Water Supply System Dataset: Non-Invasive Sensor Data for Smart Water Pumps

Carmen Cheh
Illinois Advanced Research Center at Singapore Pte Ltd
Singapore, Singapore
carmen@iarcs-create.edu.sg

Aaron Han Yen Tay
Illinois Advanced Research Center at Singapore Pte Ltd
Singapore, Singapore
aaron.tay@iarcs-create.edu.sg

Zhen Wei Ng
Illinois Advanced Research Center at Singapore Pte Ltd
Singapore, Singapore
zhenwei.ng@iarcs-create.edu.sg

Binbin Chen
Singapore University of Technology and Design
Singapore, Singapore
binbin_chen@sutd.edu.sg

Xin Lou
Singapore Institute of Technology
Singapore, Singapore
Illinois Advanced Research Center at Singapore Pte Ltd
Singapore, Singapore
lou.xin@singaporetech.edu.sg

Zaki Masood
Singapore University of Technology and Design
Singapore, Singapore
zaki_masood@sutd.edu.sg

David K.Y. Yau
Singapore University of Technology and Design
Singapore, Singapore
david_yau@sutd.edu.sg

ABSTRACT

Pumps play a central role in water supply systems, especially for high-rise buildings. The inefficiency or failure of pumps leads to high energy cost or water supply disruption. Thus, it is crucial to monitor their health. While advanced pumps have in-built sensors that help to monitor the condition of the pump, most legacy systems are not equipped with such pumps. Thus, we adopt a more practical approach by installing non-invasive sensors to measure power, vibration, water pressure and level of a water pump system in a university building and collected a year of data. In this paper, we present our dataset and detail the interesting events and failures that were captured. We demonstrate the utility of our dataset by extracting specification rules from the data and showing that application of those rules allow us to detect those captured failures.

CCS CONCEPTS

• **Computer systems organization → Maintainability and maintenance**; **Sensors and actuators**.

KEYWORDS

water supply system, dataset, faults

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

ACM ISBN 979-8-4007-0706-3/24/11
https://doi.org/10.1145/3671127.3698790

ACM Reference Format:
Carmen Cheh, Aaron Han Yen Tay, Zhen Wei Ng, Binbin Chen, Xin Lou, Zaki Masood, and David K.Y. Yau. 2024. Water Supply System Dataset: Non-Invasive Sensor Data for Smart Water Pumps. 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/10.1145/3671127.3698790

1 INTRODUCTION

Water supply systems involve the operation of transfer pumps that pump water from the mains supply at the ground level to a storage tank on the roof where the water is held and distributed to the occupants. Over time, the system components such as the pumps will age and develop faults which can cause system operation to be disrupted [2] or result in inefficient pumping operation, which incurs high energy consumption and cost. As such, it is important to monitor the system health and detect when faults occur so that it can be repaired in time before the system operation suffers.

To support such research, a lot of effort has gone into obtaining pump sensor data, especially data with pump faults. Table 1 shows the different datasets available that focus on faults in pump sensor data[1]. Many existing datasets for smart pumps that include faults rely on embedded sensors [8] which provide accurate, reliable, and high-resolution measurements. Those sensors are invasive and present a challenge for deploying on real-world legacy systems that are already operational and cannot tolerate downtime. Moreover, a large fraction of the datasets rely on capturing data from a mini-testbed pump. Then, to induce every faulty condition, the pump needs to be modified in order to generate the corresponding

[1]Note that this list is not exhaustive. However, most of the available datasets that we have seen possess the same properties as the ones listed.

sensor data [4, 7, 9, 10]. The data collected from such controlled experimental setting suffers from insufficient volume and quality. Thus, this approach is very cost-intensive and does not scale well with the different variants of potential faults.

Instead, we need to use sensor data that is derived from real-world settings [11]. Such data will reflect the behavior of the pump under both normal operation and faulty conditions. However, few water pump datasets with long measurement period from non-invasive sensors are available in pubic domain. For example, the faulty pump dataset in [1] covers a period of six months but that dataset does not provide information about the type of sensors or the faults. A lack of insight into the source of data reduces the utility of the dataset because it is difficult to translate the results obtained on the dataset to other systems. Chen et al. [6] also collected real-world sensor data from faulty water pumps. However, their sensors are invasive and only covers nine months of data. Finally, [3] collected sensor data of a water injection centrifugal pump from an offshore system. Although they have amassed huge amounts of data, that data is not publicly released.

To address the lack of real-world data for evaluating fault detection solutions, we engaged a commercial vendor to install 16 non-invasive sensors on a university building's water supply system and collected a year of data. The sensors monitor the operations of the water supply system by providing measurements of the power consumption, vibration, water pressure, and water level.

In this paper, we present our novel water pump dataset with its non-invasive sensor data and recorded faults and events. Our dataset is unique because it captures (1) sensor values from multiple points in a water supply system over an extended period of time, and (2) maintenance event and faults that happened during the system's routine operation. In particular, the faults resulted in inefficient system operation for over a month and was not captured by the water pump's original monitoring system. We provide detailed analysis of the dataset and demonstrate the utility of the dataset by showing how specification-based rules can detect faults that result in observable changes in system behavior.

2 DATASET COLLECTION: WATER PUMP MONITORING SYSTEM SETUP

We instrumented a university building water pump system with a variety of non-invasive sensors. As shown in Fig. 1, the building stores the water from the mains in a tank at the basement and pumps the water to the tank on the rooftop. Consisting of seven levels, exclusive of the basement levels, the building houses labs, classrooms, and offices. The water supply system consists of the potable water system, which supplies drinkable water, and the non-potable system, which supplies water that is meant for cleaning and flushing purposes. Each water system has separate tanks and pumps. There are two pumps per water system. The pumps are run in alternating fashion so as to share the load of pumping. Both the potable and non-potable system's storage tank on the roof have a width of 6 meters, length of 4 meters, and height of 2 meters. The effective capacity of the tank is $22m^3$. The potable and non-potable transfer pumps have a flow rate of 7.8L/s and 12.5L/s and a pump head of 42m and 44m respectively. We installed the following non-invasive sensors (shown in Figs. 1 and 2) on the system:

Table 1: Comparison of pump datasets that have faults.

Dataset	Sensor Install -ation	Sensor Types	Faults	Time Period	Public Release
Bearing Data Center [x]	Invasive	Motor rotation speed, vibration	Induced	10 secs	Yes
IEEE Dataport [2]	Invasive	Voltage, current, motor rotation speed, vibration	Induced	20 secs	Yes
Chang and Park [5]	Invasive	Acceleration, flow rate, current, pressure	Induced	10 mins	Yes
Chen et al. [8]	Non invasive	Vibration	Induced	10 mins	Email authors
Kaggle [1]	Unknown	Unknown	Real	6 mths	Yes
Chen et al. [7]	Invasive	Vibration, noise, temperature, magnetic flux, current	Real	9 mths	Email authors
Blaauw and Hovde [4]	Invasive + Non invasive	Flow rate, thrust, temperature, power, pressure, vibration	Real but not labeled	7 years	No
Our dataset	Non invasive	Vibration, power, pressure, water level	Real, labeled	1 year	Yes

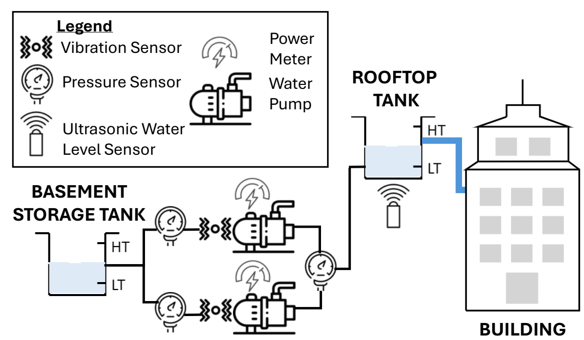


Figure 1: Model of the potable system and the installed sensors on the pumps. The non-potable system model is similar.

- **Water level sensors:** These ultrasound sensors are installed on the exterior of the water level tank and are placed at the bottom. The sensor must be placed flat so that there are no pockets of air between the sensor and the tank bottom, which will affect the quality of the water level readings. Note that these sensors will not work on concrete water tanks.
- **Power meters:** We installed four power meters to record the power consumption of the four pumps. Measurement is conducted using current transformers which are connected to a power monitoring device.
- **Vibration sensors:** These vibration sensors are installed on the connecting bracket of the pump, which is non-invasive to the workings of the pump. They record the movement of the pump as it operates.
- **Pressure sensors:** The pressure sensors are tapped from existing pressure meters located on the pipes. Those sensors are able to record the pressure within the pipe.

All sensor readings are sent to an internet gateway and uploaded to cloud. The frequency of readings is set to 20 seconds except for

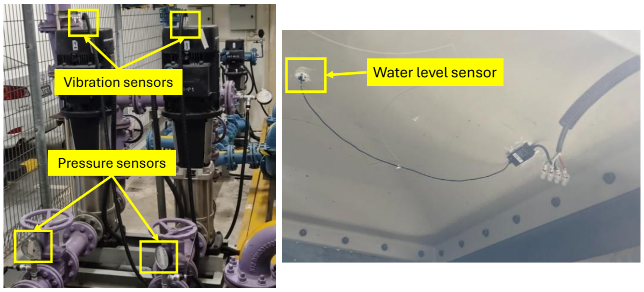


Figure 2: Sensor installation on the water supply system in a university building.

power meters which are updated every minute. The difference in frequencies is due to the limitation of certain sensors in order to conserve power. We collected one year of sensor data from July 2023 - June 2024 which is publicly available in an open repository[2], along with details of system events.

3 DATA ANALYSES AND UTILITY

In this section, we detail our analyses of the data to understand the typical system operation behavior. We also demonstrate the utility of the dataset by showing how specification rules derived from the data can be used to detect the faults recorded in the dataset.

3.1 Regular System Behavior

System thresholds. In the existing legacy system, there are sensor electrodes inside the water tank that send a signal to the control system when the water level touches the electrode. Such sensors are invasive and afford a low-level granularity reading, only allowing the system to detect when the water level has reached a particular threshold (determined by the length of the electrode). The control system is designed to trigger the pump to operate when the water level in the rooftop tank hits a low threshold and stop the pump when the water level hits a high threshold. Based on our collected data, we observe that the low threshold is set at 0.8m and the high threshold is set at 1.1m. For a tank of 2 meters, the chosen thresholds are more conservative to ensure that the tank is never empty or overflowing. However, that small range would require the pump to activate more frequently and for a shorter duration, leading to faster wear and tear.

Sensor correlation. When the pump is activated, the water level in the rooftop increases linearly. The rooftop tank will then be filled in 14 and 9 minutes for the potable and non-potable system respectively. During the pump activation, we observe that the pump's vibration increases (Fig. 3), the pressure from the pipe inlet decreases (Fig. 4a), and the pressure from the pipe outlet increases (Fig. 4b). Due to the close proximity of the two pumps in the same water system, we notice that the vibration sensor of the other idle pump also has a spike in readings (Fig. 3).

Water consumption. Using the water level readings, we can infer the water consumption of the building's occupants per hour. Fig. 5 shows the average daily water consumption for the potable and non-potable system. We see that the peak water consumption for potable system happens between 7-8am when the students and

[2]Available at: https://doi.org/10.5281/zenodo.13808085

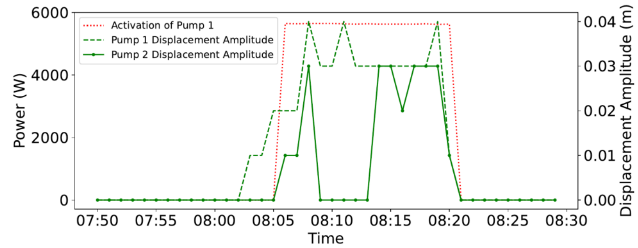


Figure 3: Vibration in pump 1 and 2 during activation of pump 1.

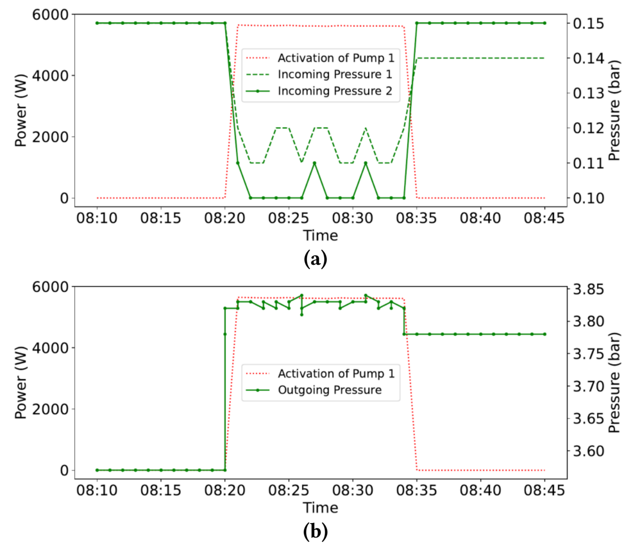


Figure 4: (a) Incoming pressure to pump 1 and 2 and (b) outgoing pressure during activation of pump 1.

staff arrive at the university and fill up their bottles with drinking water. The second smaller peak happens in the afternoon at 1-2pm during the lunch period before decreasing until most people have left the building. The lowest water consumption happens during the night. On the other hand, the non-potable system has peak water consumption during the 10am-12pm period when students have visited the washroom after finishing class and around the lunch period. In the night, the water consumption is the lowest, although it is still higher than the potable system since more water is needed to maintain the building facilities.

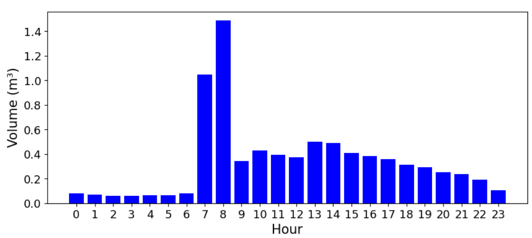
3.2 Captured Faults in Dataset

Through consultation with domain experts, we found the following events or faults have occurred in the dataset (summarized in Table 2).

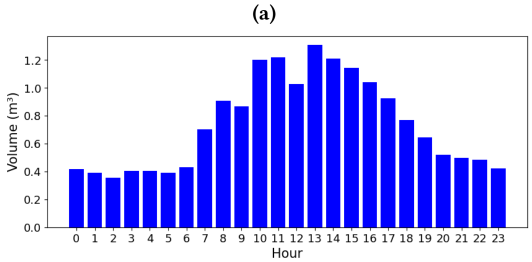
Tank washing. From 11am on 31st August to 1pm on 6th September (a period of one week), the potable tank's water level readings dropped to 0.6m before increasing slowly up to the usual 1.1m (Fig. 6). During this period, the potable pumps operated as per normal. That abnormal event is due to the periodic washing of the water tanks. Since the tank's water would be mostly drained, a connecting tank (which is linked to the rooftop tank for the purposes of providing backup storage) is used for the supply of potable water

Table 2: Summary of events and faults captured in the dataset.

Event Type	Event	Description	System	Duration
System Maintenance	Tank Washing	Drop in potable tank water level reading to abnormally low level before climbing slowly over the period of a week.	Potable	2023 Aug 31 - Sep 6
Fault	Hardware	PLC failed to receive signal to activate pump when water hits low water level threshold. Causes pump activation to kick in much later when water level is very low.	Non potable	2023 Aug 6 2023 Sep 10, 11, 12, 13, 14, 17, 19, 20, 21, 22, 23, 25, 26, 27, 28, 29, 30 2023 Oct 2, 3, 4, 5, 7, 8, 9, 10, 11, 12, 16, 17, 18, 19, 20, 21, 23, 24, 25, 26
	Pump	Pump failure causes it to stop working during activation. So pump operates for short duration and either requires multiple activations to pump the water to high level threshold or requires backup pump to kick in.	Non potable	2023 Sep 8, 11, 12, 13, 18, 20, 21, 22, 23, 25, 26, 27, 28, 29, 30 2023 Oct 1, 2, 3, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 30 2024 Mar 13
			Potable	2023 Sep 6 2023 Sep 9 2023 Dec 27



(a)



(b)

Figure 5: Average daily water consumption for (a) potable water system and (b) non-potable water system.

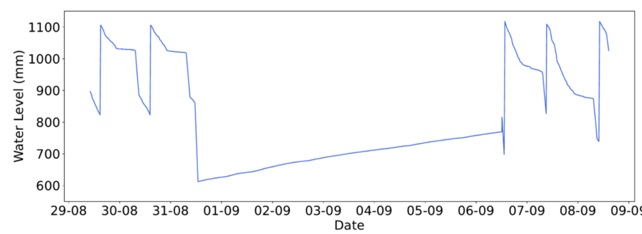
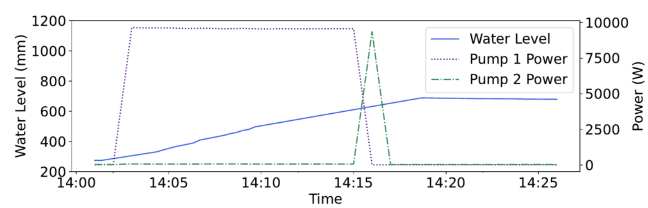
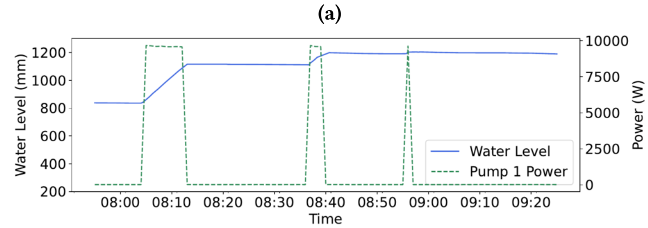


Figure 6: Washing of potable water tank that causes abnormal water level readings.



(a)



(b)

Figure 7: Examples of anomalous pumping operations. (a) Both pumps activate to cause water level to increase. (b) One pump activates multiple times to cause water level to reach high threshold.

to the occupants. The washing of the tank may have temporarily displaced or affected the ultrasonic water level sensor.

Hardware fault. The water level for the non-potable system dipped to 0.3m (Fig. 7a) during the 10th September to 26th October period, indicating that the Programmable Logic Controller (PLC) failed to receive the signal to activate the pump when the water level dipped below 0.8m. Thus, the controls to activate the pump only kicked in when the water level reached the emergency water level threshold of 0.3m. This fault also reoccurred on 6th August and 30th September.

Pump fault. Typically, the pumps operate in an alternating fashion to ensure that the pumps are not overstressed. However, during this faulty (Sep-Oct) period, we observe that the non-potable pumps are operated together, one after the other, to supply water to the rooftop tank (Fig. 7a). There are also instances of behavior where the pump operates for a short period of time, then re-activates a short time later (Fig. 7b). Such behavior indicates that the pump which started operating first had a failure, which caused it to stop. Then, the PLC would activate the other pump (or the same pump again a short time later) to finish the job of pumping the water to the high level threshold of 1.1m. This fault also occurs in the potable system on the 6th September, 9th September and 27th December.

Detecting captured faults using specification rules. Based on the daily water consumption patterns, we determine that the pump only needs to operate once a day on average. The probability that the pump needs to activate more than once in a given hour is very low since it is unlikely that the entire tank's water will be consumed by the occupants in that amount of time. As such, we

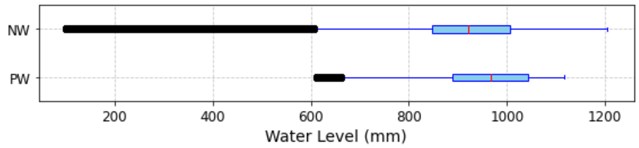


Figure 8: Range of water level readings for the potable water system (PW) and the non-potable water system (NW). The black regions denote the outlier readings.

create a specification rule that dictates the duration between any two pumping activations should be more than an hour.

Moreover, we know from the sensor data that the low water-level threshold is 0.8m and the high water-level threshold is 1.1m. However, due to the noisy sensor readings, we notice that the water level often goes outside of those bounds. Fig. 8 shows that there is a sizable percentage of readings between 0.6m and 0.8m, as well as between 1.1m and 1.2m. However, any outliers indicate that there may be a fault in the system, which may cause the tank to overflow or drain out. Thus, we create specification rules that dictate the range of water level readings should be within the inliers (0.66-1.12m for potable system and 0.61-1.20m for non-potable system).

We found that 7% and 13% of the potable and non-potable pumps' activations are flagged out as anomalies using our specification rule about inter-pumping duration. All the faults in Table 2 are flagged as anomalies. The remaining flagged anomalies are due to short one minute activations of the pump which could be attributed to noise from the power meters.

4 CONCLUSION

In this paper, we presented a real-world dataset for water pumps that consists of data from multiple non-invasive sensors installed on a water supply system on a university building. We provided detailed analyses of the data that provides insights about the regular system behavior. We also detailed the system events and faults that have been captured in the dataset. Finally, we demonstrated the utility of our dataset by showing that the data can be used to develop specification rules that are able to identify the recorded faults. Our dataset can also be used to explore multiple areas. For example, the different sensor readings can be correlated to assess the health of the pumps. Another usage of the dataset would be to employ the water consumption patterns to evaluate forecasting algorithms or for conducting experiments to optimize the pump operation [5].

ACKNOWLEDGMENTS

This research/project is supported by the National Research Foundation, Singapore, and Ministry of National Development, Singapore under its Cities of Tomorrow R&D Programme (CoT Award COT-V2-2020-1). Any opinions, findings and conclusions or recommendations expressed in this material are those of the author(s) and do not reflect the views of National Research Foundation, Singapore and Ministry of National Development, Singapore.

REFERENCES

[1] 2019. Pump sensor data for predictive maintenance. https://www.kaggle.com/datasets/nphantawee/pump-sensor-data

[2] 2024. 'Totally no water': Bukit Panjang residents affected by 3-hour water disruption. https://www.straitstimes.com/singapore/totally-no-water-bukit-panjang-residents-affected-by-3-hour-water-disruption

[3] Ane Dyveke Andenes Blaauw and Torstein Heltne Hovde. 2023. Deep Learning-Based Anomaly Detection System for Centrifugal Pumps. Available at https://ntnuopen.ntnu.no/ntnu-xmlui/handle/11250/3093895.

[4] Kyuchang Chang and Seung Park. 2023. Random forest-based multi-faults classification modeling and analysis for intelligent centrifugal pump system. *Journal of Mechanical Science and Technology* 38 (12 2023), 1–10.

[5] Carmen Cheh, Justin Albrethsen, Zhen Wei Ng, Binbin Chen, Xin Lou, Zaki Masood, and David KY Yau. 2024. Water Pump Operation Optimization under Dynamic Market and Consumer Behaviour. In *Proceedings of the 15th ACM International Conference on Future and Sustainable Energy Systems*. 335–346.

[6] Gong Chen, Lei Wang, Haoming Yang, Peifeng Wang, Jun Wei, and Jianguo Bao. 2023. Health assessment of water pumps using high-dimensional monitoring data. *Water Supply* 23, 10 (2023), 4059–4073. https://doi.org/10.2166/ws.2023.244

[7] Lei Chen, Lijun Wei, Yu Wang, Junshuo Wang, and Wenlong Li. 2022. Monitoring and predictive maintenance of centrifugal pumps based on smart sensors. *Sensors* 22, 6 (2022), 2106.

[8] Aline Elly Treml, Rogério Andrade Flauzino, Marcelo Suetake, and Narco Afonso Ravazzoli Maciejewski. 2020. Experimental database for detecting and diagnosing rotor broken bar in a three-phase induction motor. https://doi.org/10.21227/fmnm-bn95

[9] Jeong-yeong Kim, Tae-hyun Lee, Song-ho Lee, Jong-jik Lee, Won-kyun Lee, Yong-jin Kim, and Jong-won Park. 2022. A Study on Deep Learning-Based Fault Diagnosis and Classification for Marine Engine System Auxiliary Equipment. *Processes* 10, 7 (2022). https://doi.org/10.3390/pr10071345

[10] D Siano and MA Panza. 2018. Diagnostic method by using vibration analysis for pump fault detection. *Energy Procedia* 148 (2018), 10–17.

[11] Cem Ekin Sunal, Vladimir Dyo, and Vladan Velisavljevic. 2022. Review of Machine Learning Based Fault Detection for Centrifugal Pump Induction Motors. *IEEE Access* 10 (01 2022), 1–1. https://doi.org/10.1109/ACCESS.2022.3187718