Science of the Total Environment 784 (2021) 147138

Contents lists available at ScienceDirect

Science of the Total Environment

journal homepage: www.elsevier.com/locate/scitotenv

A machine learning framework to improve effluent quality control in wastewater treatment plants

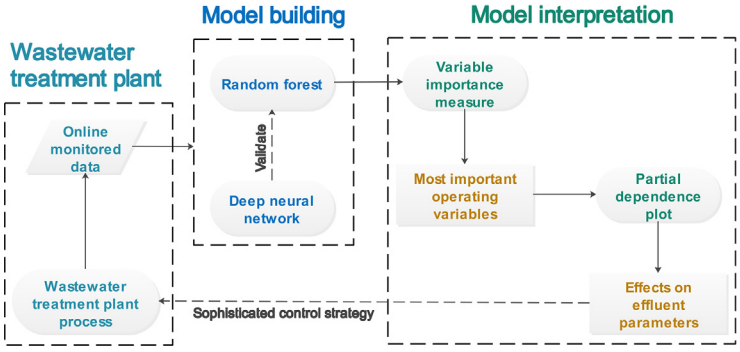
Dong Wang [a], Sven Thunéll [b], Ulrika Lindberg [b], Lili Jiang [c], Johan Trygg [a], Mats Tysklind [a,*], Nabil Souihi [a,*]

[a] Department of Chemistry, Umeå University, SE-901 87 Umeå, Sweden
[b] Vakin, Övägen 37, SE-904 22 Umeå, Sweden
[c] Department of Computing Science, Umeå University, SE-901 87 Umeå, Sweden

HIGHLIGHTS

- WWTP operational factors' effects on effluent are revealed by interpreting Random Forest (RF) models.
- Neural Network is used as reference to check RF's fitting performance.
- Time-lags among process variables are handled rigorously to ensure reliable results.
- More than 100,000 samples are used in case study to guarantee robust results.
- Results can benefit development of advanced process control strategies.

GRAPHICAL ABSTRACT



ARTICLE INFO

Article history:
Received 8 January 2021
Received in revised form 23 March 2021
Accepted 10 April 2021
Available online 16 April 2021

Editor: Dimitra A Lambropoulou

Keywords:
Wastewater treatment
Big data
Interpretable AI
Effluent quality
Process analytics

ABSTRACT

Due to the intrinsic complexity of wastewater treatment plant (WWTP) processes, it is always challenging to respond promptly and appropriately to the dynamic process conditions in order to ensure the quality of the effluent, especially when operational cost is a major concern. Machine Learning (ML) methods have therefore been used to model WWTP processes in order to avoid various shortcomings of conventional mechanistic models. However, to the best of the authors' knowledge, no ML applications have focused on investigating how operational factors can affect effluent quality. Additionally, the time lags between process steps have always been neglected, making it difficult to explain the relationships between operational factors and effluent quality. Therefore, this paper presents a novel ML-based framework designed to improve effluent quality control in WWTPs by clarifying the relationships between operational variables and effluent parameters. The framework consists of Random Forest (RF) models, Deep Neural Network (DNN) models, Variable Importance Measure (VIM) analyses, and Partial Dependence Plot (PDP) analyses, and uses a novel approach to account for the impact of time lags between processes. Details of the framework are provided along with a demonstration of its practical applicability based on a case study of the Umeå WWTP in Sweden involving a large number of samples (105763) representing the full scale of the plant's operations. Two effluent parameters, Total Suspended Solids in effluent (TSS_e) and Phosphate in effluent ($PO4_e$), and thirty-two operational variables are studied. RF models are developed, validated using DNN models as references, and shown to be suitable for VIM and PDP analyses. VIM identifies the variables that most strongly influence TSS_e and $PO4_e$, while PDP elucidates their specific effects on TSS_e and $PO4_e$. The major findings are: (1) Influent temperature is the most influential variable for both TSS_e and $PO4_e$, but it affects them in different ways; (2) $PO4_e$ depends strongly on the TSS in aeration basins – higher TSS concentrations in aeration basins generally promote PO_4 removal, but excess TSS can have negative effects; (3) In general, the impact of TSS in aeration basins on TSS_e and $PO4_e$ increases with the distances of the basin from the merging outlet, so more attention should be paid to the TSS concentration in the third or fourth aeration

* Corresponding authors.
E-mail addresses: mats.tysklind@umu.se (M. Tysklind), nabil.souihi@umu.se (N. Souihi).

https://doi.org/10.1016/j.scitotenv.2021.147138

basins than the first and second ones; (4) Returning excessive amounts of sludge through the second return sludge pipe should be avoided because of its adverse impact on TSS_e removal. These results could support the development of more advanced control strategies to increase control precision and reduce running costs in the Umeå WWTP and other similarly configured WWTPs. The framework could also be applied to other parameters in WWTPs and industrial processes in general if sufficient high-resolution data are available.

Nomenclature

ANN	Artificial Neural Network
ASM	Activated Sludge Model
BOD	Biological Oxygen Demand
CART	Classification and Regression Tree
COD	Chemical Oxygen Demand
DNN	Deep Neural Network
DO	Dissolved Oxygen
GAO	Glycogen-Accumulating Organism
MDI	Mean Decrease Impurity
ML	Machine Learning
OOB	Out-of-Bag
PAO	Polyphosphate-Accumulating Organism
PDP	Partial Dependence Plot
PI	Permutation Importance
PO4	Phosphate
R^2	Coefficient of Determination
RF	Random Forest
TS	Total Solids
TSS	Total Suspended Solids
VIM	Variable Importance Measures
WWTP	Wastewater Treatment Plant

1. Introduction

Wastewater treatment plants (WWTPs) are complex, nonlinear systems with high fluctuations in flow rate, pollutant load, chemical environment, and hydraulic conditions. Due to these complexities and uncertainties, modeling WWTP processes is challenging (Borzooei et al., 2019; Rout et al., 2021; Vučić et al., 2021). Mechanistic models such as Activated Sludge Models (ASMs) have been widely used to simulate WWTP processes and predict the behavior of certain variables (Buaisha et al., 2020; Fenu et al., 2010; Nopens et al., 2009; Wu et al., 2016). However, many simplifications and assumptions are needed to make mechanistic models tractable and computable, so they have many limitations. For example, ASMs are only valid in certain ranges of temperature, pH and alkalinity (Gujer et al., 1995; Gujer et al., 1999; Hauduc et al., 2011; Henze et al., 1987; Henze et al., 1999). In addition, coupling different mechanistic models that simulate processes in different units is difficult because of differences in approaches used to calculate state variables — for example, Total Suspended Solids (TSS) is calculated and incorporated differently in ASMs and second clarifier models (Metcalf, 2013; Volcke et al., 2006). Other intrinsic shortcomings of mechanistic models include inadequate handling of the time-varying and highly nonlinear behaviors of processes affected by various known and unknown factors, inability to comprehensively simulate various processes, high costs, and poor generalization performance (Cao and Yang, 2020; Guo et al., 2015; Liu et al., 2020; Shi and Xu, 2018; Singh et al., 2010; Verma et al., 2013). Machine Learning (ML) models avoid many of these limitations because they are based solely on extracting relationships between output and input data that enable predictions and/or facilitate decisions (Müller and Guido, 2016). An important advantage of ML models is that they reflect real reaction/process situations rather than mechanisms formulated in advance based on fundamental principles. Consequently, they are robust and comprehensive, which is important because many mechanisms involved in wastewater treatment remain unclear (Chan and Huang, 2003; Erdirencelebi and Yalpir, 2011; Faruk, 2010; Lee et al., 2002; Nadiri et al., 2018). ML modeling is therefore widely used as an alternative to mechanistic modeling of WWTPs (Cao and Yang, 2020; Guo et al., 2015; Liu et al., 2020; Shi and Xu, 2018; Singh et al., 2010; Verma et al., 2013).

However, there is a significant gap in the literature on ML modeling of WWTPs: the vast majority of published studies have focused on prediction or building soft sensors using nonlinear ML models without interpreting the models to obtain knowledge about how the studied targets can be influenced or controlled. This is true both for black-box models (e.g., ANN-based approaches) and interpretable models such as tree-based models (Corominas et al., 2018; Dürrenmatt and Gujer, 2012; Haimi et al., 2013; Meng et al., 2021; Newhart et al., 2019). One can argue that ANN-based models are difficult to interpret because the ANN structures or weights provide only minimal information about the approximated functions (Guidotti et al., 2018; Rudin, 2019). However, tree-based models are very interpretable because of their inherent structure (Breiman, 2001; Chen and Guestrin, 2016; Ke et al., 2017). Researchers should take advantage of this because knowledge about how operational factors affect effluent quality is extremely valuable in engineering scenarios to improve WWTP processes. In addition to their greater interpretability, tree-based models are comparable to ANN-based models in terms of the variation captured (often termed 'accuracy/precision' in prediction scenarios) (Ahmad et al., 2017; Chowdhury et al., 2020; Kumari and Toshniwal, 2021; Liu et al., 2013). This is important because capturing more variation increases the reliability of model interpretation and analysis. Another major gap is that models usually treat time lags between process steps in WWTP processes with insufficient rigor or neglect it entirely. This could lead to misleading or incorrect interpretation and analysis of model outputs, which in turn could cause problems when attempting to control processes based on cause-and-effect relationships identified through such analyses and interpretations.

This study is therefore motivated by three key considerations. First, as observed in a local WWTP, Umeå WWTP in Sweden, there is a clear need for developing advanced control strategies that can be used to optimize the use of energy and chemicals in WWTPs without compromising on the effluent quality. The current approaches are inefficient because they rely heavily on 'trial and error' for problem-solving. Second, there is little information on the potential benefits of using ML more extensively to understand and control processes in WWTPs rather than simply to develop soft sensors or generate predictions. Finally, there is no established way of handling time lags between process variables even though these lags must be accounted for in order to reliably and convincingly interpret the trained models.

To address these issues, a novel ML framework based on Random Forest (RF) models (representative of tree-based models), Deep Neural Network (DNN) models, Variable Importance Measure (VIM) analyses, and Partial Dependence Plot (PDP) analyses was developed and used to model WWTP processes and investigate how operational variables influence effluent quality. This paper presents a detailed description of the framework and demonstrates its applicability in engineering

scenarios through a case study on the Umeå WWTP in Sweden involving a large amount of data (105,763 samples) representing the full scale of the plant's operations.

2. Material and methods

2.1. Processes and data sources in Umeå WWTP

As shown in Fig. 1, Umeå WWTP is an activated sludge process-based WWTP. It is the biggest WWTP in Umeå municipality and receives about 13,000,000 m^3 of wastewater annually. This case study focused on the wastewater treatment section, but some sludge flows (brown pipes) were also considered because they are directly connected to the water treatment process. The first main treatment is aerated grit removal, in which organic matter is flushed off dense solids that subsequently settle and are collected. The next treatment is coagulation (in unit 10), during which a large fraction of the suspended matter coagulates in the presence of $FeCl_3$ and then flocculates before settling in a sedimentation basin (unit 11). Most of the phosphorus is removed during this process by chemical precipitation with the coagulant $FeCl_3$. Note that the sole outlet discharging sludge to the sludge system in this water process line is located in the sedimentation basin (unit 11). The water then passes to the biodegradation section (units 12–14), where most of the organic substances are degraded by microorganisms and most of the organic content is removed. After the biodegradation units, the water undergoes another coagulation process to remove residual suspended matter, some impurities, and pathogens. Unit 19, a Cl_2 contact chamber, had been installed but was not used during the data collection period of this case study.

Six kinds of online meters are installed at different points along the treatment line: flow rate, TSS, pH, phosphate (PO_4), temperature, and total solids (TS) meters. Besides them, there are multiple offline manual samplers. In Fig. 1, the online meters are indicated by round symbols while the manual samplers are indicated by the label PT. The full and abbreviated names of the online meters as well as their model numbers and properties are listed in Table S1 of the Supplementary Material. Only data from the online meters were considered in this case study because they all recorded data at the same high resolution (1 ms per sample), which substantially enhanced the reliability and robustness of the analysis. However, the raw high-resolution data from each meter were compressed by averaging over 10-minute periods to obtain time series with temporal resolutions of 10 min. This was done because the composition of the treated water does not usually change much over short periods of time and averaging in this way can alleviate the effect of sampling noise.

The output data for model development and the subsequent analyses were the quantities measured by two of the effluent variable meters (TSS_e and $PO4_e$), while the input data were the data provided by all the other online meters. The original data were embedded in multiple matrices and were very messy, with missing values, bad data cells, and unnecessary information. Therefore, the Python modules Numpy (Oliphant, 2006) and Pandas (McKinney, 2010) were used to prepare an organized 'clean' dataset for analysis. This dataset contained 105,861 samples (data points) with 34 variables, giving a matrix size of 105,861 × 34. The samples were organized in time series with 10-min intervals.

2.2. Time lag calculation

The wastewater treatment processes are dynamic and involve multiple flows of both water and sludge, so there are lags between the times that water in the process streams reaches different meters. However, the original data are in time series. Therefore, to make the ML models interpretable in terms of WWTP processes, the original time-series data must be converted into batch-series data. The conversion process is illustrated in Fig. 2, where *T*, *V*, *B* and *d* represent time points, variables, batches and data values, respectively. In this simplified case, the four data values (say, d_{11}, d_{12}, d_{13}, and d_{14}) associated with a given time point (in this case, T1) in the left-hand time series dataset represent the averaged values recorded by four different meters during the same 10-min period. However, because water takes time to flow through the WWTP and the meters are located in different parts of the plant, the values recorded by each meter at any given time point will represent different volumes or 'batches' of water. For example, if water within the plant flows from *V*4 to *V*1, then a batch of water entering the plant will first reach *V*4 at a certain time point (say, T1), generating a data value (d_{14}). Then, at some later time point (say, T2), it will reach *V*3 and generate another data value (d_{23}). It will flow further through the plant until at some later time point (say, T4) it reaches *V*2 and generates a third data value (d_{42}). Finally, it will reach V4 at a time point later still (say, T5) and generate a fourth data value (d_{54}). We therefore convert the initial time series data into a batch series (as shown on the right of Fig. 2) in which consecutive batches can be thought of as volumes of water entering the WWTP in consecutive 10-minute periods, and the data values for each batch correspond to the meter readings of that batch as it moves through the treatment stages of the WWTP. Note that only batches for which data on all variables exist should be retained; in the example shown in Fig. 2, these are batches B5–B7, which are highlighted with blue borders. The time lag between two meters may be several multiples of 10 min, and the time that it takes different batches to reach different meters will not be constant because the flow rate fluctuates. Since the flow rate is a function of time, the time lag between different meters is determined by integration. Specifically, as shown in Eq. (1), every integral between two adjacent time-points a_i and b_i (which are always 10 min apart) is summed along the dimension of time *t* to determine how many integrals

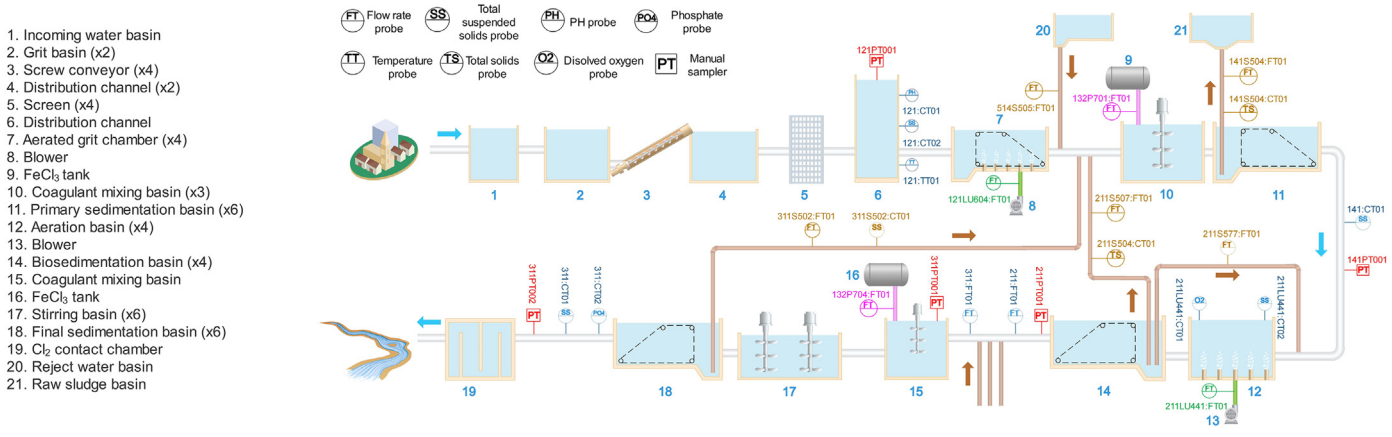


Fig. 1. Schematic depiction of process units and monitoring probes at Umeå WWTP.

	V1	V2	V3	V4
T1	d11	d12	d13	d14
T2	d21	d22	d23	d24
T3	d31	d32	d33	d34
T4	d41	d42	d43	d44
T5	d51	d52	d53	d54
T6	d61	d62	d63	d64
T7	d71	d72	d73	d74

Transformation

Time-series to Batch-series

	V1	V2	V3	V4
B1	d11			
B2	d21	d12		
B3	d31	d22		
B4	d41	d32	d13	
B5	d51	d42	d23	d14
B6	d61	d52	d33	d24
B7	d71	d62	d43	d34
		d72	d53	d44
			d63	d54
			d73	d64
				d74

Fig. 2. Illustration of the transformation of time-series data into batch-series data.

correspond to the distance in volume, VD_{AB}, between two meters A and B. In other words, the value of n in Eq. (1) must be computed to determine the time lag between pairs of meters in multiples of 10 min.

$$VD_{AB} \approx \sum_{i=1}^{n} \int_{a_i}^{b_i} f(t)dt \tag{1}$$

It should be noted that two flow meters were used to calculate the time lags before (FT_{bi}) and after (FT_{ai}) the point where the flow from the sludge system enters the water processing system.

2.3. Random Forest (RF)

As mentioned in the Introduction section, tree-based models offer comparable performance to ANN models while also being readily interpreted. That is to say, the importance of input variables and their effects on the output can be extracted from a trained RF model. RF is used in this work as a representative tree-based modeling strategy because RF models have some major advantages over alternative tree-based models; notably, they require fewer hyperparameters for tuning, their performance is robust to hyperparameter changes, and they are less likely to suffer from overfitting (Breiman, 2001; Breiman, 2002; Chen and Guestrin, 2016; Fawagreh et al., 2014; Ke et al., 2017).

RF can be described as an ensemble method in which the final result is obtained by aggregating (through averaging in the case of regression) results from multiple weak learners known as Classification and Regression Trees (CARTs) (Breiman, 2017). Each weak learner (tree) is trained on the bootstrap set, which is obtained by sampling with replacement from the original training set. For trees, the input variables are used to generate nodes. These variables are selected partially and randomly as a subset in every split, then the variable contributing to the smallest sum of impurity of two child nodes at a certain split point is chosen as the split variable. This is done repeatedly until the trees don't need to split anymore. The regression impurity of a particular node is defined by Eqs. (2), (3) and (4),

$$N_m = |\{X_i \in R_m\}| \tag{2}$$

$$\hat{y}_m = \frac{1}{N_m} \sum_{X_i \in R_m} y_i \tag{3}$$

$$I_m = \frac{1}{N_m} \sum_{X_i \in R_m} (y_i - \hat{y}_m)^2 \tag{4}$$

where R_m is the region of the node (which is indexed by m), N_m is the number of the samples in R_m, y_i is the response corresponding to X_i, $\hat{y}_m$ is the average of y_i in R_m, and I_m is the impurity of R_m.

Breiman proved that the out-of-bag (OOB) estimates based on OOB data from bootstrapping offer almost identical performance to that achieved using a test set of the same size as the training set for error estimation (Breiman, 2001). Therefore, the OOB error was used in this work as the optimization objective to select the optimal three hyperparameters, 'Tree number', 'Maximum variable subset', and 'Minimum leaf size (number of samples at the nodes)', by the *random search* (Bergstra and Bengio, 2012) method. The ranges of the hyperparameters for *random search* were predefined as: *Tree number* [200, 800], *Maximum variable subset* at every split [2, 31], and *Minimum leaf size* [2, 80].

2.4. Deep Neural Network (DNN)

Artificial Neural Network (ANN) models are used to validate the performance of RF in this study because of their superb ability to model complex information and generate accurate predictions (Oliveira et al., 2019; Ozoegwu, 2019; Parisi et al., 2019; Shabanpour et al., 2017). There are three kinds of layers in a typical feedforward ANN: the input, hidden, and output layers, each of which consists of multiple neurons (nodes). Usually, when there is more than one hidden layer, a feedforward ANN can be called a Deep Neural Network (DNN) (Sugiyama, 2019). The following equation shows how one neuron is connected to the neurons in the previous layer, and how the information from the input is fed forward.

$$x_{j+1} = f\left(\sum_i w_{i,j+1} x_{i,j} + b_{j+1}\right) \tag{5}$$

In Eq. (5), x is the neuron, j is the layer index, i is the neuron index in layer j, w is the weight between two layers, b is the bias weight term, and $f(\cdot)$ is the activation function.

It is worth noting that no activation function was applied between the last hidden layer and the output layer because DNN was used for

regression rather than classification in this work. Moreover, there was only one neuron in the output layer.

Network training is performed using the backpropagation method (Rojas, 1996). Based on the gradient descent of loss function, the weights of the neurons can be updated backward from the output layer according to Eq. (6), and the weights of biases can be updated in the same fashion.

$$w^{+}_{j,j+1} = w_{j,j+1} - \eta o_j \delta_{j+1} \tag{6}$$

In Eq. (6), w^{+} is the updated weight over w, δ_{j+1} is the derivative (gradient) of the loss function with respect to the activation function applied in the $j+1$ layer, o_j is the neuron in layer j, and η is the learning rate, which determines by how much the weight is be adjusted.

2.5. Variable Importance Measure (VIM) analysis

The two most common types of Variable Importance Measure (VIM) are the Permutation Importance (PI) and Mean Decrease Impurity (MDI) measures (Breiman, 2001; Breiman, 2002). We used the MDI measure for our RF models. MDI was adopted for the following reasons: MDI is more robust and computationally efficient than PI (Calle and Urrea, 2011; Li et al., 2019); all the variables of the data in this study are with continuous values, which means using MDI will not cause the bias issue mentioned in the literature (Boulesteix et al., 2012; Strobl et al., 2007).

MDI evaluates the importance of variable X_m by assessing the reduction in 'impurity' at the node where the variable is used to split the input data. If t_L and t_R are used to denote the two child nodes from the split of node t (s_t), N_t is used to denote the number of samples at node t, and N_{t_L} and N_{t_R} denote the numbers of samples in the left and right child nodes, respectively, then the decrease in impurity can be calculated as:

$$\Delta I(s_t, t) = I(t) - \frac{N_{t_L}}{N_t} I(t_L) - \frac{N_{t_R}}{N_t} I(t_R) \tag{7}$$

Here, $I(\cdot)$ is the impurity measure defined by Eqs. (2), (3) and (4) in Section 2.3.

At every node there is a split using a certain unique variable, and all the variables are used in the whole tree. Therefore, the MDI of a given variable in an RF model can be defined by averaging all trees and nodes where this variable is used to split. As shown in the following equation, the importance of variable X_m can be calculated by summing the weighted impurity reductions $p(t)\Delta I(s_t, t)$ at all nodes t where X_m is used and then averaging the sum over all N_T trees in the forest:

$$VI(X_m) = \frac{1}{N_T} \sum_{T} \sum_{t \in T: v(s_t) = X_m} p(t) \Delta I(s_t, t) \tag{8}$$

Here, $p(t)$ is the proportion N_T/N of samples reaching node t and $v(s_t)$ is the variable used in split s_t.

2.6. Partial Dependence Plots (PDPs)

After determining which variables are most important, their effects on the output must be examined to improve understanding of processes in the plant. This was done using Partial Dependence Plots (PDPs) (Friedman, 2001).

The partial dependence function $g_{x_s}(\cdot)$ is estimated by averaging the output over all the samples and can be expressed using Eq. (9).

$$g_{x_s}(x_s) = \frac{1}{n} \sum_{i=1}^{n} g\left(x_s, X_c^{(i)}\right) \tag{9}$$

Here, x_s is the variable for which the partial dependence function is plotted and X_c are the other variables used as inputs in the ML model: $g(\cdot)$. x_s and X_c together constitute the whole variable space X. i is the index of training data samples and n is the number of training data samples. It can be seen that the partial dependence function marginalizes the predicted output over all the variables in set C (X_c) to demonstrate the correlation between the variable x_s and the predicted outcome.

2.7. Study framework

The original data were cleaned and transformed into batch-series data as described in Section 2.1. This reduced the number of samples from 105,861 to 105,763. For both TSS_e and $PO4_e$, the data were randomly divided into a training set containing 90% of the original data and validation and test sets each containing 5% of the original data. In the FR and DNN modeling, these datasets were subjected to the processes shown in Fig. 3.

In the RF modeling, the training dataset were divided into bootstrap and out-of-bag (OOB) datasets to train the model within the predefined hyperparameter space and to validate the trained model, respectively. If the optimal OOB score (the R^2 value calculated from the observed and predicted outputs in the OOB dataset) was not good enough,[1] the hyperparameter space was adjusted. If the optimal OOB score was good enough, the corresponding model was used to predict the validation dataset to check for over-fitting and further refine the model's hyperparameters if necessary. The model was then applied to a test dataset to assess its generalization performance (ability to predict unknown datasets). The RF model's performance was evaluated alone and also compared to the DNN model's performance to determine whether the RF model was suitable for further interpretation analyses. If it was suitable, VIM was subsequently applied to the model to evaluate the importance of each variable. The three most important variables were investigated, and PDPs were generated to display their influence patterns on the output.

A non-interpretable DNN model was also generated to check whether the RF model captured sufficient variance to support VIM- and PDP-based interpretation. DNN modeling involved no division of the training dataset but required that the dataset be standardized before being used for model training. The validation and test sets were standardized as well. The validation dataset was used to prevent overfitting of the trained model, and the performance on the validation set was calculated (in terms of R^2 values) after every training epoch. If there were signs of overfitting (i.e., the prediction performance is significantly better on the training set than on the validation set, and the gap remains or increases as the training proceeds), the training was aborted. If the best performance on the validation dataset was good enough and did not vary greatly from the performance on the training dataset, the model was accepted and its performance on the test dataset was evaluated. Otherwise, the hyperparameters (e.g., numbers of hidden layers, number of neurons in the hidden layers, activation function, and/or size of mini-batches) were adjusted before the model went through the training and validation procedures again until the satisfactory performance was achieved on the validation set. This manual tuning process for acquiring optimal hyperparameters is termed *Grad Student Descent* (Gencoglu et al., 2019). The final performance on the test set was used to evaluate RF's final performance.

3. Results and discussion

3.1. Model structure & performance

This section describes the final optimal configurations of the RF and DNN models and their performance. TSS_e_RF and $PO4_e$_RF are used to denote the RF models for TSS_e and $PO4_e$, respectively. TSS_e_DNN and

[1] Here, an R^2 value larger than 0.85 was considered 'good enough'. This is a subjective criterion that only applies to this study.

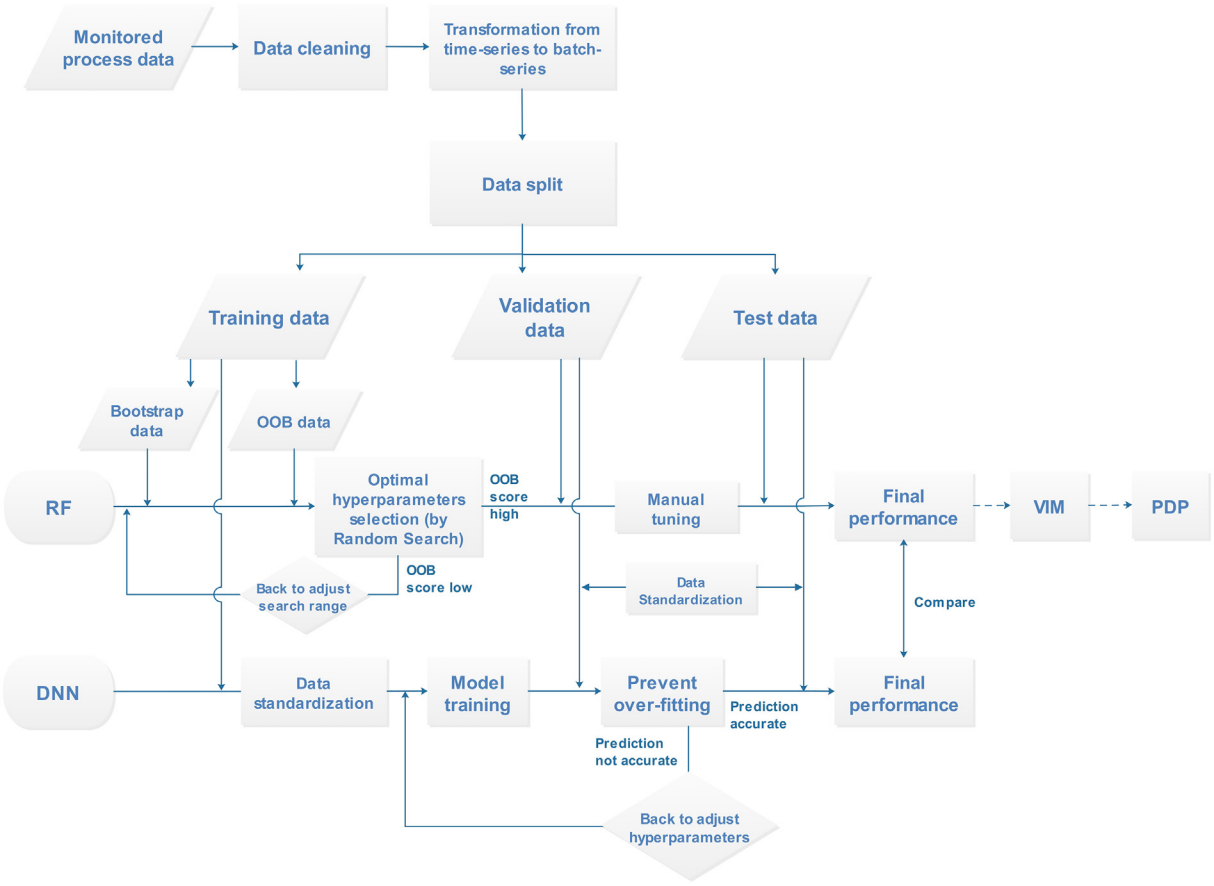


Fig. 3. Flow chart of the study. After the data are cleaned and time lags among variables have been handled, the data are divided into training, validation, and test sets. The training set are used to train the RF and DNN models, and the validation set are used to check if the trained models are overfitted. If overfitting exists, the hyperparameters are adjusted, and the models are trained again. This procedure can be repeated until there is no evidence of overfitting. The test set are then used to evaluate the models' generalization performance. The RF model's performance is evaluated alone and also compared to the DNN model's performance to determine whether the RF model is suitable for further interpretation analyses. If it is, VIM analysis is performed to identify the most influential input variables for each output. PDP analysis is then performed to investigate the effects of those variables on the outputs.

$PO4_e$_DNN are used to denote the DNN models for TSS_e and $PO4_e$, respectively.

As shown in Table 1, the optimal hyperparameter sets (specified in the order *Tree number*, *Maximum variable subset*, *Minimum leaf size*) were (253, 13, 3) for TSS_e_RF and (583, 17, 5) for $PO4_e$_RF.

The performances of these two RF models on the three datasets are shown in Table 3. For TSS_e_RF, the R^2 values of prediction on the training, validation, and test datasets are 0.934, 0.862, and 0.920, respectively, while those for $PO4_e$_RF are 0.905, 0.870, and 0.886, respectively. Both models thus capture over 90% of the variation in the training dataset, and can predict a very high percentage of the variation in the unknown dataset (92% for TSS_e_RF and 88.6% for $PO4_e$_RF). Both models also achieve excellent generalization performance — they predict the unknown dataset (test dataset) almost as well as the datasets used to train and tune them (the training and validation dataset).

Optimal hyperparameters for both DNN models are shown in Table 2. For TSS_e_DNN, the optimal number of hidden layers is three, with optimal neuron counts of 128, 256, and 128 for hidden layers 1–3, respectively. The weight initializer for each layer is the *Glorot uniform initializer* (Glorot and Bengio, 2010). For all the input and hidden layers except the last hidden layer, the activation function is *ReLU* (Agarap, 2018). The optimizer for the gradient descent operation is *Adadelta* (Zeiler, 2012). The mini-batch size is 32. All the layers are fully connected. $PO4_e$_DNN has the same optimal hyperparameters except for the numbers of neurons in the hidden layers, which are 256, 256, and 128.

The DNN models' performances on the three datasets are shown in Table 3. For TSS_e_DNN the R^2 values of prediction on the training, validation, and test datasets are 0.935, 0.892, and 0.920, respectively. For $PO4_e$_DNN the R^2 values are 0.904, 0.908, and 0.872, respectively. Both TSS_e_DNN and $PO4_e$_DNN models capture over 90% of the variation in the training dataset and can predict a very high percentage (92% for

Table 1
Optimal hyperparameters for RF models.

	Tree number	Maximum variable subset	Minimum leaf size
TSS_e_RF	253	13	3
$PO4_e$_RF	583	17	5

Table 2
Optimal hyperparameters for DNN models.

	Hidden layer number	Neuron number in 1st hidden layer	Neuron number in 2nd hidden layer	Neuron number in 3rd hidden layer	Activation function	Optimizer	Mini-batch size	Weight initializer
TSS_e_DNN	3	128	256	128	ReLU	Adadelta	32	Glorot uniform initializer
$PO4_e$_DNN	3	256	128	128	ReLU	Adadelta	32	Glorot uniform initializer

Table 3
Model performances on the training, validation, and test sets (R^2 values).

	Training set	Validation set	Test set
TSS_e_RF	0.934	0.862	0.920
TSS_e_DNN	0.935	0.892	0.920
$PO4_4$_RF	0.905	0.870	0.886
$PO4_4$_DNN	0.904	0.908	0.872

TSS_e_DNN and 87.2% for $PO4_e$_DNN) of the variation in the unknown dataset. The results on the test dataset also show that both TSS_e_DNN and $PO4_e$_DNN have excellent generalization performance.

In summary, both RF and DNN models predicted the training, validation, and test datasets extremely well and achieved excellent generalization performance for TSS_e and $PO4_e$. Moreover, TSS_e_RF and TSS_e_DNN had similar performances, $PO4_e$_RF and $PO4_e$_DNN had similar performances as well; $PO4_e$_RF even showed slightly better generalization performance than $PO4_e$_DNN. The trained RF models thus captured the relationships between the operational variables and effluent parameters well and could be reliably used for further (VIM and PDP) analysis.

3.2. VIM from RF

After acquiring the trained models, VIM analysis was used to evaluate the importance of each variable in the models. Fig. 4 shows the results the of VIM analysis for both TSS_e and $PO4_e$. The variables were categorized into three levels based on their VIM values — 'significantly important (cyan, [0.1, +∞))', 'important (orange, [0.05, 0.1))', and 'least important (green, [0, 0.05))'.

Fig. 4 shows that:

- According to the VIM values of the variables and the definition of importance given above, the important variables for TSS_e are TT_{in} (0.188), FT_{sr} (0.091), TSS_{lr} (0.091), TSS_{a3} (0.055), and FT_{ab3} (0.052). For $PO4_e$, they are TT_{in} (0.188), TSS_{a4} (0.1), TSS_{a3} (0.097), TSS_{lr} (0.089), TSS_{a2} (0.087), FT_{ab4} (0.06), DO_{a4} (0.057), and FT_{gc} (0.055).
- The VIM values of TT_{in}, TSS_{lr}, and TSS_{a3} are greater than 0.05 for both TSS_e and $PO4_e$, indicating that these three variables are important for both TSS_e and $PO4_e$. Additionally, TT_{in} is the most important variable for both TSS_e and $PO4_e$, with a VIM value of 0.188 for both. The importance of TT_{in} is consistent with the widely accepted point that temperature strongly affects both coagulation and the production of microorganisms (Nolasco, 2008; Spellman, 2013). TSS_{lr} is important because of the chain that connects TSS_{lr} and TSS_e. The source of TSS_{lr} is the sedimented solids in the final sedimentation basin, and the sedimented solids originate from the suspended solids and phosphate from the water in the final sedimentation basin, which becomes the effluent upon exiting the basin. TSS_{a3} represents the total suspended solids in the aeration basins; its importance is thus consistent with the principle that the amount of activated sludge in the aeration basins significantly affects the metabolism and reproduction of organisms that degrade organic substances or ingest phosphate (Federation et al., 2006).
- For both TSS_e and $PO4_e$, the VIM values of TSS_{a1}, TSS_{a2}, TSS_{a3}, and TSS_{a4} generally increase in that order, with the exception that TSS_{a4} has a smaller VIM value than TSS_{a3} for TSS_e. This pattern might result from the layout of the aeration and biosedimentation basins: there are four parallel 'aeration basin + biosedimentation basin' lines, and the outflows from the four lines merge at one point to enter the next unit, which is the coagulant mixing basin (unit 15). The merging point aligns with the first 'aeration basin + biosedimentation basin' line (denoted by 'a1'), and the distances between each line and the merging point increase in the order a1 < a2 < a3 < a4, meaning that the delays in reaching the merging point increase in the same order. Therefore, the order of the VIM values can be interpreted to mean that the portion merging later will have a larger impact on the effluent quality, which reveals the intrinsic 'increasing marginal utility (Kauder, 2015)' among the four parallel lines.
- TSS_{a2}, TSS_{a3}, and TSS_{a4} have VIM values of 0.87, 0.97, and 0.1, respectively, for $PO4_e$, meaning they are important for $PO4_e$. This indicates that the effluent's PO_4 concentration depends strongly on the TSS concentration of the aeration basins, which can be explained by the established knowledge: (i) the TSS concentration in the aeration basins directly reflects the concentration of activated sludge and (ii) the polyphosphate-accumulating organisms (PAOs) in activated sludge play a key role in phosphate consumption (Wiesmann et al., 2007).

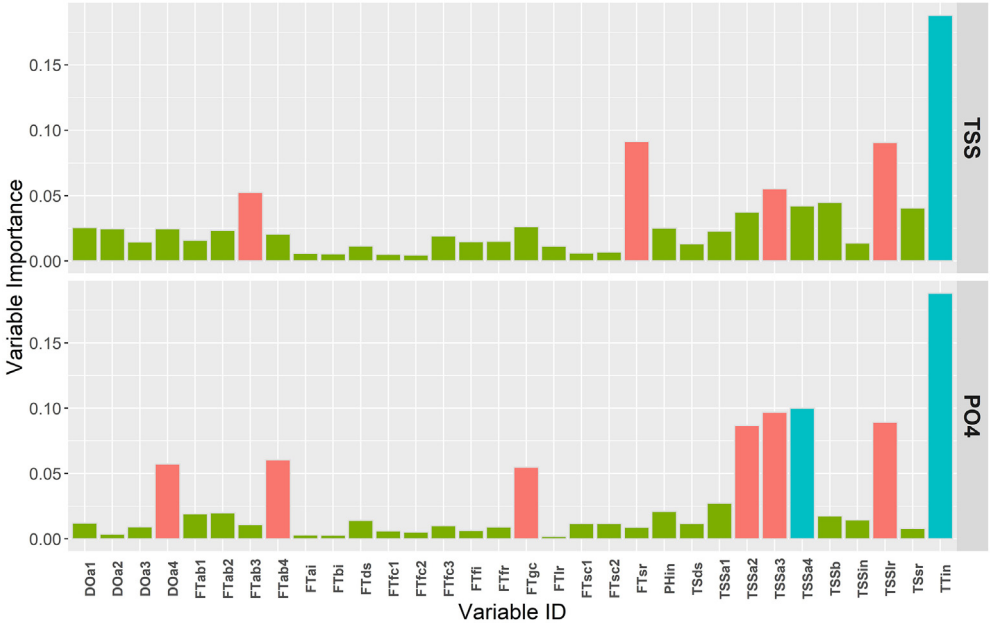


Fig. 4. Variable importance histograms for TSS_e and $PO4_e$. The subfigure titled 'TSS' is the histogram for TSS_e and the one titled 'PO4' is the histogram for $PO4_e$. The variables are categorized into three levels based on their VIM values — 'significantly important (cyan, [0.1, +∞))', 'important (orange, [0.05, 0.1))', and 'least important (green, [0, 0.05))'. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

3.3. PDP from RF

Centered PDPs (generated as described in Section 2.6) for the three most important variables in the models for TSS_e and PO_{4e} are presented in Fig. 5. In this figure, the variable values are plotted on the x-axes and the centered partial dependence values are plotted on the y-axes. The purpose of centering is to highlight the change in the partial dependence value as a variable changes from its minimum value to its maximum value. The light blue envelopes in the figure show the range from the partial dependence value minus the standard deviation to the partial dependence value plus standard deviation.

The following PDP interpretations for TSS_e and $PO4_e$ are based on discussions with engineers from Umeå WWTP, who understood the general mechanisms associated with the influence patterns but could not predict with certainty the interactive responses of the target variables to specific changes in conditions. The PDP illustrations from this case study provide potentially valuable insights into these responses, and thus into optimal operational adjustments.

3.3.1. Interpretation of PDPs for TSS_e

As shown in subfigure (a) of Fig. 5, TSS_e decreases as TT_{in} increases. This suggests that raising the temperature, at least in the 6–16 °C range, can enhance the removal of total suspended solids. It may be that this is because raising the temperature within an appropriate range can improve the physiological characteristics of microorganisms (Adams et al., 2010; Garcia-Rios et al., 2016), increase microbial growth

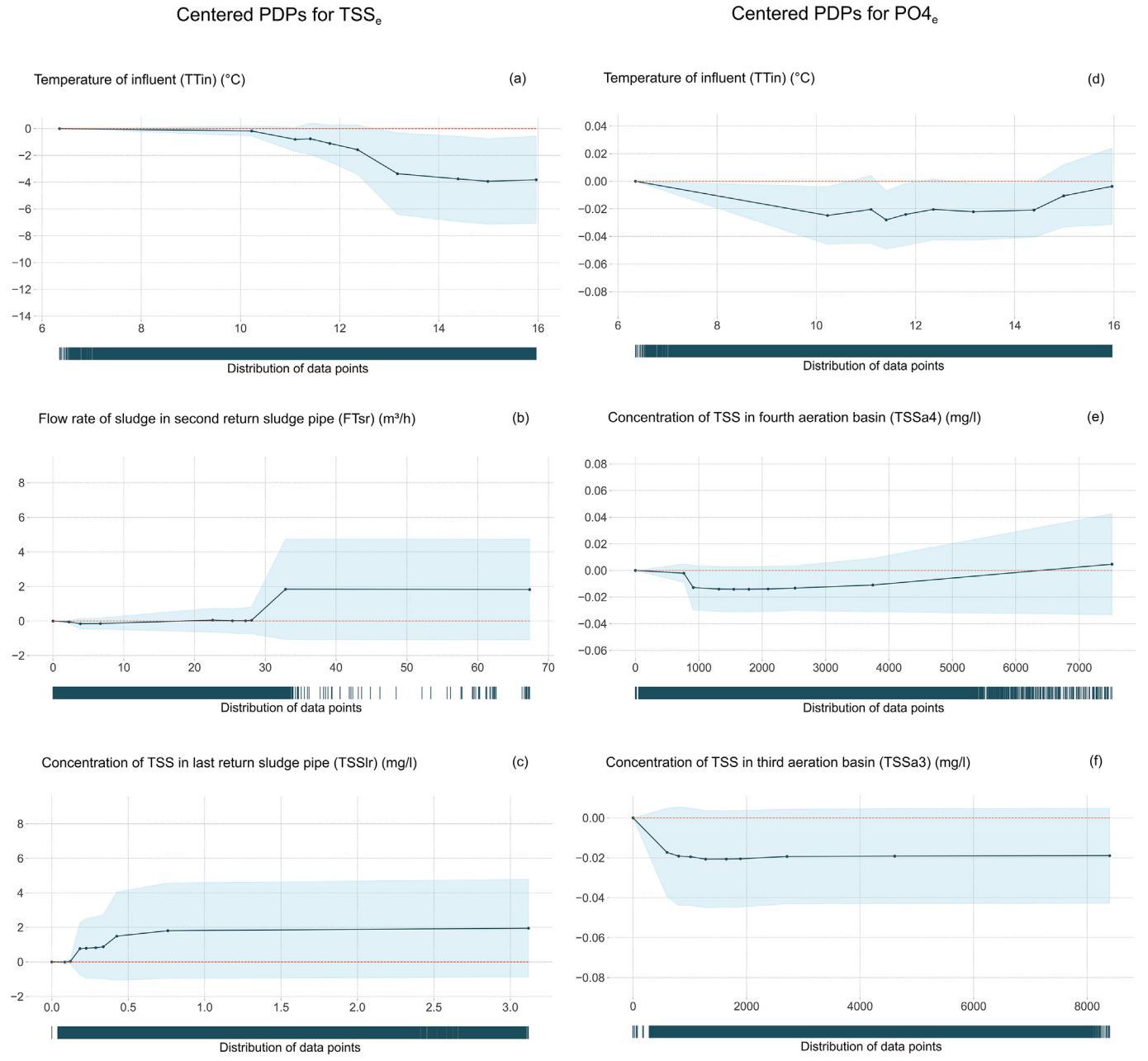


Fig. 5. Centered PDPs of the three most important variables for TSS_e (left panels — a, b, c) and $PO4_e$ (right panels — d, e, f): X axes indicate variable values, and Y axes indicate centered partial dependence values resulting from the variable values. The dark blue solid line is the line of centered partial dependence values. The centered partial dependence values are shown here instead of the original ones to highlight the variation resulting from the change of the variables' values. For clarity, the center line is indicated by an orange dashed line. The light blue envelopes in the figures span the range from the partial dependence value minus the standard deviation to the partial dependence value plus the standard deviation. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

rates (Rajeshwari et al., 2000; Spellman, 2013), boost microbial activity (Rajeshwari et al., 2000; Young et al., 2017), improve the community structure and species diversity of microorganisms (Chen et al., 2017), enhance sludge settleability (Wilén et al., 2008; Yang and Li, 2009), and promote coagulation and flocculation (Dayarathne et al., 2020).

Subfigure (b) of Fig. 5 indicates that changes in FT_{sr} have minimal effects on TSS_e when FT_{sr} is in the range of 0–28 m^3/h. However, when FT_{sr} is between 28 and 33 m^3/h, TSS_e increases as the flow rate increases. At FT_{sr} above 33 m^3/h TSS_e remains constant. However, there are very few data points in this flow rate range, so the pattern at these high flow rates is not very informative or conclusive. Nevertheless, this PDP clearly indicates that once the flow rate in the second return sludge pipe reaches a certain threshold, further increases reduce TSS removal. This can be explained by the unique structure of the process units. The second return sludge pipe (No. 14 to No. 10) and the first return sludge pipe (No. 14 to No. 12) both carry sedimented sludge from the secondary sedimentation basin. If the amount of sludge returned through the second return sludge pipe increases, the amount returned through the first return sludge pipe decreases. Once FT_{sr} reaches a certain value (generally 28 m^3/h), the amount of activated sludge returned to the aeration basin will be insufficient for biodegradation, increasing the concentration of untreated solids in downstream process units and the effluent. Additionally, when FT_{sr} exceeds this value, the first coagulant mixing basin and the primary sedimentation basin will be overloaded with suspended solids. The fraction of suspended solids not sedimented in the primary sedimentation basin will therefore flow to the aeration basin. However, this portion of sludge will have very limited biodegradation capacity because many organisms in the returned sludge will not survive passage through the anoxic units before the aeration basin (Hussain and Bhattacharya, 2019), causing a greater solids load to be introduced. This effectively doubles the negative effect of allowing FT_{sr} to rise above 28 m^3/h.

Subfigure (c) of Fig. 5 shows that as TSS_{lr} increases between 0 and 0.75 mg/L, TSS_e also increases. However, the increase in TSS_e levels off at TSS_{lr} values above 0.75 mg/L. TSS_{lr} directly reflects the concentration of sedimented solids in the final sedimentation basin, which originate from the TSS in the water in the final sedimentation basin that will become the effluent. TSS_e will therefore inevitably be positively correlated with TSS_{lr}, which explains the pattern observed when TSS_{lr} is below 0.75 mg/L. The change in pattern when TSS_{lr} exceeds 0.75 mg/L may occur because WWTP operators who see TSS_e increasing beyond a specific value may act immediately to control it (e.g. by adding extra coagulant into unit 15), thereby increasing the amount of TSS that settle in the final sedimentation basin without increasing the TSS concentration of the effluent.

3.3.2. Interpretation of PDPs for $PO4_e$

As shown in subfigure (d) of Fig. 5, as TT_{in} increases in the range 6–10.2 °C, $PO4_e$ decreases; with further increases in TT_{in}, $PO4_e$ fluctuates somewhat but is generally stable until an uptrend after 14.4 °C. Unlike the pattern observed in subfigure (a) for TSSe, $PO4_e$ does not always fall as the temperature rises. This may be because of competition for substrates between glycogen-accumulating organisms (GAOs) and polyphosphate-accumulating organisms (PAOs). PAOs have significant positive effects on PO_4 removal, but GAOs do not (Seviour et al., 2003). Additionally, PAOs are psychrophilic but low temperatures hinder the metabolism of GAOs. Therefore, PAOs dominate at lower temperatures but GAOs gain the upper hand at higher temperatures. Accordingly, cultures enriched in PAOs were observed at 10 °C, whereas cultures dominated by GAOs were observed at 15, 20, 30, and 35 °C (Lopez-Vazquez et al., 2009). A separate study confirmed that temperatures of 10 °C or less encourage PAOs' growth (Erdal et al., 2003). Thus, the beneficial effects of high temperature discussed in Section 3.3.1 are outweighed by the adverse effects of high temperatures on PAOs, leading to an overall reduction in PO_4 removal as temperatures increase.

As shown in subfigure (e) of Fig. 5, $PO4_e$ decreases as TSS_{a4} increases between 0 and 1600 mg/L, but then increases more slowly as TSS_{a4} increases beyond 1600 mg/L. Additionally, subfigure (f) of Fig. 5 shows that $PO4_e$ decreases as TSS_{a3} increases in the range 0–1200 mg/L and then levels off at TSS_{a3} values above 1200 mg/L. Both TSS_{a4} and TSS_{a3} are TSS concentrations in aeration basins. The decrease of $PO4_e$ as these variables increase can be explained in two ways. First, higher TSS concentrations in the aeration basins correspond to higher concentrations of activated sludge and thus larger populations of PAOs, which are significant phosphate consumers (Wiesmann et al., 2007). Second, higher concentrations of suspended solids make coagulation processes more effective because they increase the frequency of interactions between colloids and precipitates (Bratby, 2016; Shewa and Dagnew, 2020). However, when colloids are overabundant in the water, much of the added coagulant ($FeCl_3$) will interact with the colloids without reacting with the dissolved phosphate and inducing chemical precipitation (Bratby, 2016). This reduces the amount of phosphate removed by settling, explaining the pattern observed for TSS_{a4} values above 1600 mg/L in subfigure (e) of Fig. 5.

However, higher concentrations of TSS_{a3} and TSS_{a4} ($TSS_{a3} > 1200$ mg/L, $TSS_{a4} > 1600$ mg/L) have different effects on $PO4_e$ — the PDP of TSS_{a4} shows an uptrend, but not the PDP of TSS_{a3}. In fact, the influence of TSS concentrations in the four aeration basins on $PO4_e$ increases in the order $TSS_{a1} < TSS_{a2} < TSS_{a3} < TSS_{a4}$, according to the VIM plot shown in Fig. 4. This is explained in Section 3.2 — the contributions of tributaries further from the merging points are inherently more heavily weighted.

3.4. Significance of the study

This study provides information on the operational variables that influence TSS_e and $PO4_e$, as well as their effects, which can help operators understand the relationships between variables that interactively affect the complex processes occurring in Umeå WWTP and similarly configured WWTPs. In conjunction with traditional (chemical, biochemical, and hydromechanical) analyses, these results could be used to more effectively and reliably determine whether current operational parameters are appropriate or whether pre-emptive action is required to prevent deterioration in effluent quality. For example, the results presented here indicate that if the flow rate of sludge in the second return sludge pipe (FT_{sr}) exceeds 28 m^3/h, the operators may need to take actions such as distributing more sludge to the first return sludge pipe or adding more coagulant to the first coagulant mixing basin. Without the information provided by this study, unnecessary coagulant might be added when the FT_{sr} is below 28 m^3/h, thereby wasting chemicals. These results could thus enable the development of more advanced control strategies that could appreciably reduce running costs.

More broadly, this study explored the feasibility of using ML to comprehensively understand processes in WWTPs rather than to simply develop soft sensors or generate predictions. A framework for this purpose was developed and described at length, including details of all steps from data pretreatment to model explanation. To ensure that the explanations generated using this framework accurately reflect process properties, the time series data initially obtained from the various meters within the WWTP were transformed into batch series data, and DNN models were used to verify that the generated RF models capture sufficient variance to support robust explanations. To enable deep interpretation of the generated models, the framework incorporated both VIM and PDP analyses. This framework could in principle be applied to any other parameters of interest (and indeed to similar studies of processes in other industries) given the availability of sufficient high-resolution data, which is essential for robust and reliable analysis.

4. Conclusions

A ML framework based on RF, DNN, VIM, and PDP has been developed to model WWTP processes and investigate how plant operational

variables influence effluent quality. The proposed ML framework appears to have the potential to improve effluent quality control strategies in WWTPs, as demonstrated by a case study on Umeå WWTP involving a large dataset (105,763 samples) representing the full scale of the plant's operations. In the case study, RF models were constructed and validated using DNN models, after which VIM and PDP analyses were performed. VIM identified the variables that most strongly influence the effluent parameters (here, TSS_e and $PO4_e$), while PDP elucidated their specific effects on TSS_e and $PO4_e$. The major findings are as follows:

1) For TSS_e, the influential variables are TT_{in}, FT_{sr}, TSS_{lr}, TSS_{a3}, and FT_{ab3}. For $PO4_e$, they are TT_{in}, TSS_{a4}, TSS_{a3}, TSS_{lr}, TSS_{a2}, FT_{ab4}, DO_{a4}, and FT_{gc}.
2) Influent temperature is the most influential variable for both TSS_e and $PO4_e$, but it affects them in different ways.
3) $PO4_e$ is highly dependent on the TSS in aeration basins. Increases in the TSS concentration in aeration basins generally promote PO_4 removal but excessive TSS can have negative effects.
4) In general, TSS in aeration basins located further from the merging outlet have greater impacts on TSS_e and $PO4_e$ than TSS in more nearby basins. Thus, more attention should be paid to the TSS concentrations of the third and fourth aeration basins than the first and second basins.
5) Returning excessive amounts of sludge through the second return sludge pipe should be avoided because of its adverse impact on TSS_e.

These findings may facilitate development of more sophisticated control strategies for WWTPs that could significantly increase control precision and reduce running costs. The framework could also be applied to other effluent parameters if sufficiently abundant and high-resolution data are available. Future work will focus on assessing more algorithms and ML model interpretation systems to further improve the framework and the reliability of the results it provides.

CRediT authorship contribution statement

Dong Wang: Conceptualization, Methodology, Software, Formal analysis, Investigation, Writing – original draft, Visualization. **Sven Thunéll:** Validation, Data curation. **Ulrika Lindberg:** Validation, Data curation. **Lili Jiang:** Writing – review & editing. **Johan Trygg:** Writing – review & editing. **Mats Tysklind:** Resources, Writing – review & editing, Supervision, Project administration, Funding acquisition. **Nabil Souihi:** Resources, Writing – review & editing, Supervision.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgement

The presented work was performed as part of the Green Technology and Environmental Economics (GreenTEE) initiative at Umeå University. This involves collaboration between companies in the municipality and academic researchers, focusing on development of technologies and promotion of policy-making studies directed towards improving cities' sustainability. The authors acknowledge support from Green TEE platform for funding this project, and from Daniel Fredlander, at the VAKIN (Umeå Wastewater Treatment Plant), Sweden, for initial assistance in providing the data required for this study.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.scitotenv.2021.147138.

References

Adams, H.E., Crump, B.C., Kling, G.W., 2010. Temperature controls on aquatic bacterial production and community dynamics in arctic lakes and streams. Environ. Microbiol. 12, 1319–1333.
Agarap, A.F., 2018. Deep learning using rectified linear units (relu). arXiv preprint arXiv: 1803.08375.
Ahmad, M.W., Mourshed, M., Rezgui, Y., 2017. Trees vs neurons: comparison between random forest and ANN for high-resolution prediction of building energy consumption. Energy Build. 147, 77–89.
Bergstra, J., Bengio, Y., 2012. Random search for hyper-parameter optimization. J. Mach. Learn. Res. 13, 281–305.
Borzooei, S., Campo, G., Cerutti, A., Meucci, L., Panepinto, D., Ravina, M., et al., 2019. Optimization of the wastewater treatment plant: from energy saving to environmental impact mitigation. Sci. Total Environ. 691, 1182–1189.
Boulesteix, A.-L., Bender, A., Lorenzo Bermejo, J., Strobl, C., 2012. Random forest Gini importance favours SNPs with large minor allele frequency: impact, sources and recommendations. Brief. Bioinform. 13, 292–304.
Bratby, J., 2016. Coagulation and Flocculation in Water and Wastewater Treatment. IWA Publishing.
Breiman, L., 2001. Random forests. Mach. Learn. 45, 5–32.
Breiman, L., 2002. Manual on Setting up, Using, and Understanding Random Forests v3.1. vol. 1. Statistics Department University of California, Berkeley, CA, USA, p. 58.
Breiman, L., 2017. Classification and Regression Trees. Routledge.
Buaisha, M., Balku, S., Özalp-Yaman, S., 2020. Heavy metal removal investigation in conventional activated sludge systems. Civil Eng. J. 6, 470–477.
Calle, M.L., Urrea, V., 2011. Letter to the editor: stability of random forest importance measures. Brief. Bioinform. 12, 86–89.
Cao, W., Yang, Q., 2020. Online sequential extreme learning machine based adaptive control for wastewater treatment plant. Neurocomputing 408, 169–175.
Chan, C.W., Huang, G.H., 2003. Artificial intelligence for management and control of pollution minimization and mitigation processes. Eng. Appl. Artif. Intell. 16, 75–90.
Chen, T., Guestrin, C., 2016. Xgboost: a scalable tree boosting system. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. ACM, pp. 785–794.
Chen, Y., Lan, S., Wang, L., Dong, S., Zhou, H., Tan, Z., et al., 2017. A review: driving factors and regulation strategies of microbial community structure and dynamics in wastewater treatment systems. Chemosphere 174, 173–182.
Chowdhury, R., Rahman, M.A., Rahman, M.S., Mahdy, M., 2020. An approach to predict and forecast the price of constituents and index of cryptocurrency using machine learning. Phys. A: Stat. Mech. Its Appl. 551, 124569.
Corominas, L., Garrido-Baserba, M., Villez, K., Olsson, G., Cortés, U., Poch, M., 2018. Transforming data into knowledge for improved wastewater treatment operation: a critical review of techniques. Environ. Model Softw. 106, 89–103.
Dayarathne, H., Angove, M.J., Aryal, R., Abuel-Naga, H., Mainali, B., 2020. Removal of natural organic matter from source water: review on coagulants, dual coagulation, alternative coagulants, and mechanisms. J. Water Proc. Eng. 40 , 1–13 101820.
Dürrenmatt, D.J., Gujer, W., 2012. Data-driven modeling approaches to support wastewater treatment plant operation. Environ. Model Softw. 30, 47–56.
Erdal, U., Erdal, Z., Randall, C., 2003. The competition between PAOs (phosphorus accumulating organisms) and GAOs (glycogen accumulating organisms) in EBPR (enhanced biological phosphorus removal) systems at different temperatures and the effects on system performance. Water Sci. Technol. 47, 1–8.
Erdirencelebi, D., Yalpir, S., 2011. Adaptive network fuzzy inference system modeling for the input selection and prediction of anaerobic digestion effluent quality. Appl. Math. Model. 35, 3821–3832.
Faruk, D.Ö., 2010. A hybrid neural network and ARIMA model for water quality time series prediction. Eng. Appl. Artif. Intell. 23, 586–594.
Fawagreh, K., Gaber, M.M., Elyan, E., 2014. Random forests: from early developments to recent advancements. Syst. Sci. Cont. Eng. 2, 602–609.
Federation WE, Force WEFBNROiWTPT, Force WEFBNROiWTT, Institute WR, 2006. Biological Nutrient Removal (BNR) Operation in Wastewater Treatment Plants: WEF Manual of Practice No. 30. McGraw Hill Professional.
Fenu, A., Guglielmi, G., Jimenez, J., Sperandio, M., Saroj, D., Lesjean, B., et al., 2010. Activated sludge model (ASM) based modelling of membrane bioreactor (MBR) processes: a critical review with special regard to MBR specificities. Water Res. 44, 4272–4294.
Friedman, J.H., 2001. Greedy function approximation: a gradient boosting machine. Ann. Stat. 1189–1232.
Garcia-Rios, E., Ramos-Alonso, L., Guillamon, J., 2016. Correlation between low temperature adaptation and oxidative stress in Saccharomyces cerevisiae. Front. Microbiol. 7, 1199.
Gencoglu, O., van Gils, M., Guldogan, E., Morikawa, C., Süzen, M., Gruber, M., et al., 2019. HARK side of deep learning—from grad student descent to automated machine learning. arXiv 8 preprint arXiv:1904.07633.
Glorot, X., Bengio, Y., 2010. Understanding the difficulty of training deep feedforward neural networks. Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics, pp. 249–256.
Guidotti, R., Monreale, A., Ruggieri, S., Turini, F., Giannotti, F., Pedreschi, D., 2018. A survey of methods for explaining black box models. ACM Comput. Surv. (CSUR) 51, 1–42.
Gujer, W., Henze, M., Mino, T., Matsuo, T., Wentzel, M., Marais, G., 1995. The activated sludge model no. 2: biological phosphorus removal. Water Sci. Technol. 31, 1–11.
Gujer, W., Henze, M., Mino, T., van Loosdrecht, M., 1999. Activated sludge model no. 3. Water Sci. Technol. 39, 183–193.

Guo, H., Jeong, K., Lim, J., Jo, J., Kim, Y.M., J-p, Park, et al., 2015. Prediction of effluent concentration in a wastewater treatment plant using machine learning models. J. Environ. Sci. 32, 90–101.
Haimi, H., Mulas, M., Corona, F., Vahala, R., 2013. Data-derived soft-sensors for biological wastewater treatment plants: an overview. Environ. Model Softw. 47, 88–107.
Hauduc, H., Rieger, L., Ohtsuki, T., Shaw, A., Takács, I., Winkler, S., et al., 2011. Activated sludge modelling: development and potential use of a practical applications database. Water Sci. Technol. 63, 2164–2182.
Henze, M., Grady Jr., C.P.L., Gujer, W., GVR, Marais, Matsuo, T., 1987. Activated sludge model no. 1. IAWPRC Scientific and Technical Reports 1. IAWPRC, London.
Henze, M., Gujer, W., Mino, T., Matsuo, T., Wentzel, M.C., GvR, Marais, et al., 1999. Activated sludge model no. 2d, ASM2d. Water Sci. Technol. 39, 165–182.
Hussain, A., Bhattacharya, A., 2019. Advanced Design of Wastewater Treatment Plants: Emerging Research and Opportunities: Emerging Research and Opportunities. IGI Global.
Kauder, E., 2015. History of Marginal Utility Theory. Princeton University Press.
Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., et al., 2017. Lightgbm: a highly efficient gradient boosting decision tree. Adv. Neural Inf. Proces. Syst. 3146–3154.
Kumari, P., Toshniwal, D., 2021. Extreme gradient boosting and deep neural network based ensemble learning approach to forecast hourly solar irradiance. J. Clean. Prod. 279, 123285.
Lee, D.S., Jeon, C.O., Park, J.M., Chang, K.S., 2002. Hybrid neural network modeling of a full-scale industrial wastewater treatment process. Biotechnol. Bioeng. 78, 670–682.
Li, X., Wang, Y., Basu, S., Kumbier, K., Yu, B., 2019. A debiased MDI feature importance measure for random forests. Advances in Neural Information Processing Systems.
Liu, H., Zhang, Y., Zhang, H., 2020. Prediction of effluent quality in papermaking wastewater treatment processes using dynamic kernel-based extreme learning machine. Process Biochem. 97, 72–79.
Liu, M., Wang, M., Wang, J., Li, D., 2013. Comparison of random forest, support vector machine and back propagation neural network for electronic tongue data classification: application to the recognition of orange beverage and Chinese vinegar. Sensors Actuators B Chem. 177, 970–980.
Lopez-Vazquez, C.M., Hooijmans, C.M., Brdjanovic, D., Gijzen, H.J., van Loosdrecht, M.C., 2009. Temperature effects on glycogen accumulating organisms. Water Res. 43, 2852–2864.
McKinney, W, 2010. Data structures for statistical computing in python. Proceedings of the 9th Python in Science Conference 445. Python in Science Conference, Austin, TX, USA, pp. 51–56.
Meng, X., Zhang, Y., Qiao, J., 2021. An adaptive task-oriented RBF network for key water quality parameters prediction in wastewater treatment process. Neural Comput. & Applic. 1–14.
Metcalf, I., 2013. Wastewater Engineering: Treatment and Resource Recovery. McGraw-Hill Higher Education.
Müller, A.C., Guido, S., 2016. Introduction to Machine Learning with Python: A Guide for Data Scientists. O'Reilly Media, Inc.
Nadiri, A.A., Shokri, S., Tsai, F.T.-C., Moghaddam, A.A., 2018. Prediction of effluent quality parameters of a wastewater treatment plant using a supervised committee fuzzy logic model. J. Clean. Prod. 180, 539–549.
Newhart, K.B., Holloway, R.W., Hering, A.S., Cath, T.Y., 2019. Data-driven performance analyses of wastewater treatment plants: a review. Water Res. 157, 498–513.
Nolasco, D., 2008. Manual of Practice MOP-11, Operation of Municipal Wastewater Treatment Plants.
Nopens, I., Batstone, D.J., Copp, J.B., Jeppsson, U., Volcke, E., Alex, J., et al., 2009. An ASM/ADM model interface for dynamic plant-wide simulation. Water Res. 43, 1913–1923.
Oliphant E, T, 2006. A guide to NumPy 1. Trelgol Publishing, USA.
Oliveira, V., Sousa, V., Dias-Ferreira, C., 2019. Artificial neural network modelling of the amount of separately-collected household packaging waste. J. Clean. Prod. 210, 401–409.
Ozoegwu, C.G., 2019. Artificial neural network forecast of monthly mean daily global solar radiation of selected locations based on time series and month number. J. Clean. Prod. 216, 1–13.
Parisi, G.I., Kemker, R., Part, J.L., Kanan, C., Wermter, S., 2019. Continual lifelong learning with neural networks: a review. Neural Netw. 113 , 54–71.
Rajeshwari, K., Balakrishnan, M., Kansal, A., Lata, K., Kishore, V., 2000. State-of-the-art of anaerobic digestion technology for industrial wastewater treatment. Renew. Sust. Energ. Rev. 4, 135–156.
Rojas, R., 1996. Neural Networks: A Systematic Introduction. Springer Science & Business Media.
Rout, P.R., Zhang, T.C., Bhunia, P., Surampalli, R.Y., 2021. Treatment technologies for emerging contaminants in wastewater treatment plants: a review. Sci. Total Environ. 753, 141990.
Rudin, C., 2019. Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. Nat. Mach. Intell. 1, 206–215.
Seviour, R.J., Mino, T., Onuki, M., 2003. The microbiology of biological phosphorus removal in activated sludge systems. FEMS Microbiol. Rev. 27, 99–127.
Shabanpour, H., Yousefi, S., Saen, R.F., 2017. Forecasting efficiency of green suppliers by dynamic data envelopment analysis and artificial neural networks. J. Clean. Prod. 142, 1098–1107.
Shewa, W.A., Dagnew, M., 2020. Revisiting chemically enhanced primary treatment of wastewater: a review. Sustainability 12, 5928.
Shi, S., Xu, G., 2018. Novel performance prediction model of a biofilm system treating domestic wastewater based on stacked denoising auto-encoders deep learning network. Chem. Eng. J. 347, 280–290.
Singh, K.P., Basant, N., Malik, A., Jain, G., 2010. Modeling the performance of "up-flow anaerobic sludge blanket" reactor based wastewater treatment plant using linear and nonlinear approaches—a case study. Anal. Chim. Acta 658, 1–11.
Spellman, F.R., 2013. Handbook of Water and Wastewater Treatment Plant Operations. CRC Press.
Strobl, C., Boulesteix, A.-L., Zeileis, A., Hothorn, T., 2007. Bias in random forest variable importance measures: illustrations, sources and a solution. BMC Bioinform. 8, 1–21.
Sugiyama, S., 2019. Human behavior and another kind in consciousness. Emerg. Res. Opportun. 1–102.
Verma, A., Wei, X., Kusiak, A., 2013. Predicting the total suspended solids in wastewater: a data-mining approach. Eng. Appl. Artif. Intell. 26, 1366–1372.
Volcke, E.I., van Loosdrecht, M.C., Vanrolleghem, P.A., 2006. Continuity-based model interfacing for plant-wide simulation: a general approach. Water Res. 40, 2817–2828.
Vučić, V., Süring, C., Harms, H., Müller, S., Günther, S., 2021. A framework for P-cycle assessment in wastewater treatment plants. Sci. Total Environ. 760, 1–13 143392.
Wiesmann, U., Choi, I.S., Dombrowski, E.-M., 2007. Fundamentals of Biological Wastewater Treatment. John Wiley & Sons.
Wilén, B.-M., Lumley, D., Mattsson, A., Mino, T., 2008. Relationship between floc composition and flocculation and settling properties studied at a full scale activated sludge plant. Water Res. 42, 4404–4418.
Wu, X., Yang, Y., Wu, G., Mao, J., Zhou, T., 2016. Simulation and optimization of a coking wastewater biological treatment process by activated sludge models (ASM). J. Environ. Manag. 165, 235–242.
Yang, S.-F., Li, X.-Y., 2009. Influences of extracellular polymeric substances (EPS) on the characteristics of activated sludge under non-steady-state conditions. Process Biochem. 44s, 91–96.
Young, B., Delatolla, R., Kennedy, K., Laflamme, E., Stintzi, A., 2017. Low temperature MBBR nitrification: microbiome analysis. Water Res. 111, 224–233.
Zeiler, M.D., 2012. Adadelta: an adaptive learning rate method. arXiv preprint arXiv: 1212.5701.