

People's Water Data: Design and Validation of a Scalable Decentralized Framework for Monitoring Drinking Water Safety

Suzan Kagan, Sankar Sudhir, Ben Hamilton, Alona Maslennikov, P. Amaravathy, Timor Lichtman, Sonali Seth, Tanmayaa Nayak, Ramya Dwivedi, Ganesan Velmurugan, Hadas Mamane, Ram Fishman, and Thalappil Pradeep*



Cite This: <https://doi.org/10.1021/acssusresmgmt.6c00149>



Read Online

ACCESS |



Metrics & More



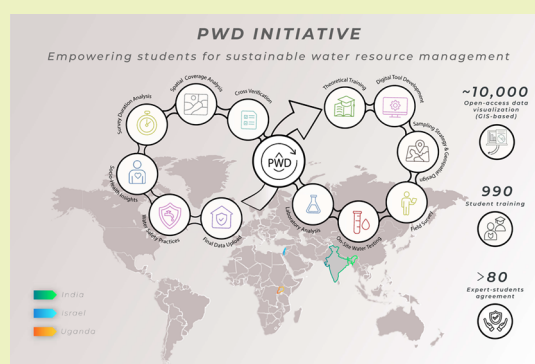
Article Recommendations



Supporting Information

ABSTRACT: Water, the mother of all resources, must be managed sustainably for global prosperity. Equitable access to clean drinking water depends on effective water-quality monitoring. Centralized drinking-water monitoring systems are essential for public health protection, yet in many regions, they lack spatial resolution and household-level coverage, allowing localized contamination risks to remain undetected between routine assessments. In this study, we present and validate a decentralized, quality-controlled drinking-water surveillance architecture at scale. In a pilot implementation, 990 university-trained participants conducted standardized assessments of over 9000 household drinking-water sources across India, Israel, and Uganda. Validation analyses demonstrated strong agreement between decentralized field measurements and expert- and laboratory-supported verification, supporting reliable data generation under structured protocols. The resulting spatial dataset identified localized areas with elevated water-quality risk indicators that may be missed by lower-resolution monitoring approaches. By expanding household-level coverage and incorporating structured quality-control and validation procedures, the framework provides early warning and supports targeted professional follow-up, while complementing, rather than replacing, certified laboratory and regulatory monitoring systems. Overall, this study establishes a transferable decentralized surveillance framework that complements existing monitoring infrastructure and strengthens sustainable water resource management in regions with appropriate training capacity, institutional support, ethical data governance, and validation infrastructure.

KEYWORDS: decentralized drinking-water monitoring, household-level water quality, monitoring density, spatial risk mapping, sustainable water resource management



INTRODUCTION

Access to microbiologically safe drinking water remains a global public health priority and an essential component of sustainable water resource management. According to World Health Organization (WHO) estimates, in 2022, 1.7 billion people relied on drinking-water sources contaminated with fecal matter.¹ WHO guidelines define microbiological safety criteria that form the basis of regulatory monitoring systems and certified laboratory analyses.² Unsafe drinking water continues to contribute substantially to preventable diseases and mortality, particularly among children under five years of age.³

Regulatory water monitoring systems primarily rely on periodic, source-based sampling and may not fully capture water-quality conditions at the point of use. In settings particularly where distribution systems are decentralized, intermittent, or aging, household-level water quality can differ from source-level measurements. This limitation has been widely documented in household drinking-water studies, including a

systematic review showing that microbiological water quality can deteriorate between the source and the point

of use during collection, transport, and household storage.⁴ Earlier studies similarly emphasized that protection of drinking water should extend beyond the source to collection, household storage, and point-of-use conditions.⁵ Large-scale household-survey evidence from low- and middle-income countries further shows that water classified as improved or basic at the source may still be contaminated at the point of use.⁶ Meta-analytic evidence further demonstrates the public-health relevance of household-level water safety, with point-of-use filtration

Received: March 10, 2026

Revised: September 6, 2026

Accepted: September 8, 2026

combined with safe storage associated with substantially reduced childhood diarrheal risk.⁷ The absence of systematic household-level monitoring creates blind spots in water-quality assurance.

Large-scale water management is constrained by the lack of field-validated, openly accessible water-quality data at high spatial resolution. Recent point-of-use evidence shows that microbial contamination is associated with household-level socio-economic conditions and water-handling practices, including storage container design, container covering, household water treatment, and hand hygiene.⁸ A population-based study in rural areas across multiple African countries similarly identifies household and socio-economic predictors of drinking-water *Escherichia coli* (*E. coli*) contamination.⁹ These findings indicate that utility-level monitoring alone may not fully capture household-level and behavioral factors influencing drinking-water safety. These limitations are especially important in rural, peri-urban, and socio-economically peripheral regions, where water sources may be heterogeneous and household-level risks may vary over short distances. Such spatial and socio-economic variability underscores the need for monitoring approaches that incorporate point-of-use conditions and household practices, alongside source-level assessments.^{8,9} This monitoring gap is particularly relevant to sustainable resource management because localized contamination at the point of use may remain undetected between routine assessments, limiting the ability of authorities, utilities, and communities to prioritize follow-up actions. Therefore, high-spatial-resolution data collected at the household level are needed not to replace certified monitoring systems, but to complement them by identifying localized risks and supporting targeted water-quality management.

Multiple studies have examined the relationship between drinking-water quality and household-level perceptions, behaviors, treatment practices, and socio-economic conditions. These studies show that household water safety, microbial contamination levels, and waterborne disease risk depend not only on source quality, but also on storage, treatment, hygiene behavior, socio-economic conditions, and point-of-use handling practices.^{4,6,8} Longitudinal evidence indicates that sustained household water-treatment practices are influenced not only by technical factors, but also by household behavior, perceived ability to maintain the practice, and the surrounding social environment.¹⁰ Earlier integrated approaches combining household surveys with point-of-use water-quality measurements were developed to capture both subjective factors (e.g., user behavior, practices, and perceptions) and objective indicators (e.g., quantitative water-quality parameters) at the community level.¹¹ However, existing studies have addressed different components of this problem separately or at limited scales, rather than integrating standardized household-level data collection, microbiological testing, quality control, and spatial analysis within a scalable monitoring framework. In particular, the relationships among household behavior, treatment-system maintenance, and independently measured microbiological outcomes remain insufficiently characterized at scale.

Studies across diverse geographic settings have illustrated both the potential and the practical limitations of community-based water-quality monitoring.^{12–14} More broadly, citizen-science approaches in hydrology and water resources have been recognized as useful for expanding environmental data generation and supporting resource-management decisions, provided

they are supported by appropriate project design, participant training, and quality-control procedures, and are integrated with expert or institutional knowledge systems.¹⁵ In Bekasi, Indonesia, 96 households conducted biweekly *E. coli* tests over six months, revealing contamination rates of 11–70% at the source and 15–44% at the point of use. While the framework improved awareness, nearly half of the households dropped out, and behavior change was inconsistent, underscoring the need for sustained support in treatment and storage to ensure long-term impact.¹² In California, the Drinking Water Tool (DWT) was developed to improve access to water-quality data among rural well users and promote environmental justice. Community engagement, particularly through organizations like the Community Water Center, helped build trust, though reliance on generalized data and limited institutional acceptance constrained its operational utility.¹³ In India, a randomized controlled trial investigated the impact of providing households with direct access to information on drinking-water quality.¹⁴ These studies suggest that community-level access to water-quality data can drive positive behavioral change and promote safer household water use through informed decision-making.¹⁴ Similarly, in rural Mexico, a community-based monitoring framework engaged residents in water-quality testing using low-cost methods for measuring pH, turbidity, and dissolved oxygen.¹⁶ The program generated locally relevant knowledge, fostered community ownership, and strengthened collaboration with local authorities. However, challenges remained regarding data reliability, the adequacy of volunteer training, and the development of institutional mechanisms to integrate community-generated data into formal water management systems.¹¹ These challenges are consistent with the broader citizen-science literature emphasizing that decentralized data generation requires quality assurance, standardization, training, and institutional pathways for data use.¹⁵

Water-quality monitoring can be organized across a continuum from centralized laboratory analysis to decentralized onsite testing, with the appropriate configuration depending on geographic, operational, economic, and institutional conditions.¹⁷ Recent studies emphasize the importance of accessible local laboratory capacity and sustained testing demand, while showing that water-quality risks and incomplete uptake of routine testing may persist even where testing infrastructure is available.^{18–21} Together, these studies support a complementary model in which decentralized field assessment expands monitoring reach and screening capacity, while laboratory-based analysis provides verification and quality assurance.

Although decentralized and community-based monitoring approaches have been explored previously, few studies have combined household-level behavioral information, field-based physicochemical and microbiological measurements, geospatial allocation, digital reporting, and independent validation within a single structured implementation framework. These aspects are considered important and constitute the specific research gap addressed in the present study.

The People's Water Data (PWD) framework addresses this structural monitoring gap by establishing a decentralized, high-density, field-validated household-level water-quality data architecture generated through standardized, university-led field implementation. As emphasized in a recent perspective in *ACS Sustainable Resource Management*, water scarcity at the planetary scale is a resource management challenge requiring credible, open-access water data supported by scientific institutions.²²

PWD operationalizes data generation under unified methodological protocols, integrating water-quality measurements, behavioral assessment, geospatial allocation, digital data collection, expert cross-verification, and open data visualization within a single operational workflow. PWD therefore functions as a quality-controlled, governance-linked monitoring model directed by academic institutions. This integrated data and training structure strengthens institutional capacity for decentralized resource management, positioning universities as knowledge intermediaries that enable transparent, scalable water security systems.

The objective of this study was therefore to design and validate PWD as a supplementary decentralized surveillance layer for drinking-water safety. Specifically, the study evaluated whether trained participants, operating under standardized university-led protocols, could generate reliable household-level water-quality and survey data, whether geospatial allocation could improve the spatial distribution of sampling, and whether field-generated data could be verified through expert assessment and laboratory analysis.

The framework was implemented across India, Israel, and Uganda to evaluate spatial variability in microbial and selected chemical indicators at the point of use. The methodology presented here includes three components: a cellular phone-based household survey, on-site water sampling and measurement of key water-quality indicators with laboratory follow-up, and integration of collected data into a unified analytical platform for spatial assessment. Together, these components support a decentralized surveillance framework that strengthens household-level quality assurance while maintaining methodological rigor. The framework is presented as transferable to other contexts where training capacity, institutional support, ethical data governance, and validation infrastructure are available, rather than as a universally applicable replacement for existing monitoring systems.

MATERIALS AND METHODS

Figure 1 illustrates the structured, multi-phased PWD workflow designed to support scientific rigor, operational consistency, and

community engagement. The process begins with theoretical training and digital tool development, followed by a sampling strategy and geospatial design. Field implementation is conducted according to a standard operating procedure (SOP) that includes installing the EpiCollect5 survey app (<https://www.epicollect.org>), an existing mobile data-collection platform developed by the CGPS Team of Oxford BDI that supports real-time and offline data entry, assembling the field kit, and implementing spatially guided, geographically distributed sampling strategies. During the field survey phase, participants conduct household surveys, perform on-site water testing, and collect sterile samples for laboratory analysis. The laboratory analysis stage includes equipment calibration, measurements of water-quality parameters, biological testing of collected water samples, and uploading the survey and measurement data to the platform. Supporting quality-control components, including cross-verification, spatial coverage assessment, and survey-duration analysis, are embedded throughout the workflow to ensure data reliability and reproducibility. Socio-health insights and water safety practices are derived from the integrated dataset generated through this process.

This paper focuses specifically on the field survey, on-site water testing, and laboratory analysis components, which constitute the decentralized supplementary surveillance layer evaluated in this study.

Training and Pre-field Preparation

Before entering the field, all participants completed a standardized theoretical preparation phase embedded within a structured training framework ("A Hybrid Course on Water Quality - An Approach to People's Water Data," IIT Madras Course ID 5011). The training combined lectures, classroom and laboratory demonstrations, and field-oriented instruction to ensure methodological consistency in water-quality assessment. Core modules addressed water-quality parameters, contamination pathways, physicochemical and microbiological indicators, analytical methods, and field protocols. The preparation phase ensured alignment with predefined sampling procedures and quality-control requirements prior to field deployment. The detailed syllabus, including learning outcomes, key references, and faculty contributions, is provided in Table S1.

Pilot Testing and Protocol Optimization

Before full-scale deployment, the PWD protocol was refined through pilot testing, local engagement, and supervised field implementation:

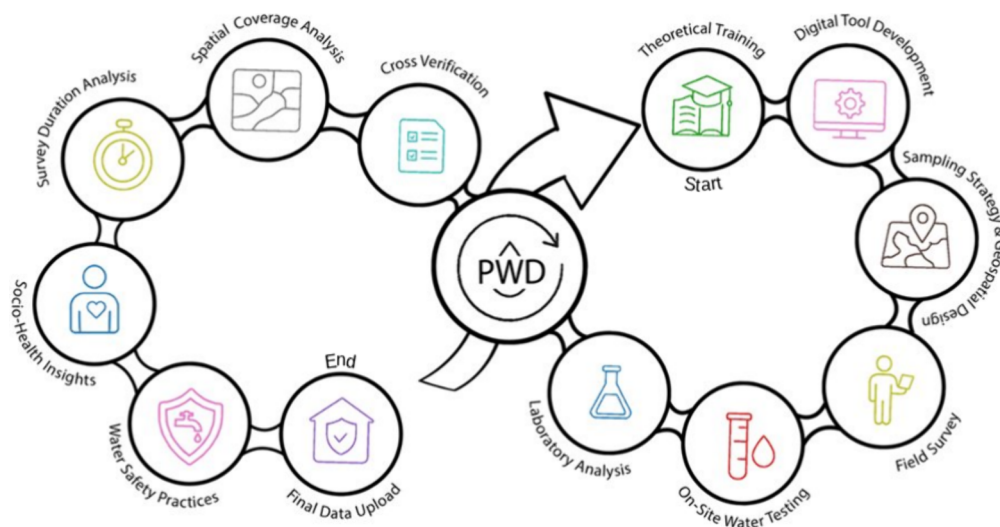


Figure 1. Structured workflow of the PWD decentralized monitoring framework, from Start at theoretical training and digital tool development to End at final data upload and integrated water-safety outputs.

- 1. Pilot field assessment:** The process began with a participatory assessment, including field visits to target areas, and pilot surveys comprising 262 surveys in Chennai (urban) and 150 surveys in Nallampatti village (rural), respectively. These pilot deployments were conducted to evaluate sampling feasibility, parameter consistency, and operational reliability under field conditions. The assessment included mapping household and community drinking-water sources, such as natural water bodies, wells, and public taps. Pilot testing was used to refine the survey and field procedures. In parallel, an internal validation assessment was conducted among the research team to evaluate the clarity, usability, and functional robustness of the digital survey tool. Feedback from this informed protocol adjustments that strengthened procedural standardization before expansion to broader geographic deployment.
- 2. Local engagement and training:** In the second stage, emphasis was placed on standardized capacity alignment of participating field teams to ensure consistent implementation across geographic locations. Classroom and laboratory sessions covered theoretical principles and applied procedures for water sampling, parameter measurement, and sample handling. This phase ensured adherence to predefined protocols and minimized inter-observer variability prior to large-scale deployment. The survey, detailed in Table S3, was translated into the local language to reflect regional cultural and environmental contexts, while maintaining the structural consistency of variables across study sites.
- 3. Supervised household survey and water testing:** In the final stage, spatially guided household sampling was implemented using a standardized field protocol. Field participants or research team members interviewed the household member responsible for water provision, storage, or use and conducted on-site water-quality tests. Sterile water samples were collected for extended laboratory validation.

Field implementation was conducted under structured supervision, with expert cross-verification integrated throughout the process to ensure protocol adherence and analytical consistency. Water-quality parameters were analyzed using portable instruments, field kits, and test strips suitable for decentralized deployment, enabling on-site measurement under defined operating conditions, as detailed in Tables S4 and S5 and Section 5 of the Supporting Information.

Digital Tool Development

A custom digital survey was implemented using EpiCollect5. In this study, it served as the deployment infrastructure for the customized PWD survey form and standardized field workflow. This configuration enabled standardized digital data capture across geographically distributed field sites while minimizing data loss in areas with limited connectivity.

The customized PWD survey architecture was developed by the research team to collect and link household-level information, water-source characteristics, field observations, on-site water-quality measurements, GPS coordinates, timestamps, and photographic documentation. Each sampled water source was assigned a unique alphanumeric identifier incorporating the participant code and timestamp. This identifier linked the household survey, field measurements, GPS location, sample collection, and laboratory results, thereby improving traceability and reducing transcription, duplication, and sample-matching errors. The survey structure and variables are provided in Table S3.

EpiCollect5 was used for standardized mobile data collection, while georeferenced field records were subsequently organized within the broader PWD digital data architecture for traceability, visualization, and data-management purposes. Further development of a dedicated PWD platform/app is ongoing to support future deployments.

Sampling Strategy and Geospatial Design

To ensure spatially guided and geographically distributed survey coverage, each study area was divided into predefined geographic polygons established by the research team, course instructors, or local coordinators to represent the habitations of interest. Large areas were subdivided into smaller operational units to maintain comparable field effort and improve spatial balance across participants. Each participant was assigned to a predefined polygon or spatial allocation unit and instructed to complete 10 surveys of household drinking-water sources or natural/source-water locations within that assigned polygon.

Within each polygon, sampling locations were selected according to a predefined spatial allocation protocol. In the initial implementation phase, participants selected households within the assigned polygon, subject to spatial constraints intended to reduce clustering and convenience-based selection. In later phases, GIS-supported candidate locations and Google Maps guidance were used to improve spatial dispersion and reduce repeated sampling of nearby households. Candidate coordinates were used as field guidance rather than fixed household addresses, allowing limited field flexibility where access, safety, or respondent availability prevented sampling at the suggested location. Each survey location was documented using GPS coordinates, timestamps, and photographic records for post-field verification, following the ethics and privacy procedures described below. If an intended household was unavailable, declined participation, or could not be accessed safely, the nearest eligible neighboring household within the same polygon was surveyed. This substitution rule was applied only when necessary to minimize non-response while preserving the intended spatial distribution within the assigned polygon. This approach reduced, but could not fully eliminate, selection bias; therefore, spatial clustering and protocol adherence were reviewed during post-field quality control using GPS locations, timestamps, and survey metadata. Detailed operational instructions, including household spacing rules, navigation guidance, non-response handling, and verification procedures, are provided in Text S3.

Field Survey and On-Site Testing

All field surveys were conducted after obtaining informed consent and under the ethical approval described in the [Ethics, Privacy, and Data Governance](#) section.

In the field phase, participants visited households selected within the assigned polygon, following the predefined spatial allocation protocol described in [Pilot Testing and Protocol Optimization](#) and detailed in Text S3. At each location, the household member responsible for water provision, storage, or use was interviewed using the standardized survey tool after obtaining informed consent. The survey documented household water-source characteristics, storage and treatment practices, user perceptions, and selected water-use behaviors relevant to point-of-use drinking-water safety. The survey questions are listed in Table S3. Additionally, on-site water testing was conducted using portable field-testing kits, including 5-in-1 dip strips (Hach 2755250 multi-parameter test strips) for measurement of hardness, alkalinity, free chlorine, total chlorine, and pH, as well as biological kits (HiMedia test strip K016) to test for *E. coli* and total coliforms. These parameters were selected because they provide rapid field-level screening of basic water chemistry, residual disinfectant status, and microbial contamination indicators relevant to household drinking-water safety. Water samples were also collected in sterile containers for subsequent laboratory analysis.

The field-testing tools were portable screening tools and were not permanently installed sensors. Field results were recorded in the digital survey form together with GPS coordinates, timestamps, and photographic documentation, supporting traceability and cross-verification. Detailed operational instructions for sampling, household substitution, sample labeling, on-site testing, photographic documentation, and verification are provided in Text S3.

The digital survey workflow also included structured modules on household water-use practices, storage, treatment, hygiene-related

behaviors, and limited self-reported health indicators; the full health-survey structure is provided in Table S6. Where household-level feedback was provided, households received a non-technical digital water-quality report with measured values, reference limits, and standardized advisory notes; the report structure is provided in Table S7.

Laboratory Analysis and Measurements

Collected water samples were transported to the laboratory within 24 h of collection to preserve sample integrity. Sample collection, preservation, transport, and handling were performed according to the Standard Methods for the Examination of Water and Wastewater.²³ For metal analysis, inductively coupled plasma mass spectrometry (Perkin-Elmer ICP-MS) was employed following established procedures,²⁴ enabling high-sensitivity detection of arsenic, lead, cadmium, chromium, and uranium. Detailed parameter ranges, testing methods, equipment, and relevant WHO/BIS standards are provided in Tables S4 and S5. This multi-parameter analytical framework enabled systematic detection of chemical and microbiological contaminants across sampled sites, as detailed in Section 5 of the Supporting Information.

Cross-verification, Validation, and Statistical Analysis

A subset of households was revisited to recollect samples and repeat measurements for inter-observer validation, enabling systematic comparison between student-generated data and expert assessments. Validation was designed to evaluate the operational reliability of the decentralized monitoring workflow under field conditions, rather than to serve as a fully randomized regulatory audit. The polygon-based framework ensured spatially matched sampling locations for these comparisons. Expert follow-up revisits were conducted for 307 previously surveyed households, of which 123 were fully verified after multiple pilot refinements and SOP-based screening; detailed agreement analyses are provided in Tables S8 and S9. Validation sites were selected based on feasibility of revisiting, household availability, geographic distribution, and the need to compare participant-generated data with expert assessment under comparable field conditions. This process allowed quantification of agreement rates and identification of measurement discrepancies between field operators. A similar analytical comparison approach has been previously reported in related validation studies.¹¹

The resulting agreement metrics were used to evaluate the consistency of participant-generated data relative to expert verification within the decentralized monitoring framework. Agreement was calculated as the proportion of matched participant and expert records that produced the same categorical response, the same presence/absence result, or the same predefined field-test category. Because several field measurements relied on categorical or semi-quantitative test-kit outputs, agreement percentages were interpreted only as indicators of operational consistency.

Spatial agreement was assessed by comparing paired GPS locations between participant and expert observations and determining whether they met the predefined spatial-matching criterion used for validation. Descriptive statistics, including means, standard deviations, ranges, frequencies, and agreement percentages, were used to summarize field measurements, laboratory results, and validation outcomes. Because this study was designed to validate a monitoring framework rather than test a causal exposure–outcome hypothesis, the analysis focused on descriptive statistics and agreement-based validation metrics rather than inferential hypothesis testing. Potential sources of uncertainty include household availability during revisits, feasibility constraints in resampling, inter-operator variability, GPS error, and the categorical resolution of field-test kits. These limitations are acknowledged when interpreting the validation results. A detailed SOP outlining each step of the validation procedure is available in Text S3. Full agreement rates for household survey variables and field-measured parameters are provided in Tables S8 and S9.

A summary of key data-quality risks and mitigation measures implemented across sampling, field testing, digital documentation, and validation is provided in Table S10.

Distance and Coverage Analysis

Implementation of the PWD model focused on semi-urban and rural regions characterized by documented variability in drinking-water quality and limited household-level monitoring coverage in India, Uganda, and Israel. Study sites were selected based on institutional capacity to support standardized field analysis and prior documentation of microbial or chemical contamination concerns. Site allocation was structured to ensure geographically distributed sampling within defined administrative boundaries. In selected areas, such as Nallampatti, the polygon-based framework enabled spatially organized survey implementation across multiple water-source types.

A village-level mapping and water-quality baseline assessment were conducted to characterize distribution patterns of sampled sites. Approximately 10% of the allocated surveys in each region were initially implemented as pilot deployments to evaluate sampling logistics, spatial distribution consistency, and contextual feasibility. Findings from these pilot deployments suggested minor refinements in the full-scale implementation protocol prior to broader deployment.

Variability in Survey Duration

A layered monitoring approach was implemented to improve data reliability and reduce the risk of fabricated or low-quality data. Time-based indicators were incorporated into the survey platform to evaluate data quality and to monitor compliance with protocols. These included the total survey duration, the time intervals between consecutive surveys conducted by the same participant, and the response time per question.

Surveys completed faster than the pre-specified minimum duration were flagged for review as potential protocol deviations, whereas extended completion times were evaluated for consistency with expected field conditions. Offline validation tests were also conducted to estimate reference survey durations under controlled conditions. Automated timestamp analysis enabled the detection of outliers in completion patterns, and flagged entries were subjected to structured review prior to final data integration. These procedures were implemented to ensure standardized survey execution across decentralized field deployments.

Ethics, Privacy, and Data Governance

All household interactions were conducted under approval from the Institutional Ethics Committee of the Indian Institute of Technology Madras (IEC/2024-02/PT/16; valid from June 21, 2024 to June 20, 2027). Respondents were informed about the purpose of the survey, the voluntary nature of participation, and the type of information collected, and informed consent was obtained before household interviews and water sampling. The survey collected household water-use information, water-source characteristics, GPS coordinates, photographs for verification, and limited self-reported health or water-related practice information.

Personal identifiers were minimized in the analytical dataset, and raw records containing household-level geolocation or potentially identifiable information were restricted to authorized research personnel for quality control, validation, and data processing. Publicly available outputs were de-identified or aggregated to prevent household identification. Where household-level water-quality reports were provided, results were communicated in non-technical language as screening information rather than regulatory certification, together with standardized guidance for safer storage, treatment, point-of-use treatment, disinfection, or follow-up testing where relevant. Further details are provided in the Ethics section of the Supporting Information (Section 4, Ethical Committee Approval).

RESULTS AND DISCUSSION

The results evaluate whether trained participants, operating under standardized protocols and expert cross-verification, can generate spatially traceable and analytically useful household-level water-quality data.

Implementation Scale and Household-Level Monitoring Density

Figure S5 summarizes the geographic distribution of the analyzed dataset across participating regions. Most records were from India ($n = 8981$), with additional implementation sites in southern Israel ($n = 54$) and Uganda ($n = 32$). The map illustrates geographic coverage, localized coverage gaps, and microregions where follow-up assessment may be prioritized.

Training and Pre-field Preparation

A total of 34 academic institutions actively hosted laboratory hubs as part of the course, providing distributed analytical support across India, Uganda, and Israel. This distributed hub structure enabled standardized deployment of laboratory and field procedures while maintaining procedural alignment across geographically dispersed sites. In the context of framework validation, the training and hub structure served as an operational quality-control layer. It supported consistent field preparation, sample handling, on-site testing, GPS documentation, and data upload across decentralized field teams. Thus, the multi-institutional deployment model demonstrates not only implementation scale but also the importance of standardized training and institutional support for maintaining reproducibility under heterogeneous field conditions.

Sampling Strategy and Geospatial Design

Figure 2A illustrates spatial correspondence between expert re-sampling locations (blue crosses) and initial household sampling points (orange circles). Spatial overlap between paired sampling points indicates alignment of georeferenced locations used for cross-validation. Agreement between student- and expert-identified sampling coordinates exceeded 80% across 1370 paired observations, demonstrating substantial spatial consistency under decentralized deployment conditions. Spatial agreement was defined as paired sampling locations falling within a predefined geographic distance of one another. Minor divergence was observed in the peripheral regions, where the student-collected data appeared more dispersed. Following protocol refinements and targeted calibration during the pilot phase ($n = 123$ households), spatial agreement improved to 99% (Figure 2B). This increase reflects procedural standardization and improved adherence to predefined geospatial allocation criteria. While the higher precision observed during pilot deployment required intensified supervision and field preparation, the broader dataset demonstrates that high spatial reliability can be maintained under routine decentralized implementation. These validation results indicate that structured decentralized sampling can achieve reproducible geospatial alignment, supporting reliable household-level monitoring density without requiring all sampling to be conducted by centralized expert teams.

To reduce convenience-based clustering and improve reproducibility of household-level sampling, a GIS-supported allocation procedure was introduced during protocol refinement. As illustrated in Figure 3, candidate sampling points were

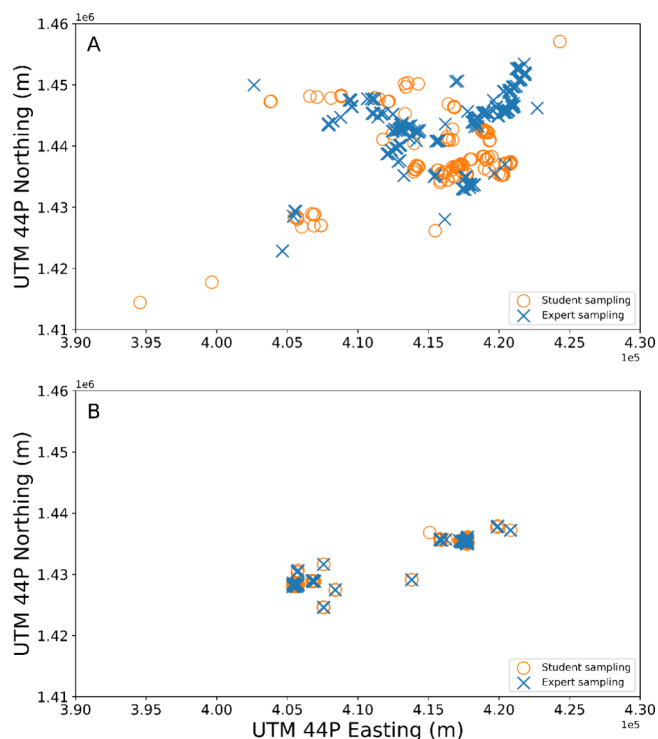


Figure 2. Spatial agreement analysis between decentralized field sampling locations and expert verification points. Axes show UTM 44P Easting and Northing coordinates in meters. (A) Geospatial comparison across 1370 observations. (B) Refined pilot phase demonstrating 99% spatial concordance across 123 validated observations.

generated within a predefined polygon using a Sobol sequence. These points were used as spatial guidance rather than as fixed household addresses, allowing field flexibility when access, safety, or respondent availability prevented sampling at the suggested location. This design was intended to preserve geographic distribution while reducing participant-driven selection bias.

Figure 3 compares two point-assignment strategies considered during protocol development. Non-clustered assignment of Sobol-generated points to participants distributed candidate locations across the polygon but often assigned widely separated locations to the same participant, increasing expected travel burden and route crossing (Figure 3A). In contrast, K-means-supported allocation grouped nearby candidate points into participant-level clusters, maintaining polygon-level coverage while producing more operationally feasible 10-site sampling assignments (Figure 3B). Thus, the final allocation strategy combined low-discrepancy spatial point generation with participant-level geographic grouping, balancing spatial dispersion with field practicality. Because candidate points were used as guidance rather than as mandatory household addresses, substitutions were permitted when necessary by sampling the nearest eligible household within the same polygon. This procedure was intended to reduce the likelihood of repeated sampling near easily accessible locations while preserving the spatial structure of the sampling design.

This geospatial design step was included to address a central limitation of decentralized field sampling: the tendency for non-expert surveyors to concentrate samples near familiar, accessible, or logistically convenient locations. Details of the

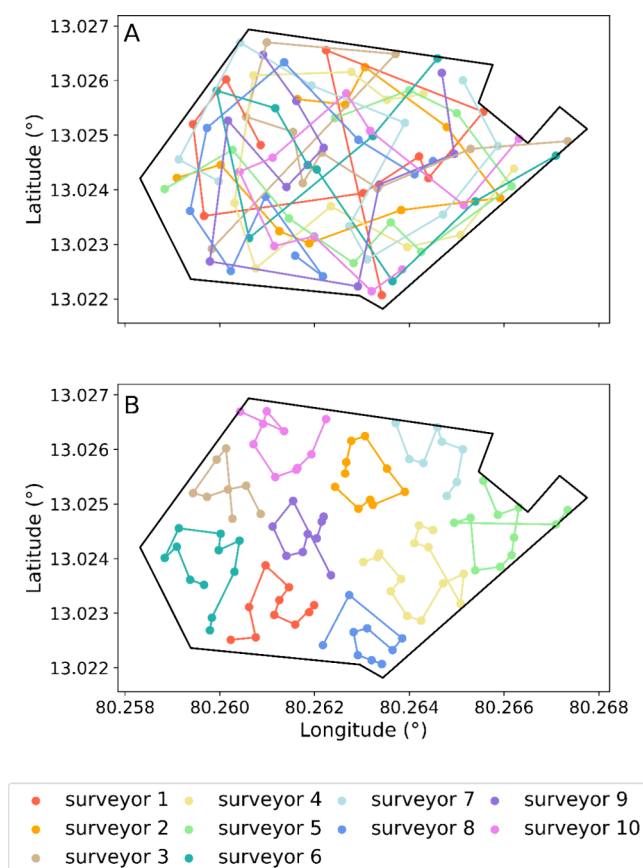


Figure 3. GIS-supported allocation of candidate sampling locations within a predefined study polygon. Candidate sampling points were generated using a Sobol sequence and assigned to field participants using either (A) non-clustered allocation or (B) K-means-supported allocation. Non-clustered allocation disperses participant-specific locations across the polygon, whereas K-means-supported allocation groups nearby candidate points into operationally feasible 10-site participant assignments while maintaining polygon-level spatial coverage. Colored lines connect assigned points for visualization of participant-level allocation patterns and do not represent measured road-network travel routes.

protocol-development sequence, phase-specific allocation procedures, and household-substitution rules are provided in the Supporting Information (Text S3).

Distance and Coverage Analysis

Figure 4 presents the 10 samples collected by each participant, with each color representing an individual field operator. The connecting lines illustrate the sequential movement between sampling sites within assigned polygons. Inter-point distances reflect the effective spatial dispersion achieved by each participant during field deployment. Figure 4A illustrates the participants' sampling routes during the first observation season ($n = 177$).

Spatial clustering and shorter inter-site distances are evident in several cases, indicating localized concentration of sampling effort within limited sub-areas of the assigned polygon. Following implementation of the structured SOP described in Text S3, sampling dispersion improved, as shown in Figure 4B ($n = 298$), with greater spacing between sampling points and reduced route overlap as visually assessed from the

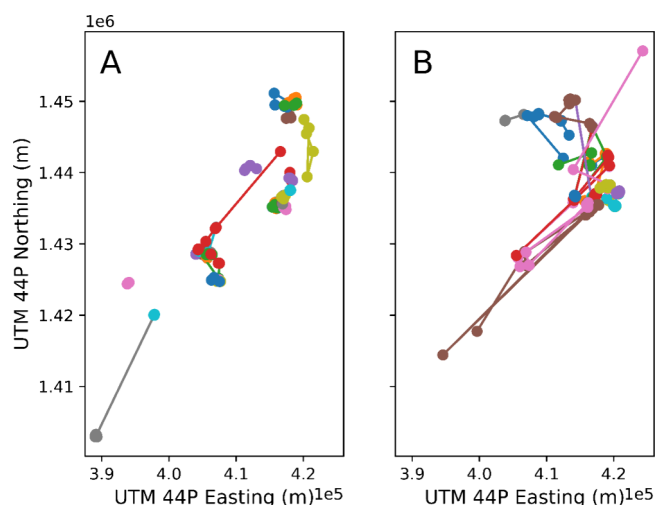


Figure 4. Household field sampling routes, with each color representing an individual field participant. Axes show UTM 44P Easting and Northing coordinates in meters. Each line indicates a 10-site sampling sequence. (A) Sampling routes from the first observation season ($n = 177$) and (B) sampling routes from the following season ($n = 298$), after implementation of the standardized operational protocol (Text S3). Substantial variability in movement patterns was observed: some participants covered extensive geographic areas, while others operated within a limited spatial range. This heterogeneity highlighted the importance of strengthening consistency in spatial deployment procedures. Uneven distribution of sampling locations during the initial phase reduced spatial representativeness in certain sub-areas of the assigned polygons. External factors such as terrain variability, accessibility constraints, and time limitations influenced participants' ability to travel between sites. In addition, participant-recorded geolocation data introduced minor spatial variation.

GPS-based sampling routes and inter-point spacing. More even spatial coverage strengthens the reliability of household-level contamination mapping, supporting better identification of localized risks relevant to decentralized water management.

These observations suggested subsequent refinements to spatial allocation protocols and the integration of GIS-based validation tools, improving coverage standardization and strengthening overall spatial data reliability within the decentralized monitoring framework. Improved spatial consistency enhances the reliability of localized contamination mapping, supporting more informed prioritization of household-level risk within decentralized water management systems.

Variability in Survey Duration

Figure 5 shows the distribution of time required to complete each survey question, based on 93,204 individual observations collected across 1194 surveys. The mean survey completion time was approximately 22 min. However, in some cases, participants completed all 10 surveys within a brief time window. This variability may reflect operational factors such as offline data entry followed by delayed upload. In addition, compressed completion times may indicate deviations from real-time field execution and were therefore treated as indicators for data-quality review rather than as direct evidence of invalid records. Figure S6, based on 3900 observations derived from 650 surveys, offers a complementary perspective by grouping questions into six main blocks. The longest median completion time was observed in Block 1, which includes the introductory and

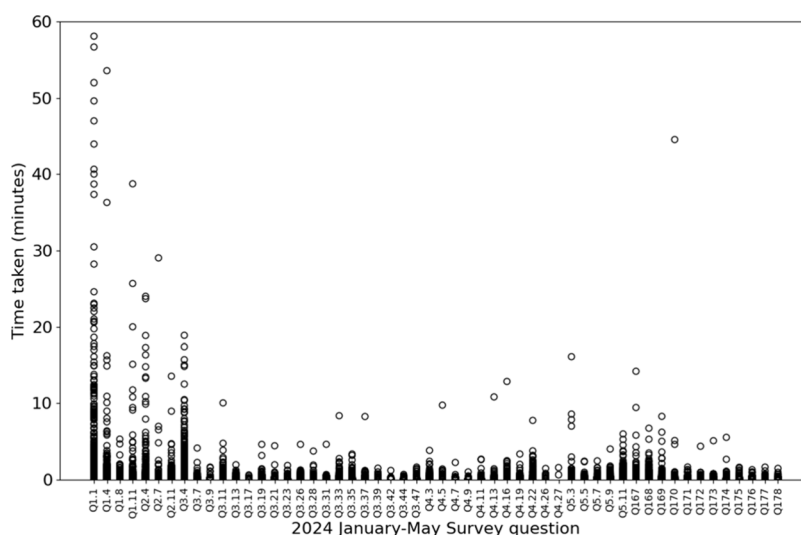


Figure 5. Distribution of survey completion times per question across 1194 surveys (93,204 observations), illustrating variability in response-time patterns.

household identification components. In contrast, Blocks 5 and 6 showed the shortest completion times, reflecting more structured and repetitive measurement-based entries. Intermediate durations were observed for Blocks 2-4.

This distribution indicates that descriptive and interaction-based sections show greater variability in completion time, whereas standardized parameter-entry sections are completed more consistently.

This block-based analysis enables examination of response-time consistency across broader conceptual sections of the survey instrument, complementing the question-level assessment presented in Figure 5. Systematic monitoring of survey duration, therefore, strengthened internal data-quality screening by helping to identify entries requiring further review within the decentralized monitoring workflow.

A comparison between the January-May and July-November survey periods showed clear differences in time management and data consistency, driven by changes in the survey system's configuration, as shown in Figure 6. During the January-May semester, participants were not subject to an automated cap on the number of surveys they could complete. As a result, survey participation was highly variable, with some participants conducting over 20 surveys. This lack of constraint contributed to substantial variability in survey duration, with some durations reaching up to 4500 min (Figure 6A). Such extended durations likely reflect operational interruptions, delayed submission of offline entries, or prolonged session activity, rather than continuous real-time survey execution.

In contrast, the July-November semester introduced a system-imposed cap, requiring each student to complete exactly 10 surveys. This modification led to a more standardized pattern of participation and reduced extreme variability in survey-duration records (Figure 6B). Most responses fell within a reasonable timeframe of 20 min or less. The cap did not directly reduce the time required to complete individual survey questions; rather, it standardized participant workload and reduced extreme completion patterns that complicated data-quality screening.

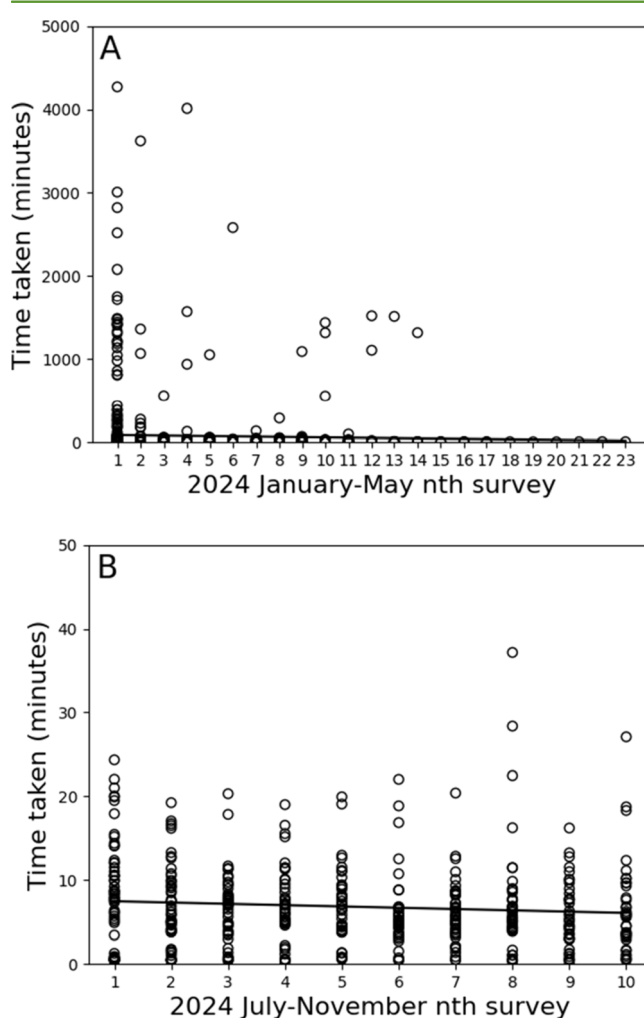


Figure 6. Comparison of survey duration before and after implementation of automated participation limits. (A) No cap on survey number ($n = 1523$). (B) Fixed cap of 10 surveys per participant ($n = 441$).

The introduction of automated participation limits improved consistency in survey completion patterns, strengthened data-quality control, and facilitated more streamlined downstream analysis within the decentralized monitoring framework.

This change made survey-duration records easier to interpret and helped flag entries requiring quality-control review.

Systematic validation and automated controls enhance data reliability by reducing individual measurement variability and improving overall consistency. Previous work on crowdsourced environmental monitoring shows that repeated assessment and comparison procedures can help identify inconsistencies and strengthen confidence in non-expert-generated data.^{25,26} These findings highlight the critical role of system-level design in shaping data-collection behavior. In the PWD framework, implementing clearly defined operational constraints and automated controls improved protocol adherence and strengthened data quality within the decentralized monitoring workflow.

Laboratory Analysis and Measurements

Tables S4 and S5 summarize all water-quality parameters tested in this study, including criteria from both the WHO and the Bureau of Indian Standards (BIS) to ensure alignment with international and national drinking-water standards.^{27,28}

Figure 7 presents the distribution of key physicochemical water parameters, including conductivity, total dissolved solids (TDS), turbidity, pH, alkalinity, hardness, and oxidation–reduction potential (ORP), across sampled households. The mean electrical conductivity was 444 ± 478 $\mu\text{S}/\text{cm}$. The mean TDS concentration was 187 ± 196 mg/L . Turbidity levels averaged 1.1 ± 1.8 NTU, ranging from 0 to 13.0 NTU, indicating that most samples met the acceptable clarity standards. The mean pH was 7.80 ± 0.30 , reflecting generally neutral to slightly alkaline conditions. ORP values averaged -5 ± 61 mV, indicating heterogeneous redox conditions across study sites. Alkalinity averaged 96 ± 71 mg/L as CaCO_3 (range: 0–240 mg/L). Hardness levels averaged 145 ± 103 mg/L as CaCO_3 (range: 0–425 mg/L). Zero values reflect the resolution limits of strip-based field measurements, which report concentration ranges rather than continuous values.

In addition to these physicochemical indicators, field-based biological testing revealed that *E. coli* was detected in 69.4% of samples, while total coliforms were detected in 88.0% of samples. Microbial contamination was observed even in households where most physicochemical parameters were within or near recommended WHO and BIS limits. These findings underscore

the importance of household-level microbiological surveillance, as compliance with conventional physicochemical indicators does not necessarily indicate microbiological safety at the point of use. This distinction supports the value of integrating microbial screening into decentralized household-level monitoring, particularly for identifying localized risks that may require confirmatory testing or targeted follow-up.

Cross-verification and Validation

Ensuring the reliability of non-expert contributions in decentralized household-level water monitoring systems requires structured validation and oversight. Previous studies found that non-expert accuracy improves with expert validation, confidence scoring, and cross-comparison with expert-labeled data.¹¹ These studies also highlighted that higher self-reported confidence correlates with greater accuracy, reinforcing the value of feedback mechanisms in crowdsourced monitoring. Other studies have demonstrated that non-experts can play a meaningful role in generating valuable data when supported by well-structured, simplified tools.¹¹ These insights were further reinforced by USAID's pilot study in Tanzania, which highlighted the promise of decentralized water monitoring while emphasizing the need for validation and standardization mechanisms.²⁹ These principles were implemented in the PWD framework through two key design strategies. First, close-ended, clearly phrased questions based on observable parameters consistently yielded higher agreement rates and supported the use of standardized survey formats in decentralized monitoring. Second, beyond training alone, methodological simplification was essential for maintaining reliability during scale-up of student- or community-led data collection.

Figure 8 presents a selection of representative questions and their agreement patterns. Details on household survey variables are provided in Table S8, and field-measured parameter agreement rates are provided in Table S9. Biological test agreement was 87% for *E. coli* and 96% for total coliforms, demonstrating strong agreement for field-based microbial screening. Adherence to GIS-tracked survey protocols increased from 60 to 90% following the implementation of real-time tracking mechanisms and the inclusion of multi-angle photographs capturing both the street environment and household entrance (Figure S2, Text S3). This improvement reflects the impact of embedded verification mechanisms on protocol compliance and spatial consistency.

Clear patterns emerge regarding which types of questions lead to high agreement and where discrepancies occur.

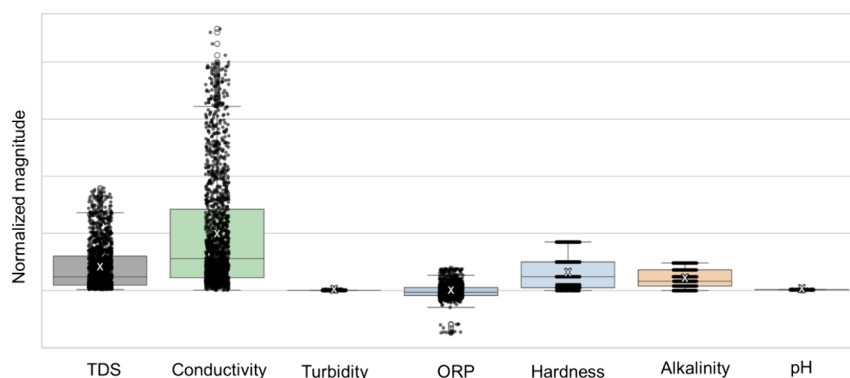


Figure 7. Distribution of key physicochemical water-quality parameters across household samples.

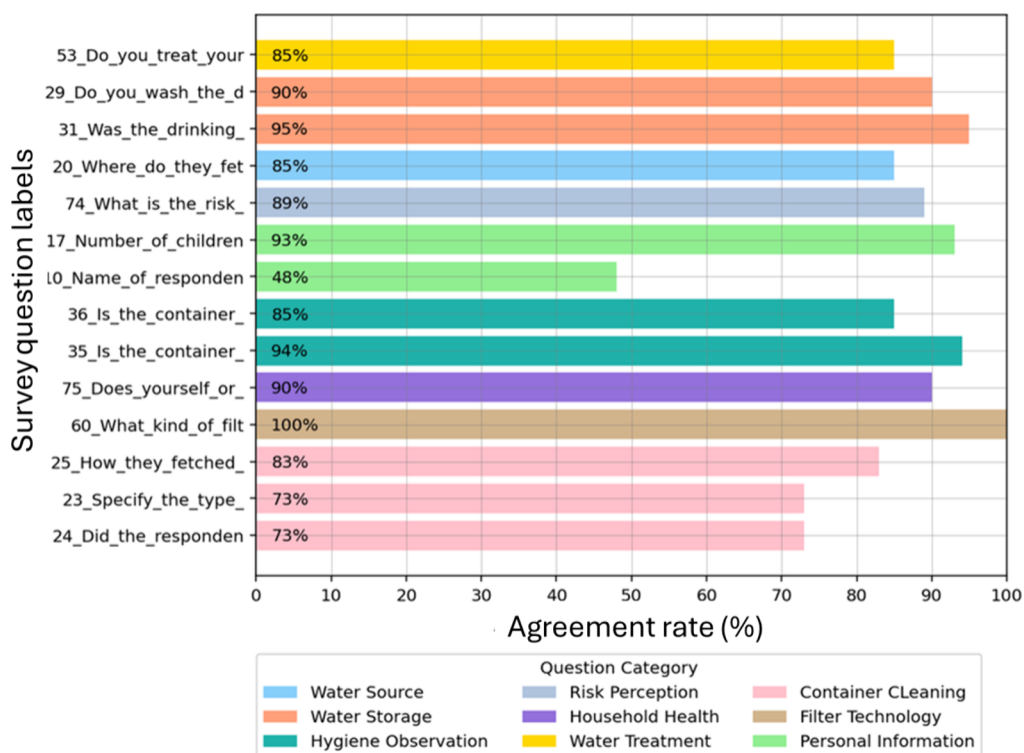


Figure 8. Agreement rates between participant and expert field assessments across key household water survey parameters, categorized by topic. Full question details are provided in Table S8.

Questions based on simple, visible observations, such as the type of water filter, whether the water is stored in a container, the container's placement, or the number of children under five, showed very high agreement rates, ranging from 93 to 100%. These findings indicate strong reliability for structured observational variables. In contrast, questions that required attention to actions or behaviors, such as how water was drawn from the container or whether the respondent's hand touched the water, had more moderate agreement levels, around 72 to 82%, reflecting interpretive variability under field conditions. The question with the lowest agreement rate (48%) was the respondent's name, which likely reflects issues with how participants recorded names, possibly due to spelling or transcription errors, or the presence of different individuals during each visit. On the other hand, a personal numerical detail, such as the number of children, was captured with high accuracy, reinforcing the idea that participants manage structured or quantitative data more effectively. Even questions related to risk perception and household health, which are more subjective, showed high agreement rates of 85 to 90%. This suggests that self-reported information could be documented consistently when questions were clearly structured. These findings suggest that future PWD deployments should evaluate whether repeated household-level measurement and feedback lead to sustained changes in water storage, treatment, and hygiene practices, using longitudinal follow-up to distinguish short-term responses from persistent behavioral change.

Overall, the findings suggest that participants collect consistent field data when supported by structured protocols, simplified survey formats, and embedded verification mechanisms, especially for factual or observational questions. However, agreement levels tend to drop when dealing with open-ended

questions, which introduce more room for variation in how responses are interpreted and recorded. In contrast, multiple-choice questions produce more consistent results and higher agreement rates. Accordingly, later survey refinement prioritized close-ended or categorized response formats where standardization was required, while open-ended responses were retained only where contextual detail was necessary. These results suggest that standardizing question formats and integrating verification mechanisms are central to maintaining data robustness at scale in decentralized monitoring systems.

Digital participation platforms have been proposed as mechanisms to expand distributed data-collection capacity for water-quality assessment, particularly in settings where systematic large-scale monitoring is limited. Previous crowdsourced water-monitoring research has shown that structured validation and comparison procedures can improve the accuracy of non-expert-generated environmental data.²⁵ The PWD framework applies these principles through standardized training, digital data capture, expert cross-verification, and structured feedback mechanisms. This approach is consistent with previous evidence showing that non-expert accuracy improves when participants receive feedback and learning opportunities.

RESOURCE EFFICIENCY AND MONITORING DENSITY IMPLICATIONS

PWD is not presented as a lower-cost alternative to certified regulatory monitoring, but as a supplementary household-level screening and prioritization layer. Based on supplier quotations and purchase records, the estimated direct equipment and consumables cost was approximately \$9.38 per household

assessment, including the microbial presence/absence test (\$6.39), multi-parameter strip (\$1.01), other consumables (\$0.22), and reusable multi-parameter and turbidity-meter costs (\$0.96 and \$0.81, respectively). Under these assumptions, approximately 107 household-level assessments, corresponding to about 1170–1180 core water-quality parameter observations, can be generated per \$1000 of documented direct equipment and consumables cost. These estimates exclude labor, transport, training, laboratory validation, institutional overhead, platform maintenance, and long-term governance costs and therefore should not be interpreted as a direct cost comparison with centralized or laboratory-based monitoring. Previous analyses have shown that the relative costs of centralized and decentralized microbial testing depend on factors including distance to laboratory facilities, system density, transport requirements, and sampling frequency, while the sustainability of local laboratory capacity also depends on sufficient routine testing demand.^{17,18} Accordingly, the resource value of PWD is best interpreted in terms of increasing household-level screening density and supporting prioritization of professional follow-up and laboratory confirmation, rather than replacing existing laboratory infrastructure. Detailed component-level assumptions are provided in Table S11 and Figure S7. Future evaluations should also examine whether decentralized screening signals translate into uptake of confirmatory testing, targeted interventions, and timely institutional responses, particularly in settings with limited routine monitoring capacity.¹⁹

CONCLUSIONS

This study evaluated the opportunities and limitations of decentralized household-level water monitoring through the PWD framework. The findings demonstrate that distributed data collection embedded within a structured validation and quality-control system offers a promising strategy for expanding the geographic scope and sampling frequency of water-quality monitoring, particularly in low-resource or under-surveyed areas. Over a two-year period, monitoring activities assessed over 9000 household water sources across India, Uganda, and Israel, demonstrating the feasibility of structured decentralized surveillance at scale.

Several operational and methodological challenges were identified. These included spatial clustering of sampling points, inconsistent adherence to protocols, and context-specific barriers such as accessibility, safety, and respondent availability. These factors can compromise data completeness and consistency and may reduce spatial representativeness within assigned sampling areas. To address such issues, the framework incorporated multiple quality assurance measures, most notably real-time GIS tracking, automated timestamping, expert cross-validation, and structured feedback loops. Data reliability was closely linked to the type and structure of survey questions. Closed-ended, clearly worded, and observable questions achieved higher agreement rates between field-generated and expert assessments, often exceeding 90%, whereas open-ended or interpretive questions yielded more variable responses. These findings underscore the importance of methodological simplification and standardized question formats when scaling decentralized monitoring systems.

In summary, the PWD framework demonstrates that when properly designed, supported, and validated, decentralized monitoring systems can generate scientifically robust, policy-relevant

data. It offers a transferable model for expanding quality-controlled drinking-water surveillance in other regional contexts, provided that appropriate institutional support, training capacity, ethical safeguards, and validation infrastructure are in place. Its longer-term value should therefore be evaluated not only by monitoring density and data reliability, but also by sustained household risk-reduction practices, use of confirmatory testing, and subsequent institutional follow-up. All data and results from this study are openly accessible on the PWD platform (<https://www.peopleswaterdata.org/>), ensuring transparency, reproducibility, and broader knowledge sharing.

ASSOCIATED CONTENT

Data Availability Statement

The training materials supporting the monitoring framework, including lecture notes, lecture videos, and student interactions across all the semesters, are available via NPTEL (<https://nptel.ac.in>). The monitoring dataset generated in this study is publicly accessible through the PWD platform, <https://www.peopleswaterdata.org> developed by Sankar Sudhir, which will be published separately. Guest lectures were delivered by experts from collaborating institutions, including faculty from IIT Madras, Tel Aviv University, and the KMCH Research Foundation, who contributed to technical oversight and domain-specific expertise within the monitoring framework.

Supporting Information

The Supporting Information is available free of charge at <https://pubs.acs.org/doi/10.1021/acssusresmgmt.6c00149>.

Additional details on the theoretical and operational training framework; household survey development and survey instruments; sampling strategy, geospatial design, and field protocols; ethical approval; laboratory analysis and water-quality measurements; health survey methodology; household feedback and reporting; implementation scale and monitoring density; survey-duration variability; participant–expert agreement analyses; data-quality risks and mitigation measures; and resource-efficiency and monitoring-density estimates, including supplementary tables, figures, and standard operating procedures (PDF)

AUTHOR INFORMATION

Corresponding Author

Thalappil Pradeep – DST Unit of Nanoscience and Thematic Unit of Excellence, Department of Chemistry, IIT Madras, Chennai 600 036, India; International Centre for Clean Water, 2nd Floor, B-Block, IIT Madras Research Park, Kanagam Road, Taramani, Chennai 600113, India; orcid.org/0000-0003-3174-534X; Email: pradeep@iitm.ac.in; Tel.: +91-44 2257 4208; Fax: +91-44 2257 0545/0509

Authors

Suzan Kagan – DST Unit of Nanoscience and Thematic Unit of Excellence, Department of Chemistry, IIT Madras, Chennai 600 036, India; School of Mechanical Engineering, Faculty of Engineering, Tel Aviv University, Tel Aviv 6997801, Israel
Sankar Sudhir – DST Unit of Nanoscience and Thematic Unit of Excellence, Department of Chemistry, IIT Madras, Chennai 600 036, India

Ben Hamilton – Faculty of Environment, Science and Economy, University of Exeter, Exeter EX4 4QF, U.K.

Alona Maslennikov – School of Mechanical Engineering, Faculty of Engineering, Tel Aviv University, Tel Aviv 6997801, Israel; School of Chemistry, Raymond and Beverly Sackler Faculty of Exact Sciences, Tel Aviv University, Tel Aviv 6997801, Israel

P. Amaravathy – Chemomicrobiomics Laboratory, Department of Biochemistry and Microbiology, KMCH Research Foundation, Coimbatore, Tamil Nadu 641 014, India

Timor Lichtman – Department of Public Policy and New Environment School, Tel Aviv University, Tel Aviv 6997801, Israel

Sonali Seth – DST Unit of Nanoscience and Thematic Unit of Excellence, Department of Chemistry, IIT Madras, Chennai 600 036, India

Tanmayaa Nayak – DST Unit of Nanoscience and Thematic Unit of Excellence, Department of Chemistry, IIT Madras, Chennai 600 036, India

Ramya Dwivedi – DST Unit of Nanoscience and Thematic Unit of Excellence, Department of Chemistry, IIT Madras, Chennai 600 036, India

Ganesan Velmurugan – Chemomicrobiomics Laboratory, Department of Biochemistry and Microbiology, KMCH Research Foundation, Coimbatore, Tamil Nadu 641 014, India; Central Research Laboratory, KMCH Institute of Health Science and Research, Coimbatore, Tamil Nadu 641 014, India

Hadas Mamane – School of Mechanical Engineering, Faculty of Engineering, Tel Aviv University, Tel Aviv 6997801, Israel;  orcid.org/0000-0001-5321-4342

Ram Fishman – Department of Public Policy and New Environment School, Tel Aviv University, Tel Aviv 6997801, Israel

Complete contact information is available at:

<https://pubs.acs.org/doi/10.1021/acssusresmgmt.6c00149>

Author Contributions

The manuscript was written through the contributions of all authors. All authors have given approval to the final version of the manuscript. S.K. initiated and led the methodological design and development of the PWD research framework, including the design, pilot testing, and refinement of core research tools. She co-developed the household survey tool with undergraduate participants from Tel Aviv University, designed and conducted the first field pilot, and co-led the subsequent pilot in Nallampatti village alongside G.V. As the course coordinator and jointly supervised Ph.D. student of H.M. and T.P., she served as the lead researcher and primary author of this manuscript, overseeing its structure, writing, and integration of field insights. S.S. developed the digital platform for PWD as part of his MS thesis work and, together with G.V., migrated the survey tool from Qualtrics to EpiCollect5, establishing the data framework of the PWD system. T.L. participated in the early pilot phases, including the Nallampatti fieldwork, and played a key role in building and refining the Qualtrics survey tool, incorporating upgrades that shaped its current version. S. Seth trained the first cohort of student participants, contributed actively to the Nallampatti pilot, and supported heavy metal measurement and analysis. T.N. joined the Nallampatti field pilot and contributed to heavy metal testing activities.

B.H. supported data analysis and field process optimization, contributing to the polygon-based randomization model and spatial design for survey allocation. R.D. coordinated the course at IIT Madras, and P.A. oversaw field operations in Coimbatore. G.V. served as a lecturer in the theoretical component, co-developed the EpiCollect5 survey and health modules and managed field logistics. R.F. supported the project from its early stages, contributing to the scientific and social framing of the framework. He provided ongoing guidance on the development of research tools and helped shape key methodological directions. A.M. coordinated fieldwork in Israel alongside S.K. and supported manuscript preparation and content refinement. H.M. supervised S.K., ensuring methodological consistency and scientific rigor. H.M. was closely involved throughout the development and implementation of the framework. She participated directly in field activities during the early pilots, provided continuous scientific support, and offered hands-on supervision and feedback across all stages. T.P. conceptualized, directed and led the PWD initiative, including its course content, structure, coordination, visualization, scale-up, funding, and coordination across institutions. T.P. supervised the project overall, integrating its scientific, educational, and societal components.

Funding

The initial pilot phase of this work was supported by the ASPER Clean Water Foundation (<https://asperfoundation.com/asper-clean-water-fund>) and the Olessia Kantor Fund (<https://www.olessia-kantor.com>), along with the existing research support of T.P. Following successful pilot implementation, the formal establishment and continued expansion of the PWD framework were supported by Petronet LNG Limited under its Corporate Social Responsibility (CSR) grant. T.P. acknowledges Petronet LNG Limited for supporting the broader objective of strengthening drinking-water monitoring capacity and advancing structured access to safe water systems. This research was also supported by the Ministry of Innovation, Science and Technology and the Planning and Budgeting Committee (PBC), Israel. Support from the Centre of Excellence on Molecular Materials and Functions through the Institute of Eminence (IoE) scheme of IIT Madras is also gratefully acknowledged.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

The framework is supported by a network of academic institutions, including the primary academic partners: the Indian Institute of Technology (IIT) Madras, Tel Aviv University, and KMCH Research Foundation. The course titled “A Hybrid Course on Water Quality - An Approach to People’s Water Data,” combining classroom instruction in hybrid mode with practical field analysis, was approved by the Institutional Ethics Committee (IEC) of the IIT Madras (Approval No. IEC/2024-02/PT/16, valid from June 21, 2024 to June 20, 2027). The course and related resources are accessible via the National Programme on Technology Enhanced Learning (NPTEL, <https://nptel.ac.in>). We extend our gratitude to the 1690 participants engaged in the broader framework across semesters, of whom 990 contributed to the dataset analyzed in this study. The collaborating institutions across India

are Dr. NGP Arts and Science College, KPR Institute of Engineering and Technology, Madras Christian College, Mar Ivanios College, Mata Gujri Mahila Mahavidyalaya, SRKR Engineering College, Stella Maris College, and multiple government colleges nationwide (a complete list is available at <https://www.peopleswaterdata.org>). The framework also works closely with key NGO and public sector organizations, including NALA Foundation (Ethiopia), Africa Innovation Foundation (Uganda), Development Alternatives (India), Engineering Without Borders (India chapter), and Jal Jeevan Mission (Assam, India). These partnerships represent coordinated institutional collaboration to support structured drinking water surveillance, sustainability objectives, and capacity development within water governance frameworks. T.L. was visiting a summer student at IITM while this work was performed.

REFERENCES

- (1) World Health Organization; United Nations Children's Fund. *Progress on Household Drinking Water, Sanitation and Hygiene 2000–2020: Five Years into the SDGs*; Geneva, 2021. <https://www.who.int/publications/i/item/9789240030848> (accessed September 10, 2025).
- (2) De Onis, M.; Monteiro, C.; Akre, J.; Clugston, G. The Worldwide Magnitude of Protein-Energy Malnutrition: An Overview from the WHO Global Database on Child Growth. *Bull. World Health Organ.* **1993**, *71* (6), 703–712.
- (3) 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.; Cumming, O. Burden of Disease Attributable to Unsafe Drinking Water, Sanitation, and Hygiene in Domestic Settings: A Global Analysis for Selected Adverse Health Outcomes. *Lancet* **2023**, *401* (10393), 2060–2071.
- (4) Wright, J.; Gundry, S.; Conroy, R. Household Drinking Water in Developing Countries: A Systematic Review of Microbiological Contamination between Source and Point-of-use. *Trop. Med. Int. Health* **2004**, *9* (1), 106–117.
- (5) Clasen, T. F.; Bastable, A. Faecal Contamination of Drinking Water during Collection and Household Storage: The Need to Extend Protection to the Point of Use. *J. Water Health* **2003**, *1* (3), 109–115.
- (6) Bain, R.; Johnston, R.; Khan, S.; Hancioglu, A.; Slaymaker, T. 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.* **2021**, *129* (9), 097010.
- (7) Wolf, J.; Hunter, P. R.; Freeman, M. C.; Cumming, O.; Clasen, T.; Bartram, J.; Higgins, J. P. T.; Johnston, R.; Medlicott, K.; Boisson, S.; Prüss-Ustün, A. Impact of Drinking Water, Sanitation and Handwashing with Soap on Childhood Diarrhoeal Disease: Updated Meta-analysis and Meta-regression. *Trop. Med. Int. Health* **2018**, *23* (5), 508–525.
- (8) Ng'andwe, S.; Ogendi, G. M.; Muoria, E.; Ngoma, J. Microbial Water Quality at the Point of Use: The Role of Socio-Economic Factors and Water Handling Practices in Kitwe District, Zambia. *Environ. Health Insights* **2025**, *19*, 11786302251337563.
- (9) Yang, A. R.; Bowling, J. M.; Morgan, C. E.; Bartram, J.; Kayser, G. L. Predictors of Household Drinking Water E. Coli Contamination: Population-Based Results from Rural Areas of Ghana, Malawi, Mozambique, Niger, Rwanda, Uganda, and Zambia. *Int. J. Hyg. Environ. Health* **2025**, *264*, 114507.
- (10) Lilje, J.; Mosler, H.-J. Continuation of Health Behaviors: Psychosocial Factors Sustaining Drinking Water Chlorination in a Longitudinal Study from Chad. *Sustainability* **2016**, *8* (11), 1149.
- (11) Ramesh, R.; Frank, E.; Padmavilochanan, A.; Barda, Y.; Eldar, I.; Wolf, H.; Pras, A.; Pousty, D.; Anita, P.; Shekar, L.; von Lieres, J. S.; Rao, R. B.; Mamane, H.; Fishman, R. Reliable Water Quality Monitoring by Women in Low-Resource Communities. *ACS ES&T Water* **2024**, *4* (9), 3832–3841.
- (12) Genter, F.; Putri, G. L.; Handayani, R.; Priadi, C.; Willetts, J.; Foster, T. Evaluation of a Participatory Citizen Approach to Monitor Microbial Water Quality of Self-Supply in Urban Indonesia. *Urban Water J.* **2024**, *21* (3), 272–284.
- (13) Pace, C.; Fencel, A.; Baehner, L.; Lukacs, H.; Cushing, L. J.; Morello-Frosch, R. The Drinking Water Tool: A Community-Driven Data Visualization Tool for Policy Implementation. *Int. J. Environ. Res. Public Health* **2022**, *19* (3), 1419.
- (14) Trent, M.; Dreifelbis, R.; Bir, A.; Tripathi, S. N.; Labhasetwar, P.; Nagarnaik, P.; Loo, A.; Bain, R.; Jeuland, M.; Brown, J. Access to Household Water Quality Information Leads to Safer Water: A Cluster Randomized Controlled Trial in India. *Environ. Sci. Technol.* **2018**, *52* (9), 5319–5329.
- (15) Buytaert, W.; Zulkafli, Z.; Grainger, S.; Acosta, L.; Alemie, T. C.; Bastiaensen, J.; De Bièvre, B.; Bhusal, J.; Clark, J.; Dewulf, A.; Foggin, M.; Hannah, D. M.; Hergarten, C.; Isaeva, A.; Karpouzoglou, T.; Pandeya, B.; Paudel, D.; Sharma, K.; Steenhuis, T.; Tilahun, S.; Van Hecken, G.; Zhumanova, M. Citizen Science in Hydrology and Water Resources: Opportunities for Knowledge Generation, Ecosystem Service Management, and Sustainable Development. *Front. Earth Sci.* **2014**, *2*, 26.
- (16) Rambau, L. D.; Mativenga, P. T.; Marnewick, A. L. Enhancing Frameworks for Utilising Sankey Diagrams in Modelling the Water-Energy Nexus and Circularity. *J. Clean. Prod.* **2024**, *456*, 142355.
- (17) Trimmer, J. T.; Delaire, C.; Marshall, K.; Khush, R.; Peletz, R. Centralized or Onsite Testing? Examining the Costs of Water Quality Monitoring in Rural Africa. *Environ. Sci. Technol.* **2024**, *58* (26), 11236–11246.
- (18) Ogisma, L.; O'Donnell, F. C.; Sawadgo, W.; Molnar, J. J.; Huluka, G.; Laguerre, E. Pricing Drinking Water Testing in Northern Haiti: Financial Sensitivity to Operating Costs, User Demand, and Economic Conditions. *Heliyon* **2024**, *10* (18), e38063.
- (19) Ogisma, L.; Sawadgo, W.; O'Donnell, F.; Molnar, J.; Huluka, G.; Saint Preux, J. F.; Louis, N.; Nord, O. Drinking-Water Quality in Neighborhood Kiosks: Suppliers' Willingness to Test Their Water in Northern Haiti. *Water Environ. Res.* **2025**, *97* (9), e70179.
- (20) Ndé-Tchoupé, A.; Crane, R.; Mwakabona, H.; Noubactep, C.; Njau, K. Technologies for Decentralized Fluoride Removal: Testing Metallic Iron-Based Filters. *Water* **2015**, *7* (12), 6750–6774.
- (21) Tepong-Tsindé, R.; Crane, R.; Noubactep, C.; Nassi, A.; Ruppert, H. Testing Metallic Iron Filtration Systems for Decentralized Water Treatment at Pilot Scale. *Water* **2015**, *7* (3), 868–897.
- (22) Pradeep, T.; Tam, M. K. C.; Serrano, J. F. Managing Water, the Mother of Resources: Thoughts on World Water Day 2024. *ACS Sustain. Resour. Manag.* **2024**, *1* (3), 368–369.
- (23) Standard Methods Committee. Collection and Preservation of Samples. In *Standard Methods for the Examination of Water and Wastewater*; Lipps, W. C., Braun-Howland, E. B., Baxter, T. E., Eds.; APHA Press: Washington, DC, 2023.
- (24) Seth, S.; Islam, M. R.; Nayak, T.; Kini, A. R.; Manna, S.; Malla, B. K.; Nagar, A.; Pradeep, T. Enhanced U(VI) Removal from Groundwater Using Amide-Functionalized Biopolymer-Based Adsorbent for a Practical Solution. *ACS Sustain. Chem. Eng.* **2025**, *13* (28), 10918–10929.
- (25) Strobl, B.; Etter, S.; van Meerveld, I.; Seibert, J. The CrowdWater Game: A Playful Way to Improve the Accuracy of Crowdsourced Water Level Class Data. *PLoS One* **2019**, *14* (9), e0222579.
- (26) See, L.; Comber, A.; Salk, C.; Fritz, S.; van der Velde, M.; Perger, C.; Schill, C.; McCallum, I.; Kraxner, F.; Obersteiner, M. Comparing the Quality of Crowdsourced Data Contributed by Expert and Non-Experts. *PLoS One* **2013**, *8* (7), e69958.

- (27) World Health Organization. *Guidelines for Drinking-Water Quality: Fourth Edition Incorporating the First Addendum*; Geneva, **2017**.
- (28) Bureau of Indian Standards. *Drinking Water Specification*, second revision, IS 10500:2012; Bureau of Indian Standards: New Delhi, **2012**.
- (29) Ajith, V.; Fishman, R.; Yosef, E.; Edris, S.; Ramesh, R.; Alookaran Suresh, R.; Pras, A.; Rahim, V.; Rajendran, S.; Yanko, M.; Amit, A.; Bhavani, R. R.; Vinodini Ramesh, M.; Mamane, H. An Integrated Methodology for Assessment of Drinking-Water Quality in Low-Income Settings. *Environ. Dev.* **2023**, *46*, 100862.