Water Practice & Technology

IWA PUBLISHING

Water Practice & Technology Vol 18 No 1, 40 doi: 10.2166/wpt.2022.171

Estimation of wastewater flowrate in a gravitational sewer line based on a low-cost distance sensor

Jani Tomperi [a],*, Pekka M. Rossi[b] and Mika Ruusunen[a]

[a] Environmental and Chemical Engineering Research Unit, Control Engineering Research Group, University of Oulu, P.O. Box 8000, Oulu 90014, Finland

[b] Water, Energy and Environmental Engineering Research Unit, University of Oulu, P.O. Box 4300, Oulu 90014, Finland

*Corresponding author. E-mail: jani.tomperi@oulu.fi

JT, 0000-0001-8850-346X

ABSTRACT

Wastewater flowrate exhibits valuable information on the conditions in a sewer line. However, existing hardware flowrate sensors are rather expensive, vulnerable to fouling and breakages, and require frequent and laborious maintenance. Therefore, they are typically only mounted in a few key locations of a sewer system, leading to a lack of important information in a major part of the network. Utilizing more cost-effective sensors and a soft sensor approach is advisable for estimating the flowrate at locations where hardware sensors are lacking. Here, the development and testing of a data-driven soft sensor for estimating the wastewater flowrate based on the water level information obtained from a low-cost ultrasonic distance sensor are presented. The research included a long-term functionality testing period of the sensor in a cold region. The soft sensor-based flowrate was applied to estimate inflow and infiltration, indicating the conditions of the sewer line. The harsh conditions inside the sewer manhole caused challenges for the reliability of the distance measurement based on an ultrasonic principle. With the developed model-based soft sensor, it seems possible to accurately estimate the wastewater flowrate. Together with additional information, it might also enable accurate monitoring of inflows and infiltrations.

Key words: cold region, modelling, online estimator, ultrasonic, virtual sensor, water level

HIGHLIGHTS

- Online wastewater level information determined by a low-cost ultrasonic distance sensor was solely utilized to estimate volumetric wastewater flowrate in a gravitational sewer line via a data-driven soft sensor.
- Developed soft sensor provided accurate estimation of flowrate and amount of regional I/I, and information on the conditions of the sewer in near real-time, automatically and economically.

1. INTRODUCTION

The volumetric flowrate of wastewater is an important parameter to monitor when assessing and managing a sewer system and wastewater treatment operation (Nguyen *et al.* 2009; Bonakdari & Zinatizadeh 2011). In a normal situation, the sewage flowrate varies in short-term and long-term time instances, changing based on the time of day, day of the week, and month of the year. Flowrate, like wastewater quality, is dependent on population density, water consumption habits, and the commercial or industrial activities in the area. Generally, domestic sewage flowrates are at their highest in the mornings and the evenings and at the lowest at night. In urban areas, sewer systems can either be combined or separate sewer lines, transporting either wastewater and stormwater (surface runoff water) or a mix of the two. The quantity of wastewater in a separate sewer system may be drastically increased by inflow and infiltration (I/I), stormwater entering the sewer from inappropriate connections, and groundwater entering the sewer via defective or broken pipes. The amount of runoff water is mostly affected by rainfall (its duration and intensity) and snow melting in colder regions but is also affected by ground undulation (Butler & Davies 2004). The excess water ending up in the sewer system due to I/I blending with the wastewater, increasing the quantity and affecting the quality of the wastewater by dilution and changing the temperature. Even in quite light rainfall, stormwater flows can multiply the overall wastewater flowrate, and, in heavy falls, the stormwater could be tens of times the average wastewater flow (Butler & Davies 2004). Hence, downstream utilities may become overloaded and plausible overflowing or damming at the sewer system, and operational challenges in the wastewater treatment process may occur.

The flowrate of an underground sewer system is typically measured at a pumping station or inside a manhole with a flowrate sensor. Online flowrate sensors are, however, rather expensive, and mounting and maintaining them is often labour intensive (Bonakdari & Zinatizadeh 2011). Therefore, online flowrate sensors are often only installed to key points of the sewer system, leading to a lack of important real-time information about the state of the sewer system in wider parts of the sewer network. Additionally, robust monitoring in sewers is challenging due to the harsh conditions, including high humidity, gases, and solid debris, and due to particular hydraulic conditions, namely the rapidly changing wastewater level and flow during heavy rain events, which reduces the functionality and reliability of the traditional flowrate sensors (Jeanbourquin *et al.* 2011; Ji *et al.* 2020). Alternatively, the volumetric flowrate can be assessed utilizing secondary measurements, wastewater flow speed, and wastewater level. The most common methods to measure the wastewater level inside a sewer system are based on water pressure, ultrasonic, or radar techniques, the last two of which are non-contact methods (Burgess 2008; Ahm *et al.* 2016). Nonetheless, this method requires information on the inside diameter of the pipe and the installation and maintenance of two sensors, at least one of which must be in constant contact with the wastewater and, thus, is vulnerable to fouling and malfunctions.

When an essential parameter cannot be automatically measured online due to high cost or workload, or at lack of proper hardware sensors available, it is advisable to utilize analytical redundancies, applying a virtual sensor approach. A virtual sensor, or a soft sensor, can be model-driven or data-driven. Model-driven soft sensors have full phenomenological and/or physical knowledge about the system background, whereas data-driven soft sensors are based on the measured data (Kadlec *et al.* 2009). A data-driven soft sensor is based on the data of available easy-to-measure variables via modelling, to estimate a variable which cannot be automatically and reliably measured at all or the measurement of which comes at a high cost and workload, is sporadic, or comes with long delays due to laboratory analysis (Souza *et al.* 2016). Generally, soft sensor applications can be set at three categories: online estimators, process monitoring and fault detection, and sensor fault detection and reconstruction (Kadlec *et al.* 2009; Souza *et al.* 2016). The soft sensor concept is highly dependent on the calibration measurements, quality, and quantity of the available data. The uncertainty analysis of the estimates is important in assessing the usefulness of the soft sensor (Viitamäki & Ritala 2018). In addition, changes and drifts in the operation environment can affect the uncertainty of the measurements.

Several techniques, including multiple linear regression (MLR), principal component analysis (PCA), partial least squares (PLS), artificial neural network (ANN), and neuro-fuzzy systems, can and have been utilized to develop soft sensors for various applications (Kadlec *et al.* 2009; Souza *et al.* 2016). Related to flowrate estimation, some approaches have been reported. Ahm *et al.* (2016) developed three different soft sensor concepts (basin flow, channel flow, and weir flow), estimating discharge volumes based on the local water level sensor and, after testing and verification, found them to be accurate. Jeanbourquin *et al.* (2011) developed an *in situ* system for sewer water flow velocity monitoring based on video images, and Ji *et al.* (2020) examined the three methods (direct visual inspection and recording, image processing, and deep learning) for measuring the flowrate by capturing images inside of a sewer using a camera and analyzing the images to calculate the water level. Nguyen *et al.* (2009), Kouyi *et al.* (2010), and Isel *et al.* (2012) have utilized water elevation measurements to estimate the water flowrate. However, these approaches are complex, often requiring hydrological and 3D vision-based modelling and complex calculations, may require multiple simultaneous measurements and commercial software, and the research cases are not located in the cold Nordic region. Reports on developing and testing a data-based soft sensor, especially for estimating the wastewater flowrate in a sewer in the colder regions based on only simple acquired water level information, seem to be rarely published.

In this research, a novel data-driven soft sensor was developed to estimate the real-time volumetric flowrate in a gravitational sewer line at a location without present flowrate measurements. There, the flowrate is an important element in monitoring the state of the sewer and for the proactive operation of a wastewater treatment plant. Soft sensor development aimed to increase the knowledge of the operational personnel on the conditions of the sewage system in real-time, automatically, and economically. The wastewater flowrate estimation is based on a novel way of utilizing online water level measurements. For this, a low-cost sensor using an ultrasonic measuring principle was mounted inside a sewer manhole to continuously measure the distance to the wastewater surface. The distance information was then automatically transmitted to the database by a server developed for the purpose. The flowrate soft sensor was developed based on data acquired in a preliminary field measurement campaign. The functionality of the soft sensor was finally verified with data achieved over a long-term test period using the distance sensor.

The contributions of this research are three-fold: (1) to inspect the functionality of the low-cost ultrasonic distance sensor mounted inside a sewer manhole and the data transmission to the server in varying environmental and seasonal conditions; (2) to develop a data-driven soft sensor for estimating the volumetric flowrate of wastewater using measured water level information and to verify its functionality over a long testing period; and (3) to utilize the soft sensor-based flowrate together with expert knowledge and measured flowrate data from the upstream pumping stations to estimate the amount of I/I in the selected area.

The rest of this paper is organized as follows. In Material and Methods and its subsections, a general description of the online and field measurements, data pre-treatment, and modelling methods is given. In Results and Discussion, the results of the research are discussed, together with practical implications. Finally, the conclusions of the research are presented.

2. MATERIAL AND METHODS

2.1. Soft sensor approach

The proposed concept for the wastewater flowrate soft sensor is illustrated in Figure 1 with the utilized measurements and the measurement locations in the sewer network (white background), the model identification stage (light grey background), and the soft sensor testing stage (dark grey background). Figure 1 also shows the data flow in the training and testing stages. In Figure 1, *n* stands for the number of manual measurement data points; PS1 and PS2 illustrate the locations of the upstream pumping stations where the online reference data were measured; S/D illustrates the location of the manhole where the distance sensor was mounted and manual field measurements were carried out; and the arrows show the direction of wastewater flow in the main sewer line(s) (solid line) and in the unmeasured minor sewer lines (dotted line).

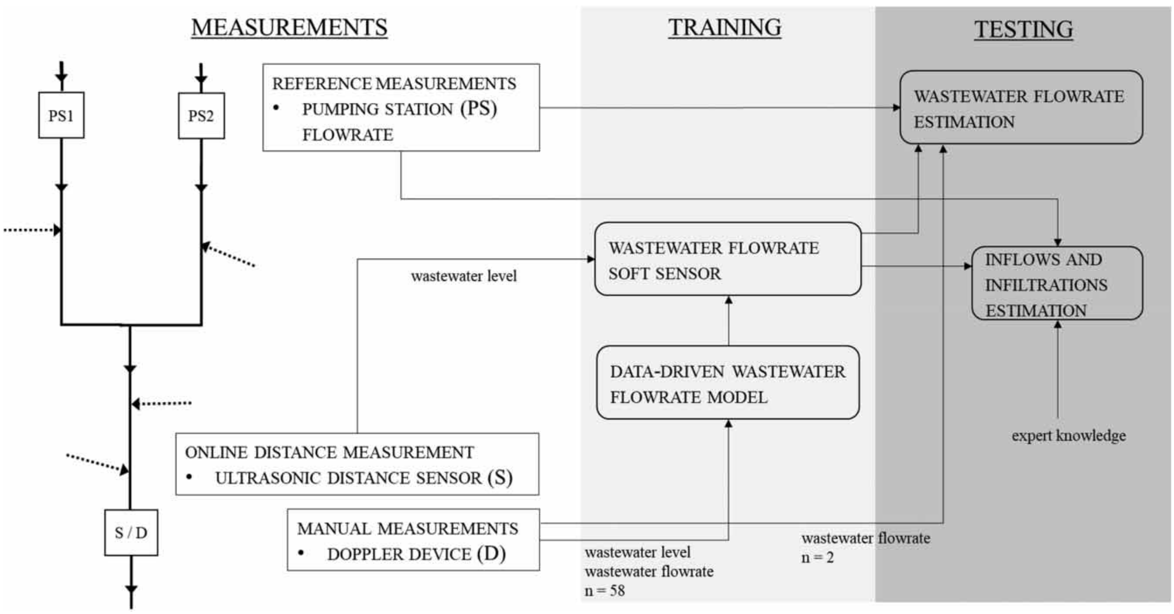

Figure 1 | The framework of the real-time soft sensor for monitoring the wastewater flowrate.

2.2. Online water level measurements

In this research, to collect the wastewater level information, a low-cost ultrasonic distance sensor was mounted right under the cover of a sewer manhole as illustrated in Figure 2. This measurement device, based on an ultrasonic technique, includes a sensor and a transmitter, which is also a datalogger. The transducer transmits short high-frequency ultrasonic pulses into the wastewater. The pulses are reflected from the surface of the wastewater back to the receiver (transducer), which calculates the travelling time of the pulses. The distance from the device

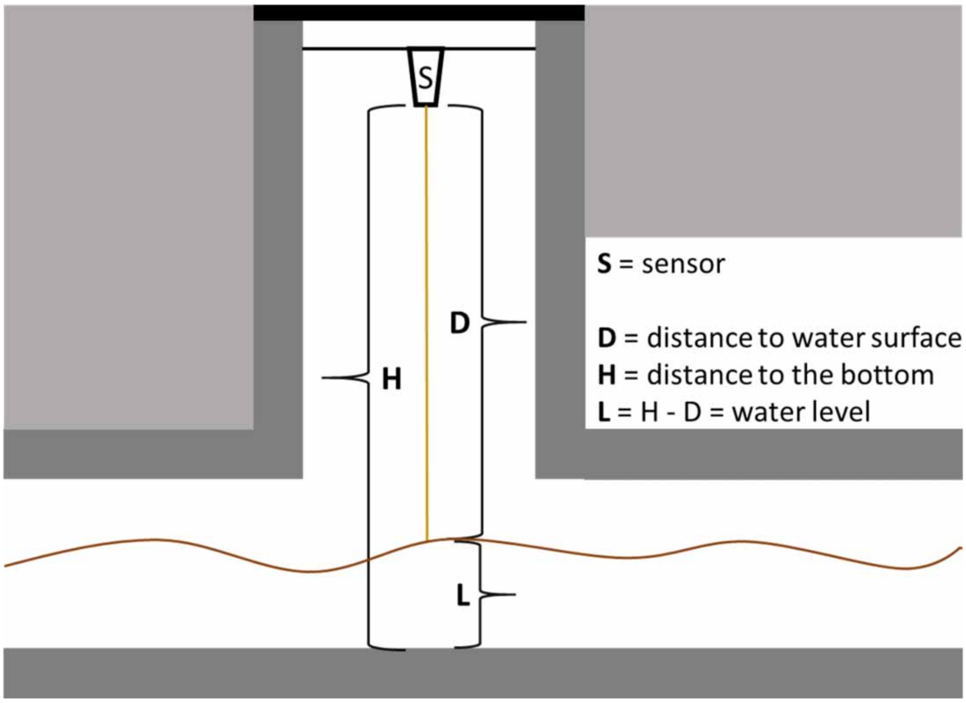

Figure 2 | The ultrasonic distance sensor located inside a sewer manhole and the principle of measuring the temporal water levels.

to the water surface is calculated utilizing knowledge about the speed of sound and the measured pulse travelling time. As the speed of sound is dependent on air density (affected by the temperature and humidity), an automatic compensation in the calculation is performed to obtain an accurate measurement. For distance measurement and data transmission, the mounted sensor required no external energy source or wiring, but it was a battery-powered solution.

As the mounted sensor measures the distance from the sensor to the wastewater surface, the measured distance information was converted to the temporal water level, as shown in the following equation:

$$L = H - D, \tag{1}$$

where L is the water level (mm), D is the distance from the sensor to the water (mm), and H is the distance from the sensor to the bottom of the sewer (mm). For the calibration of the water level measurement, the distances from the sensor to the bottom of the well (H) and to the wastewater surface (D) were first measured manually with a measuring rod. The diameter of the sewer line was 1 m with a longitudinal slope of 3.5 ‰. In total, the distance data were collected over a 1-year period. Hence, the collected data included measurement information for a wide range of environmental conditions, such as snow melting, heavy rains, and dry seasons, together with warm and cold seasons with high and low humidity inside the manhole.

2.3. Field measurement campaign

To collect data for model development, a manual field measurement campaign was carried out using a Doppler measurement device which measures the level (m), flowrate (m^3/h), velocity (m/s), and temperature (°C) of the wastewater. The measurements are based on a pressure transducer and a Doppler system, where a signal transmitted against the oncoming flow reflects from solid particles or air bubbles. The magnitude of the change in the reflected signal's frequency is proportional to the particle velocity, thus giving information on the water velocity. The field measurement data were collected over several days at different times of the day to catch the varying behaviours of the wastewater flowrate and level at the sewer system. The data acquired from the manual field measurement campaign included 58 data points for model development and two data points for soft sensor testing.

2.4. Reference measurements

Since an online sensor for monitoring the wastewater flowrate continuously and long-term inside the sewer was lacking, and a decent manual measurement campaign using the Doppler device was considered overwhelmingly laborious, the functionality of the developed soft sensor was verified by comparing the model-based flowrate (soft sensor) to the reference measurements. The reference here is the measured and combined flowrate at the two wastewater pumping stations located upstream (PS1 and PS2), as illustrated in Figure 1. The delays from the pumping stations to the location of the distance sensor were determined and considered when these flowrates were compared.

The resulting reference data were also utilized alongside expert knowledge to estimate the quantity of I/I. A couple of minor sewer lines are also connected to the main sewer lines between the pumping stations and the distance measurement location, as illustrated in Figure 1. There was no measured flowrate data available from these minor lines, instead, expert knowledge and annual consumption estimations of the area were used. According to these, the influence of the minor lines on the total flowrate at the main sewer is considered small but not negligible, and this should be taken into account when assessing the functionality of the developed soft sensor and estimating the quantities of I/I.

2.5. Data pre-treatment

The distance to the wastewater surface was measured in 10-min sampling periods and the median value of 15 distance readings was transmitted to the database server via LoRaWAN network. Outliers are defined as values that deviate from the typical or meaningful ranges of the measurement values and can be categorized as obvious outliers and non-obvious outliers (Kadlec *et al.* 2009). Obvious outliers violate the physical or technological limitations whereas non-obvious outliers do not violate any limitations but still lie outside of the typical ranges, they are, therefore, harder to identify. In this research, for soft sensor development and testing, the collected field measurement data and online distance data were first inspected, and the obvious outliers and false values were identified and removed based on expert knowledge and using MATLAB algorithms created for the purpose.

2.6. Soft sensor development

A straightforward method to develop a soft sensor is to apply MLR, where the modelled output is a linear combination of the selected input variables as described in the following equation.

$$y = b + p_1x_1 + p_2x_2 \cdots + p_nx_n \, , \qquad (2)$$

where y is the predicted output variable, $x_{1-\mathrm{n}}$ is the selected input variables, and $p_{1-\mathrm{n}}$ and b are model parameters and a bias value, respectively. The MLR was utilized in the presented research as it is easy to interpret, implement, and maintain. As the purpose of this research was to develop a soft sensor based on only the water level information, the developed model only included this information as the input variable.

The accuracy of the developed model depends greatly on the quality of the data, which should include samples that fully represent the full spectrum of possible conditions. Developing a robust model requires a long and representative subset of data for both model training and validation; however, in many real-world cases, it may be a challenge to obtain enough such data. As mentioned above, in the presented research the amount of collected data from the manual field measurement campaign was relatively low. To maximize the utilization of the potential information content of the data, its traditional split into training and test subsets was considered to be an impractical approach in this case. Instead, a k-fold cross-validation was applied. Cross-validation is an efficient resampling method to predict the fit of a model. In k-fold cross-validation, the whole data set is used to train and validate the model by dividing the dataset randomly into approximately k equal sized partitions and in turn using $k-1$ partitions to train the model and one partition for validation. The procedure is repeated k times until every partition is used once for validation. The final estimation and its uncertainty can be found by averaging the k models' results (Kohavi 1995; Arlot & Celisse 2010). In this research, the data were presumably uniformly distributed, and the value of k was selected to be 10 after testing with lower values of k.

3. RESULTS AND DISCUSSION

3.1. Characterization of the real-time distance sensor

The first target of this research was to investigate the functionality of the low-cost ultrasonic distance sensor mounted inside a sewer manhole and its data transmission capability under changing seasonal and environmental conditions in a cold region. During the long-term testing period, it was found that overall, the data transmission via LoRaWAN network was quite robust, but due to the updates in the network, there was a small number of times when the data were not saved to the server database, shown as very long gaps in the data series. Subzero temperatures or changing humidity did not affect the mechanical functionality of the sensor or the data transmission. Battery lifetime covered the research period and beyond.

The distance information from the ultrasonic sensor was converted into water level (mm) information and found to be identical to the water level values measured during the manual field measurement campaign. The collected sensor data revealed the realistic daily, weekly, and seasonal changes in the water level and the effects of heavy rain events. An example of the behaviour of wastewater flow during a week with a heavy rain event is shown in Figure 3.

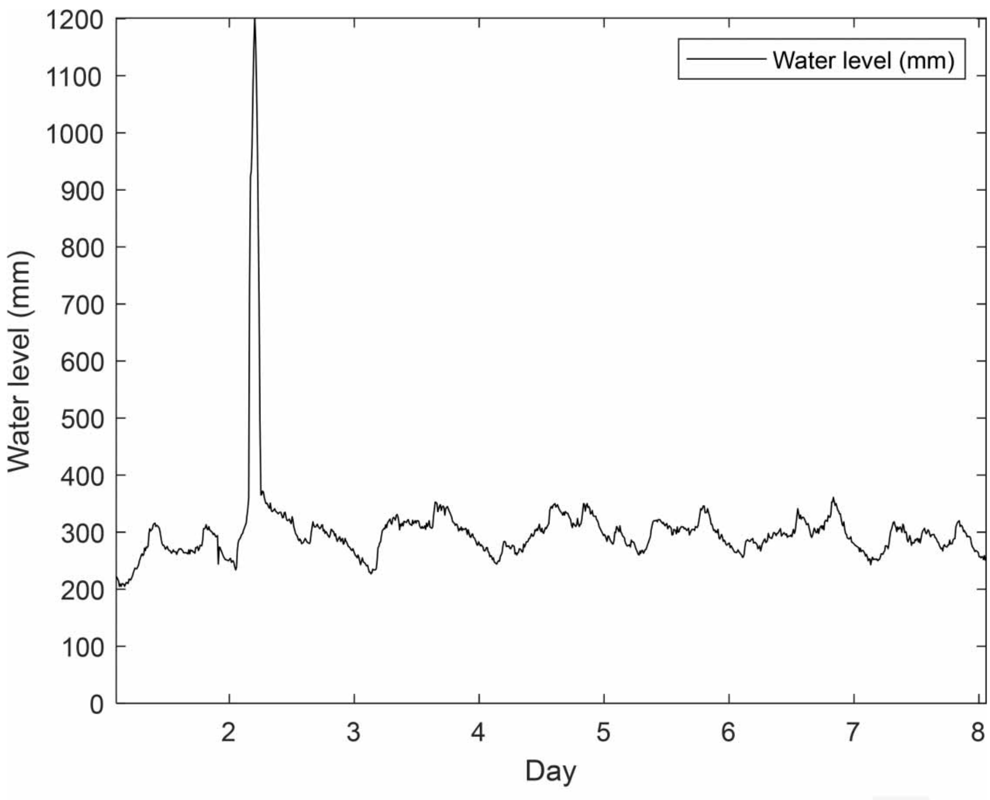

Figure 3 | An example of daily variation in measured wastewater level over a week and the effects of a heavy rain event in the early hours of Day 2.

However, excluding the summertime, the collected distance data included wide variations and a number of false measurement values. Most likely the humidity inside the manhole during the colder periods of the year caused additional uncontrolled reflections of the transmitted ultrasonic pulses and therefore unrealistic measurement values and variations were recorded. As can be seen in Figure 4, the occurrence of the outliers and variation stopped in the spring after the air temperature increased. Without pre-treatment, the data collected outside summertime cannot be utilized for data analysis or modelling purposes. Filtering removes some information from the data and the remaining values are not absolutely correct, which would also affect the results of soft sensor testing. Therefore, in the following, instead of presenting the results of testing of the developed soft sensor using the data from the whole collecting period, only the results of the selected periods when data required no filtering are presented.

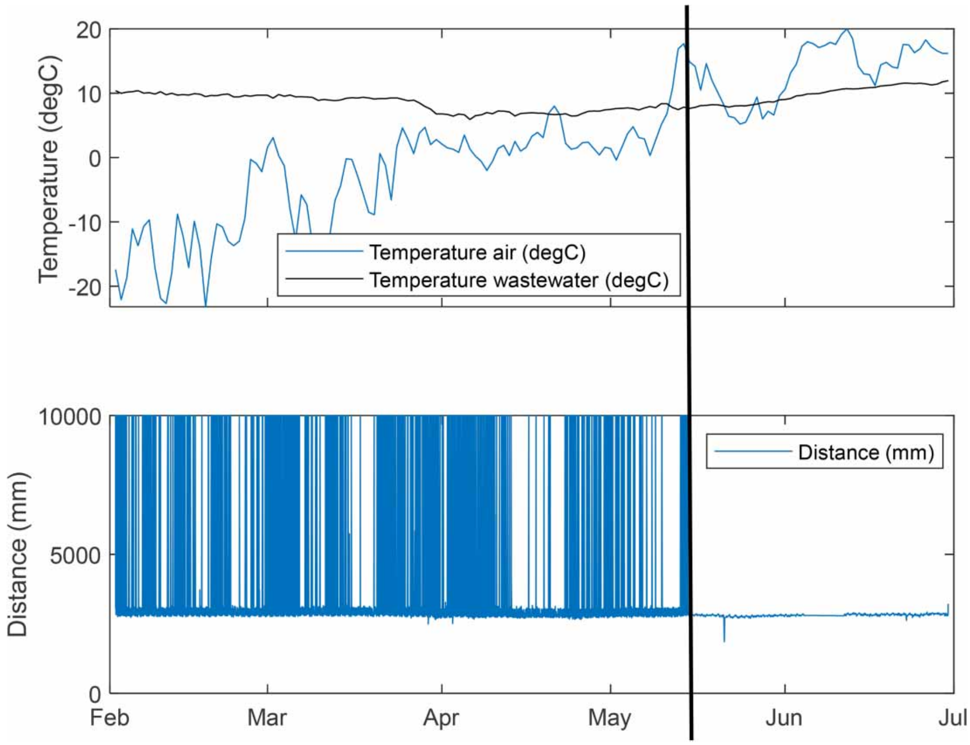

Figure 4 | The air and wastewater temperatures measured by online temperature sensors and the raw distance data measured by the ultrasonic distance sensor.

3.2. Soft sensor identification

A data-driven soft sensor was developed, as described in the subsection of Material and Methods. The linear regression model for the volumetric flowrate was developed using the data collected from the manual field measurement campaign. As seen in Figure 5, where the measured and modelled wastewater flowrates (m^3/h) with a linear fit line, the 95% confidence interval and the performance values of the developed model are presented, the accuracy of the developed model for the wastewater flowrate is excellent. There are several clusters of wastewater flowrate and only one data point clearly deviates from the linear fit line. The performance of the developed model was evaluated using the coefficient of determination (R^2), Root Mean Square Error (RMSE), and Mean Absolute Error (MAE), which were 0.92, 26.03 m^3/h, and 22.87 m^3/h, respectively. The field measurements were carried out over several days at different times of the day and different days of the week, and therefore the collected field measurement data cover almost the whole range of wastewater flowrate values at that sewer line exhibiting normal operating conditions in the daytime. However, the collected data do not include flowrate values measured at night or in any extreme conditions, which may have some effect on the model accuracy and performance of the soft sensor during the testing period.

3.3. Performance of the soft sensor

To verify the accuracy of the developed soft sensor, a comparison between the manually performed flowrate measurement (using the Doppler device), the model-based flowrate, and the reference flowrate was performed; it is presented in Figure 6. The comparison included two sampling points, illustrated as arrows in the figure. When the model-based flowrates were 207 and 205 m^3/h, the corresponding manually measured values were 212 and 207 m^3/h, and the reference flowrate values were approximately 188 and 180 m^3/h. It can be stated that in a large sewer line (diameter 1 m), the differences between manually measured flowrate and the model-based flowrate can be considered insignificant. The difference between model-based and reference flowrates is discussed in the following paragraph. Although only two verification measurements could be done, it can be stated that the model-based flowrate seems to correspond with the actual flowrate excellently when the distance sensor operates faultlessly. Hence, with the developed soft sensor, it is possible to estimate the wastewater flowrate accurately.

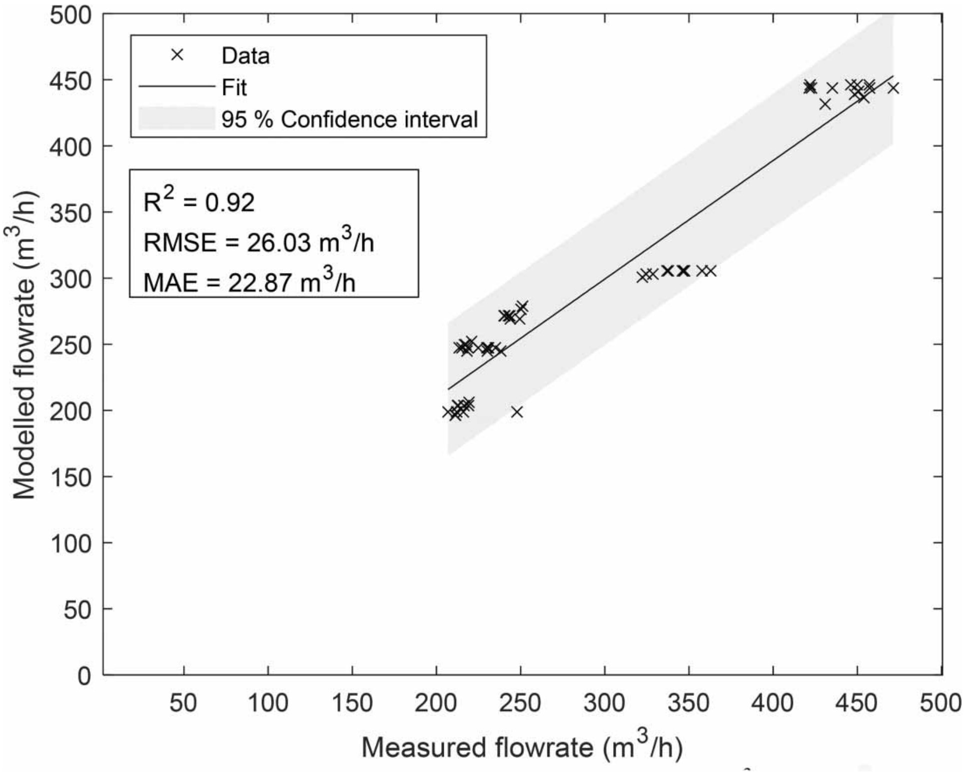

Figure 5 | The measured and modelled wastewater flowrates (m^3/h) with a linear fit line, the 95% confidence interval, and the performance values of the developed model using all the training data (58 data points) collected during the manual field measurement campaign.

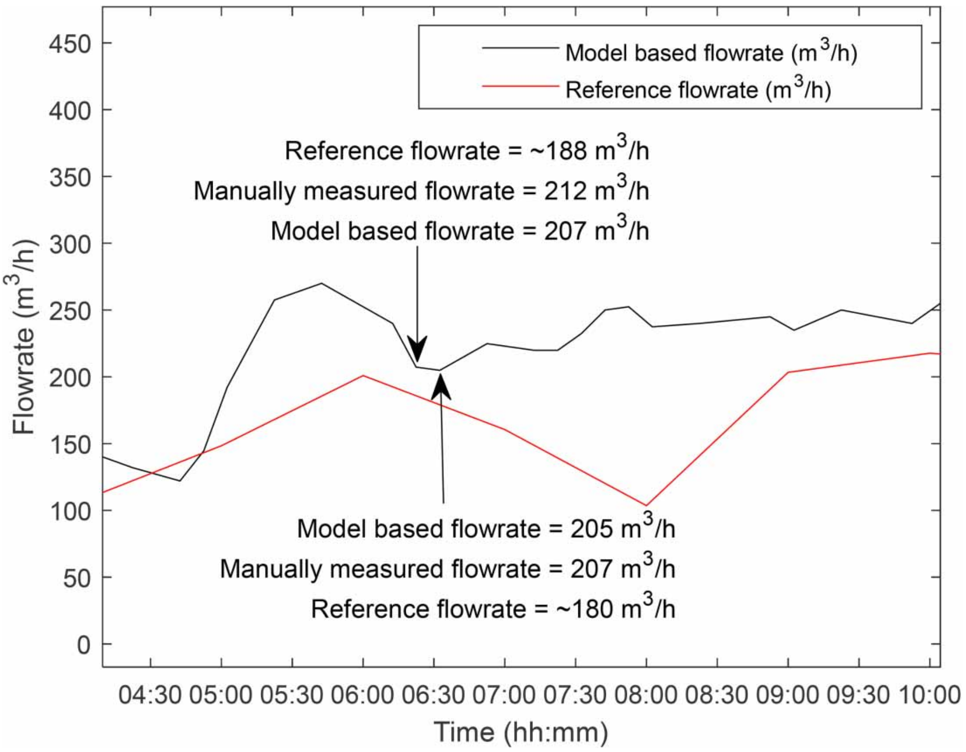

Figure 6 | The comparison of the manually performed flowrate measurement, the model-based flowrate, and the reference flowrate at two sampling points.

The long-term functionality of the developed soft sensor was tested utilizing the data recorded by the online distance sensor during the selected periods when data required no filtering. The comparison of the model-based flowrate and the reference flowrate (combined pumping stations flowrate) during a period of over a month and a period of around 2 weeks with the residual trend are presented in Figures 7 and 8, respectively. As seen in both figures, the behaviour of these flowrates is very similar; generally, the changes occur at the same point and the baseline is at the same level. There is occasionally a fairly large difference between the flowrates, which is reasonable as the reference flowrate measured at the upstream pumping stations does not include the wastewater flows from the minor sewer lines nor the I/I in the area, which is taken into account in the model-based flowrate. Based on the visual inspection and the calculated performance values of the soft sensor during the testing period (R^2 0.50, RMSE 91.16 m^3/h, and MAE 76.18 m^3/h), the developed soft sensor is also suitable for estimating the wastewater flowrate at the gravitational sewer line over longer periods.

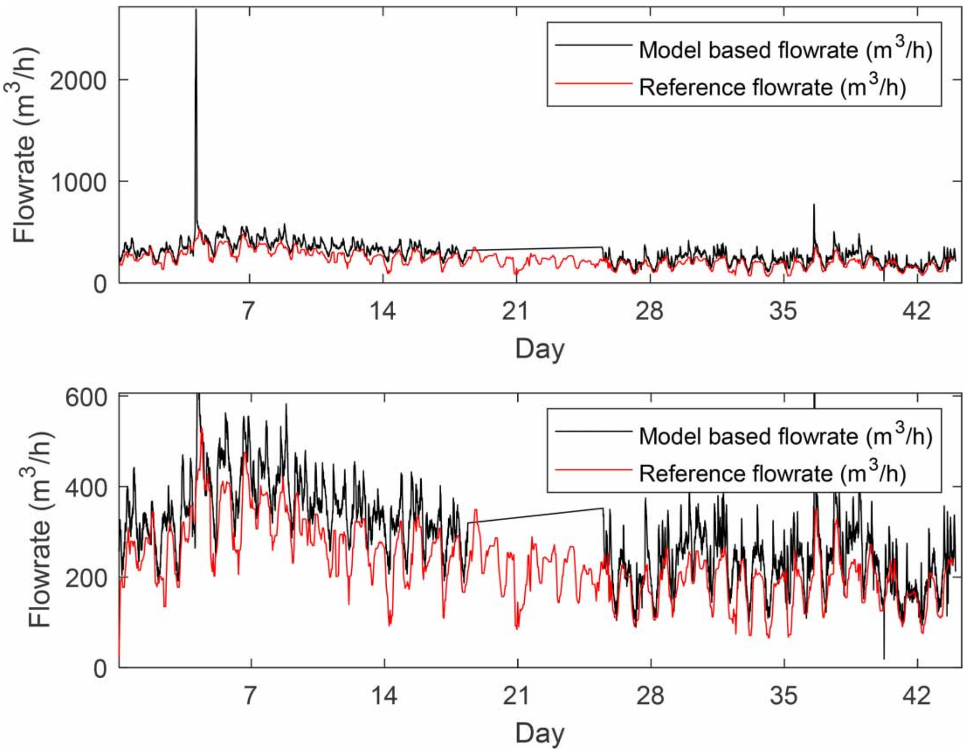

Figure 7 | The comparison of model-based flowrate and the reference flowrate during a period of over one month.

The third goal of this research was to utilize the soft sensor to estimate the quantity of I/I in the selected area. This can be carried out utilizing the same flowrate data presented in Figures 7 and 8 together with expert knowledge. As mentioned, in addition to the wastewater flows from the minor sewer lines, I/I enter the main sewer line between the pumping stations and the measurement location, explaining the varying difference between the model-based flowrate and reference flowrate. During the testing period of the soft sensor using the collected data, the mean and median differences between the soft sensor flowrate and the reference flowrate were 73.96 and 65.15 m^3/h, with a standard deviation of 53.30. According to the annual consumption data received from the local wastewater utility, the estimated quantity of unmeasured wastewater flowrate from the minor sewer lines in the area is around 67 m^3/h. Hence, the average estimated quantity of the I/I can be assumed to be roughly 7 m^3/h, based on the testing period. This can be considered to be realistic, but not necessarily absolutely correct value as the additional expert information is an annual estimation of the regional consumption. However, the developed soft sensor can also be utilized in estimating the quantity of I/I in the area.

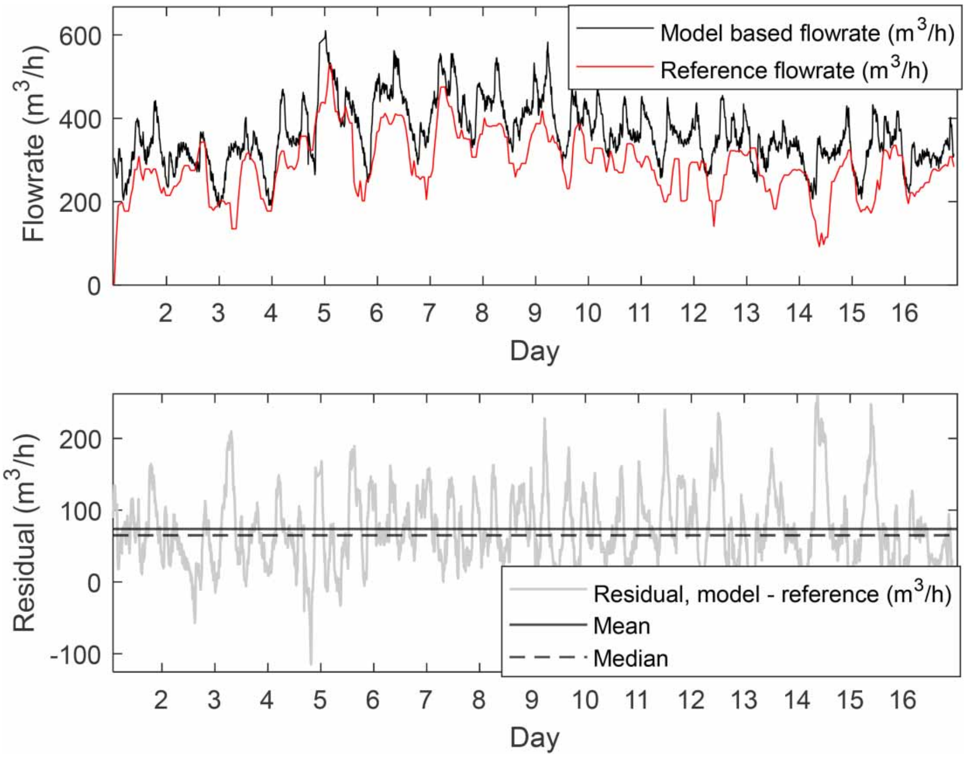

Figure 8 | The comparison of model-based flowrate and the reference flowrate over a 2-week period with a residual trend, and mean (73.96 m^3/h) and median (65.15 m^3/h) lines.

3.4. Practical implications

Based on the results presented above, it is possible to accurately estimate the volumetric wastewater flowrate in a gravitational sewer using a data-driven soft sensor that utilizes the data collected by a low-cost ultrasonic distance measurement device. This soft sensor given flowrate information can also be utilized for estimating the quantity of I/I at the sewer line. As presented, the training and validation of the soft sensor was a success, and the model-based flowrate was very close to the manually measured flowrate values, indicating the accurate functionality of the soft sensor. During a long testing period utilizing the collected data, the model-based flowrate was realistic, including similar behaviour and occasionally identical flowrate values as the collected reference data from the upstream pumping stations. The difference between the flowrates is logical considering the minor side flows with no measured flowrate information and I/I between the pumping stations and the measurement location. Therefore, the model-based flowrate is and should be larger than the reference flowrate.

The high spike in water level (Figure 3) and in the model-based wastewater flowrate (Figure 7) is due to the heavy rain event, seen in Figure 9 (on Day 5), where daily precipitation values together with daily total flowrates are presented. This momentarily high flowrate in Figure 7 is, however, slightly unrealistic. Supposedly, the heavy rain event caused, in addition to the increased flowrate, a momentary damming inside the sewer line, which is shown as an elevated wastewater level, hence giving a greater modelled flowrate. This conclusion is supported by the information on the I/I at the local wastewater treatment plant. Even though the momentary model-based flowrate value was unrealistic, the soft sensor generates valuable information on the abnormal condition in the sewer. Damming is one of the key issues in sewer operation, not only during heavy rain events but also on a daily basis due to the blockages caused by unwanted objects.

However, the development and functionality of the soft sensor also include some uncertainties. The accuracy of the soft sensor depends on the measurement data used as input, which again depends on the functionality of the measurement device and the exactness of the calibration. The calibration of the water level measurement includes uncertainties, e.g., measuring the distances from the sensor to the bottom of the well and to the surface of wavy wastewater flow may not be absolutely correct. In this research, it was ensured that there were no solid particles at the bottom of the well, but in some cases, there might be solids that may corrupt manual measurement when the measuring rod is not at the bottom of the well. Uncertainties are also included in manual measurements

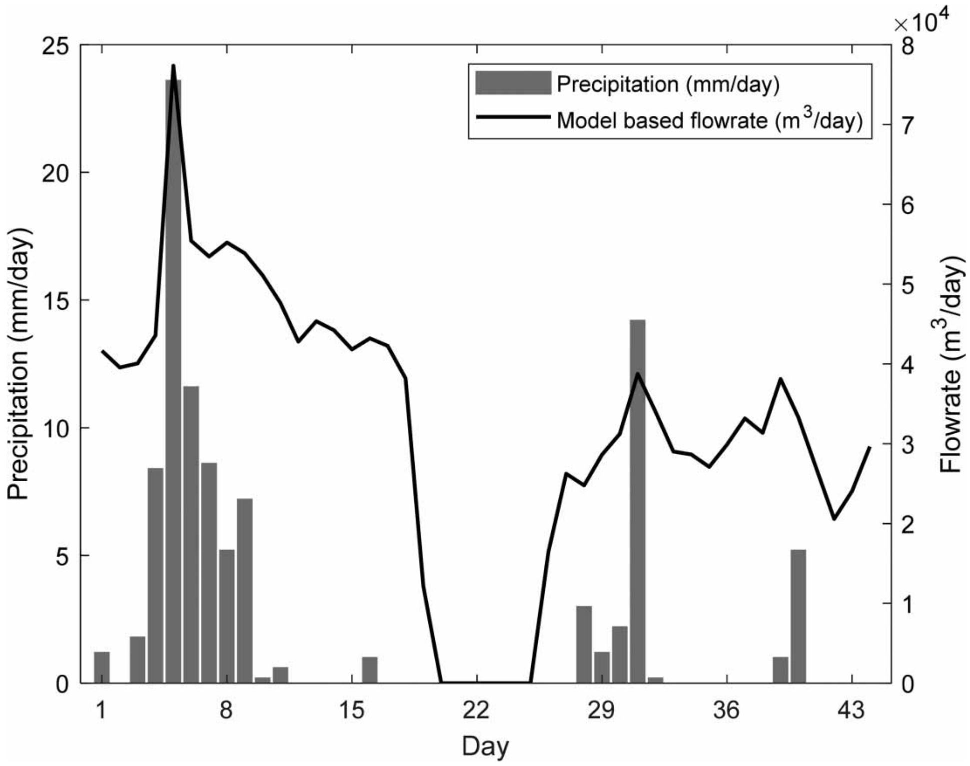

Figure 9 | Daily precipitation values and the model-based daily wastewater flowrates during the soft sensor testing period.

during the field measurement campaign and the online flowrate measurements at the pumping stations. In the study by Bonakdari & Zinatizadeh (2011), it was discovered that the type and position of the Doppler device may have an effect on measurement accuracy, and, in certain conditions, the velocity sensor does not provide representative values. Even though the normal behaviour of the wastewater flowrate in the day was captured during the manual measurement campaign, night-time and extreme events were missing from the training data, which may affect the accuracy of the soft sensor at very high or low flowrates. Also, the estimation of unmeasured wastewater from the minor sewer lines is based on annual consumption data and was converted to m^3/h value. It is known that the quantity of wastewater varies during the day, week, and season of the year; hence the values of unmeasured wastewater flowrates and quantity of I/I are rough, but realistic, estimations. Measuring the flows from these minor sewer lines would have improved the results of the research, but unfortunately, due to the limited resources, this could not be done now. However, this is something that should be executed in future research. Varying flow speed also affects the accuracy of the determination of delays.

During this research, there was wide variation and false values in the distance data collected using the distance sensor based on an ultrasonic measuring principle. Supposedly, they were due to the environmental conditions (the temperature, humidity) as they occurred only during cold season but as stated in Burgess (2008), sometimes a layer of foam may be on the surface of wastewater and affect the measurement accuracy by absorbing the emitted pulses. Here, it was confirmed that the conditions in the sewer manhole are harsher in the cold region and the measurement device for the purpose should be carefully selected, mounted, and tuned to obtain reliable measurements. To continue this research, testing a different type of device for measuring the distance to wastewater in the same manhole has already started. The functionality of this device seems promising and will be utilized in future research verifying the functionality of the developed soft sensor approach around the year in a cold region.

On the contrary to the previous research cases reported (Nguyen *et al.* 2009; Kouyi *et al.* 2010; Jeanbourquin *et al.* 2011; Isel *et al.* 2012; Ahm *et al.* 2016; Ji *et al.* 2020), which included, for instance, hydrological and 3D vision-based modelling and complex equations, required multiple simultaneous measurements or specific commercial software and were not located at the cold Nordic region, our simple data-based approach does not require any additional information on the sewer system or other measurement than distance to the water surface, and it is easy to implement in any system. In addition, it was tested in the cold and varying Nordic environment. The developed soft sensor can give valuable information to operational personnel on the sewer in near real-time:

In addition to using the soft sensor to estimate the local volumetric flowrate in a gravitational sewer where no flowrate sensor is mounted, the soft sensor can also be utilized for assessing the unmeasured flows from the side streams, or the I/I along a sewer network, or indicating abnormal conditions, for example, damming. Mounting the low-cost distance sensors all over the sewer network and utilizing the simple but efficient data-driven soft sensor approach enables monitoring of the entire sewer network in near real-time. It also enables focusing the observation and required actions on smaller parts of the network. Mounting and calibrating the distance sensor is a simple and relatively non-laborious task. A distance sensor requires minor maintenance effort and is significantly more cost-effective to purchase and maintain than a hardware flowrate sensor. It is a non-contact measurement and is not vulnerable to, for example, fouling, abrasion, or clogging. Low power consumption enables long operation time with replaceable batteries. Collecting a representative dataset for soft sensor development requires manual measurements on several occasions, and developing a soft sensor requires certain know-how, making these the most laborious tasks in the presented soft sensor approach. Even though a developed soft sensor could be copied to other locations, it is advisable to develop a unique soft sensor for each location if the operating ranges are totally different. Overall, the described soft sensor approach for wastewater flowrate is easily generalizable and recommended for indirect wastewater flowrate monitoring.

4. CONCLUSIONS

In this paper, a functionality test of a low-cost ultrasonic distance sensor mounted inside a sewer manhole; the development and testing of a novel data-driven soft sensor for estimating the wastewater flowrate; and the utilization of the soft sensor to estimate the I/I were described. The 1-year research period showed that the harsh conditions inside a sewer manhole are challenging even for non-contact measurement, and collecting a year round dataset with no errors requires careful selection and proper mounting of a measurement device. However, the development and testing of the soft sensor that utilizes only the distance measurement to estimate the wastewater flowrate in a gravitational sewer line was a success. It was verified that the soft sensor accurately showed the wastewater flowrate, and in addition to giving valuable new information on the flowrate in near real-time, it can also be used together with expert knowledge and available online measurements to estimate the I/I, providing information on the condition of the sewer lines. This information can be obtained with lower investment and maintenance costs compared to hardware sensors.

ACKNOWLEDGEMENTS

This research work was part of the Digitization of Water Utilities – Waste Water Networks Data Management and Innovative Applications (WWData) -project funded by Pohjois-Pohjanmaan liitto, which is thereby greatly acknowledged. Santeri Puhakka, Joni Koivula, and Axumawit Tesfamariam are greatly acknowledged for carrying out the field measurement campaign and providing additional data and information for the research.

DATA AVAILABILITY STATEMENT

Data cannot be made publicly available; readers should contact the corresponding author for details.

CONFLICT OF INTEREST

The authors declare there is no conflict.

REFERENCES

Ahm, M., Thorndahl, S., Nielsen, J. E. & Rasmussen, M. R. 2016 Estimation of combined sewer overflow discharge: a software sensor approach based on local water level measurements. *Water Science & Technology* **74**(11), 2683–2696.

Arlot, S. & Celisse, A. 2010 A survey of cross-validation procedures for model selection. *Statistics Surveys* **4**, 40–79.

Bonakdari, H. & Zinatizadeh, A. A. 2011 Influence of position and type of Doppler flow meters on flow-rate measurement in sewers using computational fluid dynamic. *Flow Measurement and Instrumentation* **22**(3), 225–234.

Burgess, S. 2008 Case study – level sensors in pumping stations. In: *33rd Annual Qld Water Industry Operations Workshop Indoor Sports Centre, Carrara – Gold Coast 3 to 5 June, 2008*. Available from: http://www.wioa.org.au/conference_papers/08_qld/docu-ments/scottburgess.pdf. (accessed May 2022).

Butler, D. & Davies, J. W. 2004 *Urban Drainage*, 2nd edn. Spon Press, London, UK; New York, USA, p. 543. Available from: https://vann-piseth.files.wordpress.com/2015/07/urban-drainage-butler.pdf. ISBN 0-203-34190-2.

Isel, S., Araud, Q., Philippe, R., Dufresne, M. & Vazquez, J. 2012 Assessment of the discharge in sewer pipes using two water level measurements and the relationship of gradually varying flow. In: *9th International Conference on Urban Drainage Modelling*, Belgrade, Serbia.

Jeanbourquin, D., Sage, D., Nguyen, L., Schaeli, B., Kayal, S., Barry, D. A. & Rossi, L. 2011 Flow measurements in sewers based on image analysis: automatic flow velocity algorithm. *Water Science and Technology* **64**(5), 1108–1114. doi:10.2166/wst.2011.176.

Ji, H. W., Yoo, S. S., Lee, B.-J., Koo, D. D. & Kang, J.-H. 2020 Measurement of wastewater discharge in sewer pipes using image analysis. *Water* **12**(6), 1771. doi:10.3390/w12061771.

Kadlec, P., Gabrys, B. & Strandt, S. 2009 Data-driven soft sensors in the process industry. *Computers and Chemical Engineering* **33**, 795–814.

Kohavi, R. 1995 A study of cross-validation and bootstrap for accuracy estimation and model selection. In: *Proceedings of the Fourteenth International Joint Conference on Artificial Intelligence*, Vol. 2, pp. 1137–1145.

Kouyi, G. L., Besson, F., Bier, M., Chocat, B. & Lucchinacci, P. 2010 Assessment of flow rate in a complex sewer pipe by means of a water depth sensor and modelling. In: *7th International Conference on Sustainable Techniques and Strategies for Urban Water Management*, Lyon, France.

Nguyen, L. S., Schaeli, B., Sage, D., Kayal, S., Jeanbourquin, D., Barry, D. A. & Rossi, L. 2009 Vision-based system for the control and measurement of wastewater flow rate in sewer systems. *Water Science and Technology* **60**(9), 2281–2289. doi:10.2166/wst.2009.659.

Souza, F. A. A., Araújo, R. & Mendes, J. 2016 Review of soft sensor methods for regression applications. *Chemometrics and Intelligent Laboratory Systems* **152**, 69–79.

Viitamäki, M. & Ritala, R. 2018 Soft sensor for compression and burst strength: identification and uncertainty analysis. *Journal of Science & Technology for Forest Products and Processes* **7**(1), 38–47.

First received 15 June 2022; accepted in revised form 20 December 2022. Available online 29 December 2022