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.

# Africa's Water Revenue Paradox

**Duncan McNicholl**

duncan@uptimeglobal.org

University of Oxford

**Emil Aydinsoy**

Y.H. Mammadaliyev Institute of Petrochemical Processes

**Robert Hope**

University of Oxford https://orcid.org/0000-0001-9971-9397

**Article**

**Keywords:**

**Posted Date:** March 5th, 2026

**DOI:** https://doi.org/10.21203/rs.3.rs-8878999/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.

# Abstract

Rural Africa has increasingly fallen behind global progress for provision of safe drinking water by both percentage of population and aggregate numbers. Billions of US dollars have been invested, though a lack of accurate revenue data questions the financial sustainability of these decisions. The enduring paradox is why water supply infrastructure investments fail to generate revenue to ensure service sustainability. This study offers empirical insights based on 11,957 years of verified revenue data from 6,190 piped schemes and handpumps in 11 countries between 2021 and 2024. Findings report that handpumps generated a median annual revenue of USD26 per year with non-payment and outliers excluded. In 2024, 61% of handpumps provided no revenue. Piped water schemes generated a median revenue of USD0.75 per $m^3$ of water, with a median annual consumption per person of 0.82 $m^3$. Based on these data, a revenue prediction model was developed using machine learning to illustrate an objective approach to improve public finance investment decisions for project design, tariff setting, subsidy provision, and infrastructure choice. Pathways to overcome Africa's water revenue paradox are then defined.


# Introduction

The climate resilience, economic, human rights', and public health arguments for providing safe drinking water to everyone are well rehearsed (1–3). However, the financial case to operate and maintain drinking water services is often less certain. While global progress is encouraging in terms of building infrastructure access for drinking water benefiting 2.1 billion more people between 2000 and 2022 (4), maintaining services has proved more challenging. Global estimates suggest between 2.1 and 4.4 billion people lack safe drinking water in 2023 (4, 5). Progress in Africa has stalled with an estimated 600 million or 74% of the rural population without safely managed drinking water services (5). Between 2006 and 2025, the World Bank and the African Development Bank have approved projects with financial commitments of over USD6.8 billion to contribute to rural drinking water services in 27 African countries (SI). However, one in four rural waterpoints is believed not to be working in Africa at any one time (6). This illustrates the Africa water revenue paradox: why do major public investments in rural drinking water services fail to generate sufficient revenues to sustain reliable services for vulnerable people?

Data asymmetry is one reason for the paradox. Capital expenditure budgets for infrastructure can be determined and executed with reasonable precision and transparency. Planning assumptions for future revenue flows are often estimated using a predicted daily consumption level and a standard tariff. Responsibility to then manage services and sustain revenue outcomes is usually devolved to communities in Africa (7–9). Community-based management emerged in the late 1980s as a model for newly independent countries with limited resources, capacity or regulatory reach. Over recent decades, empirical studies have illustrated weak institutional capacity, poor revenue management, and limited demand for communities to manage their water infrastructure, given a choice (10–14).

Over the last decade, professional service providers have emerged in many African countries to improve accountability with higher service levels applying results-based contracts (15, 16). Results-based contracts are specified with metrics of reliability, revenue generation, water consumption, and water safety, which are independently verified by a third party before payments are made to a service provider (17). This dataset draws on verified records of annual payment records from 6,190 piped water schemes and handpumps in 11 countries under professional management serving an estimated 2.9 million people (Fig. 1). Data from 2021–2024 represent 11,957 years of data, equivalent to 143,484 months. The model overcomes the data asymmetry problem with standardized contracts including records of revenue collection.

Standardized revenue data offer multi-country insights to a second reason for the paradox: the assumption that full cost recovery for rural drinking water services is feasible. Projects are designed and approved with the view that future revenues will be sufficient for local communities to maintain the infrastructure without a subsidy. This is in contrast to urban piped infrastructure, which are managed by professional utilities with an estimated annual global subsidy range of USD289-353 billion (18). While urban utilities benefit from high population density settlements, a natural monopoly, relatively high-income consumers, and generally adequate regulation, rural communities live in remote areas with scattered populations, face higher unit costs of access and management, and serve low income consumers with limited ability to pay (19). Predictably, rural communities with variable training or capacity struggle to manage, maintain and fund infrastructure sustainably with high service standards. The policy question is then not whether a rural water subsidy is needed but how can governments use verified revenue data to make informed public finance choices to ensure the expected return on investments from capital expenditure results in affordable, equitable and sustainable services at scale.

# Results

To understand revenue behaviors, we first empirically quantify annual revenues by population and infrastructure type to illustrate patterns in observed revenues. We then develop a machine learning model to demonstrate how revenues are predictable, even in data sparse contexts, as a pathway to resolve the water revenue paradox. We consider annual revenue by population and by infrastructure type in three ways: a) all data, b) without outliers, and c) without outliers and zero values (Fig. 2, see Methods). These different perspectives are presented to illustrate the influence of zero payments and extreme values on central tendency, without imposing subjective data exclusions.

Revenue by population (USD/person/year) has an overall median value of USD0.023 per person with 75% of values below USD0.183. Excluding both outliers and zero values raises the median to USD0.096 per person per year with 75% of values below USD0.219.

Revenue per volume (USD/$m^3$/year) normalizes piped water revenues irrespective of piped scheme size or population served. It also provides a benchmark for comparison with tariffs for the supply of drinking water from urban water utilities, which commonly fall in the range of USD1.5 per $m^3$ for OECD countries,

with an upper band of around USD8 per $m^3$ in Denmark (20). Excluding outliers and zero values, we observe a median value of USD0.767 per $m^3$ with 75% of values below USD1.028. The median volumetric water consumption is 0.821 $m^3$ per person per year, or 2.25 liters per day.

Revenues per handpump (USD/handpump/year) are influenced by a high proportion of handpumps with no annual revenue. The revenue range without outliers has a median of USD 0.00. In 2024, 1,704 handpumps (61% of total) recorded zero revenue. When zero values and outliers are excluded, the median value is USD26.16 per handpump per year, with 75% of revenues below USD48.91.

Predicting Revenues

We next apply machine learning methods to demonstrate the predictability of revenues.

Two types of data are used in the model: infrastructure characteristics reported through results-based contracts and contextual factors such as rainfall and population referenced through geolocation from public datasets. Our model applies a two-stage process. A classifier first sorts infrastructure into three predicted annual revenue ranges, then secondary regressors generate specific values. The two-stage approach is selected for its ability to improve revenue predictions across heterogeneous services and infrastructure types.

Model performance evaluation finds a high value for the classifier (ROC-AUC = 0.9030) and stable $R^2$ estimates for regressors performance in each revenue range predicted by the classifier (0.61–0.64). Evaluation of feature importance finds that more data, not more variables, is most likely to further improve predictive performance (see Methods).

We further find that the model can predict revenues with broadly stable performance across years. We apply drift analysis to assess how underlying empirical data changes from year to year (as interpreted by F1-Macro and ROC-AUC estimates). Analysis finds minimal drift from year to year but detectable drift over the full period from 2021–2024. Drift is largely driven by changes in long-tail revenue distributions, with overall stable revenue patterns for lower values. Observed drift likely reflects changes in service coverage, infrastructure composition, and operational maturation rather than instability in household payment behavior. The implication is that prior year revenue data can be a reasonable basis for predicting next year revenues (Fig. 3).

To understand how the model can be applied to new contexts, we apply a leave-one-country-out (LOCO) process and assess changes to model accuracy. Data from an excluded country is progressively re-introduced to the model to calibrate predictions for that context. Analysis considers the number of data rows required to achieve 95% classifier model performance (F1-Macro) for each country. In our dataset, annual revenue records from an average of 200 infrastructure sites were typically sufficient to restore classifier performance to within 95% of the initial benchmark with variation depending on geography and infrastructure heterogeneity.

The Revenue Predictive Model demonstrates three things. The first is that revenues can be predicted using machine learning anchored in high-resolution historical data. Models can perform well with minimal drift from one year to the next due to broadly stable revenue patterns. The second is that predictive potential can rely on a limited set of input indicators. More data, rather than more indicators, is a sufficient basis for predicting revenues. The third is that the model can be calibrated for new countries or services with a minimum quantity of revenue records from that context. Taken together, findings suggest that rural drinking water revenues in Africa are predictable and that base models grounded in empirical data can be calibrated for new contexts with minimal input requirements. Revenue predictive modeling is now a feasible and applicable tool for forecasting drinking water revenues in rural Africa as a response to the revenue paradox.

# Discussion

User payments for drinking water constitute the fundamental base of any sustainable revenue strategy. Without longitudinal data on revenue collection, policy and investment choices on infrastructure type and location, tariff design, and project prioritization will continue to operate with uncertainty and risk, potentially wasting public resources, and leaving vulnerable populations no better off. The results of this analysis provide lessons and pathways to overcome Africa's water revenue paradox.

First, almost all handpumps will require a subsidy as revenues are consistently low in all locations, generally below USD50 per waterpoint per year, and compounded by three in five users not paying anything. Importantly, handpumps serve many of the most remote and vulnerable communities in Africa. Further, handpumps are often not used consistently throughout the year as our revenue data illustrate and long-term records of daily handpump demand have shown in Kenya and Mali (11, 13, 21). Handpumps may be characterized as the waterpoint of last resort. Communities may choose to use surface water or rainwater collection until they are no longer available, at which point handpumps become the critical source of drinking water for people and livestock. In addition, when the handpump tariff is considered unaffordable, low-income households may choose unimproved and potentially unsafe water sources (22).

Therefore, a revenue strategy to increase handpump tariffs may be effective during droughts and periods of low rainfall but may price gouge and exclude vulnerable groups at their time of greatest need. The moral dilemma buffers against financial logic. Tariff strategies that have shifted from a volumetric fee to a flat fee, in the few cases where measurement is available, have resulted in three times more revenue being generated, though still not sufficient to meet a break-even point (23). With an estimated 0.5–1.3 million handpumps installed across Africa (24), there is a vast sunk cost of investment running into many billions of US dollars. In the short-term, a subsidy targeting handpumps, which are regularly used or relied upon under extreme climate events, would seem a cost effective and socially inclusive policy. Given the subsidy can be less than USD1 per person per year (25), this would be a relatively modest investment if managed under results-based contracts with independent verification of usage patterns.

Second, piped water schemes offer a stronger revenue collection case than handpumps. The revenue per cubic meter is at the lower range of OECD countries which seems promising. However, the volume of water piped water use in rural schemes is often modest in comparison to urban contexts. Rural Africans don't enjoy flush toilets, showers, dishwashers, cars to wash, or lawns to water. Therefore, the volumetric usage can be low compared to Europeans or North Americans who can use between 130 to 400 liters per person per day (20). The economic calculus then makes low usage at low tariffs financially challenging. The case for a piped water subsidy applies across the data analyzed. No doubt, there can be specific cases where a more affluent and higher density community has demand to generate sufficient revenue to break even. However, policy design must take account of the general case rather than the exception. With independent data verification, a subsidy policy can be regularly reviewed and updated to reflect changing usage patterns and opportunities.

An ancillary point to this discussion is the goal to provide inclusive service delivery at scale. Given most governments have committed to providing all citizens with safe drinking water, policy cannot single out potentially profitable piped locations and ignore the underserved majority. If a robust and long-term subsidy policy provides the right incentives within a dynamic design, professional service providers will be able to create a business case to blend handpumps and piped water service areas. There is evidence of this emerging in Africa as the legacy of devolved and unregulated service models are being challenged with new thinking driven by standardized data systems with reliable financial incentives. The transition will be challenging given vested interests can benefit from large contracts to build and devolve water infrastructure. Maintaining infrastructure incurs lower costs with lower margins and increases long-term accountability through independent verification systems.

Third, we define three pathways to overcome the paradox. First, ensure traditional large capital expenditure projects have robust and clearly funded operational expenditure to maintain infrastructure for future decades. Second, expand the knowledge and practice of results-based funding with independent verification systems. Third, acknowledge the need for evidence-based subsidies as a cost-effective approach to deliver policy goals cost-effectively and equitably.

Historically, major investments by governments and development finance institutions have focused on a narrow timeline to build infrastructure with less attention or provision for future maintenance. Bold assumptions on likely drinking water demand and payment levels substantiate project approvals. In the absence of empirical data, this approach is understandable. The Revenue Predictive Model offers a tool to make more accountable decisions for more transparent financial planning linking capital and operational expenditure. With useful insights emerging from 200 schemes with a year of observed revenue data, the potential to apply the model to refine projections and consider alternative infrastructure portfolios will allow financial pragmatism to question political or ideological choices. Public finance management will be strengthened with governments having a realistic view on the future liability of large loans based previously on assumptions of no subsidy requirement. As noted, the likelihood is a long-term subsidy will be required but the extent and nature of the subsidy can be more transparently calibrated to promote more sustainable outcomes.

Results-based funding offers one approach to link capital and operational expenditure programming. The advantages are clear in terms of a) a standardized contract design with independent verification, b) cumulative high quality generation of data to iteratively improve revenue predictions and assess program performance, c) prudent use of public finance which is disbursed based on verified results, and d) the incubation of national service providers with incentives to build local businesses and employment in the place of dependency on international organizations and increasingly unreliable and decreasing international development aid. Results-based funding also offers regulation by contract which provides an important first step to inform more comprehensive regulatory systems after institutions mature, and capacity is strengthened.

Finally, evidence-based subsidies suggest the operational costs can be affordable. While the current results-based subsidy across 16 countries is less than USD1 per person per year, this disguises sunk costs in the establishment of existing enterprises, investment in assets and supply chains, and a well-trained workforce. The positive news is most African countries have vibrant entrepreneurial sectors which could transition into this work under appropriate contracting conditions. However, contracting with governments in Africa is a risky endeavor due to weak contract enforcement, power imbalances and fiscal fragility. These important limitations can be addressed but are non-trivial issues in incubating a new market of water enterprises. In contrast, asset requirements are relatively modest and largely available in national markets other than water safety monitoring. Most African countries have limited infrastructure, variable professional expertise, and rely on expensive international supply chains for water safety monitoring from laboratory equipment to field testing consumables. Progress can still be achieved in a structured approach with pragmatic guidelines when governments are committed to achieving and sustaining results (26).

Africa’s rural water revenue paradox is challenging but resolvable. A failure to maintain drinking water services over time has led to unsatisfactory progress from over USD6.8 billion dollars of development finance commitments. Kenya is one of four of the 27 countries which has been allocated 46% of these commitments, however the 2019 census reported 23% of its citizens used surface water as the main drinking water source (27). Infrastructure is built and then quickly fails as projects close. It is no surprise that rural water uses choose not to pay for unreliable and unsafe sources. We argue that governments and donors need to recognize and be accountable for efficient and equitable subsidies for rural water services. Professional service providers can support this transition at a reasonable cost. The average annual commitment in development projects between 2006 and 2025 is USD342 million per year (SI), which would translate into a healthy annual subsidy to keep drinking water infrastructure functioning. As crude a comparison as this might seem, it reflects that the financial gap for sustainable services may not be as great if funds were efficiently redirected to verifiable performance outcomes. Overcoming Africa’s water revenue paradox will depend on increasing the flow of high-quality revenue data so financial decisions can be informed by verifiable evidence to build adaptive institutional systems that are incentivized to deliver reliable and safe drinking water to everyone sustainably.

# Declarations

# Competing interests

R.H. and D.M. are Directors of Uptime Global; R.H. is a Trustee of the Uptime Catalyst Facility.

# Author contributions

Conceptualization: R.H. and D.M. Methodology: D.M. and E.A. Modeling and Analysis: D.M. and E.A. Resources: R.H. and D.M. Writing: R.H., D.M., and E.A. Visualization: D.M., E.A.. Funding acquisition: R.H. and D.M. Project administration: D.M.

# Acknowledgements

We acknowledge support in data collation and verification from A. Armstrong, F. Trethewey, K. Nilsson, M. Naughton, collection of public project finance data from S. Furey and N. Gladstone, and graphic design from S. Ferguson. Data has been provided under results-based contracts to the Uptime Catalyst Facility by country: Burkina Faso (UDUMA), Central African Republic (Water for Good), Ghana (Safe Water Network, Project Maji, Water4), Kenya (FundiFix, Water Mission, Project Maji), Madagascar (Sandandrano), Malawi (Fisherman's Rest, Pump Aid, Self Help Africa, Water Mission), Mali (UDUMA), Sierra Leone (InterAide, Water4), Tanzania (Water Mission), Uganda (International Lifeline Fund, Ugandan Water Project, Water Compass, Water Mission, Whave Solutions) and Zambia (Water4). Funding: The results-based contracts were managed by the Uptime Catalyst Facility and funded by Aqua for All, Conrad N. Hilton Foundation, Osprey Foundation, Vitol Foundation, VO$^{x}$ Impuls Foundation, and the Waterloo Foundation.

# Data availability

The data supporting the findings of this study are available in Dryad at DOI: 10.5061/dryad.7pvmcvf7r

# Code availability

The code supporting the findings of this study is available in Zenodo at DOI: 10.5281/zenodo.18431343

# Methods

## Data sources

Development finance data from the African Development Bank (AfDB) and the World Bank has been collected from public sites with web links to wider documentation (see SI). The data represent values of financial commitments in USD with conversion from an IMF currency basket, called Special Drawing

Rights (SDR), by a standard annual calculation for equivalence. Projects were selected based on a) financial commitments made between 2006 and 2025, b) a project commitment of over USD10m, and c) a specific focus on rural drinking water services. Of the 67 projects included, 15 were funded by the World Bank and 52 by AfDB across 27 countries. Equivalent data from bilateral donors is not available in a standardized format.

Revenue data are drawn from reporting by professional service providers to the Uptime Catalyst Facility[1], a UK charity, under results-based contracts between 2021-2024 (Table 1). Service Providers were contractually required to report accurate quarterly revenue data for each “waterpoint” they maintained, defined as a handpump or piped scheme, in exchange for quarterly performance-based grant disbursements. Piped water services in Sierra Leone participated in the same data reporting standards but did not receive grant payments. Population and water volume data were also reported for each waterpoint. Water volumes were either estimated or measured. Only measured volumes were considered valid and included in volumetric analysis of revenue per cubic meter. The resulting dataset consists of individual waterpoint records for consecutive 12-month periods with associated data on infrastructure characteristics, location, population served and annual revenue.

**Table 1. Distribution of waterpoints by country**

| Country | Handpumps | Piped Schemes | Total |
|---|---|---|---|
| Central African Republic | 2191 | 39 | 2230 |
| Burkina Faso | 0 | 61 | 61 |
| Ghana | 83 | 267 | 350 |
| Kenya | 103 | 57 | 160 |
| Madagascar | 0 | 13 | 13 |
| Malawi | 600 | 26 | 626 |
| Mali | 378 | 139 | 517 |
| Sierra Leone | 502 | 35 | 537 |
| Tanzania | 0 | 25 | 25 |
| Uganda | 1415 | 23 | 1438 |
| Zambia | 149 | 84 | 233 |
| **Total** | **5421** | **769** | **6190** |

Service providers are subject to a data integrity process as part of the results-based contracting process to confirm overall data quality with a sample of individual revenue data points directly verified (*17*). It is possible that specific outliers are due to reporting errors. Population data are particularly hard to verify

due to uncertainty about how many people use the water collected and return to homes for later consumption. Heuristics may be used such as an estimated number of people in a service area or the number of households billed for water multiplied by an average household size. Neither is accurate nor can be verified on a daily or annual basis. Instead of subjective data cleaning by omitting certain values, all dataset values are reported in revenue range analysis and presented next to ranges with outliers removed.

Selection bias of service providers is a limitation. Professional service providers under results-based contracts only serve specific geographies, each with unique service models. Service providers represented in this dataset are largely international organizations based in the EU or USA that can subsidize operations, including management overheads. Only two professional service providers are entirely African-based and led. These professional service providers typically cover large service areas, ensure high waterpoint reliability by minimizing or quickly repairing breakdowns, and have developed business models with revenues from drinking water users. These service models are not representative of the wider practice of community-based management whereby local communities self-organize maintenance and payments without any requirement for standardized reporting and data management. As such, this analysis is shaped by data availability, which is confined to the specific professional service providers that have been able to consistently report high quality data through participation in results-based contracts. We do not claim representativity of the broader African context though, to our knowledge, this is the most comprehensive longitudinal data generated on this topic.

## Revenue analysis and predictive model

Quarterly revenue data were aggregated into full years for revenue analysis to reduce seasonal variation. Only sites with four consecutive quarters of revenue data were considered in analysis and model development. For the presentation of revenue ranges (Figure 2), outliers were identified using cutoff thresholds calculated from distribution quartiles: lower bound = Q1 - 1.5 * IQR; upper bound = Q3 + 1.5 * IQR.

We apply a two-stage predictive modeling framework to forecast waterpoint revenue using operational, geospatial, environmental, and neighborhood-population characteristics (Table 2). The overall workflow consists of four major components: data preprocessing, variable reduction, Stage 1 classification, and Stage 2 class-specific regression (28, 29). The selected method follows a two-stage learning paradigm: a first-stage classifier assigns each water point to a revenue segment, and class-specific XGBoost regressors are then trained within each segment. This segmented, local-expert design is closely related to mixture-of-experts and two-stage learning frameworks, where the input space is partitioned and specialised predictive models are trained for each region or cluster.

**Table 2 - Summary of predictive model input variables**

| Infrastructure characteristics | Contextual factors |
|---|---|
| *Input variables (results-based contract reporting)* | *Determined using infrastructure geolocation* |
| • Geolocation<br>• Infrastructure type: handpump or piped<br>• Number of tap stands, kiosks and household connections (piped only)<br>• Volumetric measurement type: estimated, manual or automated<br>• Estimated population served<br>• Actual revenue received from user payments (value for model training and prediction) | • Population within 500m<br>• Number of other waterpoints within 1, 5, 10 and 20km (from Infrastructure Characteristics input variables)<br>• Monthly rainfall data |

First, the dataset is cleaned by removing identifier and leakage-prone columns, handling missing values, and encoding categorical features. Missing values were handled using median imputation for numerical variables and a constant “missing” category for categorical fields (30), implemented within a unified preprocessing pipeline to ensure consistency during training and deployment. Numerical variables were standardized to promote model stability and prevent scale-sensitive bias; categorical variables were transformed through encoding (31). Multicollinearity was addressed using an iterative Variance Inflation Factor (VIF) elimination procedure, which removes redundant predictors and yields a stable, interpretable, and non-redundant feature set (32). All numerical predictors were scaled using z-score normalization to ensure that large-magnitude variables (such as population and rainfall metrics) did not dominate the learning process and to promote numerical stability across cross-validation folds. Finally, the entire preprocessing workflow was implemented through scikit-learn Pipelines and ColumnTransformers to guarantee consistent, reproducible, and leakage-free application of all transformation steps during both training and inference, while maintaining modularity and preventing human error in feature preparation.

Next, the continuous revenue variable was discretized into three ordered classes (USD0–5, USD5–500, and >USD500). These thresholds were selected based on the empirical distribution of the data, which exhibits a strong positive skew with clear operational regimes: a large group of near-zero values, a broad mid-range, and a long tail of high-revenue instances. Transforming the revenue into these three interpretable segments reduces heteroscedasticity, stabilizes variance, and enables the use of a hierarchical two-stage architecture. In Stage 1, a Decision Tree classifier was chosen because it can naturally capture non-linear decision boundaries, handle mixed feature types, and produce transparent, interpretable segmentation rules (33). The model was tuned using cross-validation and was optimized using Optuna, where an initial 80/20 train–test split preserves an independent evaluation set. The 80% training subset is internally partitioned into k-folds to obtain robust, unbiased hyperparameter estimates. For Stage 2, class-specific XGBoost regressors were optimized using Optuna, a Bayesian hyperparameter optimization framework that efficiently explores the high-dimensional parameter space,

prunes sub-optimal trials, and identifies class-tailored configurations. This two-stage structure allows each model to specialize in the statistical behavior of its revenue segment, improving predictive accuracy and stability across the highly skewed distribution. The revenue values for the highest-revenue class were treated with winsorization, where the top 1% of extreme values were clipped to reduce the influence of outliers (34). This was applied because revenue distributions in Class 2 exhibited heavy right-tail behavior, and winsorization stabilizes model training by preventing a small number of anomalously large values from dominating the regression fit while still preserving the overall distributional structure.

Finally, predicted discrete class ranges from Stage 1 route each sample to the appropriate class-specific regressor in Stage 2, producing a continuous revenue estimate. Model performance is evaluated using F1-score, Accuracy ROC-AUC, and $R^2$, RMSE, and MAE (Table 3 ). The classifier model ROC-AUC score of 0.9030 falls in the “outstanding discrimination range” (35, 36). Regressor models are trained and evaluated for Class 1 (USD5-500) and Class 2 (>USD500). The utility of a regressor model for Class 0 (USD0-5) is negligible and therefore not developed.

**Table 3 - Classifier and regressor model performance metrics**

| Model | Metric | Train | Test |
|---|---|---|---|
| **Classification (Classes 0–1–2)** | Accuracy | 0.79 | 0.76 |
| | Macro F1 | 0.8125 | 0.7866 |
| | ROC-AUC | 0.9289 | 0.9030 |
| **Regression — Class 1** | $R^2$ | 0.6914 | 0.6103 |
| | RMSE | 43.86 | 50.33 |
| | MAE | 23.79 | 27.27 |
| **Regression — Class 2** | $R^2$ | 0.7989 | 0.6360 |
| | RMSE | 1951.42 | 3284.93 |
| | MAE | 1073.25 | 1715.04 |

Model stress testing assessed the relative value of each variable with the goal of minimizing the number of required inputs without compromising model performance. Variable contributions were evaluated through Permutation Importance, SHAP values, tree-based feature importance and model-based feature selection (SelectFromModel, scikit-learn). Across methods, the most influential variables were population, nearby waterpoints and rainfall. Learning curves show continued gains in validation macro-F1 as sample size increases, while feature-importance analyses concentrate predictive signal in a small

subset of variables. Together, this indicates a data-limited regime in which model performance benefits most from more data, not more variables.

Drift analysis employed a model-centric approach, to evaluate whether the predictive relationships learned by the model remained stable across different temporal segments of the dataset (37). Dataset drift was assessed using Evidently AI's Data Drift Report, employing distribution-level tests such as Kolmogorov–Smirnov, Jensen–Shannon divergence, and Wasserstein distance. Across inter-year comparisons, the share of drifted features ranged from 19% to 40%, with 4–8 out of 20–21 variables exhibiting statistically significant changes.

Analysis of the target variable wp_revenue_usd across consecutive years (2021→2022 40% drift; 2022→2023, 20% drift; 2023→2024, 19% drift) and the long-term period (2021→2024, 40% drift) shows consistent and statistically significant drift based on median shifts and Kolmogorov–Smirnov (KS) tests. This pattern indicates that the distribution of revenue has progressively changed over time, reflecting structural shifts in the underlying environment. These changes are expected in a multi-country, multi-system dataset where revenue patterns evolve due to differences in annual data composition, economic conditions, operational performance, and reporting quality.

The presence of drift does not invalidate historical data but signals that earlier years (particularly 2021) are less representative of the most recent revenue dynamics. As a result, relying heavily on older data can introduce noise and reduce predictive accuracy for future years. Instead, the drift findings support the use of a rolling training window, prioritizing the most recent 2–3 years of data where distributions are more aligned. This approach captures current revenue behavior more effectively while still providing enough data volume for model stability. Overall, the detected drift underscores the need for regular model retraining, to ensure that predictions remain calibrated to the most recent operational reality.

The Leave-One-Country-Out (LOCO) calibration experiment quantified how many in-country samples are required to adapt a model trained on other countries to a previously unseen context. Calibration analysis identifies a characteristic saturation pattern: adding the first 150–200 country-specific rows produces the steepest improvement in macro-F1, after which the marginal gains diminish. Using the "maximum curvature" elbow criterion, the optimal calibration size was approximately 200 rows.

This finding is consistent with the country-level analysis in Table 4 . Most countries require between 150 and 400 rows to recover ~95% of the achievable performance gain (F1-Macro). High-data countries with strong baseline heterogeneity (e.g., Malawi, Uganda, CAR) demand larger calibration sets, whereas countries with simpler or more separable class structures (e.g., Tanzania, Kenya) require substantially fewer samples. Two extreme cases illustrate this variance: Madagascar reaches perfect performance with zero additional data because the existing cross-country representation already spans its feature distribution; Mali benefits almost immediately from <10 samples due to its unusually strong class separation.

Taken together, the LOCO experiment demonstrates that a practical deployment scenario would require approximately 200 calibration samples in a new country to achieve stable performance, with diminishing improvement beyond that point.

**Table 4 - Country-level leave one country out (LOCO) analysis**

| Country | n_rows | zero_shot_f1 | best_f1 | k_needed_95pct_gain |
|---|---|---|---|---|
| Burkina Faso | 177 | 0.642383 | 0.653456 | 50 |
| Central African Republic | 4556 | 0.310267 | 0.546974 | 300 |
| Ghana | 801 | 0.283414 | 0.284175 | 500 |
| Kenya | 463 | 0.508167 | 0.757153 | 150 |
| Madagascar | 24 | 1.000 | 1.000 | 0 |
| Malawi | 1151 | 0.333333 | 0.613832 | 400 |
| Mali | 856 | 0.296 | 0.969489 | 10 |
| Sierra Leone | 783 | 0.115044 | 0.532851 | 200 |
| Tanzania | 51 | 0.556522 | 0.858333 | 25 |
| Uganda | 2687 | 0.043062 | 0.556738 | 400 |
| Zambia | 406 | 0.209486 | 0.503898 | 200 |

Limitations

The modelling framework relies on observational data collected across heterogeneous national contexts, and it is likely that important socio-economic, governance, or infrastructural variables remain unmeasured. The model captures correlational structure in observed revenues rather than causal mechanisms. These latent factors may influence revenue outcomes, meaning the model captures correlations rather than causal mechanisms. Because the model reflects observed payment behaviour, it may also under-represent latent demand or unmet need among populations unable to pay, reinforcing the importance of pairing predictive tools with explicit equity objectives. As a result, residual confounding may persist, limiting the causal interpretability and policy-action validity of the predictions.

The Leave-One-Country-Out calibration analysis assumes that a small, stratified subsample from the held-out country adequately represents its revenue distribution and operational characteristics. This assumption may not hold in countries where data is sparse, heavily skewed, or operationally atypical. In such cases, the calibration curves may underestimate the amount of local data required for stable performance, leading to optimistic generalization estimates.

The use of Variance Inflation Factor–based iterative feature elimination and winsorization introduces assumptions regarding multicollinearity and outlier treatment. While these methods improve numerical stability and model interpretability, they may also remove rare but potentially informative patterns that contribute to the predictive signal. Consequently, the feature space may be constrained in ways that limit the model's ability to capture complex interactions.

The proposed two-stage architecture—consisting of a Decision Tree classifier for revenue-band prediction followed by class-specific XGBoost regressors—provides interpretability and strong performance on heterogeneous tabular data. However, this structure imposes several constraints. First, the hierarchical design can propagate classification errors into the regression stage, meaning misclassified revenue bands may cap the achievable accuracy of subsequent regressors. Second, while tree-based models effectively capture local nonlinear relationships, they are less suited for representing smooth functional transitions or subtle interactions across geospatial, environmental, and neighbourhood-population features. Third, the class-specific regressors learn independently from each subset of the data, which may limit their ability to exploit shared structure across revenue regimes.

The learning-curve analysis indicates that model performance continues to improve as additional data is added, without reaching saturation. This suggests that the system operates in a data-limited regime. As a result, predictive accuracy may degrade when the model is deployed in settings that differ substantially from the training distribution, especially in locations with sparse historical coverage or rapidly evolving operational conditions.

# References


1. Wolf J, Johnston RB, Ambelu A, Arnold BF, Bain R, Brauer M, Brown J, Caruso BA, Clasen T, Colford JM, Mills JE, Evans B, Freeman MC, Gordon B, Kang G, Lanata CF, Medlicott KO, Prüss-Ustün A, Troeger C, Boisson S, 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. Lancet 401:2060–2071
2. World Bank, High and Dry: Climate Change, Water, and the Economy (2016) ; https://doi.org/10.1596/23665
3. White GF, Bradley DJ, White AW (1972) Drawers of Water: Domestic Water Use in East Africa. University of Chicago Press, Chicago
4. WHO, UNICEF (2025) Progress on household drinking water, sanitation and hygiene 2000–2024: special focus on inequalities (978-92-4-011514–9
5. Greenwood EE, Lauber T, Van Den Hoogen J, Donmez A, Bain RES, Johnston R, Crowther TW, Julian TR (2024) Mapping safe drinking water use in low- and middle-income countries. Science 385:784–790
6. Foster T, Furey S, Banks B, Willetts J (2019) Functionality of handpump water supplies: a review of data from sub-Saharan Africa and the Asia-Pacific region. Int J Water Resour Dev 36:855–869

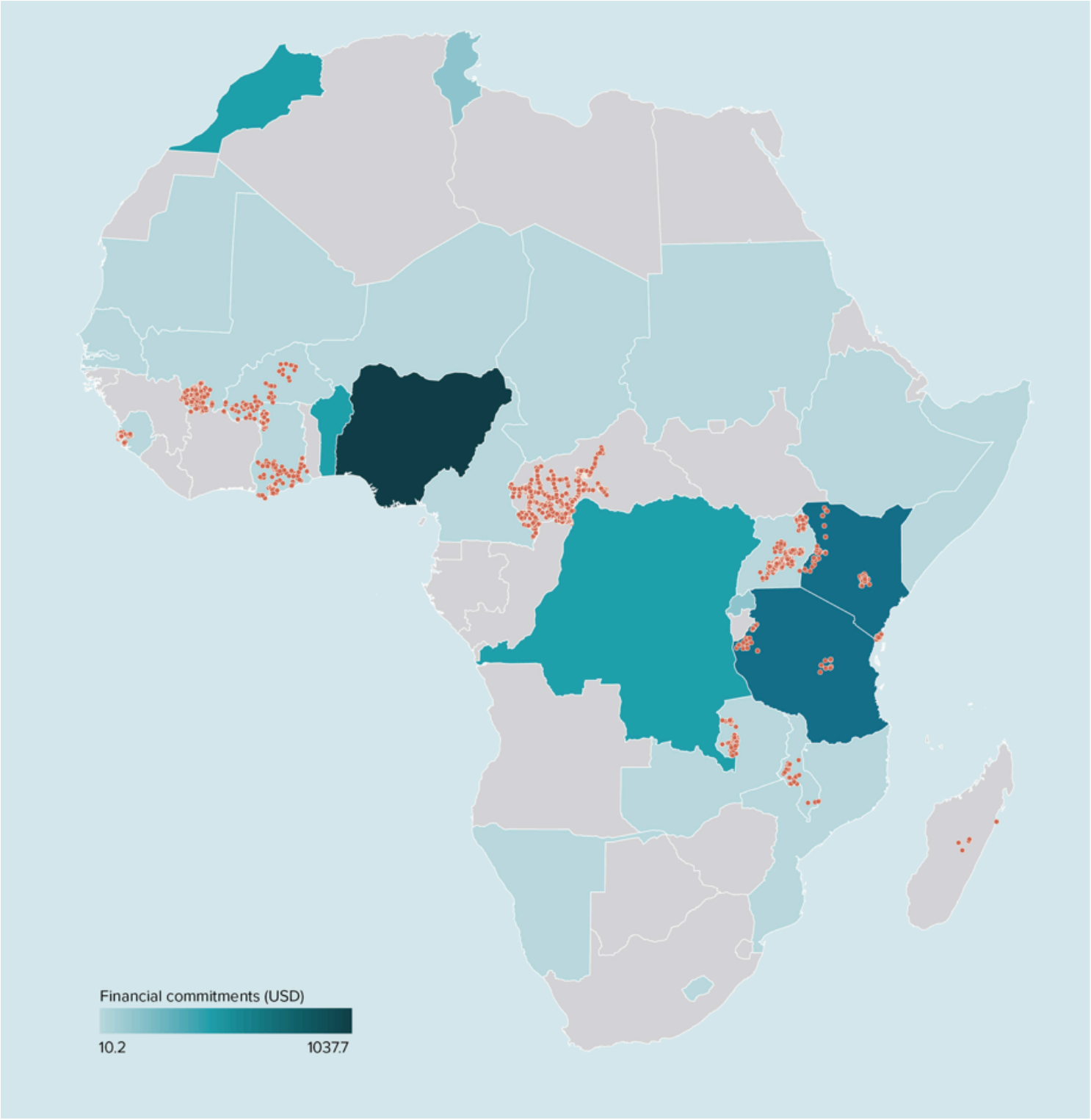
7. Carter RC (2021) Rural Community Water Supply: Sustainable Services for All. Practical Action Publishing, Rugby, UK
8. Harvey PA, Reed RA (2007) Community-managed water supplies in Africa: sustainable or dispensable? Community Dev J 42:365–378
9. Therkildsen O (1988) Watering White Elephants? Lessons from Donor-Funded Planning and Implementation of Rural Water Supplies in Tanzania. Scandinavian Institute of African Studies, Uppsala
10. Hope R, Ballon P (2021) Individual choices and universal rights for drinking water in rural Africa. Proc. Natl. Acad. Sci. 118, e2105953118
11. Foster T, Hope R (2016) A multi-decadal and social-ecological systems analysis of community waterpoint payment behaviors in rural Kenya. J Rural Stud 47:85–96
12. Foster T (2013) Predictors of sustainability for community-managed handpumps in sub-Saharan Africa: Evidence from Liberia, Sierra Leone, and Uganda. Environ Sci Technol 47:12037–12046
13. Wagner J, Merner S, Innocenti S, Geling A, Hope R (2025) Can solar water kiosks generate sustainable revenue streams for rural water services? World Dev 185:106787
14. Armstrong A, Dyer E, Koehler J, Hope R (2022) Intra-seasonal rainfall and piped water revenue variability in rural Africa. Glob Environ Change 76:102592
15. McNicholl D et al (2019) Performance-based funding for reliable rural water services in Africa. Uptime Global and University of Oxford, Oxford, UK
16. Nilsson K et al (2021) Global prospects to deliver safe drinking water services for 100 million rural people by 2030. (University of Oxford and RWSN, Oxford, UK
17. Armstrong A et al (2025) Uptime Global, Working Paper,. The Data Dividend: Strengthening data for results-based funding
18. Andres L, Thibert M, Cordoba CL, Danilenko AV, Joseph G, Borja-Vega C (2019) Doing More with Less: Smarter Subsidies for Water Supply and Sanitation. World Bank, Washington, DC
19. Hope R, Thomson P, Koehler J, Foster T (2020) Rethinking the economics of rural water in Africa. Oxf Rev Econ Policy 36:171–190
20. EurEau (2020) Europe's water in figures: An overview of Europe's water and wastewater sectors (DOI: 978-2-9602226-3–0
21. Hoque S, Hope R (2025) The Water Diaries: Living with the Global Water Crisis in Bangladesh and Kenya. Cambridge University Press, Cambridge
22. Foster R, Hope R (2017) Evaluating waterpoint sustainability and access implications of revenue collection approaches in rural Kenya. Water Resour Res 53
23. Wagner J (2024) Is volumetric pricing for drinking water an effective revenue strategy in rural Mali? npj Clean Water 7
24. Danert K, Furey S (2022) Stop The Rot - Call to action (Rural Water Supply Network (RWSN), St. Gallen

25. Garrick D, Gilson G, Hope R, Wagner J, Epstein G, Okoth EO, Silas F (2025) We must account for the results of water governance to deliver the SDGs and beyond. Environ Res Water 1:013001
26. Charles K et al (2023) Results-based funding for safe drinking water services. University of Oxford University and Uptime Global
27. Kenya National Bureau of Statistics (KNBS) (2019) Kenya Population and Housing Census 2019 – Micro-data on main source of drinking water. Government of Kenya
28. Haider P (2013) Prediction with Mixture Models, thesis, University of Potsdam, Potsdam, Germany
29. Jacobs RA, Jordan MI, Nowlan SJ, Hinton GE (1991) Adaptive Mixtures of Local Experts. Neural Comput 3:79–87
30. Alam S, Ayub MS, Arora S, Khan MA (2023) An investigation of the imputation techniques for missing values in ordinal data enhancing clustering and classification analysis validity. Decis Anal J 9:100341
31. Bolikulov F, Nasimov R, Rashidov A, Akhmedov F, Cho Y-I (2024) Effective Methods of Categorical Data Encoding for Artificial Intelligence Algorithms. Mathematics 12:2553
32. Kim JH (2019) Multicollinearity and misleading statistical results. Korean J Anesthesiol 72:558–569
33. Mienye ID, Jere N (2024) A Survey of Decision Trees: Concepts, Algorithms, and Applications. IEEE Access 12:86716–86727
34. Sharma S, Chatterjee S (2021) Winsorization for Robust Bayesian Neural Networks. Entropy 23:1546
35. Bradley AP (1997) The use of the area under the ROC curve in the evaluation of machine learning algorithms. Pattern Recognit 30:1145–1159
36. Hosmer DW, Lemeshow S, Sturdivant RX (2013) Applied Logistic RegressionWiley, Hoboken, N.J, 3. Aufl., Wiley Series in Probability and Statistics
37. Lewis GA, Echeverría S, Pons L, Chrabaszcz J (2022) Augur: A Step Towards Realistic Drift Detection in Production ML Systems (ACM, Virtual / Online, pp. 37–44

## Footnotes

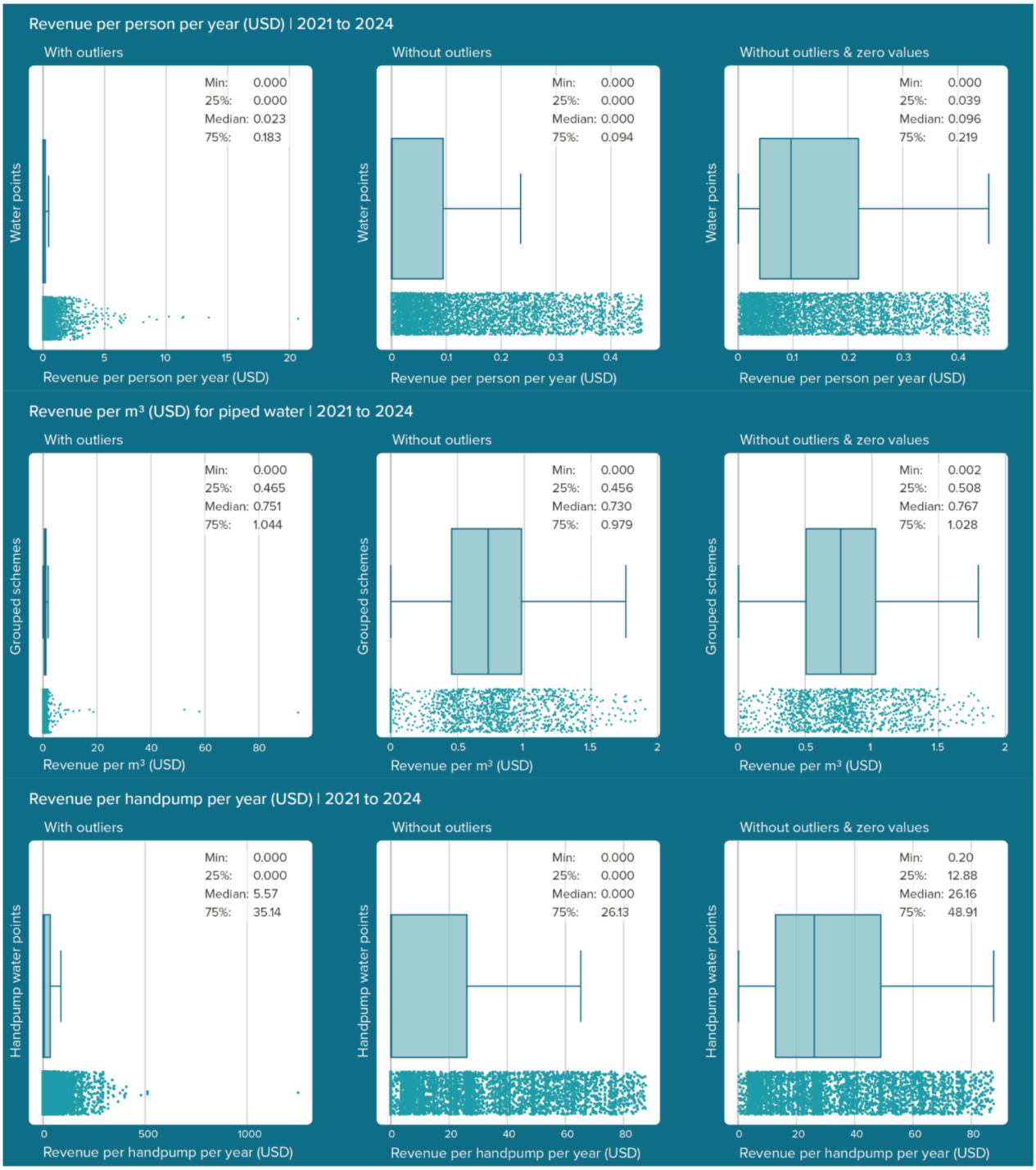
1. https://www.uptimecatalyst.org/

## Figures




## Figure 1

**Geographic distribution of water supply infrastructure locations overlaid with World Bank and African Development Bank financial commitments, 2006-2025**




## Figure 2

## Annual revenue distributions by population and waterpoint type, 2021-2024

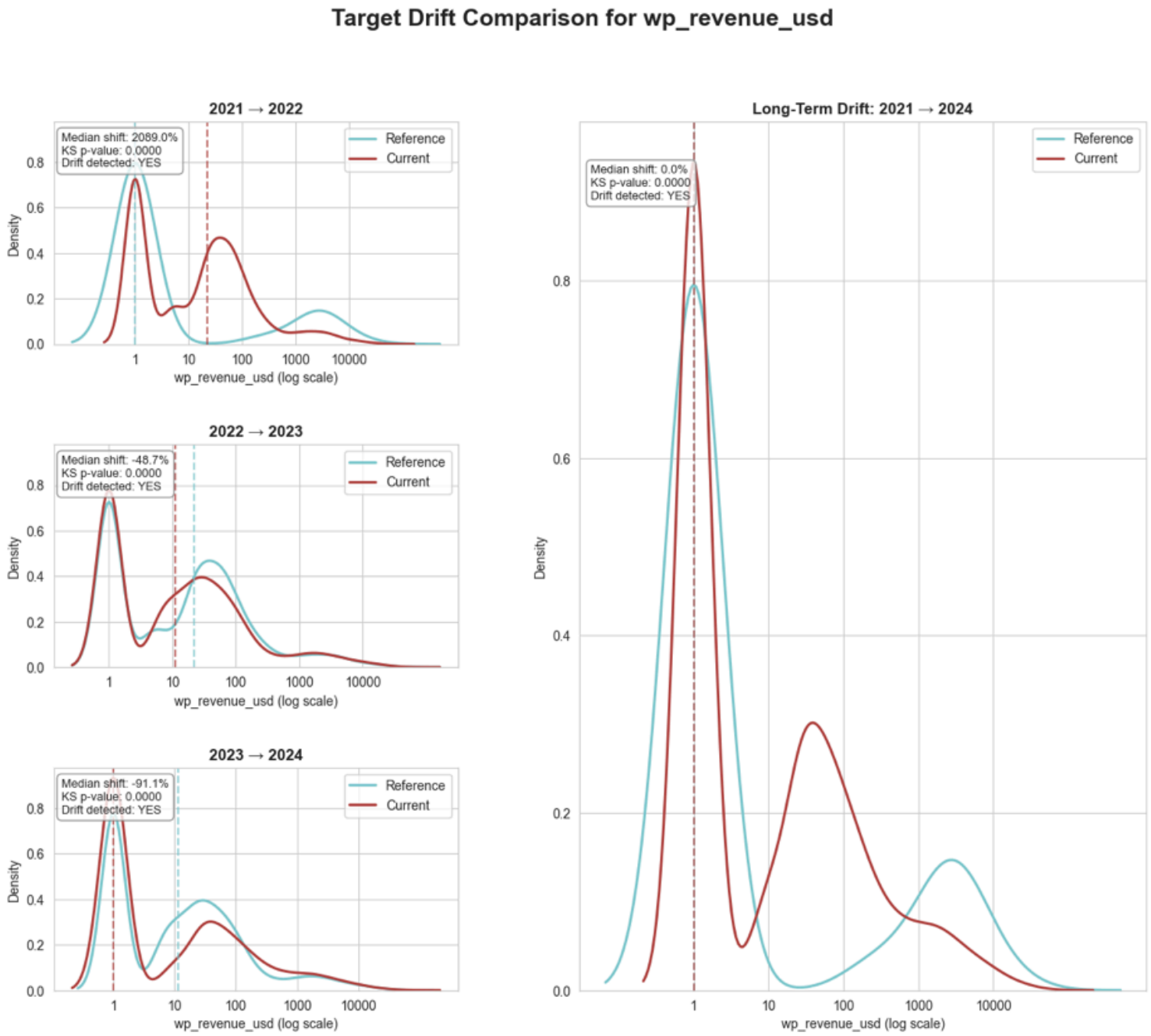



**Figure 3**

**Revenue drift analysis, 2021-2024**

# Supplementary Files

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

- nrreportingsummaryRevenueParadox260213.pdf