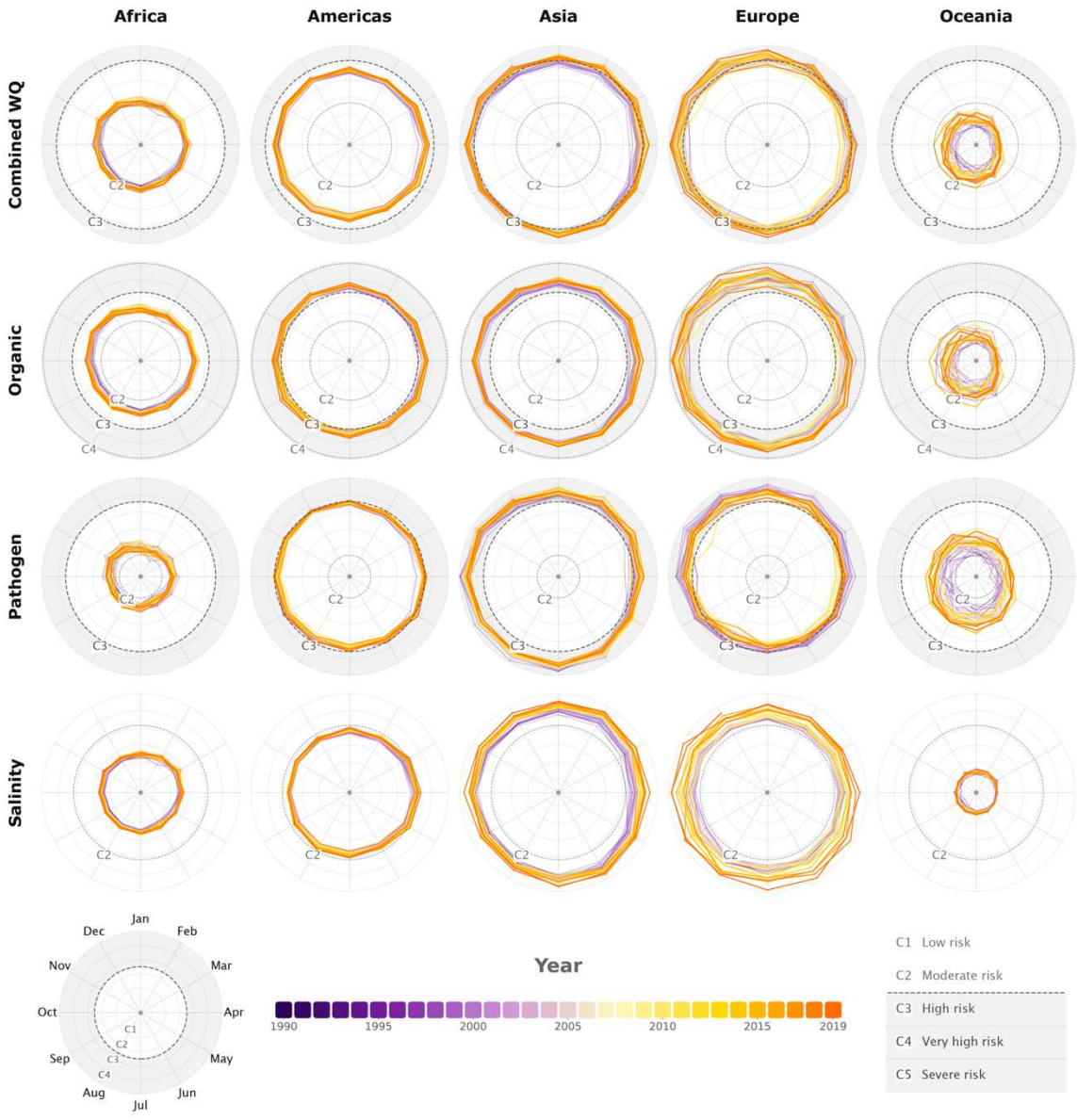


Figure 3. Water quality risk spirals showing the continental mean seasonality and evolution of risks over 1990-2019. The mean water quality indicators have been weighted using the population under surface water and unimproved service levels, and do not consider population under higher service levels.

To gain a more granular view of water quality evolution, we performed a changepoint analysis using monthly timeseries of contaminant concentrations (see Methods). The change in trends of the three water quality parameters show distinct spatial patterns. On average, we detect two distinct periods

in the timeseries across all admin areas: 2.20 periods for organic contamination, 1.97 for pathogen contamination and 2.42 for salinity. We detect no distinct periods for 42%, 29.4% and 27% of admin 1 areas for pathogen contamination, organic contamination and salinity, respectively, over the period 1990-2019. Temperate and high latitude areas have the least and tropical and subtropical areas have the largest numbers of change points (see Supplementary Figure S6). Additionally, salinity timeseries show the highest variability in terms of distinct periods.

The spatial distribution of the trend changes over the last ten years of the study period differs between the water quality indicators (Figure 4, Supplementary Figure S7). For organic contamination and salinity, approximately 35-45% and 40-55% of admin areas in every continent, respectively, experience a constant worsening trend. For pathogen contamination, constant worsening trend dominates only Africa and Oceania. Elsewhere, share of areas with no discernible trend has approximately 10-15 percentage point larger prevalence than constant worsening trend. Most positive trend changes occur in Europe; for organic contamination constant worsening and improving trends are equally prevalent, and for pathogen contamination improving constant trend is twice as prevalent than constant worsening trend. Europe, however, is the continent with the highest share of constant and accelerating salinity in approximately 60% of admin areas. Across all water quality indicators and continents, accelerating worsening water quality affects approximately 3-13% of admin areas. The most worrying trends are in Oceania, where constant and worsening water quality deterioration dwarf positive trend changes. Globally, 320 million, 272 million and 314 million people dependent on unimproved or surface water sources (in 2019) live in areas where water quality is deteriorating (positive trend, or trend changed from improving to worsening) whereas a considerably smaller number of people (149 million, 70 million and 98 million people, respectively) live in areas with improving water quality.

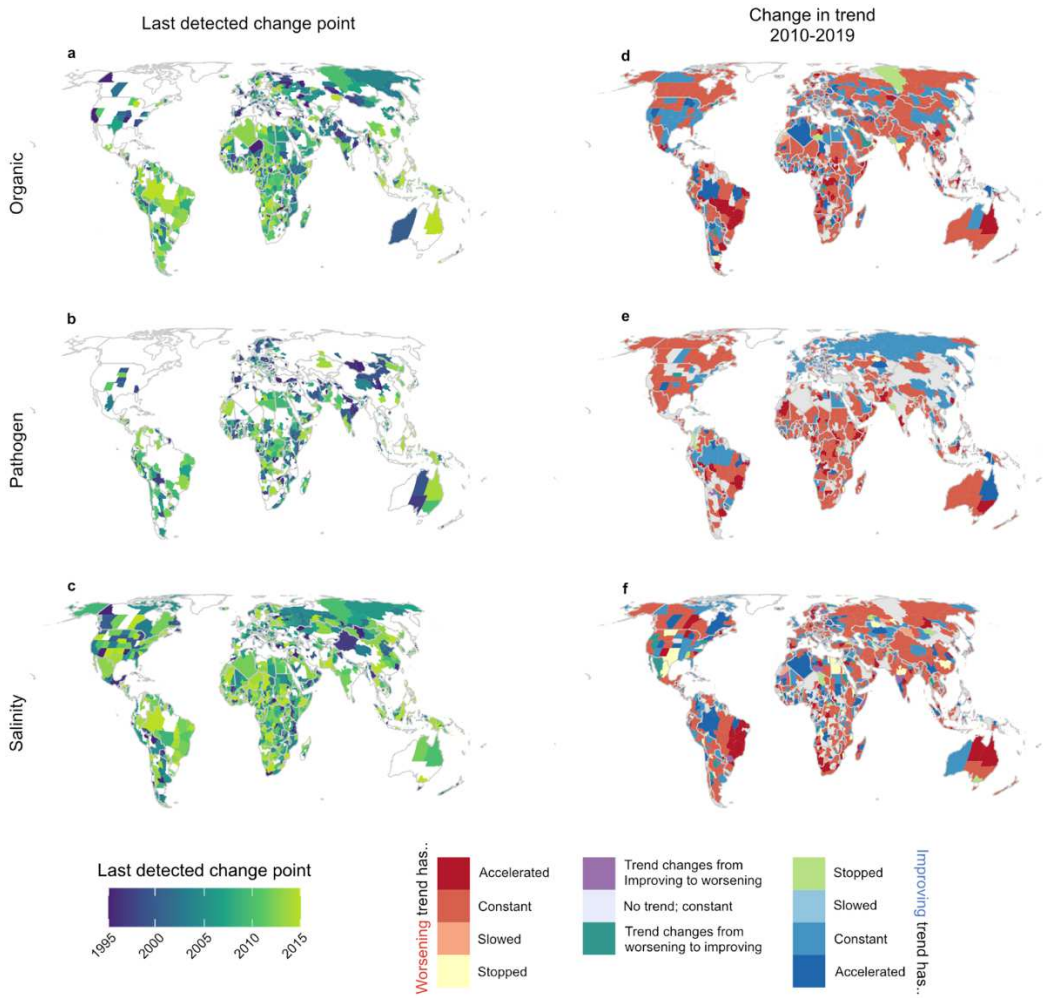


Figure 4. The year of the last detected changepoint and the classification of the recent changes in trend over the last ten years of the study period, 2010-2019, for organic contamination (a, d respectively), pathogen contamination (b, e) and salinity (c, f). Due to a minimum segment length of 5 years, the first and last possible change points are 1995 and 2015, respectively. See Supplementary Figure S7 for continental aggregations.

2.3 Population with unsafe water sources

We combined the admin level aggregated water quality data with our G-HSDW dataset and assessed the number of people experiencing the double burden of unsafe access to water with poor water quality. We also used these two datasets to map the areas under this double burden (see next section). We found that only 15% of people relying on surface water (129 million) experienced combined water quality in the 'low risk' category in 2019 (Figure 5c). When considering pathogen contamination levels alone, this was even lower at 10% (Figure 5d). These shares have remained rather stable over the past 30 years. For people with unimproved access to drinking water (378 million in 2019), the pattern was nearly identical: 14% and 11% experienced low risk for overall water quality and pathogen contamination, respectively (Figure 5a-b). Worryingly, more than half of both groups experienced overall water quality in the high, very high, or severe risk categories (52% for surface water; 60% for unimproved sources). For unimproved sources, this has declined from 71% in 1990, while for surface water it has remained roughly constant. When considering pathogen contamination levels alone, the share experiencing at least high risk was even higher: 56% for surface water and 63% for unimproved sources in 2019, down from 67% and 80% in 1990, respectively.

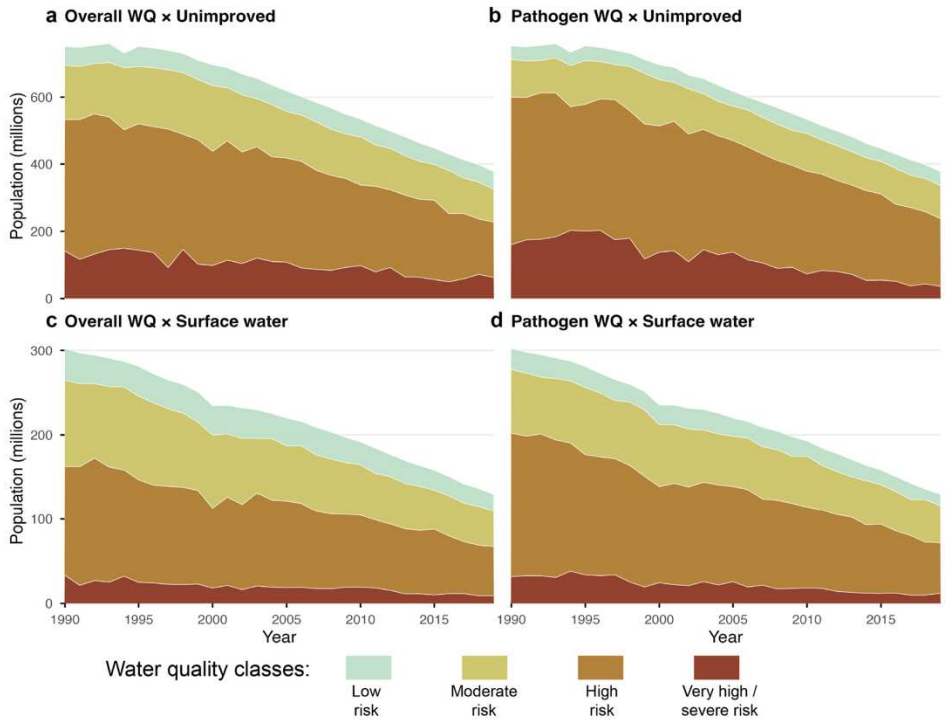


Figure 5. Population under each water quality class for unimproved and surface water access to drinking water. See water quality classes in Table 1.

2.4 Mapping dark spots and bright spots

To identify the most vulnerable areas, so called dark spot areas, we mapped (Figure 6a-b) the areas where share of population under unimproved or surface water access service levels was high and surface quality risk was high. South and East Asia, the Rift Valley and the east coast of Africa, and coastal parts of South America emerge as major regions where poor access to safe drinking water coincides with high water quality risk (level 3 or above; see Methods). In these regions, 117 million people are dependent on unimproved water sources prone to contamination, and 35 million people rely on surface water (Figure 6a-b). In China and India, the share of population dependent on surface water is relatively low (less than 2% of total population), but due to the large population of each country, the absolute number of people is substantial (96 million in 1990, 11 million in 2019). In Europe, Poland emerges as a dark spot due to the number of people relying on unimproved water sources. Sub-Saharan Africa and Insular Southeast Asia also include areas where, despite poor access to safe drinking water, the overall risk is small due to high quality of surface waters.

Finally, we conducted a bright spot analysis by cross-assessing trends of both water quality risk and drinking water service levels. Those areas where both, water quality risk and the share of population under unsafe drinking water access have decreased are considered bright spots (Figure 6c-d). These bright spots are dispersed across the globe, but only 12 million people (Figure 6c-d) with unimproved and 2.2 million with surface water sources enjoy their benefit. Worryingly, population in areas where both water quality and access to safe drinking water is deteriorating is significantly larger:33 million people for unimproved and 5.3 million people for surface water access. Areas where the water quality risk is increasing while access to safe drinking water is improving is the largest class by a large margin (196 million people for unimproved and 79 million for surface water service level), and cover large areas of South America, Sub-Saharan Africa, and Central and South India. Conversely, 0.7

million and 0.8 million people, respectively, live in admin areas with improving water quality and worsening safe drinking water access.

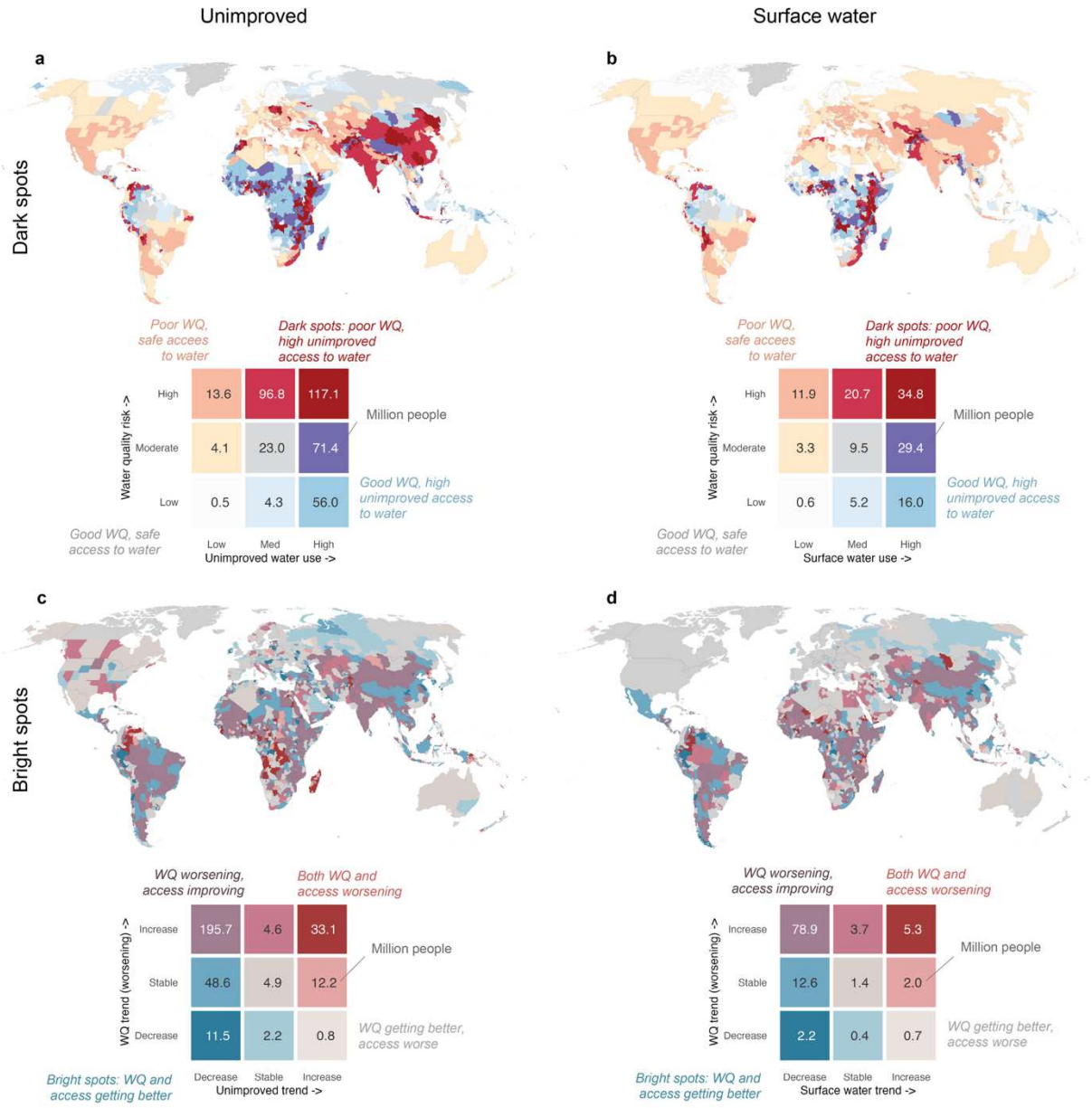


Figure 6. Bright spot and dark spot analysis using the current status and trends of combined water quality and unimproved and surface water service levels. Bivariate maps show an intersection between combined water quality and unimproved (a) and surface water (b) service levels, and trends of combined water quality and unimproved (c) and surface water (d) service levels. The bivariate legends show the population count in each category. See Supplementary Figures S8-S10 for the distribution of organic, pathogen, and salinity, respectively.

3. Discussion

In this research article we analysed, for the first time, how access to safe drinking water and water quality intersect globally using a new global subnational dataset on safe drinking water access and global water quality simulations over 1990-2019. Our G-HSDW dataset shows that the number of people under the unimproved and surface water-service levels have dropped from 1057 million to 507 million people (Supplementary Figure S1). However, while the service level has improved, water quality risk has increased in most of the admin areas.

Moreover, with G-HSDW, we created the first fully global subnational dataset covering all four service rungs of the JMP drinking water ladder. When comparing the G-HSDW dataset against two existing semi-global subnational drinking water service level products, the Local Burden of Disease (LBD) WaSH Collaborators (Deshpande *et al* 2020) geospatial estimates for 2000–2017 and the Greenwood *et al* (2024) random-forest predictions of safely managed drinking water services for 135 LMICs in 2020, we found that G-HSDW outperformed both. We validated all three products against reported national (admin 0) and subnational (admin 1) values from the JMP database that were not directly used to produce G-HSDW. For the indicators in common with LBD (surface water and unimproved), G-HSDW showed substantially lower error and stronger agreement with reported data at both spatial levels. At admin 0, G-HSDW predictions of surface water reliance achieved $R^2 = 0.995$ (RMSE = 0.64 percentage points, n = 1,890) compared with $R^2 = 0.79$ (RMSE = 3.96) for LBD, and for unimproved sources $R^2 = 0.997$ (RMSE = 0.64, n = 1,933) versus $R^2 = 0.76$ (RMSE = 6.05) for LBD. At admin 1 (n = 872 for both products), G-HSDW estimates remained closer to reported values (surface water $R^2 = 0.76$, RMSE = 5.26; unimproved $R^2 = 0.84$, RMSE = 7.01) than LBD ($R^2 = 0.52$ and 0.50; RMSE = 7.56 and 11.80, respectively). Against Greenwood *et al* (2024), for the closest comparable category, improved sources (equivalent to our combined basic + limited indicator), G-HSDW again showed closer agreement with reported data at both admin 0 ($R^2 = 0.99$ vs. 0.51; RMSE = 0.99 vs. 10.64 percentage points, n = 128) and admin 1 over 2017–2019 ($R^2 = 0.92$ vs. 0.41; RMSE = 7.19 vs. 18.78, n = 264). We note that part of the admin 0 performance advantage reflects our explicit anchoring of subnational estimates to JMP national totals, which neither alternative product imposes; the admin 1 comparison is therefore the more informative benchmark of within-country predictive skill. See Supplementary Figure S11 for G-HSDW validation plots against Deshpande *et al* (2020) and Greenwood *et al* (2024).

Our analysis is dependent on the only continuous global high-resolution water quality modelling currently available (DynQual, Jones *et al* 2023). However, while we were able to detect trends for nearly all admin areas, there are significant amount of missing data values in DynQual outputs as an uncertainty management effort. Cells with streamflow < 0.1 m^3/s are omitted since very small discharges affect concentration estimates considerably. For the monthly analysis, we fill the gaps by imputing the timeseries with monthly mean values. On average, imputation by monthly mean reduces the strength of trends, leading to a more conservative estimate. Furthermore, the spatial resolution of 5 arcminutes (~10km at the equator) and the accumulation of streamflow into larger rivers indicate that the data may not represent water quality of smaller streams within each grid cell, and water quality within each cell may significantly differ (for better or worse) from the larger streams the data represents. This may disproportionately affect the lower rung service levels, since WASH facilities likely rely on larger waterbodies.

We also note that DynQual does not include all of the core water quality indicators (see Peeters *et al* 2026), and thus we likely underestimate risk; water quality risks are widespread across all regions, but their nature varies by location (Desbureaux *et al* 2022). The water quality indicators used (organic contamination, pathogen contamination and salinity) cover three major and policy-relevant dimensions of surface-water quality risk for direct or untreated use. We are not aware of global continuous water quality datasets at comparable spatial and temporal resolution which cover other water quality dimensions (i.e. other microbiological, total organic, pharmaceutical, persistent organic pollutant, inorganic, and/or plastic contamination). Our results should therefore not be interpreted as a comprehensive water quality assessment. The implications are particularly

significant for drinking-water source across the world. In fact, a substantial fraction of the global population depends on groundwater, particularly across South Asia, Sub-Saharan Africa, the Middle East, and Southeast Asia. Salinity, arsenic, nitrate and fluoride contamination in groundwater are well-documented public health hazards (Nolte *et al* 2026, Rahman *et al* 2021, Podgorski and Berg 2022, Mukherjee *et al* 2024). Existing global groundwater datasets (e.g., Thorslund and van Vliet 2020) offering reasonable temporal coverage are geographically skewed and restricted to salinity, leaving other critical parameters and the most groundwater-dependent countries underrepresented. We call for a comprehensive global groundwater quality analysis with biogeochemical parameters, integrated with surface water assessments, to close this gap and more accurately characterize drinking water access risks in the world's most vulnerable regions. The increasing salinization trend in South Asia and China reflects the rapid expansion of groundwater-fed irrigation (Mehta *et al* 2024, Siebert *et al* 2010). Because groundwater carries higher salt concentrations than surface water, its extraction and field application introduces and accumulates salts in the soil. The interactions between surface and groundwater dynamics risk displacing populations from improved to unimproved water sources at a scale not captured here, which sometimes may lead to high levels of environmental migration (Nagabhatla *et al* 2020).

Water quality parameters are concentration-based quantities (mg or cfu per unit water volume). Under a constant pollutant load, concentrations are not static, but are modulated by the hydrological water balance of the catchment. In water-limited environments, atmospheric evaporative demand and precipitation variability (aridity index) exert a primary control on the volume of water available for dilution. Consequently, rising concentrations may reflect not an increase in pollutant mass flux, but rather a contraction of the dilution medium driven by aridification. This dual causality partitions observed concentration trends into two distinct mechanistic pathways: (i) a source-driven component, attributable to increasing pollutant loads (e.g., population growth, agricultural intensification, or expanding wastewater discharge), and (ii) a hydrology-driven component (wet vs. dry season, climate change), governed by shifts in the aridity regime of the administrative unit. Disentangling these pathways is essential for attributing observed water quality deterioration to its proximate causes, and for designing interventions that target the correct driver.

Our results show that, overall, WASH service levels and water quality are moving in opposite directions with a considerable reduction in people in the lowest rungs of the WASH service ladder and increase in risks associated with water quality. There is, however, considerable variation in the relationships of the trends, with dark spots of poor access to improved water sources and low water quality occurring in large parts of Asia, eastern Africa and coasts of Latin America. Bright spots with improving water quality and access to WASH infrastructure are dispersed across the globe but benefit only a small minority of people in the lowest WASH service levels. Our estimates of the intersection of global co-occurrence of water quality and WASH infrastructure help identify populations most in need of interventions and where the improving situation can serve as a model to accelerate positive developments and change the trajectory of trends for better elsewhere.

4 Data and Methods

4.1 Subnational drinking water service dataset

We produced annual global harmonised subnational drinking water service dataset (G-HSDW) at the first-level subnational administrative division (admin 1) for 240 countries and territories over the

period 1990–2023. Service levels follow the WHO/UNICEF Joint Monitoring Programme (JMP) classification: basic (improved source, ≤30 min round trip), limited (improved source, >30 min), unimproved (unprotected well/spring), and surface water (river, lake, pond). The four indicators are mutually exclusive and sum to 100%.

National-level data for 2000–2024 were obtained from the JMP 2025 global dataset ((WHO/UNICEF 2024), providing annual values for 220 countries. Historical national estimates (1990–1999) were obtained from the JMP for basic and surface water indicators (WHO/UNICEF JMP historical data received from JMP team through personal communication in 2024); limited and unimproved categories before 2000 were not systematically collected and were derived from the machine learning model described below. Historical and modern series were harmonised using a delta-change method: we computed the mean offset between the two datasets over the 2000–2002 overlap period and applied it additively to the 1990–1999 historical values, with subsequent normalisation to ensure the four indicators sum to 100%. For two larger countries (Kosovo and Nicaragua) and 18 small territories without national data, modelled data were used.

Subnational observations were extracted from JMP country workbooks, yielding 4,280 region × year observations from 102 countries (survey years 1997–2020, predominantly Demographic and Health Surveys (DHS) and UNICEF Multiple Indicator Cluster Surveys (MICS)). Three boundary sources were used: DHS Programme boundaries, MICS boundaries (UNICEF), and GADM 4.1 as universal fallback. For each country, the boundary source maximising spatial coverage, temporal completeness, and observation density was selected. For the machine learning model training, we used 99 countries with subnational data linked to 1,293 admin 1 polygons.

To fill missing values in observed national and subnational series, and to estimate within-country patterns where subnational observations were absent, we trained XGBoost regression models (Chen and Guestrin 2016) using ten population-weighted predictors at admin 1 resolution for 1990–2023: travel time to cities (Weiss *et al* 2018), poverty headcount ratio (World Bank 2022; harmonised and gap filled by authors), mean years of schooling (Smits and Permanyer 2019; filled and harmonised using methods from Kummu *et al* 2025), GNI per capita (Chrisendo *et al* 2025), Gini index (Chrisendo *et al* 2025), urban fraction (Niva *et al* 2023), aridity index (Trabucco and Zomer 2019), gender development index (Smits and Permanyer 2019; filled and harmonised using methods from Kummu *et al* 2025), corruption perception index (combination of VDem (Coppedge *et al* 2025) and Crombach and Smits 2024; filled and harmonised by authors), and groundwater depth (Fan *et al* 2013).

XGBoost models were trained independently per indicator using grouped cross-validation (80/20 train-test split by admin 1, stratified by country). Hyperparameters (tree depth, learning rate, regularisation) were tuned via Bayesian optimisation over 40 iterations, and final predictions were the median across a 50-model bootstrap ensemble. A post-hoc binned calibration adjustment (10 quantile bins) corrected systematic prediction bias. Test-set performances were: basic $R^2 = 0.81$, limited $R^2 = 0.55$, unimproved $R^2 = 0.61$, surface water $R^2 = 0.57$. The limited and surface water categories were inherently more difficult to predict due to their sparse occurrence (>30% zero values, median values <1%) and residual nature in the JMP ladder.

For admin 1 units with observations, missing years were linearly interpolated, with XGBoost-predicted trends anchored to the nearest observation for years before/after the survey period. For admin 1 units without observations (mainly in the 145 countries with no subnational survey data in the final dataset), XGBoost predictions were used directly. To ensure national consistency, we then

scaled admin 1 estimates so their population-weighted mean matched JMP national totals and normalised the four indicators to sum exactly to 100% within each admin 1 × year combination.

The final G-HSDW dataset includes 3,040 admin 1 areas for 240 countries. Of these, 1,085 admin areas have observed values for 95 countries, including 78% of the population without at least basic drinking water access in 2023 (57% in 1990). The G-HSDW estimates were rasterised to 5-arcminute (~10 km at the equator) global grids using the terra package (Hijmans *et al* 2026) in R (R Core Team 2024); all pixels within an admin 1 polygon receive identical indicator values. Long-term trends (1990–2023) were quantified using Theil-Sen median slopes (Sen 1968) per admin 1 unit. The number of people on each service level was computed by applying indicator percentages to GHS-POP 2023a (Schiavina *et al* 2023; 5-arcminute resolution, originally provided at 5-year intervals between 1985 and 2025, linearly interpolated to annual values).

We validated the final G-HSDW product against reported national values, and against the reported subnational data that were not directly used for the final dataset (those were used, however, to train the XGBoost model, but the admin areas differ from the ones in the final dataset as well as countries with reported admin 1 level data). We found that our national-level harmonisation worked well (Supplementary Figure S3), resulting in an R^2 of 0.99 for national-level validation across all three service levels (n = 5,157), with RMSE varying from 0.3 (limited) to 0.8 (unimproved). The R^2 for subnational validation varied from 0.69 (limited) to 0.87 (unimproved) (n = 1,027), while the range for RMSE was 5–6.7.

Several limitations apply. Predictions for the 145 countries without subnational survey data rely entirely on cross-country transferability of the XGBoost models, which was lower for limited and surface water categories. Historical estimates (1990–1999) for limited and unimproved categories are model-derived rather than observation-based and therefore have higher uncertainty than later years. Sub-admin 1 spatial variation is not captured by our 5-arcminute rasters.

4.2 Water quality data

We use water quality data modelled with DynQual (Jones *et al* 2023, 2022c). DynQual is a dynamic global process-based water quality model and integrates in global water models. DynQual's output water quality parameters are highly relevant for WASH: water temperature, biological oxygen demand (BOD; indicator of organic pollution), faecal coliform concentration (FC; indicator of faecal contamination) and total dissolved solids (TDS; indicator of salinity). It considers point and diffuse pollution loads from domestic, manufacturing, urban surface runoff, livestock, irrigation, and power production sectors and includes contaminant fate routines

The model runs we use are openly available (Jones *et al* 2022c) and are based on PCR-GlobWB2 global water model (Sutanudjaja *et al* 2018) using the Inter-Sectoral Impact Model Intercomparison Project protocols (ISIMIP, see https://www.isimip.org/protocol/3/) and include human socio-economic forcing and climate change. We specifically choose to use DynQual due to its relevant water quality parameters and a spatial resolution of 5 arcminutes which allows us to consider the correspondence between population, WASH infrastructure and water quality at a fine granularity. The temporal resolution of the available data is 1 month. We aggregate the data to the admin 1-level areas using a population weighted approach to assess water quality where people are living. We used population data from the Global Human Settlement Layer data GHS-POP R2023a (Schiavina *et al* 2023).

4.3 Water quality classification

We aggregated water quality into five classes based on the risk of consumption, from low to severe risk (*Table 1*). We adopted thresholds from World Health Organization's (WHO) TDS background document for the development of drinking water quality guidelines (World Health Organization 2003, World Health Organization 2022). WHO does not set a health-based guideline value as TDS is an aesthetic parameter. Organic pollution is expressed as BOD, for which we adopted a classification scheme which is a combination of European and United Nations schemes (Vigiak *et al* 2019). The source of BOD contamination in surface waters generally arise from sewage, livestock, or industrial sources. Elevated BOD may coincide with faecal contamination and can therefore act as a corroborating indicator of potential pathogen risk alongside FC (Vigiak *et al* 2019). FC is a faecal contamination indicator, for which we adopted thresholds based on WHO drinking water microbiological standards (World Health Organization 2017) and widely used recreational and source-water guidelines (U.S. EPA 2012, Jones *et al* 2022a). WHO standards require no detectable FC in drinking water, making Class 1 the only category consistent with safe untreated consumption. Any detection of FC suggests faecal contamination and potential pathogen exposure, while the resulting classification reflects relative severity rather than absolute safety thresholds.

Table 1. Water quality category thresholds used for 5-category classification of water quality risk levels. The justifications for the thresholds are presented in the Methods text.

Class, Risk level	TDS (mg/L)	BOD (mg/L)	FC (cfu/100 mL)	Interpretation for untreated consumption
1: Low risk	< 300	< 2	< 1	Acceptable; meets WHO microbiological standard for drinking water
2: Moderate risk	300–600	2–4	1–10	Detectable faecal contamination; palatability concerns begin
3: High risk	600–900	4–8	10–100	Clear pathogen exposure; organic pollution; objectionable taste
4: Very high risk	900–1200	8–15	100–1000	Serious disease risk; salty taste; heavy organic pollution
5: Severe risk	> 1200	> 15	> 1000	Severely contaminated; major disease/health risk

We estimate water quality between class thresholds using linear interpolation. This results in a continuous classification which allows interpreting how close or far the actual concentration is from a class threshold value. In addition to classifying each pollutant, we estimate an combined water quality by taking a geometric mean of the organic, pathogen and salinity classes. Geometric mean prevents compensability in the mean and places larger importance to poorer water quality classes. In computing the geometric mean, we temporarily inverted the classification numbers so that the more severe risk classes are emphasized by the computation, before converting the classes back to the ones shown in the table.

4.4 Associations between indicators

We estimated the associations between the water quality risks and between the pollutant timeseries using the Pearson correlation coefficient ρ (Eq. 1),

$$\rho = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum(x_i - \bar{x})^2 \sum(y_i - \bar{y})^2}}, \quad (1)$$

where x and y denote the two variables whose associations are estimated. Because ρ measures a linear *monotonic* relationship of two variables, and our risk classification [1, …, 5] is not monotonically increasing at the boundaries, we also compute the ξ_n correlation coefficient (Eq. 2; Chatterjee 2021),

$$\xi_n(X, Y) \coloneqq 1 - \frac{n \sum_{i=1}^{n-1} |r_{i+1} - r_i|}{2 \sum_{i=1}^{n} l_i (n - l_i)} \quad (2)$$

where r are the ranks of variable Y, and the denominator is a normalising factor. After sorting the pairs of X and Y using X, the function measures how much the rank of Y change from a neighbouring value X to another. If X and Y are independent and not associated with one another, the ranks r_i of Y are in random order. However, if the two variables are associated, the ranks of Y increase smoothly. ξ_n is measured between 0 (variables are independent) and 1 (X is a function of Y). It is asymmetric ($\xi_n(X, Y) \neq \xi_n(Y, X)$) and does not assume an underlying distribution or monotonicity as in Pearson correlation coefficient.

4.5 Trend estimation and identifying changes in trend

We used Sen's slope (Sen 1968) to estimate trends in pollutant load concentrations. Sen's slope estimates a linear slope that is robust against extreme values in a timeseries. It is estimated as the median of the slope between pairs of data points (whereas a standard linear slope uses the mean instead of the median).

We performed a change point analysis on the monthly pollutant concentration timeseries to find points in the data where either the mean or variance of the timeseries significantly change. For this, we use the EnvCpt R package and the Pruned Exact Linear Time algorithm (Killick *et al* 2025, Beaulieu and Killick 2018, Killick *et al* 2012). We set the minimum segment length as five years to avoid frantically changing trends. For each identified segment, we fit the trend using Sen's slope.

We then compare the slope values of the last two identified segments and classify them as having positive, negative, and no trend (annual trend within 2% of the standard deviation of the monthly timeseries). Positive and negative trends are further classified into accelerating (the latter trend has > 5% higher slope value than the earlier trend), constant (slopes are within 5% of one another), or slowing (latter trend is at least 5% smaller than earlier) trends. We also detect if the slope has changed sign from worsening to improving or vice versa.

Due to the limit in DynQual model outputs where cells with smaller than 0.1 m^3/s are assigned missing values, the timeseries is in certain cells and admin areas missing values. The change point analysis requires continuous timeseries to operate, and therefore we impute the missing values with monthly mean values from the available records. If all monthly values are missing and the imputation by monthly mean is not possible, we omit the cell and admin area from the analysis.

Acknowledgements

MKa and MKu were supported by *Maa- ja Vesitekniikan Tuki ry*. MKu was additionally supported by Research Council of Finland's Flagship Programme under project Digital Waters (grant no. 359248), Strategic Research Council (SRC) through project Water&Food (grant no. 365512), and Sven Hallin Research Foundation sr. SK was supported by the Drainage Foundation sr (*Salaojituksen Tukisäätiö*) under AdaptCrop project. VB was supported by Research Council of Finland's Profi6 T3 Co-innovating circular system funding.

Code availability: The code for the G-HSDW dataset creation and related analyses can be accessed through https://github.com/mattikummu/subnat_WASH and water quality trend and seasonal analyses through https://github.com/mkkallio/WASH_WQ.

Data availability: The output data will be made available at https://doi.org/10.5281/zenodo.20439778.

Declaration of Generative AI and AI-assisted technologies in the writing process

During the preparation of this work, author's utilized OpenAI's GPT and Anthropic's Claude tools to suggest grammar and changes in sentences where the text was unclear or could be improved. After using these tools, authors reviewed and edited the content to enhance its clarity and takes full responsibility for the publication's content. For coding, GitHub Copilot was used to support to build the scripts. After using it, authors reviewed and edited the code and take full responsibility for it.

Competing Interests

The authors have no conflicts of interest to disclose.

References

Ambel A A, Bain R, Degefu T B, Donmez A, Johnston R and Slaymaker T 2023 Addressing gaps in data on drinking water quality through data integration and machine learning: evidence from Ethiopia *npj Clean Water* **6** 63

Bain R, Cronk R, Wright J, Yang H, Slaymaker T and Bartram J 2014 Fecal Contamination of Drinking-Water in Low- and Middle-Income Countries: A Systematic Review and Meta-Analysis *PLOS Medicine* **11** e1001644

Bain R, Johnston R, Khan S, Hancioglu A and Slaymaker T 2021 Monitoring Drinking Water Quality in Nationally Representative Household Surveys in Low- and Middle-Income Countries: Cross-Sectional Analysis of 27 Multiple Indicator Cluster Surveys 2014–2020 *Environ Health Perspect* **129** 097010

Beaulieu C and Killick R 2018 Distinguishing Trends and Shifts from Memory in Climate Data *Journal of Climate* **31** 9519–43

Chambers T, Hales S, Pourzand F, Dean F, Kim A H M, Puente-Sierra M, Marek L, Hobbs M and Baker M G 2026 Drinking water quality and enteric disease: a nationwide case-crossover study (2015–2019) in New Zealand *J Expo Sci Environ Epidemiol* 1–9

Chatterjee S 2021 A New Coefficient of Correlation *Journal of the American Statistical Association* **116** 2009–22

Chen T and Guestrin C 2016 XGBoost: A Scalable Tree Boosting System *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* KDD '16 (New York, NY, USA: Association for Computing Machinery) pp 785–94 Online: https://dl.acm.org/doi/10.1145/2939672.2939785

Chrisendo D, Niva V, Hoffmann R, Masoumzadeh Sayyar S, Rocha J, Sandström V, Solt F and Kummu M 2025 Rising income inequality across half of global population and socioecological implications *Nat Sustain* **8** 1601–13

Coppedge M, Gerring J, Knutsen C H, Lindberg I, Teorell J, and et al 2025 V-Dem Country–Year Dataset v15 Online: https://v-dem.net/data/the-v-dem-dataset/

Crombach L and Smits J 2024 The Subnational Corruption Database: Grand and petty corruption in 1,473 regions of 178 countries, 1995–2022 *Sci Data* **11** 686

Desbureaux S, Mortier F, Zaveri E, van Vliet M T H, Russ J, Rodella A S and Damania R 2022 Mapping global hotspots and trends of water quality (1992–2010): a data driven approach *Environ. Res. Lett.* **17** 114048

Deshpande A, Miller-Petrie M K, Lindstedt P A, Baumann M M, Johnson K B, Blacker B F, Abbastabar H, Abd-Allah F, Abdelalim A, Abdollahpour I, Abegaz K H, Abejie A N, Abreu L G, Abrigo M R M, Abualhasan A, Accrombessi M M K, Adamu A A, Adebayo O M, Adedeji I A, Adedoyin R A, Adekanmbi V, Adetokunboh O O, Adhikari T B, Afarideh M, Agudelo-Botero M, Ahmadi M, Ahmadi K, Ahmed M B, Ahmed A E, Akalu T Y, Akanda A S, Alahdab F, Al-Aly Z, Alam S, Alam N, Alamene G M, Alanzi T M, Albright J, Albujeer A, Alcalde-Rabanal J E, Alebel A, Alemu Z A, Ali M, Alijanzadeh M, Alipour V, Aljunid S M, Almasi A, Almasi-Hashiani A, Al-Mekhlafi H M, Altirkawi K A, Alvis-Guzman N, Alvis-Zakzuk N J, Amini S, Amit A M L, Amul G G H, Andrei C L, Anjomshoa M, Ansariadi A, Antonio C A T, Antony B, Antriyandarti E, Arabloo J, Aref H M A,

Aremu O, Armoon B, Arora A, Aryal K K, Arzani A, Asadi-Aliabadi M, Asmelash D, Atalay H T, Athari S M, Athari S S, Atre S R, Ausloos M, Awasthi S, Awoke N, Quintanilla B P A, Ayano G, Ayanore M A, Aynalem Y A, Azari S, Azman A S, Babaee E, Badawi A, Bagherzadeh M, Bakkannavar S M, Balakrishnan S, Banach M, Banoub J A M, Barac A, Barboza M A, Bärnighausen T W, Basu S, Bay V D, Bayati M, Bedi N, Beheshti M, et al 2020 Mapping geographical inequalities in access to drinking water and sanitation facilities in low-income and middle-income countries, 2000–17 *The Lancet Global Health* **8** e1162–85

Fan Y, Li H and Miguez-Macho G 2013 Global Patterns of Groundwater Table Depth *Science* **339** 940–3

Greenwood E E, Lauber T, van den Hoogen J, Donmez A, Bain R E S, Johnston R, Crowther T W and Julian T R 2024 Mapping safe drinking water use in low- and middle-income countries *Science* **385** 784–90

Hijmans R J, Brown A, Barbosa M, Dyba K, Bivand R, Chirico M, Cordano E, Pebesma E, Rowlingson B and Sumner M D 2026 terra: Spatial Data Analysis Online: https://cran.r-project.org/web/packages/terra/index.html

Hofstra N, Flörke M, Jones E, Ouedraogo I, Podgorski J, Politi E, Schweden K, Simis S, Stelzer K, Strokal M, van Vliet M T H and Peeters S 2026 Water quality impacts on human health: towards an integrated, solution-oriented global assessment *Environ. Res.: Water* **2** 015009

Irianti S, Sasto I H S, Mbarep D P P, Dharmayanti I, Yunianto A, Zahra Z, Puspita T, Hidayangsih P S, Rachmat B, Anwar A and Azhar K 2024 Investigating improved drinking water quality at the point of access: Evidence from four regions of Indonesia *Journal of Water, Sanitation and Hygiene for Development* **14** 80–90

Jones E R, Bierkens M F P and van Vliet M T H 2024 Current and future global water scarcity intensifies when accounting for surface water quality *Nat. Clim. Chang.* **14** 629–35

Jones E R, Bierkens M F P, Wanders N, Sutanudjaja E H, van Beek L P H and van Vliet M T H 2022a Current wastewater treatment targets are insufficient to protect surface water quality *Commun Earth Environ* **3** 221

Jones E R, Bierkens M F P, Wanders N, Sutanudjaja E H, van Beek L P H and van Vliet M T H 2023 DynQual v1.0: a high-resolution global surface water quality model *Geoscientific Model Development* **16** 4481–500

Jones E R, Bierkens M F P, Wanders N, Sutanudjaja E H, van Beek L P H and van Vliet M T H 2022b Global hydrology and water quality data from 1980-2019, derived from the dynamical surface water quality model (DynQual) at 5 arcmin spatial resolution Online: https://zenodo.org/records/18545092

Jones E R, Bierkens M F P, Wanders N, Sutanudjaja E H, van Beek L P H and van Vliet M T H 2022c Global hydrology and water quality data from 1980-2019, derived from the dynamical surface water quality model (DynQual) at 5 arcmin spatial resolution Online: https://zenodo.org/records/18545092

Killick R, Beaulieu C, Taylor S and Hullait H 2025 EnvCpt: Detection of Structural Changes in Climate and Environment Time Series Online: https://cran.r-project.org/web/packages/EnvCpt/index.html

Killick R, Fearnhead P and Eckley I A 2012 Optimal Detection of Changepoints With a Linear Computational Cost *Journal of the American Statistical Association* **107** 1590–8

Kummu M, Kosonen M and Masoumzadeh Sayyar S 2025 Downscaled gridded global dataset for gross domestic product (GDP) per capita PPP over 1990–2022 *Sci Data* **12** 178

Mehta P, Siebert S, Kummu M, Deng Q, Ali T, Marston L, Xie W and Davis K F 2024 Half of twenty-first century global irrigation expansion has been in water-stressed regions *Nat Water* **2** 254–61

Mukherjee A, Coomar P, Sarkar S, Johannesson K H, Fryar A E, Schreiber M E, Ahmed K M, Alam M A, Bhattacharya P, Bundschuh J, Burgess W, Chakraborty M, Coyte R, Farooqi A, Guo H, Ijumulana J, Jeelani G, Mondal D, Nordstrom D K, Podgorski J, Polya D A, Scanlon B R, Shamsudduha M, Tapia J and Vengosh A 2024 Arsenic and other geogenic contaminants in global groundwater *Nat Rev Earth Environ* **5** 312–28

Nagabhatla N, Panthea Pouramin, Rupal Brahmbhatt, Fioret C, Glickman T, K Bruce Newbold and Smakhtin V 2020 Water and Migration: A Global Overview Online: https://www.researchgate.net/doi/10.13140/RG.2.2.20741.37606

Niva V, Horton A, Virkki V, Heino M, Kosonen M, Kallio M, Kinnunen P, Abel G J, Muttarak R, Taka M, Varis O and Kummu M 2023 World's human migration patterns in 2000–2019 unveiled by high-resolution data *Nat Hum Behav* **7** 2023–37

Nolte A, Bender S, Hartmann J, Baltruschat S, Moosdorf N and Reinecke R 2026 Coastal groundwater-level trends reveal global susceptibility to seawater intrusion *Nat Water* **4** 673–82

Peeters S H, Teuling F S R, Kumar R, Nkwasa A, Gosling S N, Suresh K, Micella I and Hofstra N 2026 On the value of indicators for large-scale water quality assessments *Environ. Res.: Water* **1** 042002

Podgorski J and Berg M 2022 Global analysis and prediction of fluoride in groundwater *Nat Commun* **13** 4232

R Core Team 2024 *R: A Language and Environment for Statistical Computing* (Vienna, Austria: R Foundation for Statistical Computing) Online: https://www.R-project.org/

Rahman A, Mondal N C and Tiwari K K 2021 Anthropogenic nitrate in groundwater and its health risks in the view of background concentration in a semi arid area of Rajasthan, India *Sci Rep* **11** 9279

Santos T M, Wendt A, Coll C V N, Bohren M A and Barros A J D 2023 E. coli contamination of drinking water sources in rural and urban settings: an analysis of 38 nationally representative household surveys (2014–2021) *J Water Health* **21** 1834–46

Schiavina M, Freire S and MacManus K 2023 GHS-POP R2023A - GHS population grid multitemporal (1975-2030) Online: http://data.europa.eu/89h/2ff68a52-5b5b-4a22-8f40-c41da8332cfe

Sen P K 1968 Estimates of the Regression Coefficient Based on Kendall's Tau *Journal of the American Statistical Association* **63** 1379–89

Siebert S, Burke J, Faures J M, Frenken K, Hoogeveen J, Döll P and Portmann F T 2010 Groundwater use for irrigation – a global inventory *Hydrology and Earth System Sciences* **14** 1863–80

Smits J and Permanyer I 2019 The Subnational Human Development Database *Sci Data* **6** 190038

Sutanudjaja E H, van Beek R, Wanders N, Wada Y, Bosmans J H C, Drost N, van der Ent R J, de Graaf I E M, Hoch J M, de Jong K, Karssenberg D, López López P, Peßenteiner S, Schmitz O, Straatsma M W, Vannametee E, Wisser D and Bierkens M F P 2018 PCR-GLOBWB 2: a 5 arcmin global hydrological and water resources model *Geoscientific Model Development* **11** 2429–53

Thorslund J and van Vliet M T H 2020 A global dataset of surface water and groundwater salinity measurements from 1980–2019 *Sci Data* **7** 231

Trabucco A and Zomer R 2019 Global Aridity Index and Potential Evapotranspiration (ET0) Climate Database v2 1705236666 Bytes Online: https://figshare.com/articles/dataset/Global_Aridity_Index_and_Potential_Evapotranspiration_ET0_Climate_Database_v2/7504448/3

U.S. EPA 2012 *Recreational Water Quality Criteria* (Washington, D.C., USA: U.S. Environmental Protection Agency) Online: https://www.epa.gov/sites/default/files/2015-10/documents/rwqc2012.pdf

Vigiak O, Grizzetti B, Udias-Moinelo A, Zanni M, Dorati C, Bouraoui F and Pistocchi A 2019 Predicting biochemical oxygen demand in European freshwater bodies *Science of The Total Environment* **666** 1089–105

van Vliet M T H, Jones E R, Flörke M, Franssen W H P, Hanasaki N, Wada Y and Yearsley J R 2021 Global water scarcity including surface water quality and expansions of clean water technologies *Environ. Res. Lett.* **16** 024020

van Vliet M T H, Thorslund J, Strokal M, Hofstra N, Flörke M, Ehalt Macedo H, Nkwasa A, Tang T, Kaushal S S, Kumar R, van Griensven A, Bouwman L and Mosley L M 2023 Global river water quality under climate change and hydroclimatic extremes *Nat Rev Earth Environ* **4** 687–702

Weiss D J, Nelson A, Gibson H S, Temperley W, Peedell S, Lieber A, Hancher M, Poyart E, Belchior S, Fullman N, Mappin B, Dalrymple U, Rozier J, Lucas T C D, Howes R E, Tusting L S, Kang S Y, Cameron E, Bisanzio D, Battle K E, Bhatt S and Gething P W 2018 A global map of travel time to cities to assess inequalities in accessibility in 2015 *Nature* **553** 333–6

WHO/UNICEF 2024 Downloads | JMP *The custodian of global data on Water Supply, Sanitation and Hygiene (WASH)* Online: https://washdata.org/data/downloads#WLD

WHO/UNICEF 2017 *Progress on drinking water, sanitation and hygiene: 2017 update and SDG baselines* (Geneva: World Health Organization and the United Nations Children's Fund)

WHO/UNICEF 2023 *Progress on household drinking water, sanitation and hygiene 2000-2022: Special focus on gender* Online: https://data.unicef.org/resources/jmp-report-2023/

Wolf J, Johnston R B, Ambelu A, Arnold B F, Bain R, Brauer M, Brown J, Caruso B A, Clasen T, Colford J M, Mills J E, Evans B, Freeman M C, Gordon B, Kang G, Lanata C F, Medlicott K O, Prüss-Ustün A, Troeger C, Boisson S and Cumming O 2023 Burden of disease attributable to unsafe

drinking water, sanitation, and hygiene in domestic settings: a global analysis for selected adverse health outcomes *The Lancet* **401** 2060–71

World Bank 2022 Global Subnational Poverty Atlas (GSAP) Online: https://datacatalog.worldbank.org/search/dataset/0042041/global-subnational-poverty-atlas-gsap

World Health Organisation 2003 *Total dissolved solids in Drinking-water. Background document for development of WHO Guidelines for Drinking-water Quality.* (Geneva: World Health Organisation) Online: https://www.who.int/docs/default-source/wash-documents/wash-chemicals/total-dissolved-solids-background-document.pdf

World Health Organization 2022 *Guidelines for drinking-water quality* (Geneva: World Health Organization)

World Health Organization 2017 *Guidelines for drinking-water quality* (Geneva: World Health Organization)

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- kallioetalWASHwaterqualitysupplementNatWater.pdf