Water Science & Technology

Water Science & Technology Vol 90 No 7, 1946 doi: 10.2166/wst.2024.249

Soft sensor for the dry solid content in thickened primary sludge

Hanna Molin[a,b,*], Eric Bröndum[c], Sara Nilsson[d], Per Mattsson[e], Ramesh Saagi[a], Erik Lindblom[a,b], Bengt Carlsson[e] and Ulf Jeppsson[a]

[a] Division of Industrial Electrical Engineering and Automation (IEA), Department of Biomedical Engineering, Lund University, P.O. Box 118, SE-22100 Lund, Sweden
[b] IVL Swedish Environmental Research Institute, P.O. Box 21060, SE-10031 Stockholm, Sweden
[c] Division of Theoretical Computer Science, School of Electrical Engineering and Computer Science, KTH Royal Institute of Technology, P.O. Box 24075, SE-10450 Stockholm, Sweden
[d] Stockholm Vatten och Avfall, SE-106 36 Stockholm, Sweden
[e] Division of Systems and Control, Department of Information Technology, Uppsala University, P.O. Box 337, SE-75105 Uppsala, Sweden
*Corresponding author. E-mail: hanna.molin@ivl.se

HM, 0009-0004-8299-8114

ABSTRACT

Software sensors, or soft sensors, can be a feasible option to monitor parameters that are difficult (or impossible) to measure with hardware sensors. At Henriksdal water resource recovery facility (WRRF), the operators have long experienced issues with a clogging sensor for the dry solids (DS) content in thickened primary sludge. A soft sensor was developed, and in the process, two methods were compared: long short-term memory (LSTM) network, and linear regression. The first is a recurrent neural network that can capture non-linear dynamics, whereas the latter is a linear static model. The LSTM network was the best at predicting the DS content, with a mean squared error (MSE) of 0.341 with respect to laboratory data. The linear regression model performed worse than estimating a long-time average of daily manual samples but outperformed the online sensor. Replacing the existing sensor with the developed soft sensor can open possibilities to more efficient control and operation of the thickener unit.

Key words: artificial neural network, machine learning, sludge thickening, soft sensor, wastewater treatment

HIGHLIGHTS

- Using a long short-term memory (LSTM) network and a linear regression model, we can predict dry solid (DS) content in thickened primary sludge.
- The LSTM model was able to predict the DS content with reasonably high accuracy. It also captures minute variations and can be used as a soft sensor for monitoring and controlling the process in the future.
- Linear regression performed worse than estimating a long-time average.

INTRODUCTION

Wastewater treatment inevitably results in excess sludge, or suspended solids, withdrawn, e.g., from settlers or filtering processes such as membrane bioreactors or disc filters. Sludge contains a mixture of pathogens and chemical compounds, ranging from heavy metals and micropollutants to organic carbon and nutrients including phosphorus and nitrogen (Peirce *et al.* 1998). Wastewater sludge generally has a high water content. The amount of water and the dewaterability of sludge can vary much between primary sludge and activated sludge. The water in primary sludge is mostly free or loosely bound, whereas the presence of microbial extracellular polymers in activated sludge causes the water to be tightly bound within the microbial flocs. Independent of the nature of the sludge, it is desirable to thicken or dewater the sludge before stabilization (e.g. anaerobic digestion, composting, and hydrothermal carbonization) to reduce the total volume of the sludge and increase the solid concentration (Houghton *et al.* 2001; Samanta *et al.* 2023). There is an important distinction between sludge *thickening* and sludge *dewatering*, where thickened sludge will still have characteristics closer to liquid than solids (water content 95–97%) and dewatered sludge will behave more like solids (water content 80–85%). Furthermore, thickening, in general, refers to gravitational methods (Peirce *et al.* 1998; Samanta *et al.* 2023). How well the sludge thickens depends on the properties of the sludge as well as the design and operation of the thickener. Process parameters like sludge age and retention time affect the amount of microbial extracellular polymer and, thus, the amount of free water in the sludge (Houghton *et al.* 2001). Chemical conditioning (the addition of flocculants) of the sludge can improve dewaterability (Mowla *et al.* 2013).

Both thickening and dewatering processes are difficult to model mechanistically, although efforts have been made (e.g., Severin & Grethlein 1996; Severin *et al.* 1999; Olivier *et al.* 2004). Severin & Grethlein (1996) developed a model based on Darcy's law to model the gravitational part of a belt press thickener. A belt press thickener uses gravity to drain the sludge and incorporates presses to compress the sludge cake created in the gravity zone. The model, developed based on data from laboratory tests, establishes a relationship between the kinetics of gravity drainage in belt press thickeners and operational parameters such as belt speed and solid loading (Severin & Grethlein 1996). The model inputs include information about the permeability of the sludge cake and filter cloth, respectively. Both variables will vary throughout the life cycle of the belt press (Severin & Grethlein 1996). Severin *et al.* (1999) later extended the model by adding a term to account for standing water on the sludge cake. However, the model is still dependent on inputs that will change during operation. Olivier *et al.* (2004) propose an empirical model for the dewatering of activated sludge. The model uses the initial dry solid (DS) content and the sludge loading to predict the DS content in thickened activated sludge over time. More recently, Gutin *et al.* (2018) developed a model to determine the solid capacity of the belt thickener depending on the belt speed. The model inputs include the specific resistance of the sludge, the density of the sludge, and the viscosity of the filtrate. One attempt has previously been made to model the primary sludge belt thickener at Henriksdal water resource recovery facility (WRRF), which is situated in Stockholm, Sweden. Uusitalo (n.d.) used a Bayesian approach to predict DS concentration in the incoming and thickened primary sludge. Unpublished results suggest that polymer concentration, belt speed, pressure downstream of the thickener, and the flow rate of the thickened sludge are important to determine the DS concentration in thickened sludge. Furthermore, it is proposed that more data and data with higher quality are needed to fully describe the process (Uusitalo n.d.). A recent study by Li *et al.* (2023) showed that artificial neural networks (ANNs) can be used to predict the dewaterability of activated sludge from wastewater in a pressurized filter. They used process parameters, including alkalinity, particle size, solid–liquid interface energy, and chemical dosage to predict the water content in dewatered sludge. They found that the operating conditions were more important in predicting the dewaterability than the sludge properties (Li *et al.* 2023).

Plant description

Henriksdal WRRF serves approximately 780,000 people. The primary treatment consists of screens, sand traps, and pre-settling. The biological treatment process is an activated sludge process with pre- and post-denitrification. Iron sulphate (Fe^{2+}) is dosed to the influent wastewater for chemical P-removal.

The primary sludge withdrawn from the pre-settlers is digested anaerobically, together with waste-activated sludge, to produce biogas. The primary sludge is thickened using a belt thickener. The DS content in the primary sludge varies between 3 and 5%, and, after thickening, the sludge has a DS content of 4–7%. The desired DS content in the thickened sludge is around 6%, according to the operators at Henriksdal WRRF. A low DS content involves an increased energy demand for pumping and a lower hydraulic retention time in the anaerobic digestors. If it, on the other hand, is too high, the sludge is no longer pumpable and can cause clogging in both pipes and pumps.

Polymer is added as a flocculant to enhance the thickening and to achieve the desired DS content (Figure 1). The dosing of polymer is done proportionally to the DS content in the incoming primary sludge based on the empirical experience of the operators. One reason for not using more advanced control strategies is that the sensor measuring the DS content in the thickened primary sludge works very poorly. The DS content is measured using an in-line optical sensor that is prone to clogging, causing the sensor to output erroneous signals. This makes it difficult to monitor the performance of the thickener, which has caused the operators not to use it and instead to rely solely on rather scarce lab analyses. Tests carried out in-house indicate that the sensor outputs erroneous signals just hours after cleaning and calibration.

Soft sensors

Soft sensors (or software sensors) are useful for predicting and monitoring variables that are otherwise difficult to measure, either due to physical limitations or because of cumbersome or expensive analyses, by utilizing known relations and well-monitored variables to estimate variables of interest (Haimi *et al.* 2013). The soft sensor can either be a mechanistic model that makes use of known physical, chemical, or biological relations between monitored variables to estimate

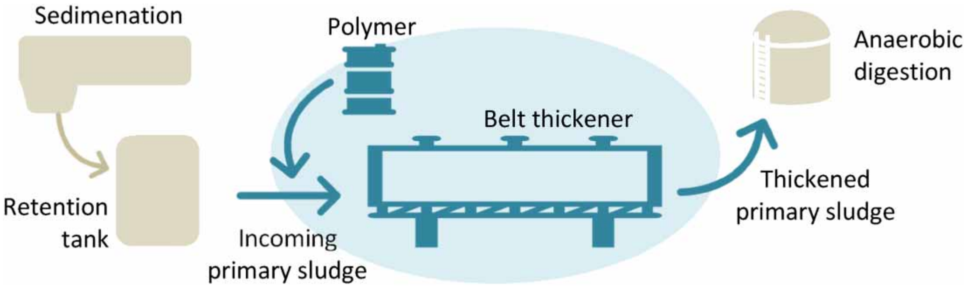

Figure 1 | A schematic overview of the primary sludge thickener unit at Henriksdal WRRF.

non-monitored variables, models developed using multivariate methods such as principal component analysis (Mali & Laskar 2020), or ANNs (Pisa *et al.* 2019), and other machine learning (ML) techniques (del Olmo *et al.* 2019). In the last decades, there has been a notable transition from mechanistic models to data-driven models with an immense addition of soft sensors developed using ML or artificial intelligence in general (Haimi *et al.* 2013; Ching *et al.* 2021). ANNs are widely used because of their ability to capture non-linear relationships (Ching *et al.* 2021). ANNs consist of a series of hidden layers and neurons, which map the input to the output variables, usually in a highly non-linear manner. There is a multitude of ANN architectures available depending on modelling objective and characteristics of the dataset. Long short-term memory (LSTM) networks are recurrent ANNs that are suitable for datasets with time dependencies. They can learn what information to remember and when to forget it by using its internal states (Hochreiter & Schmidhuber 1997). In wastewater treatment, LSTM-based soft sensors have been used to, for example, monitor treatment results (Xu *et al.* 2023), predict effluent violations (Pisa *et al.* 2019), and detect faults (Mandipoor *et al.* 2020).

Objective

The objective of this study is to develop a soft sensor to estimate the DS content of the thickened primary sludge at Henriksdal WRRF to determine whether replacing the physical sensor with a soft sensor could improve the monitoring (and in extension, control) of the process. A more reliable measure of the DS content will extend the opportunities for more efficient and robust control. An advanced control strategy, or even just adding a feedback loop to the existing controller, makes it possible to control the polymer dosing to achieve the desired DS content. This will give a more stable and predictable flow and solids load to the digestors. It will hopefully also minimize the occasions when dilution is needed, which is required occasionally since the sludge is not pumpable if it is too thick. Efficient control can reduce the polymer consumption, as well as the need for sludge transports, and the pumping energy demand. Lastly, dual measurements (soft sensor + online sensor) can be utilized to detect deviations or faults or identify maintenance needs (Darvishi *et al.* 2023; Nie *et al.* 2023).

METHODS

Data collection and pre-processing

Soft sensor output

The DS content in the thickened sludge (DS_{out}) is measured using an in-line and online optical hardware sensor. An extensive sampling campaign was therefore initiated to improve the data quality and quantity. At the start of the campaign, the DS_{out} sensor was cleaned properly and calibrated. Manual samples were collected and analysed following the European standard EN 15934:2012, with a varying frequency of 3–5 times per week (often several samples during the day of measurement). The sensor was calibrated accordingly after each sampling, with up to a few hours delay. The manual samples are shown in Figure 2, both on a time axis (top) and as individual samples (bottom). The figure highlights how poorly the online sensor works for most of the time, with large discrepancies between the online sensor and the manual samples.

The lab analyses provide a snapshot of the DS content, but since the content varies on an hourly and even minute basis, higher frequency data were needed for the soft sensor to be able to capture the dynamics of the thickening process. Therefore, it was assumed (based on empirical tests) that the online sensor provided accurate data for 24 h after calibration. Three days when the in-line sensor was working properly were then identified and used for *training* the soft sensor. The resolution of the dataset was 1 min, adding up to a total of 4,320 samples. Another dataset consisting of 1,440 samples corresponding to 1 day

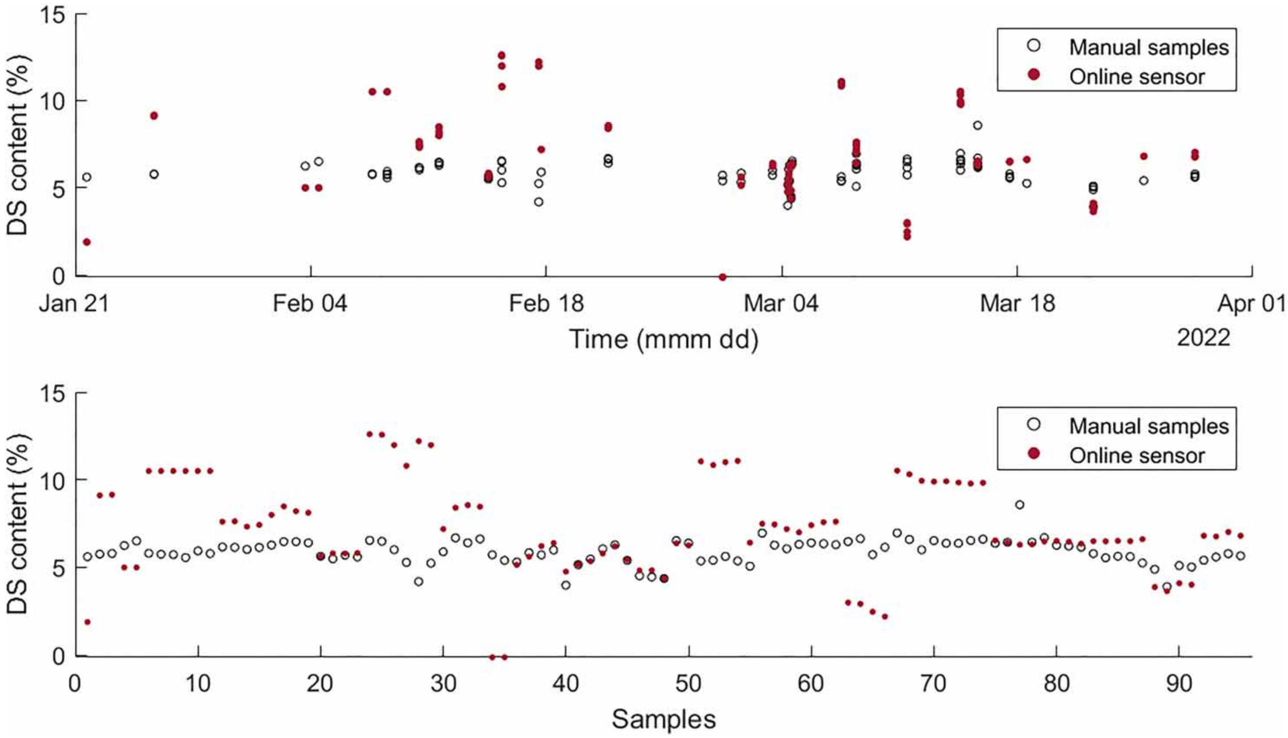

Figure 2 | Manual samples and the corresponding online sensor values show how poorly the sensor works for the majority of the time (*note*: non-equidistant samples in the lower figure).

of 1-min-data was used to *test* the trained models. The distributions of the two datasets are shown in Figure 3. The models were also tested against the manual samples, with a total of 95 samples. The datasets with online data were collected while all sensors were fully functional, and the process was operating under normal conditions. Note that the manual sampling and the enhanced maintenance on the online sensor guided the data collection. The training and testing datasets do not overlap in time. The input variables were standardized using the *Z*-score (Kreyszig 2006) before being used for model development.

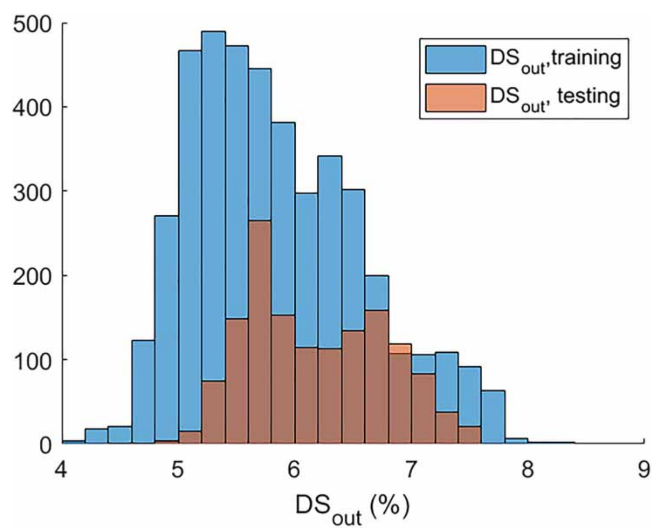

Figure 3 | Distribution of DS_{out} of the training dataset and the test dataset.

Input variables

The first set of input variables (or *features*) was selected based on expert knowledge of the process and the treatment plant. This narrowed the number of variables to 27. Further feature selection was done by analysing the features of Gini indices from a random forest model (Menze *et al.* 2009). This eliminated an additional 10 variables, resulting in a total of 17 features including pressures, flow rates, and polymer dosage (Table 1).

Figure 4 shows the relative location of the sensors used. The poorly functional online DS sensor is indicated in a darker shade than the 17 input variables. Figure 5 displays a selection of the standardized inputs over 1 day (test dataset). Note that P_2 varies less than the other pressures; hence, it is presented on a secondary *y*-axis. The spikes in the data (e.g. Q_{out}) originate from the addition of rinsing water (Q_t), as the thickener is rinsed periodically for maintenance purposes.

Table 1 | The selected input variables (features)

	Description	Unit
Q_p	Polymer flow rate	L/h
Q_{in}	Incoming primary sludge flow rate	m^3/h
Q_d	Dilution water flow rate	L/h
Q_f	Rinsing water flow rate to sludge funnel	m^3/h
Q_t	Rinsing water flow rate to thickener	m^3/h
Q_{out}	Outgoing primary sludge flow rate	m^3/h
SS	Suspended solids in reject water	mg/L
T	Temperature in the thickened sludge	°C
P_1	Pressure, pressure side	bar
GP_1	Pump, primary sludge	%
P_2	Pressure, pressure side	bar
C	Actual polymer dosing	kg/t DS
P_4	Pressure, pressure side	bar
P_3	Pressure, suction side	bar
GP_2	Pump, thickened sludge	%
C_c	Calculated polymer dose	kg/t DS
DS_{in}	DS content incoming sludge	%

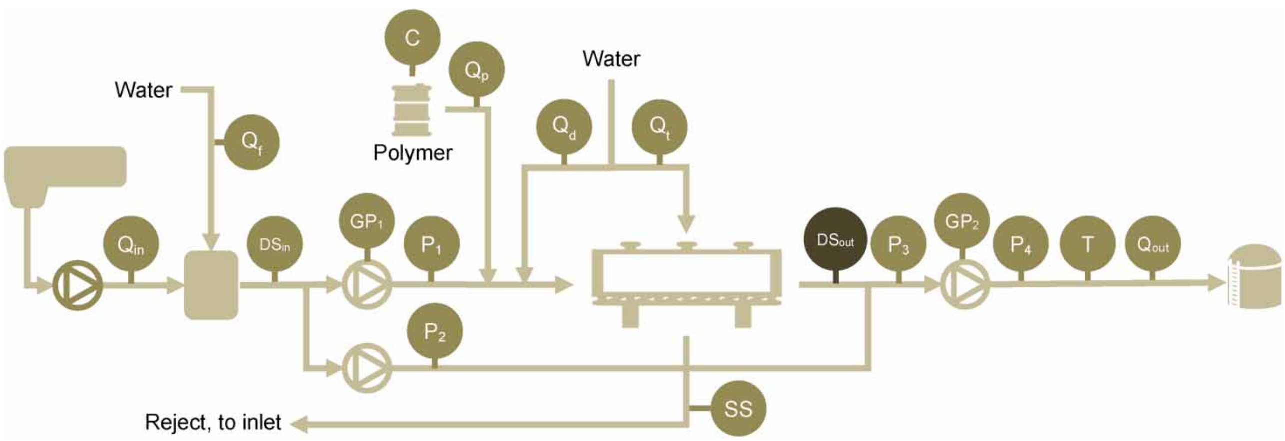

Figure 4 | Schematic overview of the primary sludge thickener unit and the placement of sensors.

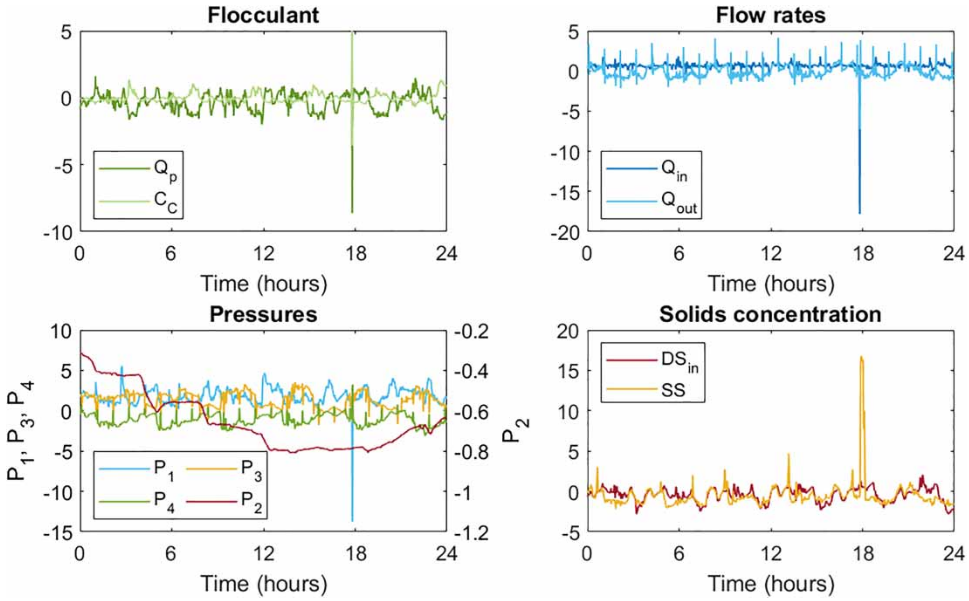

Figure 5 | A selection of standardized model inputs in the test dataset.

Model development and evaluation

The models proposed by Severin & Grethlein (1996), Severin *et al.* (1999), Olivier *et al.* (2004), and Gutin *et al.* (2018) have contributed to understanding and improving the design and operation of belt thickeners but are less suitable for real-time simulation. The proposed models rely on laboratory data, making the models less suitable to use as soft sensors. Looking at the difficulties with mechanical modelling of sludge thickening, in combination with the intended use area of the final model (soft sensor), it was deemed that a data-driven approach would be more feasible in this case. Initially, a linear regression model was developed. Linear regression models can be effective within confined intervals even if the process dynamics are non-linear and can provide valuable information about the system. Furthermore, it is a sound approach to start with simple models before going for complex solutions. The linear regression model was fitted using Matlab 2021b.

Upon evaluation (see *Results*), it was clear that the model could not capture all dynamics in the system and therefore more sophisticated methods were evaluated. Bröndum (2022) developed and evaluated various ML architectures for learning the soft sensor. The models were evaluated based on their mean squared error (MSE) with respect to the laboratory samples, a total of 95 samples. Two architectures, multi-layered perceptron (MLP) and LSTM, performed well (Bröndum 2022). The models did not show a significant difference in performance. However, the reported MSE of the LSTM network was slightly lower than that of the MLP network, and thus, it was chosen for further investigation in this study. The MLP network will be excluded henceforth in this paper.

The full description of the training of the LSTM network can be found in Bröndum (2022). A brief summary is given here. Hyperparameter tuning of the LSTM network was done using grid search and linear searches with three-fold time-series cross-validation on the training dataset (Mohammed *et al.* 2021). Time-series cross-validation is suitable for recurrent neural networks, such as the LSTM network, to avoid distorting time dependencies in data. For each iteration, a subset of the training data is used for training and validation. In the next iteration, the training data consists of the training and validation data from the previous iteration (Mohammed *et al.* 2021). An illustration is shown in Figure 6. The set of hyperparameters with the best score was later evaluated using the test dataset. Dropout regularization and early stopping regularization were applied to prevent overfitting (Srivastava *et al.* 2014). Details on the LSTM architecture can be found in Hochreiter & Schmidhuber (1997). The model was developed using primarily the TensorFlow library in Python (Abadi *et al.* 2015).

The performance of the trained LSTM network and the fitted linear regression model was evaluated based on the MSE with respect to the laboratory samples and with respect to online data (test dataset) to determine whether the models could capture the minute variations or not. A very simple model, namely estimating DS_{out} as the average of the laboratory analyses from the

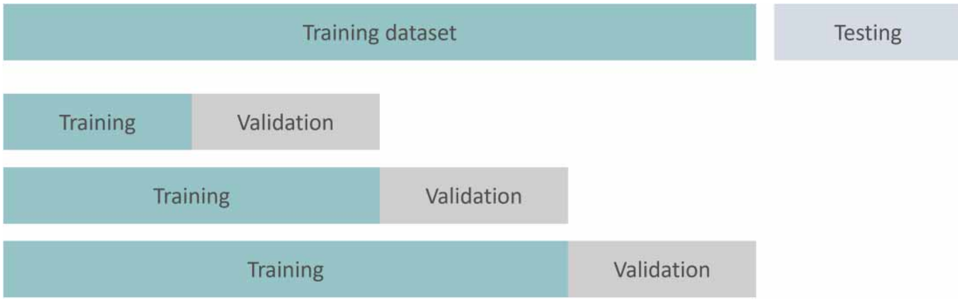

Figure 6 | Illustration of time-series cross-validation.

measurement campaign ($DS_{out} = 5.84\%$), is included in the results for comparison since the long-term average represents the monitoring currently in place.

Feature importance of trained models

The permutation feature importance was calculated for the developed models to understand what features are most important in predicting DS_{out} (Breiman 2001). Permutation feature importance is calculated by randomly shuffling each feature (one at a time, the rest kept intact) in the training dataset and calculating the MSE for each iteration. This gives a measure of the relative contribution of each feature to the output. The larger the increase in the MSE, the more important the feature (Breiman 2001). The procedure was repeated 10 times. Finally, the relative change in MSE, MSE_{irel}, was calculated as follows:

$$\mathrm{MSE}_{i_{\mathrm{rel}}} = \frac{\mathrm{MSE}_i - \mathrm{MSE}_0}{\mathrm{MSE}_0}$$

where MSE_i is the MSE obtained when permuting feature i, and MSE_0 is the training MSE.

RESULTS

The hyperparameters of the LSTM network and the parameters of the linear regression model can be found in Appendix A.

Table 2 shows the performances of the models in terms of MSE with respect to the laboratory samples and the test dataset (online sensor data). Table 2 also includes the MSE achieved during training to evaluate the models' generalizability. The training MSE is lower for the linear regression model (MSE = 0.0960) than for the LSTM network (MSE = 0.136). However, the test MSE is significantly higher for the linear regression model (MSE = 0.590) than for the LSTM network (MSE = 0.181). The difference between training and testing MSEs for the linear regression model indicates that it fails to generalize when used on a new dataset. Furthermore, the linear regression model performs worse than using the long-term average of the laboratory samples. The training MSE when using a long-term average of laboratory samples is 0.444, while the same is 0.590 for the linear regression model.

The training and test MSEs are calculated with respect to data from the online sensor. It was assumed that the sensor was fully functional during this period, but this cannot be fully confirmed due to the scarcity of the laboratory samples. Therefore, the MSEs of the models' outputs and the online sensor were calculated based on the laboratory samples. This evaluation

Table 2 | The MSEs with respect to the training dataset (training MSE), the laboratory samples (laboratory MSE) and the test dataset (test MSE)

	Training MSE	Laboratory MSE	Test MSE
LSTM	0.136	0.341	0.181
Linear regression	0.0960	1.74	0.590
Average laboratory samples	–	0.651[a]	0.444
Online sensor	–	8.98	–

[a]This number reflects the variance in the laboratory samples.

shows that the online sensor works poorly, which confirms the operators' experiences. It also shows that the relatively low MSE achieved when using the average of the laboratory samples indicates that modelling DS_{out} as a long-term average could be a 'good-enough' estimate. Note that the MSE with respect to the laboratory samples reflects the variance in the laboratory samples.

Figure 7 shows the model outputs, the average of the manual samples, and the online sensor values from the test dataset. Visually, the LSTM network gives accurate output throughout the whole time series. The linear regression model underestimates the DS content in general, but especially when the DS content decreases.

The true vs. predicted plots in Figure 8 show the same tendencies. A model with zero prediction error would have all data aligned along the black line. As seen in the figure, almost all data from the linear regression model lie below the black line, which means that the model underestimates the DS content. The LSTM network performs better but slightly underestimates DS_{out} for DS contents above approximately 6.5%. This can be a result of somewhat different distributions in the training and test datasets.

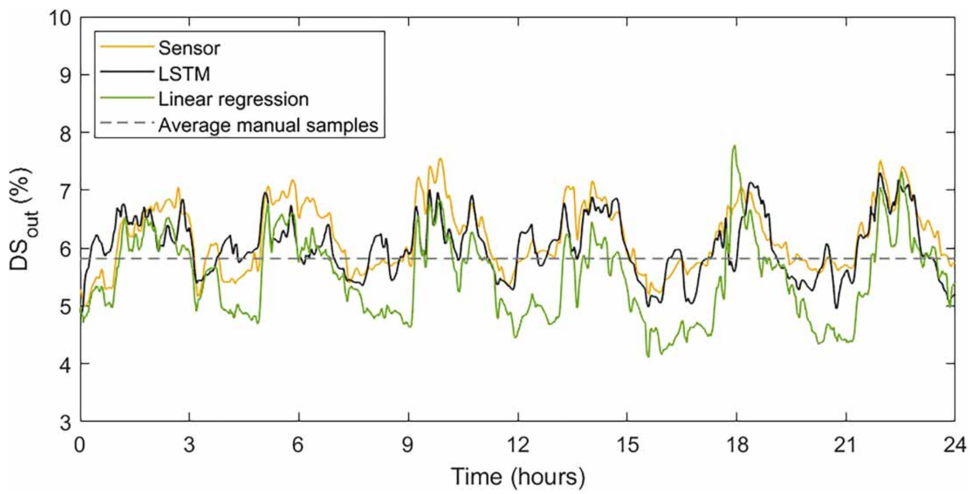

Figure 7 | Measured DS from the online sensor and estimated DS from the models based on the test dataset.

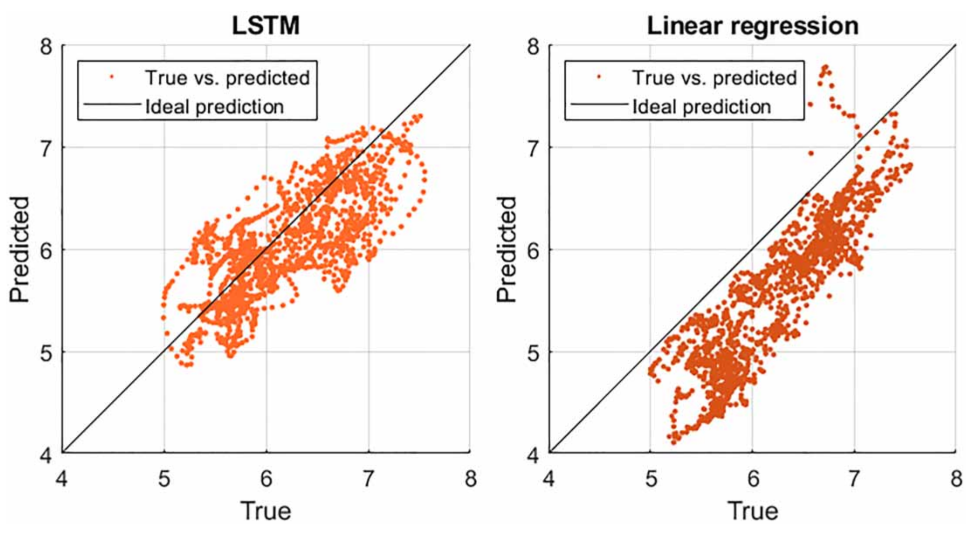

Figure 8 | True vs. predicted plots for the LSTM network and the linear regression model.

The results of the permuted feature importance are presented in Appendix A. The results indicate that the most important features in predicting DS_{out} for the LSTM network are the incoming DS content (DS_{in}), the polymer concentration (C_C), the pump rate (GP_2), and the pressure on the suction side of GP_2 (P_3). The linear regression model was most sensitive to changes in P_3 followed by GP_2, and P_2.

DISCUSSION

The two evaluated models, a LSTM network and a linear regression model, were used to predict the DS content in thickened primary sludge. Both models predict the DS content with higher accuracy than the existing in-line optical sensor. The LSTM network was superior to the linear regression model, most likely because it can capture the non-linear dynamics in the process that the linear model cannot. It could also be due to time-lags in the system that were not accounted for in the development of the linear regression model. Linear regression models are static, whereas LSTM networks are dynamic as the model uses information from previous time-steps to make the next prediction. This can partly explain why the linear regression model works poorly. However, since there is a discrepancy between the training and testing MSE of the linear regression model, it is likely that the model is overfitted to the training data and therefore fails to generalize when exposed to new data. Moreover, looking at the optimal hyperparameters of the LSTM network, something to highlight is the window size (=1). The interpretation of this is that the model only looks at the previous time step, meaning that time-steps earlier than that have a low impact on the prediction at the current time. This could be an indication that an auto-regressive or dynamic model should not be necessary and that a static model should be sufficient to describe the process. More efforts in model development or feature importance could potentially increase the performance of the linear regression model. However, since the thickening process likely has non-linear dynamics, it will probably not be able to fully describe the process in a satisfactory way.

The LSTM network underestimates DS contents above 6.5% when using the test dataset. It is possible that the training and test datasets are slightly different, and that the ML model fails to extrapolate outside the scope of the training data. Looking solely at DS_{out} (Figure 3), the datasets overlap but the test data is slightly shifted compared to the training data. The linear regression model consistently underestimates the DS content when using the test dataset. This could indicate that the training and testing datasets differ in their inputs and that the model is overfitted to the training data, or that the assumption that the online sensor is fully functional does not hold. A natural next step to improve both models would be to expand the datasets to cover more operational conditions.

The analysis of the feature importance showed that the pressure P_3 is the most important feature of the linear regression model. This pressure is measured downstream of the thickener and should physically not affect the sludge thickening. The increased pressure is instead a result of the increased DS content (thicker sludge causes higher pressure). This means that the model does not capture the dynamics of the sludge thickening process but still finds correlations in the data that can be utilized to predict DS_{out}. P_3 is one of the most important features of the LSTM network as well, but the polymer addition and the incoming DS content are the two most important features of the LSTM network. This is expected since the sludge quality and conditioning are important for the sludge dewaterability (Mowla *et al.* 2013). This also highlights that the LSTM network captures some of the intricate processes associated with sludge thickening in a way that the linear regression model fails to do. Based on this, and the MSEs of the models, the LSTM network is more suitable for modelling the thickening *process*. However, recalling the objective of this study, to determine whether a soft sensor could be a viable option to replace the physical sensor, the study shows that the LSTM network, the linear regression model, and a long-term average based on laboratory samples will give a better estimate of the DS_{out} than the physical sensor and could fill the needs of the operators. To rely on the soft sensor (i.e. the LSTM network or the linear regression model), the operators need to ensure the quality of the input data, which, with 17 features, will be laboursome. With that in mind, continuous manual sampling and laboratory analyses could be an option, with the drawback that the minutely or hourly variations in DS_{out} will not be captured and, therefore, not optimal for real-time control. Exploring the trade-off between model performance and the number of features should be considered if the goal is to use the soft sensor in real time at the facility.

An accurate sensor is essential to monitor the process performance. Replacing or complementing the online sensor with the LSTM network as a soft sensor would make it possible to explore and evaluate control strategies that potentially could lower the polymer consumption and optimize the DS content in the thickened primary sludge to improve biogas production. The same could be achieved by more frequent maintenance of the existing sensor but at the cost of increased manhours. Cleaning and calibration of the sensor are occasionally needed every 2–3 days. Dual measurements (online sensor and soft sensor) can be utilized to detect deviations or faults or identify maintenance needs (Darvishi *et al.* 2023; Nie *et al.* 2023) such as cleaning or flushing the pipes and cleaning or calibrating the sensor.

One could argue that the natural way to improve the monitoring of the thickener unit would be to replace the existing DS sensor with another hardware sensor instead of developing a soft sensor. The issue at Henriksdal WRRF is that the pipes clog, which affects the optical sensor. It is not the sensor itself that malfunctions. Installing a new sensor would therefore require extensive reconstruction in terms of new piping and relocation of pumps. A soft sensor could be a cost-efficient option.

CONCLUSIONS

- The LSTM network and the linear regression model were able to predict the DS content in the thickened primary sludge with higher accuracy than the existing in-line optical sensor.
- The linear regression model does not give better estimates of the DS content than using a long-term average of laboratory analyses.
- The developed LSTM model captures minute variations to a reasonably good accuracy.
- Replacing the existing online sensor with a soft sensor can provide better opportunities to monitor the process and can, by extension, possibly be used for feedback control of the polymer dosing or to identify maintenance needs for the existing sensor.

ACKNOWLEDGEMENTS

The research work was financed by Stockholm Vatten och Avfall AB in collaboration with Käppala Association, Syvab, IVL Swedish Environmental Research Institute, and Svenskt Vatten. The authors would like to thank the personnel at Henriksdal WRRF for providing data and lab analyses.

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

Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., Corrado, G. S., Davis, A., Dean, J., Devin, M., Ghemawat, S., Goodfellow, J., Harp, A., Irving, G., Isard, M., Jozefowicz, M., Jia, Y., Kaiser, L., Kudlur, M., Levenberg, J., Mané, D., Schuster, M., Monga, R., Moore, S., Murray, D., Olah, C., Shlens, J., Steiner, B., Sutskever, I., Talwar, K., Tucker, P., Vanhoucke, V., Vasudevan, V., Viégas, F., Vinyals, O., Warden, P., Wattenberg, M., Wicke, M., Yu, Y. & Zheng, X. 2015 *TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems*. Software Available from tensorflow.org. Available from: https://www.tensorflow.org/ (visited on 15/12/2023).

Breiman, L. 2001 Random forests. *Machine Learning* **45** (1), 5–32.

Bröndum, E. 2022 *Modeling of the Primary Sludge Thickening Process at a Wastewater Treatment Plant with the Use of Machine Learning. Master Thesis*. Available from: https://urn.kb.se/resolve?urn=urn:nbn:se:kth:diva-319876.

Ching, P. M. L., So, R. H. Y. & Morck, T. 2021 Advances in soft sensors for wastewater treatment plants: A systematic review. *Journal of Water Processing Engineering* **44**, 102367.

Darvishi, H., Ciuonzo, D. & Salvo Rossi, P. 2023 A machine-learning architecture for sensor fault detection, isolation, and accommodation in digital twins. *Sensors* **23** (3), 2522–2538. doi: 10.1109/JSEN.2022.3227713.

del Olmo, F. H., Gaudioso, E., Duro, N. & Dormido, R. 2019 Machine learning weather soft-sensor for advanced control of wastewater treatment plants. *Sensors* **19** (14), 3139.

EN 15934:2012. *Sludge, Treated Biowaste, Soil and Waste – Calculation of Dry Matter Fraction After Determination of Dry Residue or Water Content*.

Gutin, Y. V., Lavrinenko, A. A. & Golberg, G. Y. 2018 Modeling belt press filter performance in fine suspension dewatering. *Theoretical Foundations of Chemical Engineering* **52**, 554–559. https://doi-org.ludwig.lub.lu.se/10.1134/ S0040579518040383.

Haimi, H., Mulas, M., Corona, F. & Vahala, R. 2013 Data-derived soft-sensors for biological wastewater treatment plants: An overview. *Environmental Modelling & Software* **47**, 88–107.

Hochreiter, S. & Schmidhuber, J. 1997 Long short-term memory. *Neural Computation* **9**, 1735–1780.

Houghton, J. I., Quarmby, J. & Stephenson, T. 2001 Municipal wastewater sludge dewaterability and the presence of microbial extracellular polymer. *Water Science and Technology* **44** (2), 373–379.

Kreyszig, E. 2006 *Advanced Engineering Mathematics*, 9th edn. John Wiley & Sons, Hoboken, NJ, USA, p. 1018.

Li, H., Li, C., Zhou, K., Ye, W., Lu, Y., Chai, X., Dai, X. & Wu, B. 2023 Intelligent upgrade of waste-activated sludge dewatering process based on artificial neural network model: Core influential factor identification and non-experimental prediction of sludge dewatering performance. *Journal of Environmental Management* **346**, 118968. doi: 10.1016/j.jenvman.2023.118968.

Mali, B. & Laskar, S. 2020 PLS-based multivariate statistical approach for soft sensor development in WWTP. Shreesha, C. and Gudi, R.D. (eds.). *Lecture Notes in Electrical Engineering, Proceedings of the 15th Annual Control Instrumentation System Conference (CISCON)*, October 26–27, Manipal, India, pp. 123–131.

Mandipoor, B., Majd, M., Sheikhalishahi, S., Modena, C. & Osmani, V. 2020 Monitoring and detecting faults in wastewater treatment plants using deep learning. *Environmental Monitoring Assessment* **192**, 148. doi: 10.1007/s10661-020-8064-1.

Menze, B. H., Kelm, B. M., Masuch, R., Himmelreich, U., Bachert, P., Petrich, W. & Hamprecht, F. A. 2009 A comparison of random forest and its Gini importance with standard chemometric methods for the feature selection and classification of spectral data. *BMC Bioinformatics* **10** (1), 213. doi: 10.1186/1471-2105-10-213.

Mohammed, A., Khedr, A., AlHaj, D., Al Khalifa, R. & Alqaddoumi, A. 2021 Time-series cross-validation parallel programming using MPI. In: *2021 International Conference on Data Analytics for Business and Industry (ICDABI)*, December 2021, pp. 553–556. doi:10.1109/ICDABI53623.2021.9655795.

Mowla, D., Tran, H. N. & Allen, D. G. 2013 A review of the properties of biosludge and its relevance to enhanced dewatering processes. *Biomass and Bioenergy* **58**, 365–378. https://doi.org/10.1016/j.biombioe.2013.09.002.

Nie, L., Ren, Y., Wu, R. & Tan, M. 2023 Sensor fault diagnosis, isolation, and accommodation for heating, ventilating, and air conditioning systems based on soft sensor. *Actuators* **12** (10), 389. doi: 10.3390/act12100389.

Olivier, J., Vaxelaire, J. & Ginisty, P. 2004 Gravity drainage of activated sludge: From laboratory experiments to industrial process. *Journal of Chemical Technology & Biotechnology* **79** (5), 461–467. doi: 10.1002/jctb.977.

Peirce, J. J., Weiner, R. F., Vesilind, A., 1998 Chapter 9 – Sludge treatment, utilization and disposal. In: *Environmental Pollution and Control*, 4th edn. (Peirce, J. J., Weiner, R. F. & Vesilind, P. A., eds). Butterworth-Heinemann, Oxford, UK, pp. 125–135. https://doi.org/10.1016/B978-075069899-3/50010-9.

Pisa, I., Santín, I., Vicario, J. L., Morell, A. & Vilanova, R. 2019 ANN-based soft sensor to predict effluent violations in wastewater treatment plants. *Sensors* **19** (6), 1280.

Samanta, N. S., Das, P. P., Sharma, M., Purkait, M. K., 2023 12 - Recycle of water treatment plant sludge and its utilization for wastewater treatment. In: *Resource Recovery in Drinking Water Treatment* (Sillanpää, M., Khadir, A. & Gurung, K., eds). Elsevier, London, UK, pp. 239–264. https://doi.org/10.1016/B978-0-323-99344-9.00010-4.

Severin, B. F. & Grethlein, H. E. 1996 Laboratory simulation of belt press dewatering: Application to gravity drainage. *Water Environment Research* **68**, 359–369. doi: 10.2175/106143096X127802.

Severin, B. F., Nye, J. V. & Kim, B. J. 1999 Model and analysis of belt drainage thickening. *Journal of Environmental Engineering* **125** (9), 807–815. doi: 10.1061/(ASCE)0733-9372(1999)125:9(807).

Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I. & Salakhutdinov, R. 2014 Dropout: A simple way to prevent neural networks from overfitting. *Journal of Machine Learning Research* **15**, 1929–1958.

Uusitalo, L. n.d. *Empirical Modelling of Sludge Total Solids Content at the Henriksdal Wastewater Treatment Plant*. *Master Thesis*, Linköping University, Linköping, Sweden.

Xu, B., Kwek Pooi, C., Tan, K. M., Huang, S., Shi, X. & Ng, H. Y. 2023 A novel long short-term memory artificial neural network (LSTM)-based soft-sensor to monitor and forecast wastewater treatment performance. *Journal of Water Process Engineering* **54**. doi: 10.1016/j.jwpe.2023.104041.

First received 15 January 2024; accepted in revised form 8 July 2024. Available online 22 July 2024