PressureML: Modelling Pressure Waves to Generate Large-Scale Water-Usage Insights in Buildings

Tanmaey Gupta
Microsoft Research India
v-tanmgupta@microsoft.com

Anupam Sobti
Plaksha University
anupam.sobti@plaksha.edu.in

Akshay Nambi
Microsoft Research India
akshay.nambi@microsoft.com

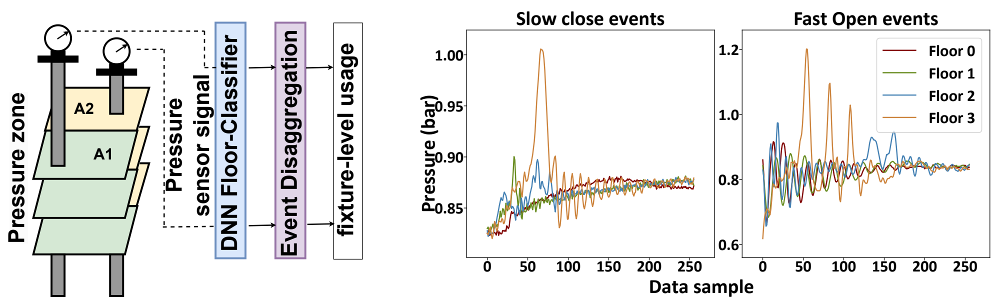
Abstract

Several studies have indicated that delivering insights and feedback on water usage has been effective in curbing water consumption, making it a pivotal component in achieving long-term sustainability objectives. Despite a significant proportion of water consumption originating from large residential and commercial buildings, there is a scarcity of cost-effective and easy-to-integrate solutions that provide water usage insights in such structures. Furthermore, existing methods for disaggregating water usage necessitate training data and rely on frequent data sampling to capture patterns, both of which pose challenges when scaling up and adapting to new environments. In this work, we aim to solve these challenges through a novel end-to-end approach which records data from pressure sensors and uses time-series classification by DNN models to determine room-wise water consumption in a building. This consumption data is then fed to a novel water disaggregation algorithm which can suggest a set of water-usage events, and has a flexible requirement of training data and sampling granularity. We conduct experiments using our approach and demonstrate its potential as a promising avenue for in-depth exploration, offering valuable insights into water usage on a large scale.

1 Introduction

Efficient water use is essential for long-term sustainability goals. To achieve this, it's crucial to provide individuals and organizations with tools to optimize water usage effectively. One such approach is offering water usage insights, which informs decisions about water use and infrastructure design. Studies show that real-time feedback through smart meters and digital solutions can reduce water consumption by up to 30% [1, 2]. Beyond reducing consumption, water usage insights can also lower energy usage in water systems, ensure fair billing, and detect leaks. While a significant percentage of water consumption can be attributed to commercial and residential buildings due to high occupant density, there is a lack of viable solutions which provide these insights on such large-scales. In this work, we present `PressureML`, an end-to-end and scalable approach leveraging DNN to generate real-time water usage feedback in buildings. Delivering such feedback involves two stages: *1. Recording water usage data at an appropriate level of detail in time and space*, and *2. Using disaggregation techniques to identify specific fixture or appliance-level usage.*

Tackling Climate Change with Machine Learning: workshop at NeurIPS 2023.



(a) `PressureML` architecture

(b) Test setup pressure waveforms

Figure 1

Climate impact Our work aims at reducing water consumption and enabling sustainable water management practices. According to the UN World Water Development Report [3], this directly translates to reduction in energy used in treatment, transportation and waste-management of water, ultimately reducing carbon emissions. Further, it reduces the supply pressure on natural ecosystems and groundwater, thus protecting them for a sustainable future.

Related work Current methods for recording total water usage primarily involve installing electromagnetic or ultrasonic-pulse water meters at the main inlet of individual houses [4]. They are, however, (i) Difficult to use in a multi-inlet setting, (ii) Economically expensive to install for every house or area, (iii) Challenging to retrofit by cutting pipes, and (iv) Require high maintenance due to sensitive components. HydroSense[5] takes a different approach by calculating volume usage from pressure sensors for a set of fixtures in a house, but requires extensive labelled data and calibration efforts. Our approach, however, is able to get volume usage for multiple floors simultaneously, and does not require fixture-level training data and calibration. Water-usage event disaggregation research efforts [6, 7, 8] primarily rely on identifying and encoding distinct features and usage patterns associated with different fixtures or appliances. However, current approaches face one or more of the following challenges: (i) Focus on high-frequency input data, (ii) Requirement of environment-specific training data, and (iii) Requirement of additional sensing infrastructure. `PressureML` addresses the aforementioned challenges by presenting a cost-effective solution which can be easily integrated in various water system configurations in residential and commercial complexes.

In this work, we focus on high-rise residential apartments, where each house's areas (such as bathrooms and kitchens) on multiple floors (divided into zones) [9] are connected via dedicated pipelines supplied by a gravity-based distribution network, fed by overhead tanks (OHT). `PressureML` proposes a novel approach that shifts from collecting individual water-usage readings for all areas within each house to obtaining data for each area across all houses. The overall architecture of the solution is presented in figure 1a. A high-resolution pressure sensor is installed at the top of each vertical pipeline within a zone. When an end-user valve is opened or closed, the velocity of water in the pipeline changes rapidly, and transient pressure wave called water hammer is generated and recorded by the pressure sensor. The signature and magnitude of the wave can be characterised by valve-dependent factors such as the rate of change of velocity of water, distance of the source to the sensor, and reflections in the pipeline [10]. Utilizing DNN architectures, we model these time-series waves to identify the specific floor of water usage. While this modelling inherently requires minimal training data due to significant inter-class variability, cross-deployment adaptability, and potential use of physics-based simulations, we employ data augmentation techniques and transfer learning to enhance performance, for which we explore and evaluate multiple pre-training architectures and datasets. The aggregate water consumption for each floor and area is then calculated by analyzing the duration of pressure drop events. Subsequently, this data is input into the second stage of `PressureML`, where it is disaggregated into valve-level events. We describe this stage in detail in section 3.

2 Pressure wave classification

Experiment setup and data To develop and evaluate the pressure wave-classification approach, we simulated a 4-floor pressure zone of a building using a test setup consisting of a network of pipes and user-operated taps installed across 4 floors, fed by an OHT. A pressure sensor with a sampling frequency of 500Hz was installed at the outlet valve of the OHT. While collecting data, the tap at each floor was opened and closed at 2 different speeds - fast (0.5-1 sec), and slow (1.5-3 sec). 1 experiment iteration at a floor consists of 4 events: Open and Close at 2 speeds. Experiments being manual, a small dataset of 4 such iterations per floor was recorded, leading to a total of 32 open samples and 32 close samples across 4 classes (floors). A few recorded waveforms are shown in figure 1b.

Event extraction First, the transient section of a pressure wave is extracted from a buffered or continuous stream of data. After passing through a low-pass filter to eliminate noise and highlight the underlying pattern of the signal, rapid pressure changes (denoting open/close events) are found using a forward difference quotient as: $\Delta p(t) = \frac{p(t+k)-p(t)}{k}$. To accommodate different rates of change in pressure due to different valve-operation speeds, multiple such quotients are computed with varying $k(50, 100, 500)$, whose value is determined in accordance with the sampling frequency of the sensor. Event starting points are marked when the quotient with the largest average value in an event exceeds an empirical threshold (0.01 in our case), while event end points are marked when majority of the quotients are below a threshold (0.005 in our case) for k continuous values. All such events extracted are upsampled/downsampled to have a uniform length of 256 points.

Data augmentation To prevent DNN models used for event classification from over-fitting due to limited training data, we employ data augmentation, which has previously shown benefits even in time-series classification (TSC) tasks [11]. A recursive and incremental augmentation approach is used, involving transformations suited for our TSC task. These include *i. Jittering [11]*, which adds Gaussian noise, *ii. Window warp [12]*, which stretches and constricts a time-window, *iii. Magnitude warp [12]*, which randomly stretches and constricts across magnitude dimension, and *iv. DGW-sD [12]* , which uses discriminative shape descriptors for guided warping. Using these augmentations, we increase our dataset size four times to 128 open samples and 128 close samples.

Transfer learning To classify pressure waves into floors of their origin, we use a Transfer Learning approach, which, after vision-based tasks, has recently shown performance gains in time-series classification tasks(TSC) as well [1]. We explore this approach across two axis : state-of-the-art TSC models, and time-series datasets, used for pre-training. We pre-train each model on each of the source dataset and then finetune it on our collected pressure dataset. We handpick 4 datasets from the UCR Time Series archive [13], which have >1000 labelled samples, 1 feature dimension, vary in terms of number of classes, sequence length and source of data, and naturally would fit for our downstream task. For models with sequence length-dependent blocks like LSTM and Attention, we resample the source length to match with the target length of 256. We also include in this analysis an additional modified dataset (PressureLD, with sequence length =256 and number of classes =4) consisting of pressure sensor signals recorded to detect leaks in a water distribution system [14]. This helps to consider the effect of pre-training on data from the same modality as that of the target task.

Aggregate volume calculation A particular water usage instance is bounded by an open(start) and close(end) event. Based on these floor-wise classified events, a square waveform is computed for each floor with y-axis representing pressure and x-axis representing time. Area under the curve (AUC) is calculated, with static pressure level as the base. Entire timeline is then broken into windows of width w, and scaled area falling in that window is denoted as its total water volume consumed. This approach works for overlapping events as well, wherein greater pressure drop will give higher AUC value for that time-frame. Since water-usage insights do not require real-time updates, we analyse data and generate insights at the end of each day. This allows employing statistical usage heuristics for disaggregation, and can provide more context for classification.

3 Disaggregating Events

This section describes the second stage of our proposed solution, which provides fixture-level water usage insights from aggregate volume usage of a particular area (like bathroom or kitchen). To make the solution adaptable to diverse usage patterns, varying sampling frequencies and fittings in different water systems, the proposed water disaggregation algorithm doesn't require labelled training data to learn differentiating features, but rather takes as input only general specifications about the system configuration, which is mostly known by domain experts, or can easily be found or measured. This includes the number of fixtures N (different operating modes of each fixture counted separately), their corresponding volume consumed in one usage $v^{(n)}$, and approximate flow-rates $r^{(n)}$. The proposed algorithm uses combinatorial optimization to find the best subset of fixtures for a given usage instance of continuous non-zero flow. For a usage event θ with total volume consumed V, we solve:

$$\arg max_{k^{(n)}} \prod_n p(k^{(n)}v|\theta) \text{ such that } (1-\sigma)V \leq \sum_n k^{(n)}v \leq (1+\sigma)V \text{ where } k^{(n)} \in \mathbb{Z}, n \in [1, N]$$

where σ is the error margin, and $p(k^{(n)}v|\theta)$ is calculated using general heuristics such as possibility of a fixture's contribution given its flow rate ($r^{(n)}$) and actual water used in each window w of θ,

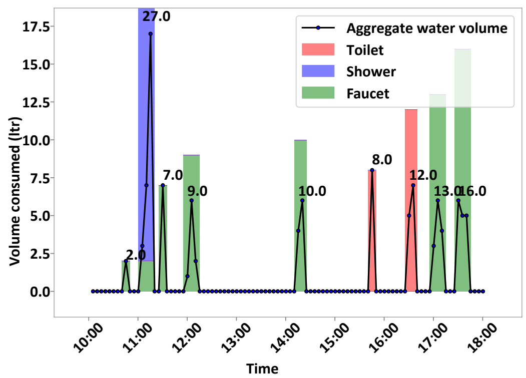


Figure 2: Disaggregated events for total water consumption of a bathroom in a residential apartment.

Table 1: Fine-tuning F1 score (Open/Close events) for different models(columns) and datasets(rows)

	ResNet[15]	InceptionTime[16]	TST[17]	LSTM-FCN[18]	ConvTran[19]
FordA	1.0/1.0	1.0/0.937	1.0/0.876	1.0/0.968	1.0/0.811
StarLightCurves	1.0/1.0	1.0/1.0	0.934/0.846	1.0/0.968	1.0/0.938
ECG5000	1.0/1.0	0.968/0.937	0.746/0.875	1.0/0.937	1.0/0.968
UWaveGestureLibraryAll	1.0/1.0	0.968/0.937	0.938/0.685	1.0/0.937	0.968/1.0
UWaveGestureLibraryAll	1.0/1.0	0.968/0.937	0.938/0.685	1.0/0.937	0.968/1.0
PressureLD	1.0/1.0	0.968/0.968	0.938/0.841	1.0/0.968	1.0/0.938

and the approximate number of large events occurring per day. Additional heuristics can be used to prune the search space and build a better probability distribution. The proposed method is suitable for overlapping events as well as events split across multiple w. As we deal with individual areas, the search space of fixtures is small, making this approach feasible.

4 Results and conclusion

Table 1 presents that the proposed transfer learning approach is able to classify pressure waves with high F1 scores and can enable classification with less training data for `PressureML`. CNN based models perform better as compared to RNN or Attention-based models, possibly due to less inter-dependence on long and short term context, but a significant presence of patterns in the time-series data. Also, the influence of source dataset used for pre-training is small if chosen reasonably, while model used plays a more crucial role in downstream task performance. The results are promising as the best-performing ResNet model achieves an F1 score of 1 for both open and close events. This calls for a more extensive evaluation study, which we plan to do in future work. We do preliminary evaluation of the presented water disaggregation algorithm on an aggregate water consumption data of a house in residential building collected at an interval of 5 min, where labelled data of fixture-level usage in a bathroom was available for a duration of 1 week. Overall accuracy of top-3 suggestions of presented method was 78%. Figure 2 presents the temporal water meter data showing total consumption and disaggregation output of our proposed method for a time window of 8 hours.

This paper introduces `PressureML`, an end-to-end method for deriving water usage insights in large-scale buildings, presenting a two-stage solution by first determining aggregate water consumption per floor area by analyzing pressure waves, and then employing an innovative disaggregation algorithm for fixture-level insights. Our approach demonstrates promising potential for accuracy and scalability, as evidenced by our evaluations. Future work will extend these evaluations, and explore important avenues such as handling overlapping waves, utilizing physics-based simulations for improved feature learning, testing cross-deployment adaptability, and optimizing the disaggregation algorithm's search space. Our contributions aim to advance the understanding and implementation of water usage insights in real-world scenarios.

References

[1] Hassan Ismail Fawaz, Germain Forestier, Jonathan Weber, Lhassane Idoumghar, and Pierre-Alain Muller. Transfer learning for time series classification. In *2018 IEEE international conference on big data (Big Data)*, pages 1367–1376. IEEE, 2018.

[2] Andrea Cominola, Matteo Giuliani, Andrea Castelletti, Piero Fraternali, Sergio Luis Herrera Gonzalez, Joan Carles Guardiola Herrero, Jasminko Novak, and Andrea Emilio Rizzoli. Long-term water conservation is fostered by smart meter-based feedback and digital user engagement. *npj Clean Water*, 4(1):29, 2021.

[3] UN Water. Water and climate change. *The United Nations World Water Development Report*, 2020.

[4] Davide Carboni, Alex Gluhak, Julie A. McCann, and Thomas H. Beach. Contextualising water use in residential settings: A survey of non-intrusive techniques and approaches. *Sensors*, 16(5), 2016.

[5] Jon E Froehlich, Eric Larson, Tim Campbell, Conor Haggerty, James Fogarty, and Shwetak N Patel. Hydrosense: infrastructure-mediated single-point sensing of whole-home water activity. In *Proceedings of the 11th international conference on Ubiquitous computing*, pages 235–244, 2009.

[6] Bingsheng Wang, Zhiqian Chen, Arnold P Boedihardjo, and Chang-Tien Lu. Virtual metering: An efficient water disaggregation algorithm via nonintrusive load monitoring. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 9(4):1–30, 2018.

[7] Vijay Srinivasan, John Stankovic, and Kamin Whitehouse. Watersense: Water flow disaggregation using motion sensors. In *Proceedings of the Third ACM Workshop on Embedded Sensing Systems for Energy-Efficiency in Buildings*, pages 19–24, 2011.

[8] Laura Pastor-Jabaloyes, FJ Arregui, and R Cobacho. Water end use disaggregation based on soft computing techniques. *Water*, 10(1):46, 2018.

[9] Designing domestic water systems for high-rises. `https://www.phcppros.com/articles/16415-designing-domestic-water-systems-for-high-rises`, 2022.

[10] Water hammer waves. `https://web.eng.fiu.edu/arleon/courses/Transient_flows/Lectures/Waterhammer.pdf`, 2022.

[11] Bo Liu, Zhenguo Zhang, and Rongyi Cui. Efficient time series augmentation methods. In *2020 13th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI)*, pages 1004–1009, 2020.

[12] Brian Kenji Iwana and Seiichi Uchida. An empirical survey of data augmentation for time series classification with neural networks. *PLOS ONE*, 16(7):1–32, 07 2021.

[13] Hoang Anh Dau, Anthony Bagnall, Kaveh Kamgar, Chin-Chia Michael Yeh, Yan Zhu, Shaghayegh Gharghabi, Chotirat Ann Ratanamahatana, and Eamonn Keogh. The ucr time series archive. *IEEE/CAA Journal of Automatica Sinica*, 6(6):1293–1305, 2019.

[14] Mohsen Aghashahi, Lina Sela, and M Katherine Banks. Benchmarking dataset for leak detection and localization in water distribution systems. *Data in Brief*, 48:109148, 2023.

[15] Fawaz et al. Deep learning for time series classification: a review. *Data mining and knowledge discovery*, 33(4):917–963, 2019.

[16] Fawaz et al. Inceptiontime: Finding alexnet for time series classification. *Data Mining and Knowledge Discovery*, 34(6):1936–1962, 2020.

[17] George Zerveas, Srideepika Jayaraman, Dhaval Patel, Anuradha Bhamidipaty, and Carsten Eickhoff. A transformer-based framework for multivariate time series representation learning. In *Proceedings of the 27th ACM SIGKDD conference on knowledge discovery & data mining*, pages 2114–2124, 2021.

[18] Fazle Karim, Somshubra Majumdar, Houshang Darabi, and Shun Chen. Lstm fully convolutional networks for time series classification. *IEEE access*, 6:1662–1669, 2017.

[19] Navid Mohammadi Foumani, Chang Wei Tan, Geoffrey I Webb, and Mahsa Salehi. Improving position encoding of transformers for multivariate time series classification. *arXiv preprint arXiv:2305.16642*, 2023.