Predicting sludge settleability in large wastewater treatment plants: a deep learning time series perspective

François Guichard, Madiha Nadri-Wolf, Rachid Ouaret, Antonin Azaïs

Predicting sludge settleability in large wastewater treatment plants: a deep learning time series perspective

François Guichard[1,2], Madiha NADRI-WOLF [2,] Rachid OUARET[3] and **Antonin AZAÏS**[1,*]

[1] INRAE, UR REVERSAAL, 5 Rue de La Doua, 69625, Villeurbanne, France

[2] LAGEPP UMR 5007, 43 bd du 11 novembre 1918, 69100, Villeurbanne, France

[3] LGC, CNRS, 118 Route de Narbonne, 31062, Toulouse, France

[*] Corresponding author, antonin.azais@inrae.fr

Keywords: Sludge settleability, Time series, Machine learning, Neural Network.

Abstract:

This study investigates the application of machine learning tools to predict sludge decantability in wastewater treatment. Machine learning covers all the methods used to design models from experimental data. Focusing on the activated sludge treatment process, we use five years of monitoring data from a large Wastewater Treatment Plant (WWTP). The aim is to develop models using automatic water quality monitoring data to estimate the quantities characterising the sludge's settling properties (SV30, Sludge Volume Index (SVI) and Filamentous Index (FI)). Indeed, measuring these properties presents a number of operational difficulties and financial constraints. Through the application of different pre-treatment methods, we develop two dynamic models based on Recurrent Neural Networks (RNN) that could provide estimates for a 100 days horizon with an uncertainty level similar to that of real measurements.

1. INTRODUCTION

To ensure the effective elimination of the domestic pollution, WWTPs are finely controlled, generally requiring control loops, supervision, online sensors, in addition to laboratory measurement campaigns (off-line), which take time to carry out. Challenges persist in terms of instrumentation, data acquisition and analysis. Issues as fouling and biofilm, ionic interference and corrosion have already been identified in the literature (**Ching et al. 2021**). Furthermore, the limited use of time series acquired throughout the treatment processes does not allow taking advantage of the current data massification (**Newhart et al. 2019**). In order to minimise equipment and maintenance costs, data-driven soft-sensors are a relevant alternative. This approach uses readily available measurements (for example : T°, pH, flow, COD, etc.) to estimate output variables that are difficult to measure, using reduced knowledge models, neural networks or fuzzy logic (**Corominas et al. 2018**). Although this research topic has been studied to predict various parameters as BOD_5 or E. Coli concentrations **(Li and Zhang, 2020 ; Foschi et al. 2021)**, to our knowledge, no study has deployed this methodology in biological reactors to estimate sludge settlability ((SV30, Sludge Volume Index (SVI) and Filamentous Index (FI)), while it is one of the most limiting factors for the operation of large WWTPs. Sludge settling depends on many factors as operating conditions but also sludge characteristics (bioflocculation properties, expolymeric susbtances content and nature, microbial communities, etc.) and cannot be predicted empirically.

We propose to apply two learning based models, the GRU (Gated Recurrent Unit) and the LSTM (Long Short-Term Memory) to estimate these two operating parameters. Both these models are variants of recurrent neural networks and in the literature obtain similar performances. While more

recent work in deep learning of time series is based on self-attention and transformers (**Shaw et al. 2018**), we will opt for recurrent models as they allow to explicitly model the process dynamics in a latent space (**Chung et al. 2014**).

2. DATA AND METHODOLOGY

The global methodology is summarized in the Fig. 1 and is composed in three main parts : i) the data collection, ii) the data pre-processing and finally iii) the modelling including the training, test and validation steps.

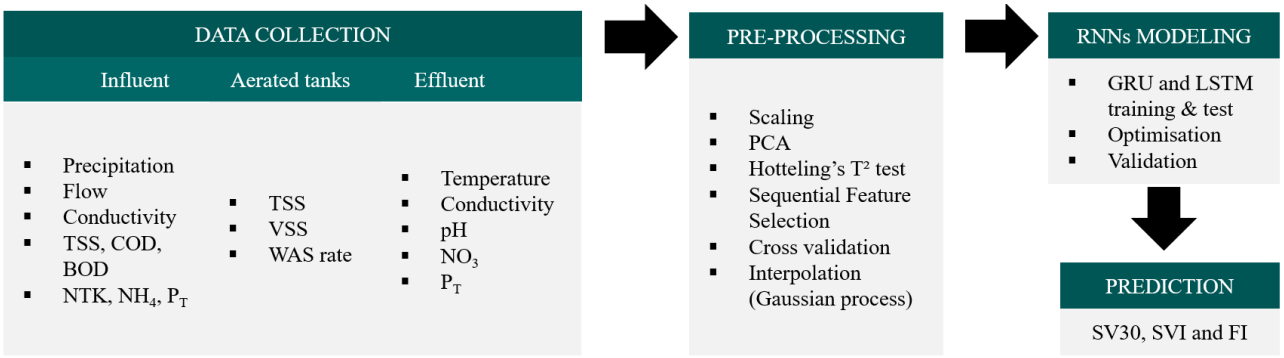

Figure 1. Study methodology : from data collection to prediction.

- *Data collection and pre-processing*

Data pre-processing is a fundamental step to ensure that data-driven models operate correctly and provide reliable predictions. A combination of techniques and expertise is used to address common issues such as missing values, outliers, errors, and inconsistencies while preparing the data for effective use in the learning process. In this work, the data originates from self-monitoring of a wastewater treatment plant with a capacity of around 900,000 p.e., five years of data were collected on three biological nutrient removal (BNR) lines (nitrification - denitrification). A total of 526 SVI and 283 FI measurements were selected, with inlet and outlet measurements daily for more than 1,800 measurements over the five-years. On these data sest, various pre-processing methods were used in this work including normalisation, principal component analysis, Hotteling's T^2 test, using Taskesen's method (**Taskesen et al. 2020**). Gaussian process interpolation was used to consistently complete the time series **(Rasmussen et al. 2006**).

- *Model training, optimisation and validation*

For model training, the k-fold cross validation methodology proposed by *Anguita et al*. (**Anguita et al. 2012**) was used . The training set is divided into k blocks. The model is then trained k times, with k-1 blocks used for training and one block used to test the models. The neural network models are optimised by stochastic gradient descent using ADAM algorithm (**Kingma et al. 2014**). The use of ADAM leads to three hyper-parameters (epochs number or training length, the size of the batches and the learning rate) for which, the most relevant values will be optimized on a hold-out set. The calculated k-scores are averaged. Finally, mean absolute error (MAE) was evaluated as the average difference between the observations (true values) and model output (predictions).

3. RESULTS AND DISCUSSION

- *Data pre-processing*

The method of detecting extreme values using Hotteling's T2 with PCA enabled us to remove 122 outliers from the influent dataset, 14 values from the aerated tanks and 27 values in the times series of WWTP effluent. Interpolation using a Gaussian process was used to obtain a large dataset for TSS, V30, sludge index and even filamentous index values. After interpolation, we obtained 5,876 points for the V30 and BVI and 4,878 for the filamentous index for the three basins combined. From the PCAs, we have determined the quared cosines of variables (Cos2), known as the quality of the representation of variables on an axis) and the percentage of variance explained by the principal components. Beyond PC_4, the components account for less than 4% of the variance in our data, and the cumulative figure is 80% at PC_5. Finally, the explanatory variables for decantation to be taken into account are TSS and VSS influent concentrations, and temperature with N-NO_3 concentration in the effluent.

- *Data-driven modeling*

The models used (GRU and LSTM) are trained to generate estimates, considering both self-monitoring measurements and the previous day's estimate. The chosen models are trained to forecast over a specific number of consecutive days, referred to as the prediction horizon. The RNNs models operate with hidden states that accumulate information, making it advisable to extend the number of initial days when the actual measurement is available. This period is referred to as the 'delay'. To evaluate the performance of the developed models in forecasting and determine the optimal required data quantity, the models are trained with varying memory horizons. **Table 1** presents the main results in terms of MAE for different prediction horizons, highlighting the errors.

Table 1: Comparison of MAE between GRU and LSTM for estimating IVB over different delays and prediction horizons.

Horizon	Delay							
	1		5		10		20	
	GRU	LSTM	GRU	LSTM	GRU	LSTM	GRU	LSTM
1	12,6	12,6	12,2	12,2	12,4	12,3	12,2	12
5	15,1	15,6	15	14,9	15,2	15	15,4	15,3
10	17,7	18,5	17,4	16,9	17,8	17,2	17,6	17,3
20	20,8	20	19,6	19,1	19,9	19,1	20,3	18,3
50	20,6	20,7	19,6	19,1	19,7	20	19,3	20,8
100	23	23,3	22	21,4	21,4	22,5	20,5	20,9

In terms of comparison, predictive capacities of both models are quite similar. The results obtained for prediction on a one-day horizon are comparable to the uncertainty of +/-10 mL/g of a real measurement. However, with a forecasting on horizon of up to 100 days, the average error is at most doubled. A memory of more than one day systematically outperforms a one-day delay. For a prediction horizon spanning 5 to 20 days, the findings suggest that, beyond a 5-day delay there is no significant decrease in MAE. Consequently, we can infer that the optimal MAE is achieved with a five-day delay. Indeed, five days of measurements are sufficient to identify the characteristics of the system for a short-term horizon. For a long-term horizon, the graphic representation allows us to see that it is in fact a decrease in error of up to 15 mL/g on the estimates for the first few days. However, even if the prediction error over a short horizon is small for a given delay, forecasting over a long horizon entails an accumulation of errors over time, necessitating a compromise.

4. CONCLUSION AND PERSPECTIVES

The GRU and LSTM have shown good capacity to model the dynamics of the system with average errors of 15 to 25 mL/g. Their hidden state should therefore contain all the information needed on the state of the system to predict its evolution and the quantities of interest. In our applications, a single MLP is used to decode the hidden state and provide all three estimated quantities (TSS, VSS, SV30/IVB). It might be interesting to test other implementations, for example using one MLP per variable to be estimated. In addition, the model could seek to predict all the variables measured in the basins and at their outlets. Another possibility would be to estimate the settleability variables at different treatment plants, which vary in size and conditions of use, to produce models with better generalisability. In this way, we hope to develop versatile models that can be used in wastewater treatment plants. Finally, a hybrid model combining mechanistic models (ASM type) and these data-derived models would be an attractive development. This model can be then used to develop an observer for the estimation of non-measured states.

Acknowledgments

We are grateful for financial support from the ANR (French National Research Agency) via the Carnot Eau & Environnement.

References

Ching, P.M.L., So, H.Y.R., Morck. T., 2017. Advances in soft sensors for wastewater treatment plants: A systematic review. Journal of Water Process Engineering 44, 102367.

Newhart, K.B., Holloway, R.W., Hering, A.S., Cath, T.Y., 2019. Data-driven performance analyses of wastewater treatment plants: A review. Water Research 157, 498-513.

Corominas, Ll., Garrido-Baserba, M., Villez, K., Olsson, G., Cortes, U., Poch, M., 2018. Transforming data into knowledge for improved wastewater treatment operation: A critical review of techniques. Environmental Modelling & Software 106, 89-103.

Li, W., and Zhang, J., 2020. Prediction of BOD concentration in wastewater treatment process using a modular neural network in combination with the weather condition. Applied Sciences 10 (21), 7477.

Foschi, J., Turolla, A., Antonelli, M., 2021. Soft sensor predictor of E. coli concentration based on conventional monitoring parameters for wastewater disinfection control. Water Research 191, 116806.

Shaw, P., Uszkoreit, J., & Vaswani, A. (2018). Self-attention with relative position representations. arXiv preprint arXiv:1803.02155.

Chung, J., Gulcehre, C., Cho, K., & Bengio, Y. (2014). Empirical evaluation of gated recurrent neural networks on sequence modeling. arXiv preprint arXiv:1412.3555.

Taskesen, Erdogan (oct. 2020). pca : A Python Package for Principal Component Analysis. Version 1.8.4. url : https://erdogant.github.io/pca

Carl Edward Rasmussen and Christopher K. I. Williams (2006) Gaussian Processes for Machine Learning, The MIT Press, . ISBN 0-262-18253-X.

Anguita, D., Luca Ghelardoni, Alessandro Ghio, L. Oneto, and Sandro Ridella. 2012. "The 'K' in K-fold Cross Validation." In The European Symposium on Artificial Neural Networks.

Kingma, D. P., & Ba, J. (2014). Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980