Postprint of: Bagherzadeh F., Shojaei Nouri A., Mehrani M. J., Thennadil S., Prediction of energy consumption and evaluation of affecting factors in a full-scale WWTP using a machine learning approach, PROCESS SAFETY AND ENVIRONMENTAL PROTECTION, Vol. 154 (2021), pp.458-466, DOI: 10.1016/j.psep.2021.08.040

Prediction of energy consumption and evaluation of affecting factors in a full-scale WWTP using a machine learning approach

Faramarz Bagherzadeh[1*], Amirreza Shojaei Nouri[2], Mohamad-Javad Mehrani[3,4], Suresh Thennadil[5]

1- Faculty of Mechanical and Ship Engineering, Gdansk University of Technology, Narutowicza Street 11/12, 80-233 Gdansk, Poland
2- Faculty of Civil and Environmental Engineering, University of Stuttgart, 70049 Stuttgart, Germany
3- Faculty of Civil and Environmental Engineering, Gdansk University of Technology, Narutowicza Street 11/12, 80-233 Gdansk, Poland
4- Department of Urban Water- and Waste Management, University of Duisburg-Essen, Universitätsstraße 15, 45141, Essen, Germany
5- Energy and Resources Institute, College of Engineering, Information Technology and Environment, Charles Darwin University, Darwin, Northern Territory 0909, Australia

*Corresponding Author: Faramarz Bagherzadeh (Email: s179532@student.pg.edu.pl)

Link: https://doi.org/10.1016/j.psep.2021.08.040

Abstract

Treatment of municipal wastewater to meet the stringent effluent quality standards is an energy-intensive process and the main contributor to the costs of wastewater treatment plants (WWTPs). Analysis and prediction of energy consumption (EC) are essential in designing and operating sustainable energy-saving WWTPs. In this study, the effect of wastewater, hydraulic, and climate-based parameters on the daily consumption of EC by East Melbourne WWTP was investigated based on the data collected over six years (2014-2019). Data engineering methods were applied to combine features from different resources. To this end, four various feature selection (FS) algorithms were used to reveal the relations among those variables and to select the most relevant variables for training the machine learning (ML) models. Further, the application of artificial neural networks (ANN) and two decision tree algorithms of Gradient Boosting Machine (GBM), and Random Forest (RF) were studied to predict EC records followed by a 95% confidence interval assessment. Results of FS algorithms revealed that total nitrogen, chemical oxygen demand (COD), and inflow-flow had the highest impact on WWTP energy consumption. Moreover, GBM had the best performance prediction among all other

regression algorithms. 95% of confidence interval showed a reasonable prediction error band ($\pm 68\ MWh/Day$).

Keywords: *Machine learning; energy consumption; power-grid prediction; WWTP; feature selection; wastewater characteristics*

1 Introduction

Recent developments in data science have provided the opportunity to predict prerequisite phenomena to improve operations, energy-saving, and costs reduction. The appropriate wastewater collection, treatment, and safe discharge from WWTPs are important issues for protecting the environment and public health, requiring a significant amount of energy (Hernández-Chover et al., 2018; Molinos-Senante et al., 2015). WWTPs collect raw sewage from the sewer networks and use different operational unit processes to remove pollutants to meet stringent effluent standards for reuse or discharge of wastewater (Newhart et al., 2019).

The energy is not only used inside WWTPs (like reactors, aeration supply, and electricity for pump devices), but also used for transportation, and production of different chemicals used in treatment processes (Longo et al., 2016), and energy consumption (EC) can be accounted up to the 48% of the WWTP operation cost (Wang et al., 2020). Hence, forecasting of EC in the WWTPs can be one of the fundamental subjects for obtaining sustainable development, power management, improved decision-making, monitoring various operational functions, and environmental protection (Ahmad and Chen, 2018).

Long-term prediction of EC in WWTPs provides a better understanding of required energy for various operational strategies, which eventually can lead to the minimization of the energy. It can also allow the wastewater practitioners to know which form of energy is mainly used and try to change the consumption pattern (Torregrossa et al., 2016). Different variables, primarily caused by the aeration processes in activated sludge, wastewater quality, weather condition, and operational circumstances, influence the amount of energy consumed in WWTPs which makes the forecasting of EC in WWTPs, a complex task (De Gussem et al., 2014; Panepinto et al., 2016).

Feature selection is a method for investigating the effect and importance of independent factors on the dependent variable, and these methods are used for the selection of the best input data to enhance prediction accuracy (Bagherzadeh et al., 2021; Kazemi et al., 2021). Artificial intelligence (AI) is used in a variety of fields to predict natural and artificial processes. As a subset of AI, Machine Learning (ML) is the method of recognizing a specific pattern giving the required data for prediction or classification (Géron, 2019).

ML techniques have recently grown in popularity in many fields such as wastewater components prediction and optimization, owing to their high accuracy without requiring a detailed understanding of the underlying mechanisms which is needed for building mechanistic models (Picos-Benítez et al., 2020; Zhao et al., 2020). In addition, in many cases, such understanding at the level required for accurate modeling is not available, thus making data-driven approaches using machine learning is preferable.

Some studies have been recently published on WWTPs energy consumption modeling and prediction with different statistical and ML algorithms. For instance, Żyłka et al. (2020) evaluated an application for the least square linear regression model for electricity consumption forecasting for a Polish dairy WWTP which caused optimization of the energy usage. Also, the impact of the air temperature and biological load as effective parameters on EC was observed. Yang et al. developed a regression model for WWTPs annual EC under different influent conditions. It was reported that influent flow rate and COD concentration are two main correlated features with EC of bioreactors (Yang et al., 2020).

The weather-based cluster analysis of historical influent data coupled with the different aeration strategies was implemented by (Borzooei et al., 2020) to optimize the EC of a large-scale WWTP in Italy. Furthermore, an EC model was proposed using ANN in WWTP with considering environmental and biological characteristics which reported an application of ANN model developed based on environmental parameters and wastewater characteristics in a WWTP (Oulebsir et al., 2020). Ahmad and Chen presented four different ML forecasting models for the EC requirement of pumps using in the water source sector, and environmental and power usage data of the water source pumps were used as input data on a weekly and monthly basis (Ahmad and Chen, 2018). Energy consumption based on a weather-based dataset in Egypt with a hybrid method (random vector functional link and artificial ecosystem-based optimization) was predicted (Essa et al., 2020). Yu et al. estimated the EC of WWTP using Bayesian semi-parametric quantile regression. Wastewater parameters such as BOD, COD, pH, and total nitrogen (TN) were input parameters to predict the EC, and correlation analysis results showed the highest relationship between the target with COD (0.96) and BOD (0.86), respectively (Yu et al., 2019).

In this study, the goal is to predict the long-term daily EC of the Melbourne East WWTP with four ML algorithms by evaluating the effectiveness and importance of various parameters (time, climate, hydraulic flow, and wastewater characteristics) on the target to enhance the performance. Until now, just a few studies have investigated the forecasting of EC in the WWTP using ML technologies, and to our best of knowledge, the combination of wastewater characteristics and climate conditions for a full-scale WWTP EC prediction has not been studied yet. Furthermore, applying the GBM for this aim is another novelty of this work.

Moreover, obtained results would be useful for WWTPs across the world for operational monitoring, electricity consumption management, and prediction toward the global sustainable development policy.

2 Methodology

2.1 Case study information

Melbourne city has two main WWTPs, including Western Treatment Plant (WTP) and Eastern Treatment Plant (ETP), which are situated in the suburbs of Werribee and Bangholme, Melbourne, respectively. Both WWTPs are operated by Melbourne Water. The ETP was built to reduce the pressure on the western WWTP of Melbourne to meet the needs of the south-eastern growing population. ETP size is almost one-tenth the size of the western treatment plant (WTP) of Melbourne, but it treats nearly half of Melbourne sewage (2.5 million householders) (Melbourne Water, 2021). Around 330 million liters of sewage per day from the southeastern suburbs of Melbourne flow through the trunk sewerage network to the ETP and turns into class A recycled water. This plant consists of a mechanical plant, tanks, effluent basins, sludge drying pans, and holding areas for dried sludge, as shown in Figure 1. The treatment procedure starts with physical cleaning of raw wastewater such as removing debris, separating large particles, and grit filtering, then moving further for the biological treatment section.

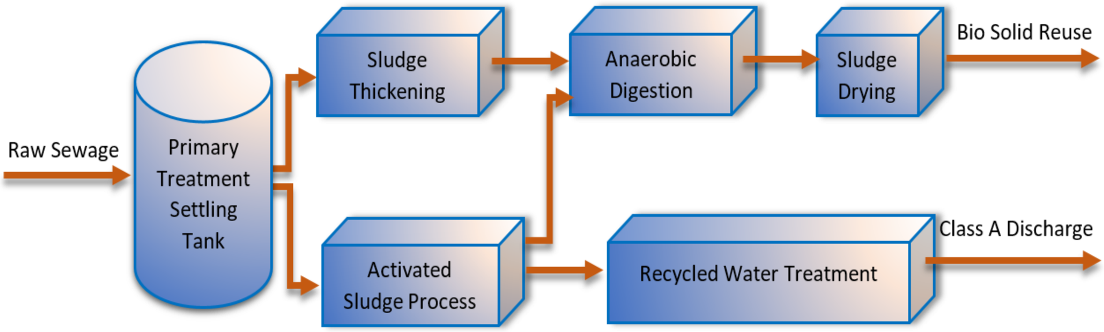


Figure 1.The Eastern Treatment Plant (ETP) of Melbourne (Melbourne Water, 2021)

The influent composition data was gathered at various time intervals using continuous sensors and quarterly/daily lab reports. The standard of sampling, treating, and effluent management of ETP follows the national water quality management strategy (National water quality, 1997).

2.2 Data collection

A data set consisting of 1000 records (almost 6 years between 2014-2019) was collected from the Melbourne water (open) database. Under the Victoria Government's open data policy this data can be accessed used and shared by anyone (Melbourne water database, 2021). The total daily electricity consumption data was collected via revenue quality meters with a cycle of 15 minutes and represented in a daily form. The wastewater characteristics were recorded by sampling and sensors. Furthermore, climate reports were collected from the Melbourne airport weather station (Melbourne airport weather station, 2021), which is the nearest weather station to the ETP. To join the data, the hourly power consumption was replaced by the average daily power consumption with the corresponding date. Then, the climate table was obtained by web scraping methods (with BeautifulSoup python library). As there are many unfilled rows in biology data (lack of sampling due to holidays and weekends), all six datasets were joined with an inner-joint operation on the column of the record date. In the end, several records with Null values, very low power consumption (near zero), extremely high consumptions, irrelevant feature values, and outlier data values (data point that differs significantly from other observations) were removed. The number of removed records was less than 5% of the dataset.

The overall dataset consists of daily total grid power consumption, Ammonia (NH_4-N), total nitrogen (TN), biological oxygen demand (BOD), chemical oxygen demand (COD), average Temperature (T_{avg}), Maximum temperature (T_{max}), Minimum temperature (T_{min}), atmospheric pressure (AP), average relative humidity (H), total rainfall and/or snowmelt (Pr), average visibility(VIS), average wind speed (WS_{avg}), and Maximum wind speed (WS_{max}), The dataset was split into a training set which comprised of the, first 48 months (75%) of the selected data and an unseen test set comprising of the subsequent 18 months (25%) of the data. Furthermore, as there are many features in the dataset, comprehensive feature engineering was done to remove the redundant variables and train the ML models with the best subset of variables (Julián Luengo et al., 2020; Ranjan et al., 2021).

To perform the calculations, Python 3.8 programming language via the Anaconda platform (Jupyter notebook), and libraries such as Pandas (data engineering), Scikit-Learn for FS, and ML models, MatplotLib, and Seaborn for visualization were implemented.

2.3 Feature selection (FS)

Efficiently decreasing the input data dimensions helps to simplify the model and obtain better accuracy. FS methods aim to score the independent variables based on specific criteria and introduce the best subset of features (Luíza da Costa et al., 2021). In this study, mutual information (MI) (Gao and Wu, 2020; Gonzalez-Lopez et al., 2020), Pearson correlation (PC) (Michalak and Kwasnicka, 2006), backward elimination (BE) (Fernando Jimeneza et al., 2020), and RF (Masmoudi et al., 2020) were used as FS methods. More details about each method can be found in the supplementary information (SI).

2.4 Modeling approaches

2.4.1 Neural networks (ANN, RNN-LSTM)

In the simplest format of neural networks, each neuron in each layer is fully connected with all neurons of the previous layer and the next layer, which is called ANN. This type of network architecture usually consists of three layers: input, hidden, and output. Mostly, the size of the input layer is equal to the number of independent input variables. Similarly, the output layer has the number of neurons equal to some dependent features (target features). With the increasing complexity of the model (dimensionality), the hidden layer becomes larger and more complicated. Selecting the size and number of hidden layers is still challenging and requires a trial and error method to obtain the optimal hyper-parameters (P. Raut and A. Dani, 2020; Tosun et al., 2016).

Similar to ANN, the Recurrent neural network (RNN) is a sequence of neurons. RNNs can utilize their internal memory to process input data with variable lengths. A large number of RNN neurons lead to the issue of vanishing gradient. The Long Short-Term Memory (LSTM) network was suggested to solve the problem of vanishing gradient by incorporating the data-dependent controls to the RNN unit. Also, LSTM can learn the short-range relations between features (Sherstinsky, 2020).

2.4.2 Decision tree (RF, GBM)

Utilizing a combination of algorithms to increase prediction accuracy is called ensemble learning. Bootstrap aggregation is a method for combining the result of different algorithms in an ensemble technique (Sharafati et al., 2020). RF generates many decision trees (a forest) based on random subsampling of the dataset, and then it aggregates the results of trees to predict the output value (Lakshmanaprabu et al., 2019). Similarly, GBM uses the decision tree algorithms but with a different strategy for generating the forest. GBM creates a new tree to decrease the error of estimation by considering the target value. As it adds more trees, the prediction error shrinks (Natekin and Knoll, 2013).

In this study, four ML algorithms were employed. The best hyper-parameters were selected for each type of model through trial and error on the training dataset (Table 1).

Table 1

Hyperparameters for each model

Model	Parameter	Value
GBM	Number of estimators	3000
	Learning rate	0.001
	Min sample leaf	20
	Min sample split	30
	Max depth	15
RF	Number of estimators	2000
	Min sample leaf	5
	Min sample split	5
	Max depth	30
	Max features	Auto
ANN	Input layer	10
	Hidden layer 1 (Relue)	10
	Hidden layer 2 (Relue)	10
	Hidden layer 3 (Relue)	4
	Output layer	1
	Loss	MAE
	Optimizer	Adam
RNN-LSTM	Input layer	(None,10,1)
	Hidden layer 1 (LSTM)	10
	Hidden layer 2 (LSTM)	10
	Output layer	1
	Loss	MSE
	Optimizer	Adam

2.5 Model construction

As shown in Figure 2, the model construction was as follows: After joining datasets from different sources, a comprehensive set of data engineering such as removing outliers, analyzing multi-collinearity, normalization, and FS was performed as described in section 2.2. Then, four different ML algorithms (ANN, LSTM, RF, GBM) were trained to predict total grid power consumption. Finally, the accuracy of each model was calculated based on model metrics to report the best model.

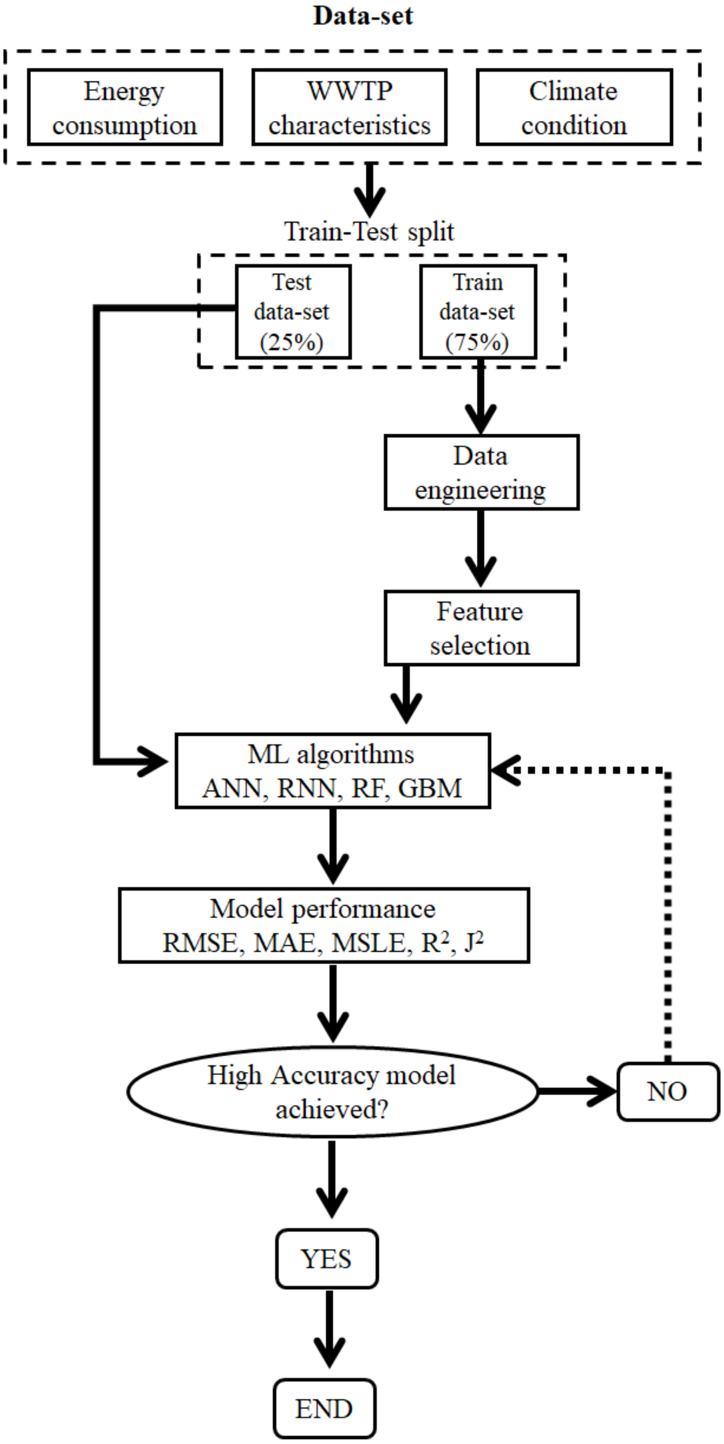


Figure 2. Modeling and prediction procedure construction

2.6 Model performance evaluation

The performance of the model determines the model's goodness. In this study, Root means square error (RMSE) (Eq.1), Mean absolute error (MAE) (Eq.2), and Mean square logarithmic error (MSLE) (Eq.3) (Bhagat et al., 2021; Camacho et al., 2020) were calculated. MSLE indicates the relative difference between actual and predicted values while treating large and small errors differently. It generates a percentage difference which is more meaningful in the case of large numbers. Also, it penalizes underestimated values more than overestimated ones. Coefficient of determination (R^2) was used to measure the goodness-of-fit of a model that considers the changes in the target variable related to the features with a linear relationship. A model with accurate prediction will possess higher values (up to 1.0) of R^2 (Rosenthal, 2011; Warner, 2013).

$$\mathrm{MSE} = \frac{1}{\mathrm{n}} \sum_{\mathrm{i}=1}^{\mathrm{n}} \left(\hat{y}_{\mathrm{i}} - y_{\mathrm{i}}\right)^2 \; ; \; \mathrm{RMSE} = \sqrt{\mathrm{MSE}} \tag{1}$$

$$\mathrm{MAE} = \frac{1}{\mathrm{n}} \sum_{\mathrm{i}=1}^{\mathrm{n}} \left|\hat{y}_{\mathrm{i}} - y_{\mathrm{i}}\right| \tag{2}$$

$$\mathrm{MSLE} = \frac{1}{\mathrm{n}} \sum_{\mathrm{i}=1}^{\mathrm{n}} (\log_e(1 + y_{\mathrm{i}}) - \log_e(1 + \hat{y}_{\mathrm{i}}))^2 \tag{3}$$

$$R^2 = 1 - \frac{RSS}{TSS} \tag{4}$$

$$J^2 = \frac{RMSE_{test}^2}{RMSE_{train}^2} \tag{5}$$

Where $i = 1,2,..n$ is the number of observations, and n is the total number of records. Considering $\hat{y}_i$ for model output (prediction) and y_i as real values. For Eq. 4, RSS is the sum of squares of residuals, and TSS is total sum of squares.

After finding the best model, the confidence interval (CI) for predicted values will be calculated with Eq. 6, where $\bar{X}$ is the sample mean, z is confidence level value, σ is the standard deviation, and n is the total number of observations. As there is a large number of observations and the error distribution is approximately normal (as will be seen in Figure 7), the z-score was used to determine the CI.

$$CI = \bar{X} \pm z\frac{\sigma}{\sqrt{n}} \tag{6}$$

3 Results and discussion

3.1 Data - Statistical information

A brief description of elementary statistical properties is given in Table 2 which shows all the features of the joined dataset. Total Grid Consumption is the dependent variable (prediction target), and the rest are independent variables.

Table 2

Statistical description of the data set

Parameters (Abbreviation)	Unit	Max	Mean	Min	SD
Average Inflow (Q_{in})	m^3/s	19	4.5	2.6	1.4
Average Outflow (Q_{out})	m^3/s	7.9	3.9	0.1	1.2
Energy Consumption (Ec)	MWh	398	275	116	44
Ammonia (NH_4-N)	mg/L	93	39	13	7
Biological Oxygen Demand (BOD)	mg/L	850	382	140	86
Chemical Oxygen Demand (COD)	mg/L	1700	846	360	145
Total Nitrogen (TN)	mg/L	92	62	40	3.6
Average Temperature (T_{avg})	◌̊C	35.5	15	0	5.4
Maximum temperature (T_{max})	◌̊C	43.5	20.5	0	7.1
Minimum temperature (T_{min})	◌̊C	28	10	-2	4.7
Atmospheric pressure (AP)	(hPa)	1022	3.7	0	61
Average humidity (H)	%	97	63	0	14
Total rainfall and / or snowmelt (Pr)	mm	18	0.2	0	1.3
Average visibility (VIS)	Km	512	9	0	16
Average wind speed (WS_{avg})	Km/h	49	19	0	7.1
Maximum wind speed (WS_{max})	Km/h	83.5	35.4	0	11.6
Year (year)	-	2019	-	2014	-
Month (month)	-	12	-	1	-
Day (day)	-	31	-	1	-

3.2 Summary of the feature selection procedure

The main objective of FS methods was to reduce the number of input variables by removing non-informative and/or redundant predictors in a predictive model. To this end, two statistical-based, as well as two intrinsic and wrapper methods, were implemented in the FS step. Pearson correlation and Mutual information were used for numerical and categorical variables to measure dependence between variables of the dataset. Further, the RF algorithm was used to describe the contribution of each feature in the prediction of the target values. To facilitate the comparison among different methods, calculated scores were presented as a percentage by considering the maximum score as 100% (Figure 3).

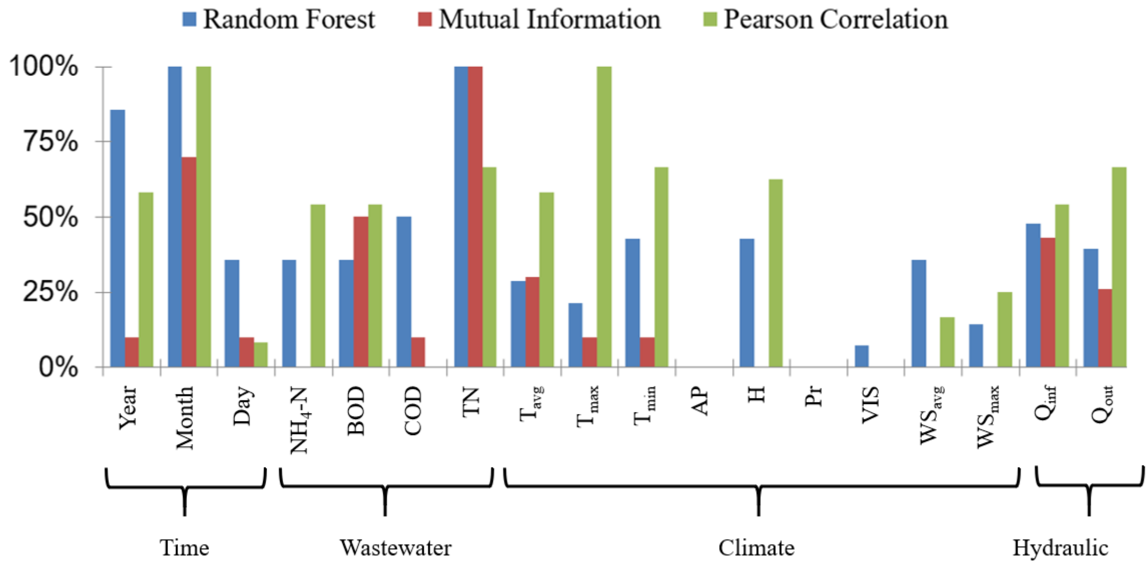


Figure 3. Feature selection scores for each group affecting on the target

As shown in Figure 3, features are divided into four groups of Time, wastewater, climate, and Hydraulic parameters. In the first group (Time) all the FS methods indicated that the *month* is an important variable followed by the *year*. Melbourne Water has a constant development policy to improve the plant as the population of the city increases every year, explaining why annual power consumption changes. However, the total EC relation with *month* seems to be due to the seasonal and weather changes as there is a 0.22 correlation between the *month* and daily maximum temperature (T_{max}) (See SI, Figure S2). *Day*, *month*, and *year* can be considered as a time variable with different levels of smoothing. *Year*, with the highest smoothing, will have reduced sensitivity to daily and seasonal variations since for the same year, the daily energy consumption will differ according to seasons while the value of the year is the same. Furthermore, the energy consumption on a particular day of the year can vary from year to year. On the other hand, the *day* is the time variable with the lowest level of smoothing. In this case, days within a season can vary significantly from year to year in addition to seasonal impacts varying from year to year. Thus a low level of time smoothing could lead to random fluctuations obfuscating the time correlation of EC. The data and analysis presented here indicate that amongst *day*, *month* and *year*, the level of smoothing afforded by using the *month* as the time variable leads to the best correlation with daily consumption. It may be possible to determine an optimal block of time (other than the *day*, *month*, or *year*) that provides maximum correlation with EC. Additionally, although the *month* is indicating a higher correlation with EC compared to *day* and *year*, there might be other factors hidden in this feature (*month*) that we have not investigated yet. Further investigation and additional data would be required to obtain a more comprehensive understanding of the effect of time on EC.

In the second group (wastewater parameters) *TN* and *BOD* received higher scores in correlation with the target (100% and 50%, respectively). Also as shown by the PC heat-map graph (Figure S2), there is a high correlation between *COD* and *TN* (0.68), as well as *COD* and *BOD* (0.52). Thus, it is reasonable to ignore *COD.* Wang et al (2020) found the relation between reduction of COD and ammonia with the energy consumption. In the third group (Climate parameters), variables such as T_{avg}, T_{max}, T_{min}, and H were more effective. Considering the correlation between these values H with T_{avg} (0.55), T_{min}, with T_{max} (0.76), and T_{avg} with T_{max} (0.92) utilizing all of them for ML training can lower our model accuracy. As the correlation values of more than 0.8 are known for "very highly correlated" and more than 0.7 is known for "highly correlated" among T_{avg}, T_{max}, and T_{min} only one is chosen in the final subset to be used for the training process (Yoon et al., 2013). Several climate features such as AP, Pr, VIS, WS_{avg}, and WS_{max} are irrelevant to the target feature, and they should not be considered for training the model.

In hydraulic parameters, average daily influent and effluent flowrates are highly correlated (0.54), and it is logical to select only one of them for the final subset of features. Considering all the FS methods used in this study, Q_{inf} with a higher score (55%) was considered in the final subset.

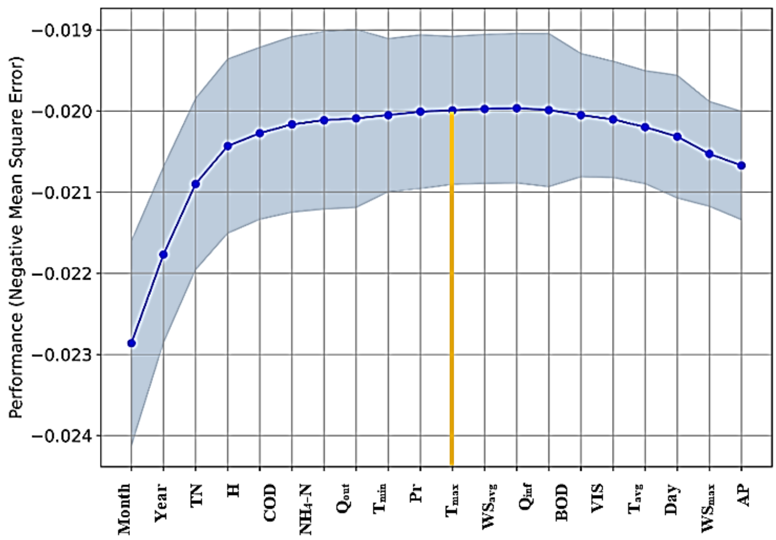


Figure 4. Sequential backward elimination

The backward elimination result (Fig. 4) displays the maximum number of variables that can improve the model. It shows that with the first ten features, the model can reach the maximum possible accuracy, and adding more independent variables may lead to overfitting issues. Considering all these steps (FS results), and trial and error in training the models, the following input parameters were selected finally: Month, TN, NH4-N, BOD, T_{max}, H, Pr, and Q_{inf}, and the Energy Consumption as model output. This subset of features was used to train and test all the ML models during this study.

3.3 Energy consumption Prediction

The forecasting accuracy of the trained models applied to test datasets was evaluated in terms of model metric functions, namely RMSE, MAE, MSLE, $R^{2,}$ and J^2, which are summarized in Table 3. Comparing all the performance measures listed in Table 3, it can be seen that overall, GBM provides the best performance. Decision tree models (GBM and RF) had the lowest RMSE and MAE error values. Among all models, RF and RNN showed the highest and lowest J^2, respectively, indicating the presence of overfitting issues in the RF model as well as poor calibration of RNN. Therefore, GBM was revealed as the best prediction model.

Table 3

Evaluation of each forecasting approach by the model metrics

Model Type	Dataset	RMSE	MAE	MSLE	R^2	J^2
GBM	Test	33.9	26.9	0.02	0.18	1.66
	Train	29.9	23.7	0.01	0.53	
RF	Test	34.8	27.7	0.02	0.14	4.16
	Train	24.3	18.6	0.01	0.69	
ANN	Test	39.8	32.1	0.02	0.00	1.45
	Train	36.3	28.1	0.02	0.31	
RNN	Test	37.3	29.3	0.02	0.01	0.76
	Train	39.9	31.8	0.02	0.17	

Furthermore, to better understand model metrics, a multi-feature representation was developed as shown in Fig.5. This figure indicates the prediction errors (MAE and MSE) after inverse scaling based on training set scale, in a scatter plot for various ML algorithms. The color bar shows the coefficient of determination (R^2), and the shape of points defines the type of dataset (train or test). Overall, decision tree algorithms (GBM and RF) showed better performance than neural networks (ANN, and RNN). Although RF had the lowest errors on the training dataset

(MAE=18.6 and RMSE=24.3 MWh), the result on the test data set was found to be far from train data performance indicating overfitting. GBM performed better for train and test data with MAE=23.8 MWh, RMSE=29.9 MWh, R^2=0.53, and MAE=26.9 MWh, RMSE=33.9 MWh, R^2=0.18 for train and test data sets respectively. On the other hand, neural networks, including RNN and ANN, showed the least level of R^2 (near zero) on the test dataset and substantial errors.

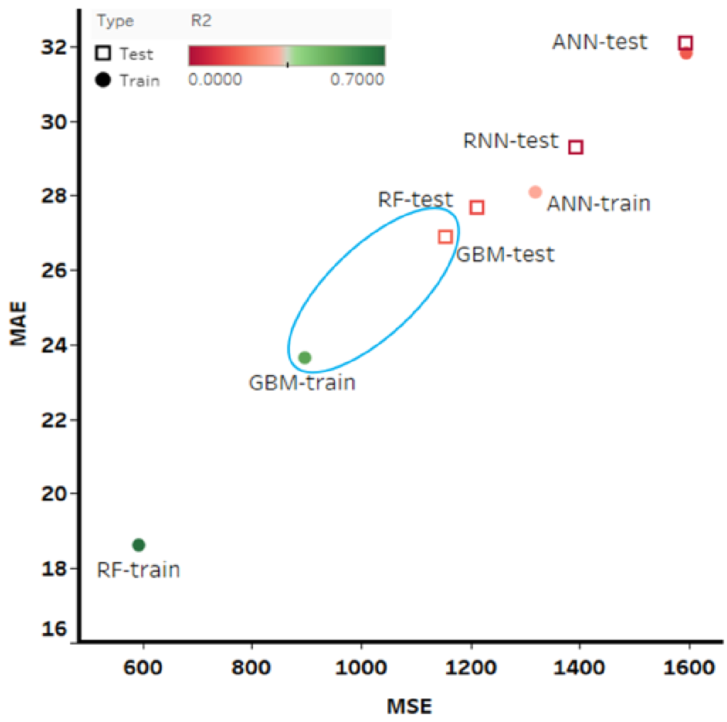
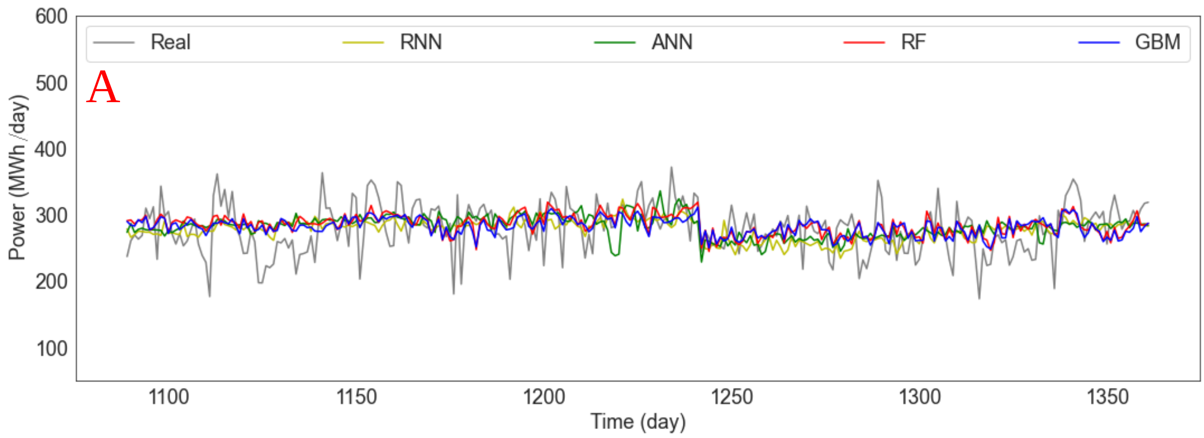


Figure 5. Model performance by all algorithms

Figure 6 illustrates model predictions versus real values for the period of the test dataset (1/1/2018 – 1/7/2019). The daily power consumption actual records are fluctuating dramatically almost every week. Although models have failed to capture every peak and valley of the target value, they could predict the trend and approach of the target curve relatively accurately.

It is noticeable that the power consumption has a flat trend in the test set with a sudden drop in the mean consumption from around day 1250. The models capture this drop in the mean. A close inspection of Figure 6(A) indicates that the model based on GBM is slightly better in capturing the finer fluctuations compared to other models. Another point to note is that the power consumption in the test set has relatively larger fluctuations around the mean than in the training set, which has much fewer sharp day-to-day fluctuations. One reason for the model not capturing these fluctuations could be due to not enough “examples” in the training set for the model to be trained to pick up these fluctuations.

Wastewater plants undergo frequent changes with multiple upgrades to the equipment and process over the years. Also, requirements can change due to population changes. Since future changes that manifest in the test set data cannot be taken into account when building the machine learning models, degradation in model performance when applied to data from subsequent months can be expected. Thus, some form of retraining the model will have to be developed to account for process and equipment changes to ensure the model can accurately predict power consumption. Due to a large number of records of the training dataset, and visualization limitations, Fig 6(B) only shows a part of the training dataset.



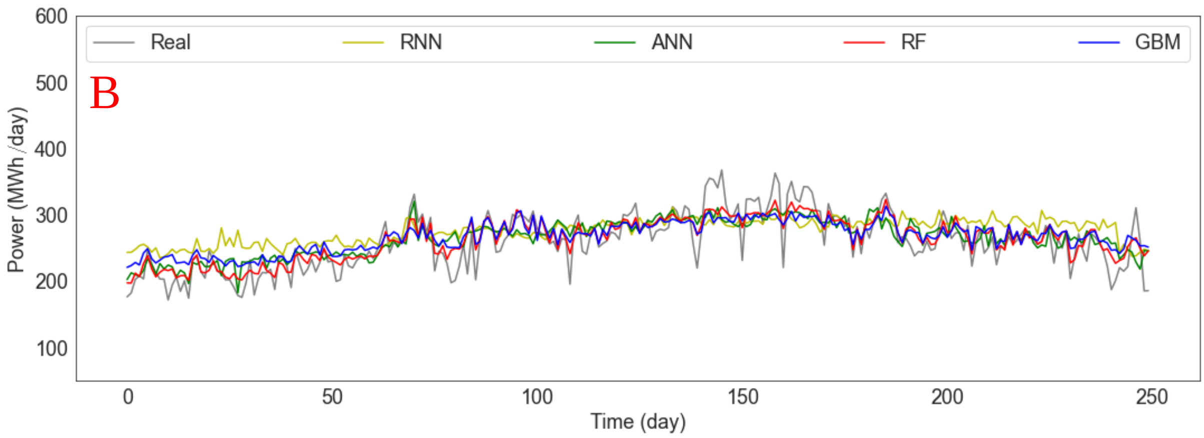


Figure 6. Total grid power consumption prediction of Melbourne east WWTP. A: unseen test dataset (1/1/2018 – 1/7/2019), and B: a part of the training dataset (1/1/2014 – 1/1/2015),

The residual histogram in Fig.7 confirms that the natural (Gaussian) distribution of errors. Errors are symmetrically distributed around a mean value of -2.2 MWh. Considering 95% CI, the lower limit is -68 MWh, and the higher limit is 64 MWh for GBM predictions on the test dataset as shown in Fig.7.

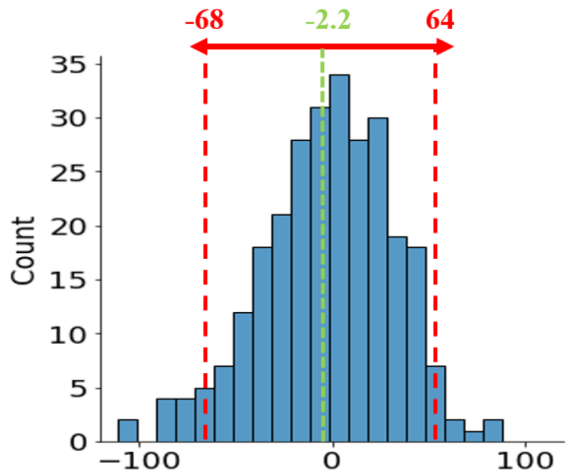


Figure 7. Residual distribution of GBM model with 95% confidence interval

To better visualize the model prediction and 95% confidence interval, Figure 8 was developed, showing that the given CI covers the majority of the data points from the unseen test dataset to validate the model predictions by the GBM model. As there are a high number of observations and error distribution is normal, the z-score was used to determine the CI.

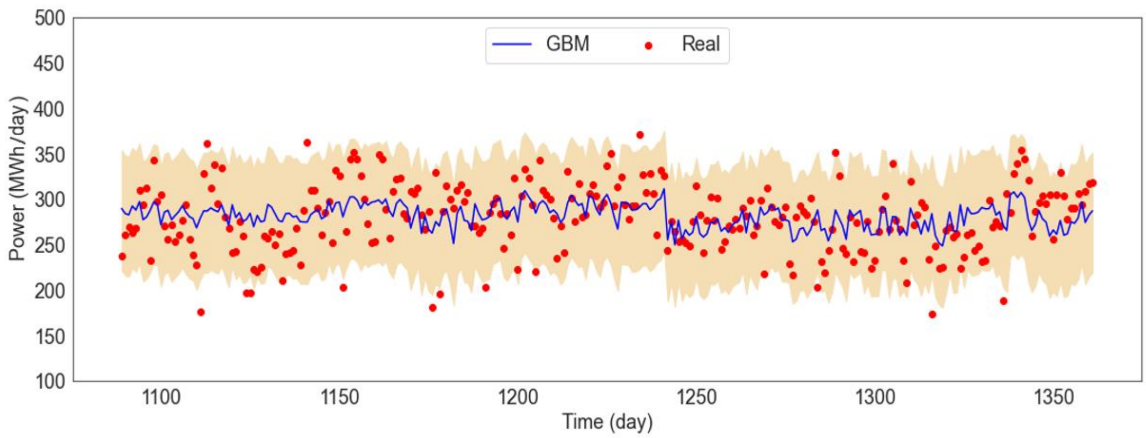


Figure 8. Prediction result by GBM model with 95% confidence interval
[Predicted value -68 MWh/day , Predicted value +64 MWh/day]

As summarized in Table 4, various ML models and benchmarking methods were used in previous studies to predict the power consumption in WWTPs and define correlated parameters on the target.

Table 4. Summary of different studies on the WWTPs power consumption prediction using ML methods

Features	Prediction Algorithm	Performance metric	Remarks	Dataset	References
TN, TP, BOD, COD, T	LSLR	R^2_{Train}=0.912	Air temperature and biological load had effective parameters on energy consumption. Prediction performance was not evaluated using a separate test set.	Features were collected from 3 different points of the system consists of 95 series of measurements over 30 month	(Żyłka et al., 2020)
Q_{inf} ,T, BOD, TN,	RNN (GRU and LSTM)	RMSE=509 kWh/day, MAE=389.2	The presented model can be used in optimization scenarios to provide data-driven solutions for regular WWTP activity. R^2 values were not provided.	Training data were collected daily from 2010 till 2017 and one year (365) records were used as a test dataset	(Cheng et al., 2020)
pH, BOD, COD, SS, Chrom, TP, TN, NH3,	Bayesian semi-parametric quantile regression .	R^2=N/A	the highest relationship between the energy consumption with COD and BOD was observed. Regression analysis was done for 3 different energy consumption levels for investigating the effects of parameters on consumption. Energy prediction performance was not evaluated.	Daily records, 363 samples (from December 2015 to December 2016)	(Yu et al., 2019)
COD, BOD, SS, NH4, T, Flowrate	DNN	R^2_{Test} =0.74 RSR=0.33-0.52	Pollution indicators are efficient estimators for the prediction and optimization of power consumption	A total number of 318 records were used from 2006 till 2016. Two selection steps, which significantly reduced the number of data points, were used before model building and testing. The final number of data points used was not given. 20% of the selected data points were used for testing the models utilizing key performance indicators of WWTP	(Oulebsir et al., 2020)
COD, BOD, TP, TN, Flowrate,	ANN RF	R^2_{Train} =0.6-0.9 R^2_{Test} =0.4-0.8	Increasing the number of neurons doesn't necessarily improve the ANN models. In case of overfitting issues, RF had better results than ANN	317 WWTPs using CAS technology, and located in northwest Europe. The test dataset (112 records) was selected randomly from the database. Models were built for predicting yearly energy consumption.	(Torregrossa et al., 2018)
Months, TN, NH_4-N, BOD, T_{max}, H, Pr, and Q_{inf}	GBM RF ANN RNN	R^2_{train}=0.53 R^2_{test}=0.18	TN, ammonia, BOD, temperature, humidity, and influent flow were among the highest correlated parameters with energy consumption of ETP based on three FS methods.	Nearly 1000 records of data from ETP Melbourne were collected after data engineering during the years (2014-2019). Dataset was a result of inner joining between weather, wastewater characteristics, and energy consumption parameters. Models were built for predicting daily energy consumption.	**This study**

Total Nitrogen(TN), Total Phosphorus (TP), Chemical Oxygen Demand (COD), Biological Oxygen Demand (BOD), Temperature (T), List Square Linear Regression (LSLR), Chlorine (Cl), Suspended Solids (SS), RMSE-observations standard deviation ratio (RSR), Recurrent Neural Network (RNN), Long Short Term Memory (LSTM), Gated Recurrent Unit (GRU), Dense Neural Network (DNN), Gradient Boosting Machine (GBM), Random Forest (RF), Artificial Neural Network (ANN), Conventional Activated Sludge (CAS), Ammonia (NH_4-N), Maximum Temperature (T_{max}), Minimum Temperature (T_{min}), Average relative humidity (H), Total rainfall and/or snowmelt (Pr), Eastern Treatment Plant (ETP)

4 Conclusions

In this study, at first, the importance and effectiveness of different characteristics like hydraulic, climate, time, and wastewater parameters were studied on the EC of Melbourne east WWTP. Then, four ML algorithms were applied regarding the best-achieved input dataset to predict the power consumption forecasting. Results of the current study can be useful for decision-making and energy saving of WWTPs. By estimating of needed EC, the influential parameters can be better adjusted and cause savings in cost and energy by the operational systems. Furthermore, the following conclusions can be derived:

- Melbourne east WWTP energy consumption positively relates to weather situations such as temperature (average, max, min) and humidity.
- Precepetations, atmospheric pressure, and wind speed did not have considerable effects on the target.
- TN, BOD, ammonia, daily temperature, humidity, and influent flow had the highest impact on the EC in Melbourne east WWTP.
- GBM algorithm revealed the best performance for prediction among other algorithms showing its prediction power in non-linear irregular patterns.
- A confidence interval study showed a reasonable error interval band ($\pm 68\ MWh/Day$), which can be used for further optimization of the ETP energy consumption.

Acknowledgment

The authors gratefully acknowledge Melbourne East Wastewater Treatment Plant and Melbourn Water organization. Especially thanks to Mr. James Lloyd for his technical support and for providing useful open-source data.

Data available

The merged data (wastewater characteristics and climate parameters) used in this paper is accessible at: https://data.mendeley.com/datasets/pprkvz3vbd/1

References

Ahmad, T., Chen, H., 2018. Utility companies strategy for short-term energy demand forecasting using machine learning based models. Sustainable Cities and Society 39, 401-417.

Bagherzadeh, F., Mehrani, M.-J., Basirifard, M., Roostaei, J., 2021. Comparative study on total nitrogen prediction in wastewater treatment plant and effect of various feature selection methods on machine learning algorithms performance. Journal of Water Process Engineering 41, 102033.

Bhagat, S.K., Tiyasha, T., Awadh, S.M., Tung, T.M., Jawad, A.H., Yaseen, Z.M., 2021. Prediction of sediment heavy metal at the Australian Bays using newly developed hybrid artificial intelligence models. Environmental Pollution 268, 115663.

Borzooei, S., Miranda, G.H.B., Abolfathi, S., Scibilia, G., Meucci, L., Zanetti, M.C., 2020. Application of unsupervised learning and process simulation for energy optimization of a WWTP under various weather conditions. Water Science and Technology 81, 1541-1551.

Camacho, J., Smilde, A.K., Saccenti, E., Westerhuis, J.A., 2020. All sparse PCA models are wrong, but some are useful. Part I: Computation of scores, residuals and explained variance. Chemometrics and Intelligent Laboratory Systems 196, 103907.

Cheng, T., Harrou, F., Kadri, F., Sun, Y., Leiknes, T., 2020. Forecasting of Wastewater Treatment Plant Key Features Using Deep Learning-Based Models: A Case Study. IEEE Access 8, 184475-184485.

De Gussem, K., Fenu, A., Wambecq, T., Weemaes, M., 2014. Energy saving on wastewater treatment plants through improved online control: case study wastewater treatment plant Antwerp-South. Water Science and Technology 69, 1074-1079.

Essa, F.A., Abd Elaziz, M., Elsheikh, A.H., 2020. Prediction of power consumption and water productivity of seawater greenhouse system using random vector functional link network integrated with artificial ecosystem-based optimization. Process Safety and Environmental Protection 144, 322-329.

Fernando Jimeneza, Pham, B.T., Nguyen-Thoi, T., Ly, H.-B.N., M.D.,, Al-Ansari, N.T., V.-Q.,, Le, T.-T., 2020. Extreme Learning Machine Based Prediction of Soil Shear Strength: A Sensitivity Analysis Using Monte Carlo Simulations and Feature Backward Elimination. Sustainability 12.

Gao, L., Wu, W., 2020. Relevance assignation feature selection method based on mutual information for machine learning. Knowledge-Based Systems 209, 106439.

Géron, A., 2019. Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow,. O'Reilly Media, Inc., 1005 Gravenstein Highway North,

Sebastopol, CA 95472,, Canada.

Gonzalez-Lopez, J., Ventura, S., Cano, A., 2020. Distributed multi-label feature selection using individual mutual information measures. Knowledge-Based Systems 188, 105052.

Hernández-Chover, V., Bellver-Domingo, Á., Hernández-Sancho, F., 2018. Efficiency of wastewater treatment facilities: The influence of scale economies. Journal of Environmental Management 228, 77-84.

Julián Luengo, Diego García-Gil, Sergio Ramírez-Gallego, Salvador García, Francisco Herrera, 2020. Big Data Preprocessing. Springer.

Kazemi, P., Bengoa, C., Steyer, J.-P., Giralt, J., 2021. Data-driven techniques for fault detection in anaerobic digestion process. Process Safety and Environmental Protection 146, 905-915.

Lakshmanaprabu, S.K., Shankar, K., Ilayaraja, M., Nasir, A.W., Vijayakumar, V., Chilamkurti, N., 2019. Random forest for big data classification in the internet of things using optimal features. International Journal of Machine Learning and Cybernetics 10, 2609-2618.

Longo, S., d'Antoni, B.M., Bongards, M., Chaparro, A., Cronrath, A., Fatone, F., Lema, J.M., Mauricio-Iglesias, M., Soares, A., Hospido, A., 2016. Monitoring and diagnosis of energy consumption in wastewater treatment plants. A state of the art and proposals for improvement. Applied Energy 179, 1251-1268.

Luíza da Costa, N., Dias de Lima, M., Barbosa, R., 2021. Evaluation of feature selection methods based on artificial neural network weights. Expert Systems with Applications 168, 114312.

Masmoudi, S., Elghazel, H., Taieb, D., Yazar, O., Kallel, A., 2020. A machine-learning framework for predicting multiple air pollutants' concentrations via multi-target regression and feature selection. Science of The Total Environment 715, 136991.

Melbourne airport weather station, 2021. https://en.tutiempo.net.

Melbourne Water, 2021. https://www.melbournewater.com.au/about/strategies-and-reports/annual-report.

Melbourne water database, 2021. https://data-melbournewater.opendata.arcgis.com.

Michalak, K., Kwasnicka, H., 2006. Correlation-based Feature Selection Strategy in Neural Classification, Sixth International Conference on Intelligent Systems Design and Applications, pp. 741-746.

Molinos-Senante, M., Hanley, N., Sala-Garrido, R., 2015. Measuring the CO2 shadow price for wastewater treatment: A directional distance function approach. Applied Energy 144, 241-249.

Natekin, A., Knoll, A., 2013. Gradient boosting machines, a tutorial. Frontiers in Neurorobotics 7.

National water quality, 1997. https://www.waterquality.gov.au/guidelines/sewerage-systems#effluent-management.

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.

Oulebsir, R., Lefkir, A., Safri, A., Bermad, A., 2020. Optimization of the energy consumption in activated sludge process using deep learning selective modeling. Biomass and Bioenergy 132, 105420.

P. Raut, A. Dani, 2020. Correlation Between Number of Hidden Layers and Accuracy of Artificial Neural Network. Springer, Singapore.

Panepinto, D., Fiore, S., Zappone, M., Genon, G., Meucci, L., 2016. Evaluation of the energy efficiency of a large wastewater treatment plant in Italy. Applied Energy 161, 404-411.

Picos-Benítez, A.R., Martínez-Vargas, B.L., Duron-Torres, S.M., Brillas, E., Peralta-Hernández, J.M., 2020. The use of artificial intelligence models in the prediction of optimum operational conditions for the treatment of dye wastewaters with similar structural characteristics. Process Safety and Environmental Protection 143, 36-44.

Ranjan, K.G., Prusty, B.R., Jena, D., 2021. Review of preprocessing methods for univariate volatile time-series in power system applications. Electric Power Systems Research 191, 106885.

Rosenthal, J., 2011. Statistics and Data Interpretation for Social Work. Springer publishing.

Sharafati, A., Asadollah, S.B.H.S., Hosseinzadeh, M., 2020. The potential of new ensemble machine learning models for effluent quality parameters prediction and related uncertainty. Process Safety and Environmental Protection 140, 68-78.

Sherstinsky, A., 2020. Fundamentals of Recurrent Neural Network (RNN) and Long Short-Term Memory (LSTM) network. Physica D: Nonlinear Phenomena 404, 132306.

Torregrossa, D., Leopold, U., Hernández-Sancho, F., Hansen, J., 2018. Machine learning for energy cost modelling in wastewater treatment plants. Journal of Environmental Management 223, 1061-1067.

Torregrossa, D., Schutz, G., Cornelissen, A., Hernández-Sancho, F., Hansen, J., 2016. Energy saving in WWTP: Daily benchmarking under uncertainty and data availability limitations. Environmental Research 148, 330-337.

Tosun, E., Aydin, K., Bilgili, M., 2016. Comparison of linear regression and artificial neural network model of a diesel engine fueled with biodiesel-alcohol mixtures. Alexandria Engineering Journal 55, 3081-3089.

Wang, S., Zou, L., Li, H., Zheng, K., Wang, Y., Zheng, G., Li, J., 2020. Full-scale membrane bioreactor process WWTPs in East Taihu basin: Wastewater characteristics, energy consumption and sustainability. Science of The Total Environment 723, 137983.

Warner, R.M., 2013. Applied Statistics: From Bivariate through Multivariate Techniques. SAGE Publications, Thousand Oaks.

Yang, X., Wei, J., Ye, G., Zhao, Y., Li, Z., Qiu, G., Li, F., Wei, C., 2020. The correlations among wastewater internal energy, energy consumption and energy recovery/production potentials in wastewater treatment plant: An assessment of the energy balance. Science of The Total Environment 714, 136655.

Yoon, B., Jang, Y.-C., Jung, C., Lee, W., 2013. Covariance fitting of highly-correlated data in lattice QCD. Journal of the Korean Physical Society 63, 145-162.

Yu, Y., Zou, Z., Wang, S., 2019. Statistical regression modeling for energy consumption in wastewater treatment. Journal of Environmental Sciences 75, 201-208.

Zhao, L., Dai, T., Qiao, Z., Sun, P., Hao, J., Yang, Y., 2020. Application of artificial intelligence to wastewater treatment: A bibliometric analysis and systematic review of technology, economy, management, and wastewater reuse. Process Safety and Environmental Protection 133, 169-182.

Żyłka, R., Dąbrowski, W., Malinowski, P., Karolinczak, B., 2020. Modeling of Electric Energy Consumption during Dairy Wastewater Treatment Plant Operation. Energies 13, 3769.