A Critical Review of Household Water Datasets

Justus Breyer
Chair of Communication and Distributed Systems
RWTH Aachen University
Aachen, Germany
breyer@comsys.rwth-aachen.de

Maximilian Petri
Chair of Communication and Distributed Systems
RWTH Aachen University
Aachen, Germany
maximilian.petri@rwth-aachen.de

Muhammad Hamad Alizai
SBA School of Science and Engineering
LUMS
Lahore, Pakistan
hamad.alizai@lums.edu.pk

Klaus Wehrle
Chair of Communication and Distributed Systems
RWTH Aachen University
Aachen, Germany
wehrle@comsys.rwth-aachen.de

Abstract

The increasing prevalence of water scarcity has spurred a surge in research efforts globally. To test and validate new concepts and ideas, readily accessible data is crucial for thorough evaluation. In this study, we examine recently published household water datasets and position them within the context of a prior review, revisiting trends theorized four years ago. Furthermore, we assess the impact these datasets have had on the research community and observe that, among more than 400 related publications, none involve an external research group reusing the published data. To provide guidance for future data collection efforts, we identify five factors that may influence external research groups' decisions on whether to utilize published data in their studies.

CCS Concepts

• **General and reference** → **Surveys and overviews**; • **Human-centered computing** → Ubiquitous and mobile computing; Accessibility.

Keywords

urban water consumption, water demand, water data accessibility, data resolution

ACM Reference Format:
Justus Breyer, Maximilian Petri, Muhammad Hamad Alizai, and Klaus Wehrle. 2024. A Critical Review of Household Water Datasets. In *The 11th ACM International Conference on Systems for Energy-Efficient Buildings, Cities, and Transportation (BUILDSYS '24), November 7–8, 2024, Hangzhou, China.* ACM, New York, NY, USA, 5 pages. https://doi.org/10.1145/3671127.3698793

1 Introduction

Climate change has gradually emerged as one of the most significant challenges facing humanity in recent decades. With rising global temperatures and increasingly extreme weather phenomena, such as droughts, water is becoming an increasingly valuable resource [7, 31]. Consequently, the conservation of water and the enhancement of its efficient use have become central topics for a growing research community [1]. A key area for improvement is water consumption in residential households, which are among the largest consumers of potable water [16]. Given the large number of affected consumers, even minor changes in household water use can lead to substantial reductions in overall water consumption.

To understand, analyze, and test new methodologies related to water consumption, a substantial amount of reliable data is essential. The process of collecting such data is often labor-intensive, requiring specialized expertise across various stages, from sensor setup to post-processing. To reduce the burden on other researchers and to facilitate comparability and reproducibility, there has been a growing trend of publishing datasets collected as part of research studies [8]. However, unlike in the related field of household electricity consumption—where datasets and methods like Non-Intrusive Load Monitoring (NILM) [11] have gained widespread acceptance [26]—the field of household water consumption research has seen only limited adoption of published datasets [8]. As a result, most research on household water consumption is conducted with little comparability.

This review aims to identify the challenges that currently hinder the advancement of interconnection and comparability in the field of household water consumption research. To achieve this, we make the following contributions:

- **Comprehensive dataset overview:** We provide the most up-to-date and comprehensive overview of household water datasets by integrating 18 newly published datasets from multiple regions into an existing collection and analysis of 31 datasets [8], uncovering increased interest of regions where water scarcity is becoming an issue. This extension offers a broader understanding of global water consumption trends at the household level.
- **Identification of key barriers to data reuse:** Through an analysis of over 400 related publications referencing the datasets, we uncover that the published data sees no reuse. By investigating the accessible datasets themselves, we identify five critical factors—data quality, type, transferability, duration and documentation—that prevent the reuse of household water datasets by external researchers. These insights

BUILDSYS '24, November 7–8, 2024, Hangzhou, China

ACM ISBN 979-8-4007-0706-3/24/11
https://doi.org/10.1145/3671127.3698793

highlight actionable areas for improving dataset utility and adoption in future research efforts.
- **Recommendations for future dataset creation:** We propose recommendations to improve the design and sharing of household water datasets. By addressing issues such as dataset size, temporal resolution, and transferability, our work offers practical guidelines to enhance dataset reusability and foster cross-validation, ultimately promoting a more interconnected research community.

2 Background and Motivation

Research on water consumption has garnered increased attention in recent years, largely due to significant advancements in metering technology [8], which have made it feasible to collect data for subsequent analysis. Consequently, there have been efforts to consolidate insights and publications related to this research domain. Previous reviews, such as those by Abu-Bakar et al. [1], have primarily focused on methodological advancements in sensor selection, technology, placement, and subsequent data processing.

There have also been discussions on the availability and quality of datasets, which are closely related to our work. Notably, di Mauro et al. [8] provided a comprehensive overview of the datasets cited and used in the literature until 2020, introducing several categories and scales for comparison. Importantly, they introduced the spatial scales of data resolution (*city, district, household* and *end-use*), finding that out of 92 datasets reviewed in total, 31 were on the household scale. They identified trends such as an increase in the number of datasets published each year and a rise in open-access datasets. Mazzoni et al. [17] examined cross-regional similarities in water consumption by aggregating and comparing findings from existing publications. While they also analyzed related datasets using scales similar to those of di Mauro et al., their focus was on the findings reported by the publications rather than on providing a comprehensive list or quality assessment of the datasets.

Van Tuyl and Whitmire [32] assessed datasets based on four main criteria—Discoverability, Accessibility, Transparency, and Actionability—and found that nearly all examined publications fell significantly short. However, their analysis was limited to publications funded by the NSF at Oregon State University.

Both di Mauro et al. and Van Tuyl and Whitmire highlighted several issues regarding the accessibility of published data. The objective of our contribution is to determine whether these concerns have been acknowledged by the research community and to what extent the current situation has improved. Furthermore, we examine datasets with open-access policies to identify potential issues with their utility, as cross-validation (e.g., in [20]) has received limited attention in the field to date.

3 Methodology

To provide an updated perspective on di Mauro et al.'s comprehensive dataset review [8], we adopted a similar research methodology. Using databases such as *Google Scholar*, *Mendeley*, *Mendeley Data*, and *data.world*, we searched for keywords including "Household water", "water consumption", "household water consumption", "water demand", "water meters", "smart water meters", and "household water meters". Consistent with our focus, and as indicated by our keyword choices, we excluded datasets at the *district scale* [8], opting only for those with granularity at the building level. This is because research on district-level water consumption typically concentrates on the supplier rather than the consumer.

We compiled a list of the identified datasets and excluded all those already mentioned by di Mauro et al. Subsequently, we cross-referenced our list with Mazzoni et al.'s analysis [17], which explicitly aimed to include datasets more recent than those reviewed by di Mauro, and we extended our findings accordingly. Given that we retrieved more datasets than those covered in Mazzoni et al.'s extension, our contribution, in conjunction with the work of di Mauro et al., should represent the most comprehensive overview of existing household water datasets. We are in the process of submitting a merge request to update the repository provided by di Mauro et al, effectively increasing their list of household water datasets by 58%. The complete list of newly retrieved datasets and their respective metadata is presented in Table 1.

4 Dataset Characteristics

We aim to contextualize the datasets presented in Table 1. To achieve this, we categorize and compare them according to the proposed and used categories in the set reviewed by di Mauro et al. [8].

Firstly, on the **spatial scale**, we examine the geographical locations where the datasets were collected. The distribution of household datasets in the review of di Mauro et al. [8] and the newly included datasets, as well as the combination of both, is depicted in Table 2. Positive trends have been marked in teal, whilst negative trends are marked in red. Note that due to the shorter timespan of our investigation, an absolute decrease (or constant number) of published datasets can still result in a positive trend.

While in previous decades, just over 60% of the datasets were collected in the USA and EU alone, this proportion has decreased to 44% in recent years. The distribution within these major contributors has shifted significantly: previously, the United States contributed 39% and the EU 26%. In recent years, the contribution from the USA has decreased sharply, with only 2 new datasets (11%), while the EU's share has increased to 33%. Other regions (as categorized in [8]) have shown substantial changes in their contributions: Australia/New Zealand dropped to 0%, whilst Asia and Africa at least tripled their total released datasets over a short timespan. South America, of which previously no datasets have been reported, has entered with one contribution.

Notably, Australia, a former hotspot, no longer contributes, while Asia and Africa have significantly ramped up their efforts. A closer look at the contributing countries reveals that, with few exceptions, they either have a highly active and broad research community (e.g., USA, EU) or face a high water stress index, meaning they are already dealing with or are on the brink of serious public water supply challenges (e.g., Libya, Botswana, India). Aside from the decline in dataset creation in Australia/Oceania, the overall distribution aligns with intuitive expectations.

On the **temporal scale**, we observed that the sampling rates range from 1 second to one month, similar to previous findings [8]. Comparing the distribution of sampling rates to older datasets (cf. Figure 1), we noticed a slight trend towards higher time resolutions. This aligns with di Mauro et al.'s prediction of an increase in higher

Table 1: Household Water Datasets not included in [8]

Authors	Month/ Year	Location	Dataset Size	Time Frame	Sampling Rate	Access Policy
Nasser et al. [22]	08/20	Cairo, Egypt	20 households	1 year	10min	Open
Melville-Shreeve et al. [19]	02/21	Exeter, England	7 univ. buildings	7/2019 - 12/2019	1min	Restricted
Abu-Bakar et al. [2]	02/21	England	20,000 households	20 weeks	1d	Restricted
Dini and Kapitsaki [10]	04/21	Cyprus	1 household	12/2020 - 4/2021	10s	Open
Motho et al. [21]	03/22	Ngamiland, Botswana	497 households	-	1m	Restricted
Zozmann et al. [35]	03/22	Pune, India	50 households	1 week	$\frac{1}{3}$d	Restricted
Otaki et al. [23]	04/22	Sri Lanka	127 households	6/2017 - 11/2018	7d	Open
Arsene et al. [4]	07/22	Romania	5 households	2 weeks	1min	Open
DiCarlo and Berglund [9]	08/22	North Carolina, USA	16,000 smart meters	10/2018 - 11/2018	1h	Restricted
Alharsha et al. [3]	08/22	Sirte, Libya	380 households	12/2017, 2/2018	<1h	Restricted
Qin et al. [27]	10/22	China	41,649 households	1/2010 – 5/2019	1d	Restricted
Heydari et al. [12]	10/22	USA	1 household	4 weeks	1s	Open
Parra-Orobio et al. [25]	04/23	Cesar, Colombia	137 households	7/2022	-	Open
Mazzoni et al. [18]	06/23	Netherlands	9 households	7/2019 - 2/2020	1s, 2s	Restricted
Wilhelm et al. [33]	06/23	Bavaria, Germany	17 households	9/2021 - 9/2022	12s	Open
Zhou [34]	11/23	China	90 households	100 days - 1 year	1d	Open
Schaffer et al. [28]	02/24	Aalborg, Denmark	10,765 households	5 years (2018-2022)	1h	Restricted
Palacios-García [24]	03/24	Belgium	2 households	12/2023 - 1/2024	5min	Open

Table 2: Spatial Distribution of Household Water Datasets

Region	[8]	New	Total
USA	12 (38.7%)	2 (11.1%)	14 (28.6%)
EU	8 (25.8%)	6 (33.3%)	14 (28.6%)
AUS/NZL	6 (19.3%)	0 (00.0%)	6 (12.2%)
Asia	2 (06.5%)	4 (22.2%)	6 (12.2%)
UK	2 (06.5%)	2 (11.1%)	4 (08.2%)
AFR	1 (03.2%)	3 (16.7%)	4 (08.2%)
LATAM	0 (00.0%)	1 (05.6%)	1 (02.0%)
Canada	0 (00.0%)	0 (00.0%)	0 (00.0%)
Total	31 (100%)	18 (100%)	49 (100%)

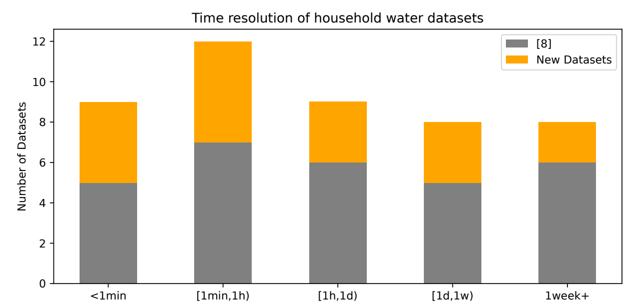

Figure 1: The time resolution of household datasets presented in [8], expanded by the datasets reviewed in this work.

sampling rate data due to the growing adoption of smart water meters. This trend is promising, as the closely related field of energy disaggregation has seen significant improvements in algorithm performance with higher sampling rates [14, 15, 26, 29], suggesting that water consumption analyses could benefit similarly.

Di Mauro et al. [8] also identified a strong inverse correlation between time resolution and dataset size, specifically in terms of the number of monitored buildings and the duration of the time series. While this correlation generally still holds, the boundaries have shifted or become less rigid. Previously, datasets covering several hundred homes only offered daily or monthly time resolutions [8]. However, more recent datasets now include measurements from several thousand households on an hourly basis (e.g., [9, 28]). Nonetheless, it remains true that no dataset involving more than 20 households features a sub-hourly sampling rate, except for a water diary from Libya [3]. Conversely, while datasets with short measurement periods typically consist of only a few houses with sub-hourly time resolution, there are now datasets with longer durations (>1 year) that also maintain high-resolution data (e.g., [33]).

Lastly, concerning **data accessibility**, di Mauro et al. [8] identified two distinct trends: first, a more-than-linear increase in the overall number of datasets, and second, a rising number and proportion of open-access datasets. We integrated their metadata with ours to revisit these predictions. As shown in Figure 2, both trends are reaffirmed, particularly the sharp increase in the number of open-access datasets in the domain of household water consumption in recent years.

It was further theorized that the restricted access to the majority of datasets was primarily due to privacy concerns associated with monitoring a large number of households [8]. Upon closer examination of the datasets in question (cf. Table 1), we find that none of the datasets covering more than 150 homes is openly accessible. Conversely, only one dataset with fewer than 50 buildings is not openly accessible. Thus, regarding the accessibility of datasets, we can reaffirm all previously identified trends and correlations.

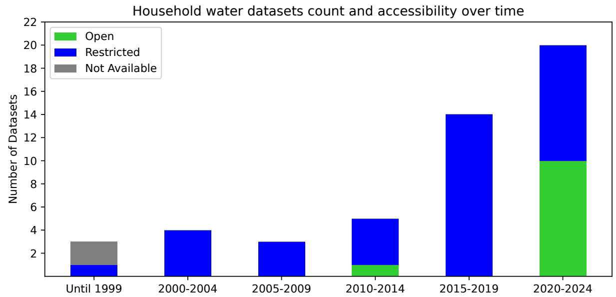

Figure 2: The count and accessibility over time of household datasets presented in [8], expanded by the datasets reviewed in this work.

5 Dataset Usage

The increasing trend of published and accessible datasets reflects a growing interest within the water research community. Di Mauro et al. identified several categories of research typically conducted using datasets of this scale: water demand forecasting, water demand pattern recognition, water conservation and customer awareness, and water end-use disaggregation, with the latter being particularly dependent on higher time resolutions [8]. They also anticipated a rise in publications in these areas, driven by a higher number of accessible datasets.

We screened more than 400 publications that reference datasets listed in Table 1 by analyzing the context in which these datasets were referenced to evaluate their popularity and explore the reasons behind their varying levels of usage. To our surprise, we discovered that none of these datasets have been reused by an independent group of researchers, despite their availability for several years. Only two datasets [4, 12] have follow-up work published [5, 13], and all of these are from the original research groups. Di Mauro et al. [8] performed a similar analysis of dataset reuse only for the *end-use* scale and found only two datasets (GOLD COAST [30] and SEQREUS [6]) to hold provable value for further analyses.

Given these findings, we endeavored to access all open access datasets in Table 1 to identify any barriers preventing the research community from effectively reusing the available data. The number of publications referencing these datasets suggests that not all challenges related to the research objectives have been addressed.

During our investigation, we identified the following five factors:

Data Quality. Many datasets were created to test a single idea or implementation and were published primarily to ensure the verifiability of the original work, rather than to serve as a resource for other researchers. As a result, the effort required to generalize and prepare the data (e.g., providing raw measurements and ground truth) often outweighs the benefits for the originating research groups. Typically, these processes require a distinct approach from the outset of data collection, making them challenging to implement retrospectively. Consequently, these datasets are ill-suited for cross-validation and are rarely reused by the research community.

Data Type. Closely related to data quality is the type of data that has been collected. The datasets we analyzed contain a variety of data types, including aggregated water usage, water usage over specified time spans (e.g., daily), general statistics, water pipe flow rates, and images of supply water counters. The additional effort required for a research group to convert the provided data type into the format they have collected or need for their analysis poses a significant barrier to the quick adoption of published datasets in follow-up studies by external researchers. Household electricity datasets are known to face a similar issue [15, 26]; however, they often at least provide either active power or current/voltage data. In our findings, the data types in water datasets are more varied, making it harder to find a second dataset with exactly the same data type and resolution.

Transferability. For similar reasons as before, but now focusing on the scope of monitored households rather than the type of data provided, many datasets feature only a small number of monitored households. To be attractive for cross-validation purposes, datasets should cover a substantial number of buildings. This is especially important for stand-alone analyses, where algorithms trained on one subset of households can be validated on another. This requirement has also been highlighted in the field of energy disaggregation [29]. For comparison, the SEQREUS dataset [6], which has been highlighted by di Mauro et al. [8] for seeing some reuse in the community, includes 250 houses at high resolution, collected in 2010 and 2011. As previously mentioned, nearly all recent datasets that cover a significant number of households have restricted access, creating an additional barrier for researchers looking to utilize them.

Duration. In addition to the number of houses monitored, the scope of a dataset must also be sufficient in terms of its duration, i.e., the length of the time series. For robust testing and validation, it is essential to be able to split the time series into different segments, each of which should still contain a sufficient amount of data for each phenomenon being studied to draw statistically sound conclusions. This should also include data from different seasons to capture possible related shifts in consumption behavior.

Metadata & Documentation. Many of the open-access datasets we examined lack proper documentation regarding their creation or provide no metadata altogether. Without adequate documentation, users of the dataset have no information about any post-processing steps that may have been applied, such as balancing the dataset or removing biases. This is problematic, as post-processing can introduce or eliminate patterns that could be mistaken for genuine patterns in the data, potentially skewing model results during training and testing. Faced with uncertainty about the conditions under which a dataset was created, researchers may choose to exclude such data rather than risk drawing incorrect conclusions.

Almost every dataset in our list could enhance its reusability by addressing one or more of these factors. However, as mentioned earlier, many of these improvements are challenging to implement retrospectively and should ideally be considered before data collection begins. Our contribution is not meant to criticize the potential shortcomings in the arguably complex and labor-intensive work of other researchers; therefore, we refrain from making direct or indirect references, such as statistics over the dataset collection. Instead, our aim is to provide constructive insights for those planning to create new datasets for the research community and/or to maintain and enhance the quality of existing datasets.

6 Conclusion

In this review, we analyzed the latest household water datasets, updating and expanding on the work of di Mauro et al. [8]. Our findings show a significant increase in datasets, particularly from water-stressed regions like Africa and Asia. However, challenges persist, especially in terms of high-resolution data for larger datasets and the absence of sub-hourly sampling in studies covering many buildings.

Despite the growth in available datasets, external reuse remains limited. We identified five factors—data quality, type, transferability, duration, and documentation—that hinder widespread adoption. Addressing these issues is crucial for improving cross-validation and fostering collaborative research.

To ensure greater dataset utility, future efforts should focus on improving documentation, ensuring consistency in data types (e.g., always including flow rates), and expanding temporal and spatial scales. These improvements will enhance the reuse and impact of household water datasets, supporting global efforts to address water conservation and demand management.

References

[1] Halidu Abu-Bakar, Leon Williams, and Stephen Henry Hallett. 2021. A Review of Household Water Demand Management and Consumption Measurement. *Journal of Cleaner Production* 292 (2021), 125872.

[2] Halidu Abu-Bakar, Leon Williams, and Stephen H Hallett. 2021. Quantifying the Impact of the COVID-19 Lockdown on Household Water Consumption Patterns in England. *NPJ Clean Water* 4, 1 (2021), 13.

[3] Iman Alharsha, Fayyaz A Memon, Raziyeh Farmani, and Wa'El A Hussien. 2022. An Investigation of Domestic Water Consumption in Sirte, Libya. *Urban Water Journal* 19, 9 (2022), 922–944.

[4] Diana Arsene, Alexandru Predescu, Bogdan Pahonțu, Costin Gabriel Chiru, Elena-Simona Apostol, and Ciprian-Octavian Truică. 2022. Advanced Strategies for Monitoring Water Consumption Patterns in Households based on IoT and Machine Learning. *Water* 14, 14 (2022), 2187.

[5] Diana Arsene, Alexandru Predescu, Ciprian-Octavian Truică, Elena-Simona Apostol, and Mariana Mocanu. 2023. Decision Support Strategies for Household Water Consumption Behaviors based on Advanced Recommender Systems. *Water* 15, 14 (2023), 2550.

[6] Cara Beal, Rodney Stewart, T Huang, and E Rey. 2011. *South East Queensland Residential End Use Study.* Urban Water Security Research Alliance Brisbane, Australia.

[7] Benjamin I Cook, Justin S Mankin, and Kevin J Anchukaitis. 2018. Climate Change and Drought: From Past to Future. *Current Climate Change Reports* 4 (2018), 164–179.

[8] Anna Di Mauro, Andrea Cominola, Andrea Castelletti, and Armando Di Nardo. 2021. Urban Water Consumption at Multiple Spatial and Temporal Scales. A Review of Existing Datasets. *Water* 13, 1 (2021), 36.

[9] Morgan Faye DiCarlo and Emily Zechman Berglund. 2022. Using Advanced Metering Infrastructure Data to Evaluate Consumer Compliance with Water Advisories during a Water Service Interruption. *Water Research* 221 (2022), 118802.

[10] Alexia Dini and Georgia M. Kapitsaki. 2021. *Smart Home Data Set - Temperature, Humidity, Motion, Sound, and Smart Water Meter Data.* https://doi.org/10.5281/zenodo.4718373

[11] George William Hart. 1992. Nonintrusive Appliance Load Monitoring. *Proc. IEEE* 80, 12 (1992), 1870–1891.

[12] Zahra Heydari, Andrea Cominola, and Ashlynn S Stillwell. 2022. Is Smart Water Meter Temporal Resolution a Limiting Factor to Residential Water End-use Classification? A Quantitative Experimental Analysis. *Environmental Research: Infrastructure and Sustainability* 2, 4 (2022), 045004.

[13] Zahra Heydari and Ashlynn S Stillwell. 2024. Comparative Analysis of Supervised Classification Algorithms for Residential Water End Uses. *Water Resources Research* 60, 6 (2024), e2023WR036690.

[14] Jana Huchtkoetter and Andreas Reinhardt. 2019. A Study on the Impact of Data Sampling Rates on Load Signature Event Detection. *Energy Informatics* 2 (2019), 1–12.

[15] Hafiz Khurram Iqbal, Farhan Hassan Malik, Aoun Muhammad, Muhammad Ali Qureshi, Muhammad Nawaz Abbasi, and Abdul Rehman Chishti. 2021. A Critical Review of State-of-the-art Non-intrusive Load Monitoring Datasets. *Electric Power Systems Research* 192 (2021), 106921.

[16] Diana Ivanova, Konstantin Stadler, Kjartan Steen-Olsen, Richard Wood, Gibran Vita, Arnold Tukker, and Edgar G Hertwich. 2016. Environmental Impact Assessment of Household Consumption. *Journal of Industrial Ecology* 20, 3 (2016), 526–536.

[17] Filippo Mazzoni, Stefano Alvisi, Mirjam Blokker, Steven G Buchberger, Andrea Castelletti, Andrea Cominola, Marie-Philine Gross, Heinz E Jacobs, Peter Mayer, David B Steffelbauer, et al. 2023. Investigating the Characteristics of Residential End Uses of Water: A Worldwide Review. *Water Research* 230 (2023), 119500.

[18] Filippo Mazzoni, Stefano Alvisi, Marco Franchini, and Mirjam Blokker. 2023. Exploiting High-resolution Data to Investigate the Characteristics of Water Consumption at the End-use Level: A Dutch Case Study. *Water Resources and Industry* 29 (2023), 100198.

[19] Peter Melville-Shreeve, Sarah Cotterill, and David Butler. 2021. Capturing High-Resolution Water Demand Data in Commercial Buildings. *Journal of Hydroinformatics* 23, 3 (2021), 402–416.

[20] Bettina E Meyer, Khoi Nguyen, Cara D Beal, Heinz E Jacobs, and Steven G Buchberger. 2021. Classifying Household Water Use Events into Indoor and Outdoor Use: Improving the Benefits of Basic Smart Meter Data Sets. *Journal of Water Resources Planning and Management* 147, 12 (2021), 04021079.

[21] Manga Motho, Oluwatoyin D Kolawole, Moseki R Motsholapheko, and Patricia K Mogomotsi. 2022. Influence of Household Demographic and Socio-Economic Factors on Water Demand in Ngamiland District, Botswana. *Water Science* 36, 1 (2022), 48–59.

[22] Ahmed Abdel Nasser, Magdi Z Rashad, and Sherif E Hussein. 2020. A two-layer Water Demand Prediction System in Urban Areas based on Micro-Services and LSTM Neural Networks. *IEEE Access* 8 (2020), 147647–147661.

[23] Y Otaki, M Otaki, T Chaminda, T Matsui, R Bokalamulla, R Thathsarani, and T Joganathan. 2022. Demand-side Water Management Using Alternative Water Sources based on Residential End-use. *Water Practice & Technology* 17, 4 (2022), 949–959.

[24] Emilio José Palacios-García. 2024. PREMISES: Smart Meter Measurements. https://doi.org/10.48804/UNFBX7

[25] Brayan Alexis Parra-Orobio, Jonathan Soto-Paz, Anulfo Ramos-Santos, Keiner Fernando Sanjuan-Quintero, Rossember Saldaña-Escorcia, Isabel Cristina Dominguez-Rivera, and Antoni Sánchez. 2023. Assessment of the Water Footprint in Low-Income Urban Neighborhoods from Developing Countries: Case Study Fátima (Gamarra, Colombia). *Sustainability* 15, 9 (2023), 7115.

[26] Lucas Pereira and Nuno Nunes. 2018. Performance Evaluation in Non-intrusive Load Monitoring: Datasets, Metrics, and Tools—A Review. *Wiley Interdisciplinary Reviews: data mining and knowledge discovery* 8, 6 (2018), e1265.

[27] Ping Qin, Shuai Chen, Jie-Sheng Tan-Soo, and Xiao-Bing Zhang. 2022. Urban Household Water Usage in Adaptation to Climate Change: Evidence from China. *Environmental Science & Policy* 136 (2022), 486–496.

[28] Markus Schaffer, Martin Veit, Anna Marszal-Pomianowska, Martin Frandsen, Michal Zbigniew Pomianowski, Emil Dichmann, Christian Grau Sørensen, and Jesper Kragh. 2024. Dataset of Smart Heat and Water Meter Data with Accompanying Building Characteristics. *Data in Brief* 52 (2024), 109964.

[29] Changho Shin, Seungeun Rho, Hyoseop Lee, and Wonjong Rhee. 2019. Data Requirements for Applying Machine Learning to Energy Disaggregation. *Energies* 12, 9 (2019), 1696.

[30] Rodney A Stewart, Damien Giurco, Kriengsak Panuwatwanich, Bill Capati, and Rachelle Willis. 2009. Gold Coast Domestic Water End Use Study. *Water: Journal of the Australian Water Association* 36, 6 (2009), 84–90.

[31] Richard G Taylor, Bridget Scanlon, Petra Döll, Matt Rodell, Rens Van Beek, Yoshihide Wada, Laurent Longuevergne, Marc Leblanc, James S Famiglietti, Mike Edmunds, et al. 2013. Ground Water and Climate Change. *Nature climate change* 3, 4 (2013), 322–329.

[32] Steven Van Tuyl and Amanda L Whitmire. 2016. Water, Water, Everywhere: Defining and Assessing Data Sharing in Academia. *PLoS One* 11, 2 (2016), e0147942.

[33] Sebastian Wilhelm, Jakob Kasbauer, Dietmar Jakob, Benedikt Elser, and Diane Ahrens. 2023. Exploiting Smart Meter Water Consumption Measurements for Human Activity Event Recognition. *Journal of Sensor and Actuator Networks* 12, 3 (2023), 46.

[34] Wenqing Zhou. 2023. Water and Electricity Usage Dataset for Multi-Meter Integration Demonstration Area. https://doi.org/10.21227/hj78-c619

[35] Heinrich Zozmann, Christian Klassert, Bernd Klauer, and Erik Gawel. 2022. Water Procurement Time and its Implications for Household Water Demand — Insights from a Water Diary Study in five Informal Settlements of Pune, India. *Water* 14, 7 (2022), 1009.