Article

Exploiting Smart Meter Water Consumption Measurements for Human Activity Event Recognition

Sebastian Wilhelm *, Jakob Kasbauer, Dietmar Jakob, Benedikt Elser and Diane Ahrens

Deggendorf Institute of Technology; Technology Campus Grafenau, 94481 Grafenau, Germany;
jakob.kasbauer@th-deg.de
* Correspondence: sebastian.wilhelm@th-deg.de; Tel.: +49-8552-975620-25

Abstract: Human activity event recognition (HAER) within a residence is a topic of significant interest in the field of ambient assisted living (AAL). Commonly, various sensors are installed within a residence to enable the monitoring of people. This work presents a new approach for human activity event recognition (HAER) within a residence by (re-)using measurements from commercial smart water meters. Our approach is based on the assumption that changes in water flow within a residence, specifically the transition from no flow to flow above a certain threshold, indicate human activity. Using a separate, labeled evaluation data set from three households that was collected under controlled/laboratory-like conditions, we assess the performance of our HAER method. Our results showed that the approach has a high precision (0.86) and recall (1.00). Within this work, we further recorded a new open data set of water consumption data in 17 German households with a median sample rate of 0.08$\bar{3}$ Hz to demonstrate that water flow data are sufficient to detect activity events within a regular daily routine. Overall, this article demonstrates that smart water meter data can be effectively used for HAER within a residence.

1. Introduction

Human activity recognition (HAR) is a field that has garnered significant research attention, as the automated detection of human activities has numerous practical applications, including surveillance, health monitoring, human–machine interaction, entertainment, and intelligent environments [1,2]. Current approaches for human activity recognition (HAR) are either vision-based or sensor-based [2]. Vision-based approaches offer high accuracy but have several drawbacks, such as privacy issues, limited monitoring range, and high cost. Sensor-based approaches, on the other hand, are more diverse. They can be further divided into three categories, wearable sensing, object-tagged sensing, and dense/infrastructure-mediated sensing, each with its advantages and disadvantages.

However, it is usually expensive and time-consuming to fully equip residences with suitable sensors. Furthermore, a disadvantage of most existing systems is that their components cannot be completely inconspicuously integrated into the home environment, often making residents feel monitored [3].

The objective of our work is to identify human activities on a more general level. We would rather not distinguish between the individual activities but determine the occurrence of any activity within a residence. Thus, instead of HAR, we consider the broader human activity event recognition (HAER) problem in this work. The information about activity events in the household shall later be used for emergency detection, especially in the context of ambient assisted living (AAL) systems, for example, by detecting atypically long periods of inactivity, which potentially indicate an emergency [4,5].

To address the HAER problem, we propose a novel dense sensing approach that exploits data from smart water meters instead of introducing additional sensors into the

households. Smart water meters are already present in many German households and provide detailed information in particular on cumulative water consumption, which can be used to infer human activity. Using smart water meter data has several advantages, including the fact that water consumption is vital to human survival and can be easily measured in almost all households [6]. It is easy to access and prepare the data without major effort and cost. Further, the water meter measurements contain only a low level of noise. Possible causes of noise in a water sensor are dripping taps, leaking pipes, and/or quantization noise.

The literature describes that smart water meter data are used to optimize water supply systems [7–9] or provide residents with accurate information about their water consumption [7,8,10–15]. Using smart water meter data for HAR/HAER or in the AAL domain has not been widely explored. When using water meter measurements in the context of an AAL system, it is crucial to optimize the precision of the HAER because the precision affects the ability to detect potential emergencies. In contrast to the HAER problem, a single false positive classification for the water supply optimization problem is acceptable. Optimizing the recall is of subordinate importance for emergency detection using an inactivity profile since emergency detection only considers the absence of activities; false negative (FN) classifications, therefore, do not influence the general ability of emergency detection, as long as there are still some regular true positive (TP) classifications. FN classifications only affect the granularity of how quickly emergencies can be detected.

We aim to develop a methodology for HAER that can run on an edge system in real time directly inside the residential environment (e.g., on a single-board computer). Thus, the water consumption data do not need to be sent to third parties, improving privacy. Additionally, the system does not require a continuously active internet connection, which increases the system's availability. Only in the case of an emergency would an external communication be required.

The main contribution of this article is to present a new approach for human activity event recognition (HAER) within a residence using smart water meter measurements. By using smart water meter measurements, non-intrusive monitoring of human activity can be performed using existing data sources. Our method is designed for (near-)real-time activity detection in a streaming setting using edge computing. Further, this article presents a new publicly available data set of water consumption measurements from 17 German households, which we collected using commercially available smart water meters.

The remaining paper is structured as follows: we review the literature on analyzing water consumption in private households in Section 2 and provide an overview of the different features used for water consumption analysis and the purposes for which water consumption is analyzed. Since we could not identify suitable data sets with water consumption data measured using a smart water meter in the literature, we collected a new data set in 17 German households as part of this work. We present the method of collection, as well as the gathered data set itself, in Section 3. Subsequently, in Section 4, we address the necessary data pre-processing steps before introducing a novel approach for HAER based on measurements from smart water meters in Section 5. We then evaluate this approach in Section 6. The paper ends with a discussion in Section 7 and a conclusion and outlook in Section 8.

2. Related Work

One of the earliest works in household water consumption analysis can be traced to Bennett and Linstedt [16]. In 1975, the authors investigated wastewater treatment and disposal systems and extensively analyzed household water consumption. The authors collected water consumption data in five households using analog flow meter recorders on wax discs and manual activity records. Based on these records, Bennett and Linstedt then analyzed the water use of individual appliances and inferred various household water-use characteristics to understand wastewater generation better [16]. For example,

the authors examined the typical water consumption within a day and the average daily water consumption per person [16].

Later, De Oreo et al. [10] presented in 1996 one of the first modern approaches to disaggregate water consumption data using digital data loggers. In their work, the authors spotted specific water flow patterns that could be assigned to activities or individual appliances (toilet, dishwasher, washing machine) [10].

The focus of most related work is to analyze water consumption measurements either to manage and optimize water supply networks [7–9], to help residents and planners conserve water by obtaining insight into the households' water consumption [7,8,10–15], for leak detection [9], or for HAR or activity of daily living (ADL) detection [6,17–19]. The papers by Salomons et al. [20] and Cardell-Oliver et al. [21] critically address the disaggregation of water measurements or smart water metering by noting privacy issues [20,21].

In the literature review, we identified no related work that exploits water consumption data to detect activity events in the sense of a binary classification problem. We can assume that HAER is a sub-problem of the HAR problem. In other words, if we can perform HAR based on water consumption data, this implies HAER. However, the reverse is not true in general.

Several publications on the analysis of water consumption data are attributed to the research group of the Australian professor Cardell-Oliver. The authors use various analytical methods on a data set of smart meter readings from Western Australia to conclude residents' routine behavior to motivate them to conserve water. The researchers also raise privacy concerns arising from smart meters [9,14,21,22].

Instead of analyzing water flow rates, the works by Froehlich et al. [17,23] and Thomaz et al. [13] analyze water pressure. Froehlich et al. addressed water data disaggregation for HAR. They developed a method called "HydroSense" that aims to detect the unique pressure waves that occur when fixtures (e.g., a tap) are opened or closed. Using a dynamic Bayesian approach, the authors combined template matching, grammar, and prior probabilities to determine the probability of a specific event. They achieved an overall accuracy between 76% and 96% for households [17,23]. Thomaz et al. attempted to identify human activities, especially in bathrooms and kitchens. For this, they used a vector space model (VSM)-based method on water pressure measurements. The evaluation with 23 participants achieved an overall accuracy of 70% in recognizing eleven high-level activities in the kitchen and bathroom [13].

Chen et al. investigated activity disaggregation on high-resolution synthetic water flow data (and additional medium-resolution smart-meter measurements). The study aimed to provide residents with a more in-depth insight into their water use. Chen et al. focused on detecting toilets, showers, and washing machines. The authors used different approaches: HMM, kNN-GMM, ANN-GMM, and SVM-GMM. For shower recognition, they achieved a precision of around 95% with a recall between 77–81%. For the washing machine, they achieved a precision between 78 and 86% with a recall between 89–96% [11]. Using a combination of hidden Markov model (HMM), artificial neural networks (ANN), and dynamic time wrapping (DTW) algorithms, Nguyen et al. tried to detect event patterns in water consumption and thus achieve CO_2 savings in the water network (especially at the pumps) by providing optimizations in households, such as determining when to water the garden [8].

Koketsu et al. used water consumption data to identify the rhythm of the daily life of elderly people, determine deviations, and contact relatives if necessary. They concluded that although regularities in daily water consumption can be identified, more research is needed to implement comprehensive monitoring of the elderly [6]. The design of a similar system of a "daily-use water monitoring platform for the elderly" is also described by the authors Yanyan et al. However, the paper omitted a discussion of the actual implementation [19].

Hu et al. presented another approach for monitoring the daily activities of the elderly based on data from a three-axis accelerometer mounted on the surface of the main water pipe in the house. They describe the detection of different ADLs (wash, cook, flush, bath)

using a support vector machine (SVM) or fusion model. However, the precision and recall of the detection vary widely based on the individual ADLs [18]. Moreover, based on vibration data, Kim et al. presented a "Nonintrusive Autonomous Water Monitoring System" [12].

The existing approaches for analyzing water consumption measurements differ primarily in the data (sensor types, features, temporal resolution, and measurement resolution) on which the analysis or disaggregation is conducted. A comprehensive overview of the water consumption data used in related studies, their particular features and resolution, and the respective collection method is presented in Table 1.

The majority of studies use flow rate sensors [6,8–11,16,19,21,22,24,25]. However, there are also studies using other features, such as water pressure [13,17] or measurements of a three-axis accelerometer mounted on the surface of the main water pipe in the house [12,18].

Another essential point of differentiation for studies based on water flow is the temporal resolution of the measured values. Here, the literature distinguishes among high, medium, and low resolutions. There is no final specification for the distinction between high, medium, and low resolutions. Following Clifford et al. [7] and Cardell-Oliver [22], a high resolution means values measured within seconds (≤1 min), a medium resolution means values measured up to once a day (>1 min; ≤1 day), and a low resolution means lower resolutions than once a day (>1 day). Research on water disaggregation mainly uses medium- [6,9,11,14,19,22] or high-resolution [8,10,11,24] data. Low-resolution data are mainly used for general flow analysis, seasonal changes, or billing purposes.

A further point to note is that none of the approaches with the goal of HAR or ADL detection uses measurements from commercial smart meters as the data basis; the authors either used special flow meters [6,19] or even used different features than water flow [17,18]. An extensive discussion of the various resolutions of smart meter water measurements can be found in Clifford et al. [7].

Table 1. Overview of aggregated water consumption data used in related work.

Author	Feature/Resolution	Collection Method
Bennett and Linstedt [16]	water flow/n.a.	recording on wax discs
Cardell-Oliver [14]	water flow/1 h	smart meter
Cardell-Oliver [22]	water flow/1 h	smart meter
Cardell-Oliver and Carter-Turner [21]	water flow/1 h+	smart meter
	water flow/10 s	synthesized
Chen et al. [11]	water flow/15 min+	smart meter
	water flow/10 s	data logger
De Oreo et al. [10]	water flow/10 s	data logger
Froehlich et al. [17]	water pressure/500 Hz	pressure sensor (in pipe)
Hu et al. [18]	three-axis accelerometer/85 Hz	three-axis accelerometer (on pipe)
Kim et al. [12]	three-axis accelerometer/100 Hz	three-axis accelerometer (on pipe)
Koketsu et al. [6]	water flow/5 min	flow meter
Nguyen et al. [8]	water flow/5 s	smart meter
Srinivasen et al. [24]	water flow/0.5 s	flow meter
Thomaz et al. [13]	water pressure/200 Hz	pressure sensor (in pipe)
Wang et al. [9]	water flow/1 h	smart meter
Willis et al. [25]	water flow/10 s	data logger
Yanyan et al. [19]	water flow/2 h	flow meter

Another aspect highlighted in our literature review is the absence of commonly used benchmark datasets for water use in the scientific literature. Instead, researchers generally collect their own data sets (except for a group led by Cardell-Oliver [9,14,21,22]), making it difficult to compare results across studies. Furthermore, the primary data sets are typically not available as open data. This is in contrast to the disaggregation of electrical power consumption data, where publicly available datasets (e.g., *REDD* [26] or *UK-DALE* [27])

are common. Notably, the scientific literature pays less attention to the disaggregation of water consumption data compared to the disaggregation of power consumption data.

3. Data Recording

As described in Section 2, there are no publicly available open data sets on water consumption that have been collected using commercial smart meters. Therefore, as part of this work, we recorded a new data set from 17 German households. The new data set is available as open data on *Zenodo* at https://doi.org/10.5281/zenodo.7506076.

In the following, the data set's collection, the data set's structure, and the collected data's quality are described in detail.

3.1. Methodology of Data Collection

3.1.1. Measuring Instruments

We recorded the cumulative household water consumption over time using the commercially available smart meters *Hydrus 1.3, DN 20,00* from *Diehl Metering*. The installation of the water meters was carried out by the local water utility. These smart meters feature a wireless M-Bus communication interface and periodically transmit the current cumulative meter reading in an interval of at least 8 s via the open metering system (OMS) protocol on 868 MHz.

The discretization accuracy of the readings is 1 liter. Further technical information can be found in the datasheet (see: https://www.diehl.com/cms/files/HYDRUS-1.3.pdf (accessed on 5 January 2023)).

3.1.2. Data Transmission and Storage

The data from the smart meters can be received in the local vicinity of the water meter, depending on the signal strength of the wireless M-Bus. To continuously monitor the cumulative water consumption of each household, we installed a Raspberry Pi with an appropriate antenna capable of receiving the OMS signals at 868 MHz in each of the test households. These signals were decoded and decrypted directly on the Raspberry Pi. We annotated the sample with the current timestamp from an NTP-synchronized clock source. The measurements were temporarily stored on the Raspberry Pi. Using a secure VPN connection, the measured values were regularly transferred to a central database on a server in a data center.

Figure 1 shows the software architecture and data flow for acquisition and transmission.
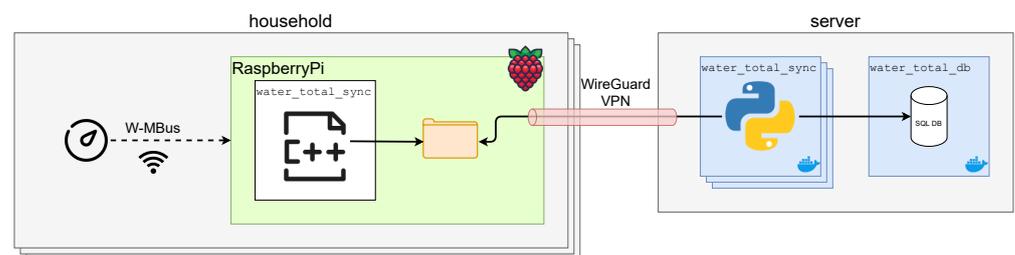
Figure 1. Software architecture for acquiring and transmitting cumulative water flow measurements from private households.

3.1.3. Ethics and Privacy Considerations

Privacy and ethics are a focus of the project, as sensitive data relating to subjects' daily lives will be collected and processed. To follow ethical norms and values and be as transparent as possible, a reflection on the project, especially on how the data were collected and processed, was conducted in advance based on a guideline from the Ethics Commission DGP e.V. [28].

To ensure data security, all data in our project are stored on internal servers. All measured data were stored pseudonymized by default. Furthermore, we implemented data privacy management according to Jakob et al. [29].

3.2. Data Set Description

In the period from September 2020 to September 2022, the following data were collected in a total of 17 households in Germany:

$$household = (location, \#residents)$$
$$measurement_{total} = (household, timestamp, total)$$

where:

$$location = \text{German postcode of the location of the household}$$
$$\#residents = \text{Number of residents usually living in the household}$$
$$timestamp = \text{NTP-synchronized UNIX timestamp (in ms)}$$
$$total = \text{Cumulative water consumption (in } m^3)$$

The *#residents* parameters are based on information provided by the residents at the beginning of the collection. There was no change in the statistical *#residents* parameters during the collection period.

The *timestamp* parameter describes the NTP-synchronized timestamp generated on the Raspberry Pi when receiving the measured values. Since (1) the Raspberry Pi does not have a real-time clock and (2) transmission delays and processing delays are not considered, minor time shifts may occur.

In the following, the time series of the value *total* is denoted as $V(t)$, where $[V] = m^3$.

To ensure that no household can be de-identified by the *#residents* parameter, we *k*-anonymized this parameter [30]. Therefore, we divided up the information about the number of residents living in the household into pairs of four groups (1, 2, 3, and $\geq$4). This results in a 3-anonymous data set. In addition, we performed a vertical shift of the measured value time series $V(t)$ in a household-specific range between 0 and 100 m^3 by adding a constant value to each measurement.

The root folder of the *Zenodo* archive contains the documents `readme.md`, `info.txt`, and `info.json`, as well as the 17 folders for the individual households. The `readme.md` provides basic information on the data set and its use. The documents `main.txt` and `main.json` contain the metadata for the measurements of all households in both human- and machine-readable format.

For each test household, there is a separate folder with the documents `info.txt` and `info.json`. These contain household-specific metadata. The total water consumption measurements are stored in `smartmeter.csv`.

The statistical characteristics of each test household are shown in Table 2.

3.3. Data Quality Reflection

In the following, we consider the quality of the gathered data set.

3.3.1. Recording Gaps

First, it can be noted that there are long recording gaps (several days/weeks) in the measurements. The gaps are mainly caused by technical failures, especially the failure of the Raspberry Pi or the receiving antennas. Furthermore, measurement gaps may have been caused by unstable radio signals or interference signals on the transmission frequency of the smart meters.

Due to household access restrictions, access to the test hardware was limited during the survey period, so technical problems could only be solved late in some cases. Therefore, long gaps must be expected when using the data set. The longest continuous timespans without recording gaps are shown in Table 3, split by test household and by the expected measurement frequency.

Table 2. Statistical characteristics of the test households.

	Location	#Residents	#Samples	First Date	Last Date
WM-A-01	GER-94151 Mauth	2	1,388,374	2020-12-15	2021-06-24
WM-A-02	GER-94151 Mauth	3	1,314,504	2021-03-02	2022-03-05
WM-A-03	GER-94151 Mauth	3	926,611	2021-03-02	2021-10-22
WM-A-04	GER-94151 Mauth	2	145,515	2021-03-11	2022-08-25
WM-A-05	GER-94151 Mauth	3	2,861,708	2021-02-09	2022-06-01
WM-A-06	GER-94151 Mauth	2	1,484,185	2021-02-25	2021-10-26
WM-A-07	GER-94151 Mauth	2	1,341,377	2021-02-09	2021-08-13
WM-A-08	GER-94151 Mauth	≥4	3,896,681	2020-09-04	2022-09-30
WM-A-09	GER-94151 Mauth	≥4	944,727	2020-10-21	2022-04-22
WM-A-10	GER-94151 Mauth	2	1,956,478	2021-02-09	2021-12-26
WM-A-11	GER-94151 Mauth	3	4,658,361	2020-09-23	2022-09-30
WM-A-12	GER-94151 Mauth	1	1,952,055	2021-03-02	2022-01-18
WM-A-13	GER-94151 Mauth	1	643,652	2021-05-03	2021-08-13
WM-A-14	GER-94151 Mauth	1	428,090	2021-03-02	2022-08-05
WM-A-15	GER-94151 Mauth	≥4	2,103,119	2021-03-02	2022-09-01
WM-A-16	GER-94151 Mauth	≥4	480,232	2021-03-02	2021-08-13
WM-A-17	GER-94151 Mauth	3	1,988,162	2021-02-20	2022-07-15

Table 3. Maximum continuous timespans without recording gaps per household, with at least one measured value per 10 s, per 1 min, per 1 h, and per 1 day.

|	Periods without Measuring Gaps with at Least			
	10 s	1 min	1 h	1 Day
WM-A-01	0 d 00:00:25	105 d 22:59:09	189 d 10:24:08	191d 03:54:11
WM-A-02	0 d 00:12:19	34 d 22:17:03	60 d 08:29:20	60 d 08:29:20
WM-A-03	0 d 00:03:11	8 d 00:15:09	37 d 13:39:33	93 d 02:14:56
WM-A-04	0 d 00:00:24	2 d 10:35:21	14 d 01:51:08	16 d 00:48:39
WM-A-05	0 d 00:02:38	220 d 05:14:59	292 d 10:09:52	354 d 09:25:47
WM-A-06	0 d 00:14:15	56 d 15:33:10	59 d 06:16:03	59 d 06:16:03
WM-A-07	0 d 00:04:52	101 d 23:59:00	184 d 20:59:15	184 d 20:59:15
WM-A-08	0 d 00:16:41	24 d 11:13:08	90 d 20:51:10	239 d 09:09:18
WM-A-09	0 d 00:02:22	7 d 10:01:01	19 d 11:29:28	76 d 23:11:07
WM-A-10	0 d 00:02:28	96 d 21:43:07	96 d 21:43:07	97 d 12:45:10
WM-A-11	0 d 00:15:14	90 d 06:29:56	112 d 22:06:01	217 d 21:30:35
WM-A-12	0 d 00:08:48	91 d 23:10:27	111 d 01:39:03	111 d 01:39:03
WM-A-13	0 d 00:00:17	29 d 18:26:06	29 d 18:56:09	29 d 18:56:09
WM-A-14	0 d 00:04:00	9 d 00:54:31	16 d 00:59:35	71 d 17:46:06
WM-A-15	0 d 00:01:32	41 d 09:00:42	45 d 09:25:08	80 d 10:49:11
WM-A-16	0 d 00:04:07	17 d 21:50:31	35 d 17:41:13	35 d 17:41:13
WM-A-17	0 d 00:01:56	7 d 16:34:20	48 d 07:47:18	149 d 00:28:20

3.3.2. Faulty Timestamps

From the smart meter's perspective, a private household is always a water consumer and not a water producer. Thus, by definition, the time series of measured values of cumulative consumption must be monotonically increasing. However, we find this is not the case when analyzing the data records. The reason for this may be out-of-order samples with obviously faulty timestamps. Two examples of faults are shown in Figure 2.

On the one hand, it can be seen that the last sample before a gap sometimes has an obviously incorrect timestamp (e.g., in Figure 2a). On the other hand, as shown in Figure 2b, some out-of-order samples occur even in gap-free sequences.

Another point to be considered in the context of the reflection on the data quality is the temporal resolution of the measured values. Theoretically, this should be at least 0.125 Hz–i.e., at least one sample every 8 s. However, if we look at the data from the recordings, we can observe that the temporal resolution is subject to major variation and is less than 0.125 Hz. In the entire data set, we obtain a median sample rate of $0.08\bar{3}$ Hz (one

sample every 12 s). Figure 3 shows the time intervals between two consecutive samples for an example period of two days.

(**a**) Faulty timestamp   (**b**) Out-of-order samples
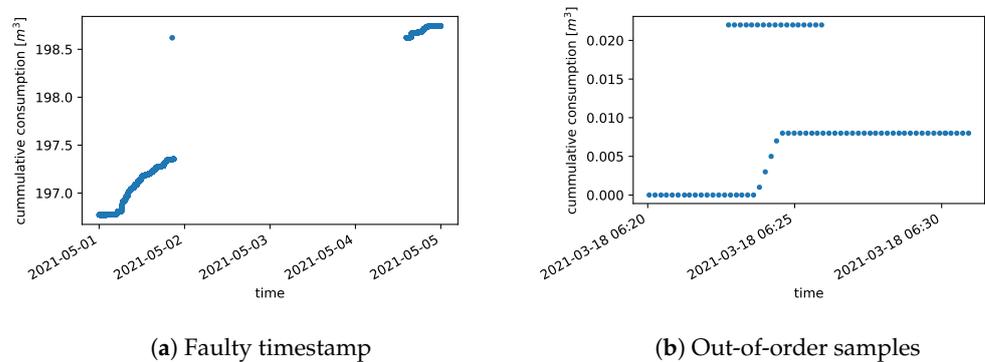
Figure 2. Examples of faulty timestamps and out-of-order samples in the raw measurements. (Household: WM-A-08).
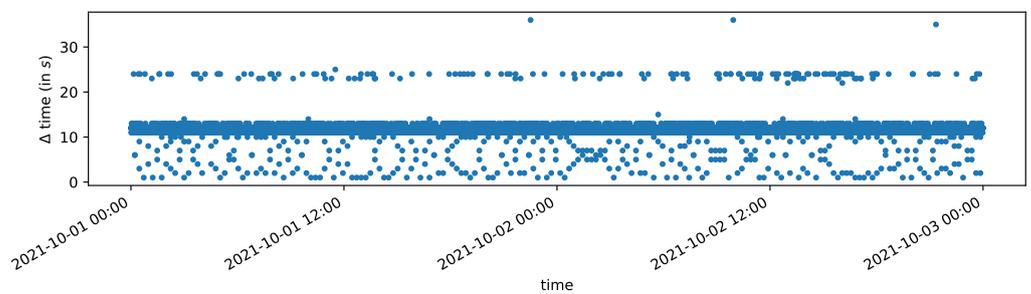

Figure 3. Time intervals between two consecutive samples in an example period of two days. (Household: WM-A-08).

4. Data Pre-Processing

We must first perform pre-processing steps to detect human activity from the smart water meter data. These steps include removing faulty samples from the raw data (Section 4.1), resampling the data (Section 4.2), and interpolating gaps (Section 4.3). With smoothing, we finish the data pre-processing in Section 4.4.

4.1. Data Cleaning

In Section 3.3, we showed that the raw data contain faulty samples. In particular, faulty time stamps or out-of-order samples occur in a few cases. To clean these faulty out-of-order samples, we use the monotonic property of the cumulative water consumption series and the property that the water supply of a private household is limited.

4.1.1. Monotonic Increasing

Some measurement samples contain a faulty timestamp and arrive out of order, resulting in transient spikes in the cumulative water consumption time series, as shown in Figure 4a. These samples violate the monotonic constraint and should be removed.

A naive method for handling these samples is to replace the time series with the cumulative maximum of the time series; however, this also modifies all correct samples after a spike, as shown in Figure 4b.

To address this issue, we propose a different approach. First, we reverse the time series, then compute the cumulative minimum, and reverse the result again. Finally, we remove the sample at the spike location, depicted with an $\times$ in Figure 4c. This method effectively removes the faulty samples without affecting the correct samples.

(**a**) Faulty raw data containing out-of-order sample

(**b**) Naive approach using $CumMax(y)$

(**c**) Adjusted approach using $Rev(CumMin(Rev(y)))$
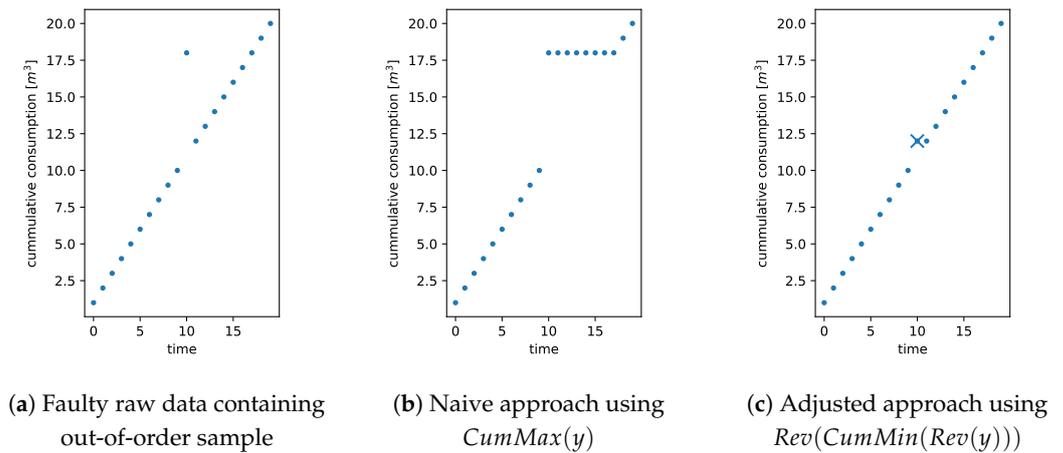
Figure 4. Approach for removing out-of-order samples that violate the monotonic constraint.

4.1.2. Filter Unrealistic High Water Flow Rates

During the data cleaning process, we use the physical constraints of the household water supply system to filter out unrealistic, high-consumption samples that are likely to be faulty data. Specifically, we remove all samples where the flow rate is higher than 30 L per minute.

Over the entire data set, we notice 0.011 % of faulty samples that violate the plausibility constraints, monotonicity, and maximum flow rate.

4.2. Resampling

As shown in Section 3.3, the measurement frequency is not constant at 0.125 Hz. To standardize the frequency, we perform downsampling using the *max* resampling rule because $V(t)$ is a cumulative series.

For this work, we use a measurement resolution of $0.0\bar{6}$ Hz, i.e., one measurement per 15 s. By using a resolution slightly smaller than the median sample rate of the entire data set ($0.08\bar{3}$ Hz), minor variations in the time axis can be well-compensated for and there are fewer gaps.

4.3. Gap Interpolation

The resampled data may contain gaps due to measurement gaps or removed samples from the pre-processing steps from Section 4.1. Some of these gaps can be interpolated to improve the data continuity. However, to ensure that the interpolation does not influence the subsequent activity-recognition steps, we only interpolate gaps where there have been no changes in the cumulative consumption between time points.

To determine which gaps can be interpolated, we use the following criterion: if $V(t_f - 1) = V(t_l + 1)$ for a gap in the interval $[t_f, t_l]$, we interpolate the values for the gap by setting $V(t) = V(t_f - 1)$ for all $t \in [t_f, t_l]$. This gap criterion ensures that the interpolated values accurately reflect the underlying consumption.

We limited interpolation to gaps of no more than 1 h.

4.4. Smoothing

The water meter outputs the measured values with quantization errors. A constant flow rate of 0.1 liter per timestep produces the following sequence of volume readings: $[0.9, 1.0, 1.1]$ liter. The quantization to 1.0-liter precision transforms the readings to the sequence $[0.0, 1.0, 1.0]$ liter. To compute the flow rate, we perform the first derivative of this sequence. The result is the faulty sequence $[0.0, 1.0, 0.0]$ liter per timestep, instead of the correct sequence $[0.1, 0.1, 0.1]$.

To remove the transient spikes of flow rates caused by the quantization error, we smooth the time series of flow rates with a rolling median filter. We choose the rolling

median filter because the median filter preserves the location and shape of edges, whereas a rolling mean filter would smear the edges.

We use a rolling median with a window size of 3 samples. The result of the rolling median filter is a majority vote for each sliding window. A negative consequence of the filter is the removal of short water consumption activities, such as filling a glass of water.

5. Detecting Human Actions Using Smart Water Meter Measurements

The main objective of our work is to use smart water meter measurements for HAER within a residence. Previous research has shown that it is possible to identify human actions by observing their interactions with specific objects or devices [31–33]. We are considering the interaction with household water tapping points (e.g., taps, showers, toilet flushes, washing machines). Therefore, we have to consider different types of water consumption in private households:

Type I   Direct water usage, where the resident actively starts and stops water use (e.g., tap, shower).

Type II   Water consumption that is partially autonomous but requires prior direct user interaction (e.g., washing machine, dishwasher).

Type III   Water consumption that does not require user interaction (e.g., automated garden watering).

When inferring human actions from household water consumption, it must be ensured that only the start and stop of water consumption for Type I consumption and only the start of water consumption for Type II consumption are interpreted as human action. Type III consumption must not be interpreted as a human activity at all. Since Type III consumption is rare in Germany, we ignore this category. Moreover, automatically controlled or programmable Type II consumers are not considered in this article.

To quantify the intensity of human water usage of Type I and Type II consumption, we convert the pre-processed cumulative water consumption time series $V(t)$ into a time series of volume flow rates $q(t)$. The conversion process computes the difference between consecutive samples regarding the timespans between the consecutive samples.

Figure 5 shows, for an exemplary data extract over one day, both the cumulative measured values $V(t)$ and the volume flow rates time series $q(t)$ derived from the measurements.
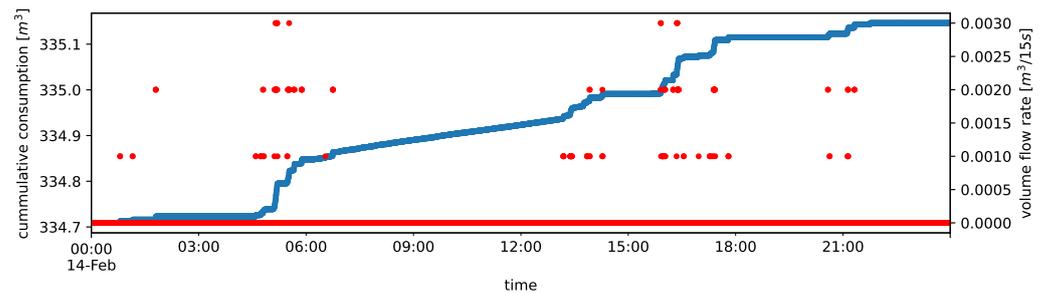

Figure 5. A sample extract of the series of measured values after pre-processing $V(t)$ over one day (blue), as well as the derived volume flow rates series $q(t)$ (red). (Household: WM-A-08; Day: 2022-02-14).

If we assume that when the pipe system in the household is completely sealed, the incoming volume flow rate at time t is equal to the sum of the individual outgoing volume flow rates of all water tapping points in the residence at time t. The following applies:

$$q(t) = \sum_{p \in P} q_p(t) \tag{1}$$

where P is the set of all water tapping points in the residence and $q_p(t)$ is the individual volume flow rate of the tapping point $p \in P$ at time t, with the same temporal and quantitative resolution as $q(t)$. From this definition, it follows that:

$$q(t) \neq 0 \implies \exists p \in P : q_p(t) > 0 \tag{2}$$

This means that at any time t where water is being used throughout the house, water must be used at least at one tapping point $p \in P$.

For Type I consumption, per the definition, at least the start and end of any consumption is the result of human activity. For Type II consumption, at least the start of consumption must result from human action. In summary, we can consider a human activity at any time t if the following conditions are fulfilled:

(i) $q(t) > 0$ and
(ii) $q(t-1) = 0$

This means that we consider those times t with human activity at which a water flow ($q(t) > 0$) takes place and at the previous time $t - 1$ there was no water flow ($q(t-1) = 0$). This assumption is because, for example, a constantly running water tap does not imply constant human activity, but only turning on the tap requires human action.

In practice, however, it cannot be assumed that water pipes and water tapping points are completely sealed and that measurements are always accurate. Furthermore, minor leaks, such as a dripping tap, cannot be measured since the 1-liter quantization resolution of the series $V(t)$ is smaller than the leaking volume. Consequently, we have to further adjust Equation (1) by adding a noise term $n(t)$ that describes the leaks or measurement deviations at time t:

$$q(t) = n(t) + \sum_{p \in P} q_p(t) \tag{3}$$

On this assumption, the following now applies:

$$q(t) > n(t) \implies \exists p \in P : q_p(t) > 0 \tag{4}$$

If $n(t) > 0$, the implication in Equation (2) does not apply anymore without restrictions. Therefore, to determine when water is consumed, at least one tapping point, $q(t) > n(t)$, must apply. Since the term $n(t)$ depends on multiple factors (e.g., the pipe system, leaking tapping points) and is not constant, it cannot be specified clearly. Therefore, we define a global fixed threshold value *thres*, for which $thres > max(n(t))$ holds.

Further, we assume that we consider only such times t as human actions for which both of the following apply :

(i) $q(t) \geq thres$ and
(ii) $q(t-1) = 0$

In Figure 6, the times for which the two conditions mentioned above are fulfilled using a sampling frequency of $0.0\bar{6}$ Hz with $thres = 1\ l$ are plotted with orange lines. At each of these time points, human activity can be assumed.
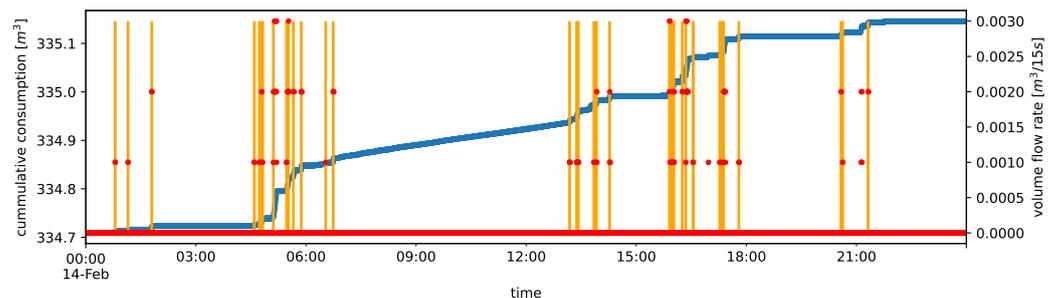

Figure 6. From the series of measured values, $V(t)$-derived times of human activity events (orange lines) in the analogous time slot to Figure 5.

6. Evaluation

In Section 5, we introduced a novel approach to infer human activities based on smart water meter data. In the following section, we evaluate this approach.

The evaluation is divided into three parts: First, in Section 6.1, we utilize a separately collected evaluation data set to perform a threshold sweep. The result is a precision–recall Curve over the threshold sweep. Then, in Section 6.2, also for a separate evaluation data set, we consider only those periods when the residents were out of the house and focus on possible false positive (FP) classifications. FP classifications are the most severe errors for the use-case of emergency detection. With Section 6.3, we conclude the evaluation by detecting activity/inactivity periods as a cross-check to the absence evaluation based on the overall data set.

6.1. Precision–Recall Curve of Activity Detection

The first evaluation step is to investigate the precision and recall of activity detection. We compute the precision–recall curve over a threshold sweep to make the evaluation independent of the hyperparameter *thresh*. Since the data set collected in Section 3 does not contain any labels, we first collected a separate evaluation data set in which the individual activities are annotated.

6.1.1. Evaluation Data Collection

To evaluate the activity detection's precision and recall, we collected a separate evaluation data set in three test households (households WM-A-01, WM-A-08, and WM-A-11). During defined timespans L_p, we performed two typical household activities associated with water consumption and recorded the time at which they were performed. Specifically, we turned on various taps for 30 s and flushed the toilets. We avoided concurrent water consumption and always waited for at least 5 min between two activities. In addition to the cumulative water consumption $V(t)$, we thus obtain an additional annotation data set L for the timespans $[l_b : l_e] \in L_p$ containing:

$$timestamp = \text{UNIX timestamp (in ms) of an activity/water consumption}$$
$$note = \text{Note/category of an activity/water consumption}$$

It can be assumed that for all $t \in L_p$, if $t \notin L$, then there was no activity/water consumption because we ensured that no residents were present during data recording.

The activity label data set is available as a supplement to this article on *Zenodo* at https://doi.org/10.5281/zenodo.7585347.

6.1.2. Evaluation Metric

For calculating the precision–recall curve of activity detection, we consider the annotated timespans L_p. For $t \in L_p$ and $thres \in \mathbb{R}_+$ we can define any classification as:

true positive (TP), if:	$t \in L$	and	$q(t) \geq thres$	and	$q(t-1) = 0$	
true negative (TN), if:	$t \notin L$	and	$q(t) < thres$	or	$q(t-1) > 0$	
false positive (FP), if:	$t \notin L$	and	$q(t) \geq thres$	and	$q(t-1) = 0$	
false negative (FN), if:	$t \in L$	and	$q(t) < thres$	or	$q(t-1) > 0$	

The annotated time stamps contain a temporal inaccuracy caused by the manual annotation and caused by the time delay of some water consumers' mechanisms, such as flushing the toilet and the subsequent refilling. We introduce a temporal tolerance for the activity-detection timestamps to accommodate the temporal uncertainty. We define the parameter *tol* as a tolerance timespan that allows some flexibility in the timing of the detected action relative to the annotated timestamps. With this assumption, we classify any time point $t \in L_t$ as

TP, if: $\quad t \in L : \exists t' \in [t - tol : t + tol] : q(t') \geq thres \quad$ and $\quad q(t'-1) = 0$

TN, if: $\quad t \notin L : \exists t' \in [t - tol : t + tol] : q(t') < thres \quad$ or $\quad q(t'-1) > 0$

$$\text{FP, if:} \quad t \notin L : \exists t' \in [t - tol : t + tol] : q(t') \geq thres \quad \text{and} \quad q(t' - 1) = 0$$
$$\text{FN, if:} \quad t \in L : \exists t' \in [t - tol : t + tol] : q(t') < thres \quad \text{or} \quad q(t' - 1) > 0$$

For the precision–recall curve, we sweep over the value *thres* and calculate the *precision* and *recall* for each *thres* separately.

On the time axis, analogous to the sampling frequency, we consider 15 s sections within each interval $\in L_t$.

6.1.3. Evaluation Results

The evaluation data set contains 428 min of three households and includes 66 activities. Figure 7 shows the precision–recall curve by sweeping over the value *thres* in the interval from 0 to 5 (step: 1.0) with *tol* = 30 s.
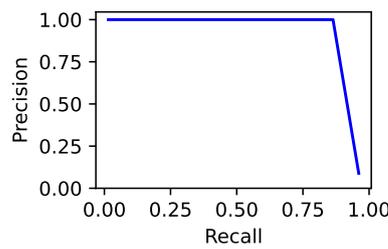
Figure 7. Precision–recall curve by sweeping over the value *thres* in the interval from 0 to 5 (step: 1.0) with *tol* = 30 s.

Since there are no TP or FP classifications in the evaluation data set for a *thres* ≥ 4 ($\#TP = 0$ and $\#FP = 0$), it is impossible to calculate a precision above this threshold.

With a precision of 0.86 and a recall of 1.00, *thres* = 1 liter can be identified as the most appropriate threshold value. For a more detailed description of the quality of the classification, Table 4 shows the recall of the recognition, categorized by individual activities/water consumptions with *thres* = 1.

Table 4. Recall depending on individual activities/water consumers with *thres* = 1 in the evaluation data set.

	#Samples	# TP	# FN	Recall
toilet flush	38	37	1	0.97
water tap (30 s)	28	20	8	0.71

6.2. FP Classifications during Absence Periods

In the second step of the evaluation, we focus on possible FP classifications, as these represent the most critical errors for the use-case of emergency detection. To achieve this, we consider periods when it is known that no human activity could have occurred because nobody was at home. Any activity detected during these periods can be directly categorized as an FP classification.

6.2.1. Evaluation Data Collection

A separate data set is collected in two test households (households WM-A-08 and WM-A-11) for this evaluation. Besides the readings from the smart water meters, the absence times of the residents were collected using a diary method.

We handed the residents a log form in which they recorded those time points when no residents were in the house. The residents were further asked to note whether a washing machine or a dishwasher was running during a period of absence.

In addition to the cumulative water consumption $V(t)$, we thus obtain an additional set A containing:

$[ta_l : ta_b]$ = Period of absence from ta_l to ta_b as UNIX timestamp (in ms)

i = Note on active washing machine/dishwasher appliances during the period

Data collection occurred in household WM-A-08 from 2022-11-09 to 2022-12-15 and in household WM-A-11 from 2022-11-09 to 2022-12-03.

The data are available as a supplement on *Zenodo* at https://doi.org/10.5281/zenodo.7585347.

6.2.2. Evaluation Metric

We restrict our analysis to periods when we know from the diary entries that nobody was in the house and no washing machine/dishwasher was running. During these periods, it is expected that no activity will be detected. Thus, by definition, it must apply for each period $[ta_l : ta_b] \in A$:

$$\forall_{t \in [ta_l:ta_b]} : q(t) < thres \quad \text{or} \quad q(t-1) > 0 \tag{5}$$

Consequently, all times $t \in [ta_l : ta_b]$ are to be considered as FP, for which applies:

$$q(t) \geq thres \quad \text{and} \quad q(t-1) > 0 \tag{6}$$

6.2.3. Evaluation Results

Out of two households, a total of 3603 min could be used for the evaluation from data set A with the above criteria (WM-A-08: 2011 min; WM-A-11: 1592 min). Within this period, no activities were detected in both households. Thus, there was no FP classification ($\#FP = 0$).

6.3. Maximum Periods of Inactivity

In the third step of our evaluation, we will show that the developed approach for human action detection can detect regular activity events in a normal daily routine. To achieve this, we utilize the data set from Section 3 and calculate an inactivity profile for each household. Therefore, we calculate the time that elapses between two detected activity events. We consider the maximum inactivity time (*max*) as well as the 99% quantile ($Q_{0.99}$) of the inactivity times to exclude any prolonged absences, such as those caused by travel. The results are shown in Table 5, broken down by individual households.

Table 5. Maximum periods of inactivity, as well as 99% quantiles ($Q_{0.99}$) of inactivity per household.

	Times of Inactivity	
	max	$Q_{0.99}$
WM-A-01	12:24	9:50
WM-A-02	1 d, 0:33	5:57
WM-A-03	14:57	10:44
WM-A-04	6:11	5:36
WM-A-05	9:20	5:03
WM-A-06	2 d, 3:08	8:51
WM-A-07	2 d, 17:47	1 d, 6:30
WM-A-08	8:45	4:28
WM-A-09	7:03	5:31
WM-A-10	11:34	6:34
WM-A-11	12:40	6:48
WM-A-12	1 d, 2:45	16:52
WM-A-13	1 d, 12:41	1 d, 4:45
WM-A-14	10:51	7:52
WM-A-15	13:35	6:58
WM-A-16	8:20	5:29
WM-A-17	13:50	8:30

In $Q_{0.99}$, we observe that only two of the 17 households had an inactivity time of more than 1 day: one household had a maximum inactivity time in $Q_{0.99}$ of 16:57 and one of 10:44:00. All the others (n = 13) did not have an inactivity time of more than 10 h in the 99 th quantile. The median inactivity time for all 17 households in $Q_{0.99}$ is 06:58.

7. Discussion

Our research focuses on HAER based on water consumption readings. We use the flowrate feature ($q(t)$), which is common in the related literature [6,8–11,16,19,21,22,24,25], but unlike many other studies [6,17–19], we consider the HAER problem as a binary classification problem. In other words, we do not distinguish between different consumption methods or types of activity, but rather determine whether there is an activity at a given time t or no activity. Thus, instead of HAR, the presented approach is suitable for HAER, which is novel in the research field for using smart water meter data.

Our approach showed an increased precision (*precision* $= 0.86$) compared to previous work on HAR, which was our main research goal. However, it should be noted that the evaluation data set was collected in actual households but under controlled/laboratory-like conditions, with no overlap of different activities and a limited range of activity types. As a result, the recall (*recall* $= 1.00$) of our approach may be biased. In future work, the recall should therefore be re-examined with a more comprehensive evaluation dataset.

As an extension to the precision–recall curve, we have performed a further evaluation in Section 6.2 in two households to investigate whether FP classifications occur over an extended period of time. This evaluation again demonstrates the high precision of the HAER with a total of zero FP classifications. However, it should be noted that periods of absence during which a washing machine or a dishwasher was active were not included in the evaluation. In general, it must be stated that our method for HAER is potentially prone to errors caused by Type II or Type III consumption. This means that if, for example, a washing machine is running, HAER performance may be impaired during this time. In particular, it must be considered that, for example, a washing machine intakes water several times during a washing cycle, but generally only the first intake process is directly related/close in time to the human activity of switching on the washing machine; all other intakes occur with a time delay. The same applies to further Type II consumers.

A further limitation is that Type II devices may be started with a timer or via a smart home, and thus be considered equally a Type III consumer. This cannot be distinguished from a direct human activity with the presented approach.

From a practical standpoint, one major benefit of using smart water meter data for HAER is that it is a fully non-intrusive way of monitoring people within the household. Since smart water meters are already common in many households, this method of HAER does not require any additional sensors to be installed, but solely works with re-using existing data. However, it should be noted that not all commercially available smart water meters offer the high temporal resolution of the smart meter used in this work.

Since water consumption is vital for humans, we can show that re-using the smart water meter data provides a fine-grained activity profile. With Section 6.3 we showed that the median overall household $Q_{0.99}$ of inactivity time was 06:58; for 13 out of 17 households, the $Q_{0.99}$ of inactivity time was less than 10 h. This finding is particularly relevant when further work is conducted to investigate using the activity information from the smart water meter for an AAL system in emergency detection. Using household-specific inactivity thresholds (e.g., 10 h) it is possible, using the disaggregated HAER information from the smart water meter, to detect potential emergencies within the household by detecting atypically long periods of inactivity [4]. In particular, it must then be considered that inactivity also results from the absence of residents.

One of the major benefits of the proposed approach is its ability to perform HAER using only a short timespan of water flow data. With a sample rate of 15 s, the cumulative water consumption of the last 60 s is sufficient to perform the pre-processing of the data as presented in Section 4 and then to perform the classification according to Section 5. This makes it possible to cache the measurements locally in the household (e.g., using a ring buffer) and to perform HAER on an edge system (e.g., a single-board computer) in a streaming mode. This edge computing eliminates the need to transmit water consumption data to an external third party, which increases privacy. There is also no need for a constantly

active internet connection, which can increase the system availability/reliability. Only dedicated emergency situations require external communication to notify third parties.

8. Conclusions and Further Work

On the one hand, this article outlines the data collection of a comprehensive data set of water consumption measurements. We used commercial smart meters in 17 German households between September 2020 and September 2022 to record water consumption measurements with a median sample rate of $0.08\overline{3}$ Hz, which is the second-highest among data sets collected using smart meters in the related literature [8,9,11,14,21,22]. Only Nguyen et al. [8] describe that their system can achieve a higher sample rate. However, our data set also contains long gaps and faulty records, which may negatively impact its quality, as discussed in detail within Section 3.3. A further significant impediment to the possibility of reusing the dataset is the absence of label data. The collected data set is now publicly available on *Zenodo* for further research.

On the other hand, we have presented a novel approach for HAER using water consumption measurements collected from commercial smart meters. This noninvasive approach facilitates the monitoring of human activity events within a household without the need for installing additional sensors. Given that water is a vital necessity for humans and is regularly consumed, such as in toilet use, our findings demonstrate the potential of utilizing smart water meter data to generate fine-grained activity profiles of household residents.

Our approach assumes that a change in the water flow rate from no flow ($q(t) = 0$) to a positive flow above a certain threshold ($q(t) \geq thres$) indicates human activity. Our method requires several data pre-processing steps before human action detection can be performed.

By using separately collected labeled data from three households over a period of 428 minutes, we could show that our approach can detect human activity events with high precision (*precision* = 0.86) and recall (*recall* = 1.00). In addition, we could also show that activity events are regularly detected during a typical daily routine.

In future work, the approach should be verified with additional data sets and benchmarked against other approaches for HAR/HAER, such as those based on water pressure measurements. Furthermore, it should be investigated how potential errors/bias can be compensated for, primarily caused by Type II or Type III consumption.

Using advanced algorithms, the inferred information about human activities within the residence can later be used for intelligent AAL systems. For example, information about human activity derived from water consumption could replace the acknowledgment function of home emergency call systems or be used for emergency detection without proprietary sensors.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

Glossary/Abbreviations

Edge system	A decentralized computing infrastructure that processes and analyzes data at or near the source of data generation, reducing the latency and enhancing real-time data processing capabilities.
Interpolation	The method of estimating or calculating intermediate values between known data points based on mathematical algorithms.
Precision–recall curve	A graphical representation that illustrates the trade-off between precision and recall for a binary classification model or algorithm. It aids in evaluating the model's performance and determining an optimal threshold for decision-making.
Real-time clock	A clock device that provides accurate and precise timekeeping, typically used in systems that require synchronized timing for various operations or data logging. It enables events and actions to be timestamped and synchronized with other devices or processes.
Resampling	The process of changing the sampling rate of a signal or data by interpolating or decimating the existing samples. It allows for adjusting the temporal resolution of data to match a desired rate or to facilitate further analysis.
Rolling median	A statistical method that calculates the median value within a sliding window or moving interval of data points. It helps to smooth out variations and outliers, providing a robust measure of central tendency over time or spatial domains.
Smoothing	A technique used to reduce noise or fluctuations in data by applying mathematical filters or algorithms.

The following abbreviations are commonly used in this manuscript:

AAL	ambient assisted living
ADL	activity of daily living
HAR	human activity recognition
HAER	human activity event recognition
FN	false negative
FP	false positive
NTP	network time protocol
TN	true negative
TP	true positive
VPN	virtual private network

References

1. Ann, O.C.; Theng, L.B. Human activity recognition: A review. In Proceedings of the 2014 IEEE International Conference on Control System, Computing and Engineering (ICCSCE 2014), Penang, Malaysia, 28–30 November 2014; pp. 389–393. [CrossRef]
2. Hussain, Z.; Sheng, M.; Zhang, W.E. Different Approaches for Human Activity Recognition: A Survey. *arXiv* **2019**, arXiv:1906.05074.
3. Eldib, M.; Deboeverie, F.; Philips, W.; Aghajan, H. Behavior analysis for elderly care using a network of low-resolution visual sensors. *J. Electron. Imaging* **2016**, *25*, 041003. [CrossRef]
4. Wilhelm, S. Exploiting Home Infrastructure Data for the Good: Emergency Detection by Reusing Existing Data Sources. In *Advances in Intelligent Systems and Computing*; Springer International Publishing: Berlin/Heidelberg, Germany, 2021; pp. 51–58. [CrossRef]
5. Floeck, M.; Litz, L. Inactivity patterns and alarm generation in senior citizens' houses. In Proceedings of the 2009 European Control Conference (ECC), Shangai, China, 16–18 December 2009. [CrossRef]
6. Koketsu, T.; Ohno, Y.; Ishihara, T.; Nishimoto, Y.; Kobayashi, K. Monitoring Living Activities of the Elderly Living Alone Using a Lifeline. *Jpn. J. Appl. Healthc.* **2018**, *13*, 12–19. [CrossRef]
7. Clifford, E.; Mulligan, S.; Comer, J.; Hannon, L. Flow-Signature Analysis of Water Consumption in Nonresidential Building Water Networks Using High-Resolution and Medium-Resolution Smart Meter Data: Two Case Studies. *Water Resour. Res.* **2018**, *54*, 88–106. [CrossRef]

8. Nguyen, K.A.; Sahin, O.; Stewart, R.A.; Zhang, H. Smart Technologies in Reducing Carbon Emission. In Proceedings of the Proceedings of the 9th International Conference on Machine Learning and Computing, Singapore, 24–26 February 2017. [CrossRef]

9. Wang, J.; Cardell-Oliver, R.; Liu, W. Discovering routine behaviours in smart water meter data. In Proceedings of the 2015 IEEE Tenth International Conference on Intelligent Sensors, Sensor Networks and Information Processing (ISSNIP), Singapore, 7–9 April 2015. [CrossRef]

10. DeOreo, W.B.; Heaney, J.P.; Mayer, P.W. Flow trace analysis to access water use. *J. Am. Water Work. Assoc.* **1996**, *88*, 79–90. [CrossRef]

11. Chen, F.; Dai, J.; Wang, B.; Sahu, S.; Naphade, M.; Lu, C.T. Activity analysis based on low sample rate smart meters. In Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining-KDD'11, San Diego, CA, USA, 21–24 August 2011; ACM Press: New York, NY, USA, 2011. [CrossRef]

12. Kim, Y.; Schmid, T.; Charbiwala, Z.M.; Friedman, J.; Srivastava, M.B. NAWMS: Nonintrusive Autonomous Water Monitoring System. In Proceedings of the 6th ACM Conference on Embedded Network Sensor Systems-SenSys'08, Raleigh, NC, USA, 5–7 November 2008; ACM Press: New York, NY, USA, 2008. [CrossRef]

13. Thomaz, E.; Bettadapura, V.; Reyes, G.; Sandesh, M.; Schindler, G.; Plötz, T.; Abowd, G.D.; Essa, I. Recognizing water-based activities in the home through infrastructure-mediated sensing. In Proceedings of the 2012 ACM Conference on Ubiquitous Computing-UbiComp'12, Pittsburgh, PA, USA, 5–8 September 2012. [CrossRef]

14. Cardell-Oliver, R. Discovering water use activities for smart metering. In Proceedings of the 2013 IEEE Eighth International Conference on Intelligent Sensors, Sensor Networks and Information Processing, Melbourne, Australia, 2–5 April 2013. [CrossRef]

15. Wagner, M. Einblicke in den Wasserverbrauch: Charakterisierung des Wasserverbrauchs mittels Machine Learning. In Proceedings of the INFORMATIK 2020, Gesellschaft für Informatik, Bonn, Germany, 28 September–2 October 2021; pp. 345–368. [CrossRef]

16. Bennett, E.R.; Linstedt, K.D. *Individual Home Wastewater Characterization and Treatment*; Environmental Resources Centre, Colorado State University: Fort Collins, CO, USA, 1975.

17. Froehlich, J.; Larson, E.; Saba, E.; Campbell, T.; Atlas, L.; Fogarty, J.; Patel, S. A Longitudinal Study of Pressure Sensing to Infer Real-World Water Usage Events in the Home. In *Lecture Notes in Computer Science*; Springer: Berlin/Heidelberg, Germany, 2011; pp. 50–69. [CrossRef]

18. Hu, L.; Chen, Y.; Wang, S.; Jia, L. A Nonintrusive and Single-Point Infrastructure-Mediated Sensing Approach for Water-Use Activity Recognition. In Proceedings of the 2013 IEEE 10th International Conference on High Performance Computing and Communications & 2013 IEEE International Conference on Embedded and Ubiquitous Computing, Zhangjiajie, China, 13–15 November 2013. [CrossRef]

19. Yanyan, W.; Xiaoyan, Y.; Qinyun, G.; Qiang, Z. Design of daily-use water monitoring platform for the elderly living alone in intelligent community. In Proceedings of the 2021 International Conference on Big Data Analysis and Computer Science (BDACS), Kunming, China, 25–27 June 2021. [CrossRef]

20. Salomons, E.; Sela, L.; Housh, M. Hedging for Privacy in Smart Water Meters. *Water Resour. Res.* **2020**, *56*. [CrossRef]

21. Cardell-Oliver, R.; Carter-Turner, H. Activity-aware privacy protection for smart water meters. In Proceedings of the 8th ACM International Conference on Systems for Energy-Efficient Buildings, Cities, and Transportation, Coimbra, Portugal, 17–18 November 2021. [CrossRef]

22. Cardell-Oliver, R. Water use signature patterns for analyzing household consumption using medium resolution meter data. *Water Resour. Res.* **2013**, *49*, 8589–8599. [CrossRef]

23. Froehlich, J.E.; Larson, E.; Campbell, T.; Haggerty, C.; Fogarty, J.; Patel, S.N. HydroSense. In Proceedings of the 11th international conference on Ubiquitous Computing, Orlando, FL, USA, 30 September–1 October 2009. [CrossRef]

24. Srinivasan, V.; Stankovic, J.; Whitehouse, K. FixtureFinder: Discovering the existence of electrical and water fixtures. In Proceedings of the 12th International Conference on Information Processing in Sensor Networks-IPSN '13, Philadelphia, PA, USA, 8–11 April 2013; ACM Press: New York, NY, USA, 2013. [CrossRef]

25. Willis, R.M.; Stewart, R.A.; Giurco, D.P.; Talebpour, M.R.; Mousavinejad, A. End use water consumption in households: Impact of socio-demographic factors and efficient devices. *J. Clean. Prod.* **2018**, *60*, 107–115. [CrossRef]

26. Kolter, J.Z.; Johnson, M.J. REDD: A public data set for energy disaggregation research. In Proceedings of the Workshop on Data Mining Applications in Sustainability (SIGKDD), San Diego, CA, USA, 21–24 August 2011; Volume 25, pp. 59–62.

27. Kelly, J.; Knottenbelt, W. The UK-DALE dataset, domestic appliance-level electricity demand and whole-house demand from five UK homes. *Sci. Data* **2015**, *2*, 150007. [CrossRef] [PubMed]

28. Ethikkommission DGP, e.V. Fragen zur Ethischen Reflexion. Available online: https://dg-pflegewissenschaft.de/wp-content/uploads/2017/05/FragenEthReflexion.pdf (accessed on 23 August 2019).

29. Jakob, D.; Wilhelm, S.; Gerl, A. Data Privacy Management (DPM)—A Private Household Smart Metering Use Case. In Proceedings of the Mensch und Computer 2020-Workshopband, Magdebug, Germany, 6–9 September 2020; Hansen, C., Nürnberger, A., Preim, B., Eds.; Gesellschaft für Informatik e.V.: Bonn, Germany, 2020. [CrossRef]

30. Sweeney, L. k-anonymity: A model for protecting privacy. *Int. J. Uncertain. Fuzziness-Knowl. Based Syst.* **2002**, *10*, 557–570. [CrossRef]

31. Perkowitz, M.; Philipose, M.; Fishkin, K.; Patterson, D.J. Mining models of human activities from the web. In Proceedings of the 13th Conference on World Wide Web-WWW '04, New York, NY, USA, 17–20 May 2004. [CrossRef]
32. Philipose, M.; Fishkin, K.; Perkowitz, M.; Patterson, D.; Fox, D.; Kautz, H.; Hahnel, D. Inferring Activities from Interactions with Objects. *IEEE Pervasive Comput.* **2004**, *3*, 50–57. [CrossRef]
33. Wyatt, D.; Philipose, M.; Choudhury, T. Unsupervised activity recognition using automatically mined common sense. In Proceedings of the AAAI, Marina del Rey, CA, USA, 1–3 June 2005; Volume 5, pp. 21–27.