Journal of Environmental Management xxx (2016) 1–10

Contents lists available at ScienceDirect

Journal of Environmental Management

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

Research article

Wastewater treatment aeration process optimization: A data mining approach

Ali Asadi, Anoop Verma*, Kai Yang

Department of Industrial and Systems Engineering, Wayne State University, Detroit, MI 48202, United States

ARTICLE INFO

Article history:
Received 26 April 2016
Received in revised form
13 July 2016
Accepted 14 July 2016
Available online xxx

Keywords:
Data-driven modeling
Energy optimization
Effluents
Aeration process
Data-mining

ABSTRACT

Being water quality oriented, large-scale industries such as wastewater treatment plants tend to overlook potential savings in energy consumption. Wastewater treatment process includes energy intensive equipment such as pumps and blowers to move and treat wastewater. Presently, a data-driven approach has been applied for aeration process modeling and optimization of one large scale wastewater in Midwest. More specifically, aeration process optimization is carried out with an aim to minimize energy usage without sacrificing water quality. Models developed by data mining algorithms are useful in developing a clear and concise relationship among input and output variables. Results indicate that a great deal of saving in energy can be made while keeping the water quality within limit. Limitation of the work is also discussed.

1. Introduction

In order to clean wastewater from certain contaminants, wastewater treatment includes different methods and processes that energy intensive. Across USA, wastewater treatment facilities collect, treat, and release about 4 billion gallons of treated effluent per day from about 26 million homes, businesses, and recreational facilities nationwide (Electric Power Research Institute and Inc. (EPRI, 2002). Such moving and treating processes accounts for more than 4% of the US electricity consumption. Minimizing the energy use of WWTPs by just 10% could lead to an annual savings of $400 million or more (http://water.epa.gov/infr). Due to the environmental regulations, wastewater industries are primarily concerned with water quality. The energy consumption in WWTPs is mainly attributed to their heavy mechanical systems, such as the pump and air support systems which are responsible for moving and treating wastewater (Singh et al., 2012; Zhang et al., 2016). The air support system consists of a group of air blowers that provides oxygen to the aeration tanks for removing organic compounds and converting ammonia. Pump system and the air support system are typically 0.5-MW class mechanical equipment and accounts for more than 70% of the electricity consumption of WWTPs.

Traditionally, WWTP operations and designs are based on kinetic models or simulated data (Flores-Alsina et al., 2008; Sin et al., 2009). While such models have provided promising results, it requires some expert knowledge about different systems and subsystems within the process. Moreover, modeling of such systems heavily depends on the design of WWTPs and hence cannot be efficiently generalized.

In wastewater treatment plants, much effort and money is invested in operating and maintaining dense plant-wide measuring networks which is often untouched. With the proliferation of information technologies (IT), it is now possible to perform long term data archiving for analysis. The steadily growing amount of plant data fosters the avenues for plant wide analysis. Over the past few years, data-mining algorithms have gained tremendous popularity in industrial engineering sector consisting of numerous process and sub-processes. Successful applications of data-mining are visible in many domains such as semiconductor manufacturing (Kusiak, 2000; Tan et al., 2006), fault prognosis and diagnosis (Bae et al., 2003), information retrieval (Seo et al., 2001), transportation systems (Long and Li, 2015; Mashayekhy et al., 2015) and renewable energy (Krioukov, 2011; Lu et al., 2005). Few applications of data-mining algorithms in wastewater treatment industry have also been reported. In this regard, Maurice, et al. (Dixon et al., 2007) implemented a set of data-mining algorithms namely regression,

* Corresponding author.
E-mail address: anoop.verma@wayne.edu (A. Verma).

http://dx.doi.org/10.1016/j.jenvman.2016.07.047
0301-4797/

neural network (NNs), rule induction and visual analysis on TELEMAC project datasets to analyses and predict anaerobic digestion in the wastewater treatment plants. Garcia and Gonzalez (Garcıa and González, 2004) applied self-organized maps (SOM) and *k*-means clustering to develop wastewater supervision techniques in acidic chromic wastewater treatment plant. Researchers applied data mining in specific industrial wastewater process namely alcoholic beverage production (Dixon et al., 2007) and metallic industry wastewater (Garcıa and González, 2004). However, being a rather consistent combination of pollutants, and steady and predictable wastewater production, there approach cannot be generalized to other wastewater treatment plants. Gernaey et al (Gernaey et al., 2004). provided a comprehensive list of white box models for municipal wastewater systems. Authors also suggested using more advance data analytics techniques particularly artificial intelligence (AI) to understand and improve performance of wastewater treatment facilities. For improving the prediction accuracy, Chen and Chang (Chen et al., 2003) developed a hybrid control algorithm combining neural network (NN), Genetic Algorithm (GA) and Fuzzy Logic (FL). The control algorithm developed in their work can be used for optimizing and controlling systems when coping with online upset conditions. While their models provide good results, it demands a field expert to set up the fuzzy rules and also demands the full control of the system to be able to be implemented. Hernandez-del-Olmo (Hernández-del-Olmo et al., 2012; Hernandez-del-Olmo et al., 2012) applied AI techniques in order to improve the performance of wastewater plant. Authors utilized model-free reinforcement learning to minimize operational cost while keeping the quality of water within acceptable level. Despite, their methodology seems promising, the model is tested on simulated data with assumption of 70% sunny, 20% raining and 10% stormy days in a year. Tay and Zhang (Tay and Zhang, 1999) attempted to simulate a lab-scale anaerobic wastewater treatment system utilizing lab scaled wastewater treatment and simulated data. Villez et al. (Villez et al., 2008). used a two stage process to aiming to remove nitrogen and phosphorus from a pilot-scale SBR. Authors applied a multi-way principal component analysis (MPCA) process first to clean the data, and then they utilized LAMBDA based clustering method. Their method claims to converge fast but relies on visual inspection to detect outliers and erroneous data. Later, Verma et al. (Verma et al., 2013; Kusiak et al., 2013), utilized data-mining algorithms to predict total suspended solids and carbonaceous biochemical oxygen demand (CBOD) of an industrial wastewater treatment facility. Kusiak and Wei (Kusiak and Wei, 2012; Wei et al., 2012; Wei and Kusiak, 2015) developed a multi-objective model to optimize the activated sludge process in a WWTP and a significant energy saving was observed.

The literature review above indicates lack of large scale, real studies on plant wide aeration process which do not need the full control of the system in order to be implementable as well as being accurate while keeping analysis understandable and explainable to decision makers. Even the published work in the literature that utilizes real world data has simplified models, i.e. effect of suspended phosphorous and dissolved phosphorous etc. is not analyzed (Kusiak and Wei, 2012; Wei et al., 2012; Wei and Kusiak, 2015). The research developed here aims to bridge the gap in the literature by performing analysis on aeration process of a treatment facility and developing easy to use and implementable data-driven models without scarifying the process accuracy.

The paper is organized as follows. In section 2, the description of the aeration process and related dataset is presented. Section 3 describes the proposed solution methodology along with the formulation of the optimization models. In section 4, results obtained from different optimization models are provided. Finally section 5 concludes the present analysis.

2. Data description

The industrial data used to perform the analysis was obtained from Detroit Water and Sewerage Department (DWSD), located in Detroit, MI. DWSD is the largest single-site wastewater treatment facility in the United States. It serves approximately 35% of the population of the State of Michigan, providing treatment of wastewater. DWSD distribute, treat and collects approximately 1.5 billion Gallons of water and wastewater per day (BGD) to be finally discharged into Detroit River. A generic flow diagram of the wastewater treatment process is shown in Fig. 1.

The collected wastewater enters the plant and passes through bar screens. Large items, such as rags and sticks, are screened out for later disposal. After screening, the influent wastewater enters a wet well and then is pumped to primary clarifiers. After a retention time of 1–2 h, scum floats to the surface where it is removed by a skimmer. Then, the wastewater is delivered by intermediate pumps to adjacent aeration tanks. In each aeration tank pure oxygen is provided by centrifugal blowers through bottom of thank. During normal operations, a required quantity of the sludge from the secondary clarifiers, called Returned Activated Sludge (RSL), enters the aeration tanks through sludge pumps. When the RSL and the wastewater are mixed, microorganisms in the activated sludge use oxygen provided by the fine bubble diffusers located on the bottom of the aeration basins to break down the organic matter. The remaining sludge from the secondary clarifiers and the sludge from the primary clarifiers are either pumped to the anaerobic digesters to produce biogas or fed to the incineration process and the final remaining is transported to the land field. The wastewater then enters cylindrical clarifiers for the secondary treatment. The settled sludge is returned back to the aeration basins for continuous supply of microorganisms. The water after being treated from secondary clarifiers is disinfected through chlorination and then discharged into the River.

The analysis presented here aims to improve the aeration process of the DWSD and hence the corresponding three years' worth of data is collected from the plant. The available data for the analysis was collected for the period of September 2012 to October 2014 (see Table 1). Data includes influent flow rate, influent pollutants, effluent pollutants, and aeration process parameters. The data is recorded at 1 h frequency, out of which two years of data is used for building the models and the last year data is used for model testing and validation. Despite the availability of advanced supervisory control and data acquisitions systems, the archiving of numerous parameters is done manually on a shift by shift basis. This poses issues in data quality, including, missing, and invalid values. In this study, the missing values are imputed based on the values recorded in previous time-periods.

3. Solution methodology

In this section, DWSD data (described earlier) is used to model the aeration process with an aim to optimize water quality and energy consumption. In the analysis, the dissolved oxygen (DO) is used as a controlled variable, whereas, influent flow rate, carbonaceous biochemical oxygen demand (CBOD), total suspended solids (TSS), total dissolved phosphorous (TDP), total suspended phosphorous (TSP) and air flow rate were uncontrollable. Due to strong correlation between Chemical Oxygen Demand (COD) and Biochemical Oxygen Demand (BOD) of municipal wastewater under normal operating condition (excluding big storms and flood), COD is not considered as an independent variable. DO is used as an indicator of energy consumption as most of the energy consumed and associated costs in the aeration process is derived from processes which results in increase DO. These processes may include

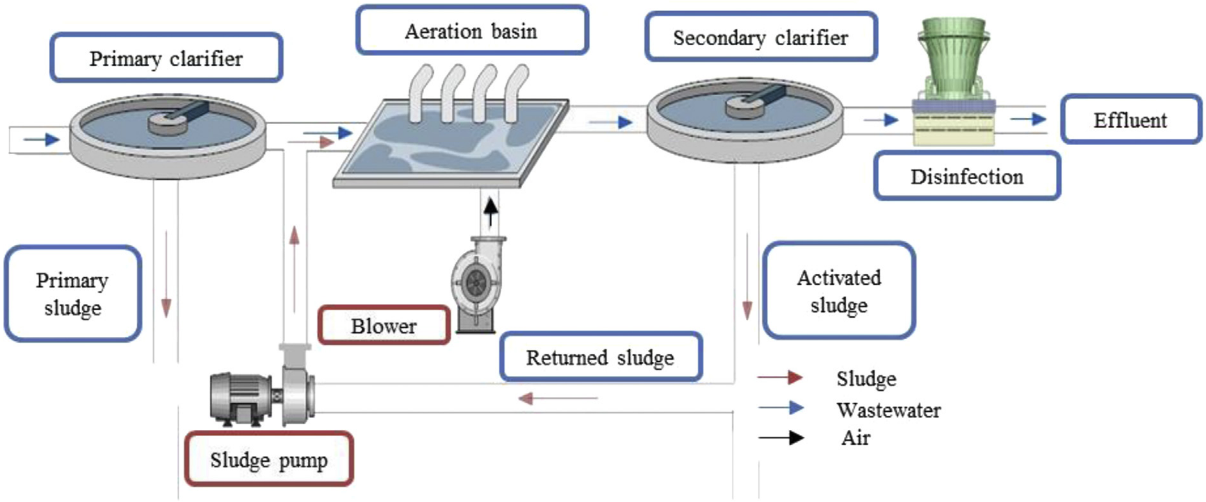

Fig. 1. Wastewater treatment process flow diagram (adopted from (Wei and Kusiak, 2015)).

Table 1
Dataset used in the analysis.

No	Description	No. of instances	Time period
1	Training dataset	Hourly 4368 data points	2012 and 2013
2	Testing dataset	Hourly 2544 data points	2014

generation of oxygen or compressed air and blowing them into the liquid mix therefore in this study DO is used as an indicator of energy consumption. On the other hand, effluent CBOD, effluent TSS, effluent TDP, and effluent TSP are used as an indicator of water quality (a well-known metric). All five metrics (i.e. DO, effluent TSS, effluent CBOD, effluent TDP, and effluent TSP) are minimized by formulating a multi-objective model. Fig. 2 illustrated the salient features of our modular approach. As described in Fig. 2, the process starts with analyzing data and fixing outliers. The initial values are then standardized (0 mean and 1 standard deviation) for algorithms to treat the variables equally. The data mining models are then built on transformed and cleaned data. A ranking approach is used to identify best data-mining algorithm which is fed into the optimization routine. The analysis of numerous energy saving scenarios is then performed. Below sub-sections describes the salient features of our approach in details.

3.1. Feature selection and parameter description

For the analysis, over 35 input parameters were available. In order to reduce the computational load and get better generalized model, only relevant parameters are selected in the modeling process. For the particular task, a boosting tree algorithm is used to evaluate the relative importance of the process variables w.r.t. target output variables. Influent flow rate, returned sludge flow rate, DO concentration, and airflow rate, influent CBOD, effluent TSS, temperature and pH in the aeration tank were among the selected input variables to construct the models of the airflow rate and effluent pollutant concentrations by a data-mining algorithm (Table 2). Errors and outliers in the dataset were removed to improve the accuracy of the model. The airflow rate essentially provides a measure of the energy consumed, which is one of the objectives of this study. With less air flowing into aeration tanks, the quality of the effluent is degraded, which is a matter of concern as it is desirable to maximize the quality of the effluent to meet federal and state requirements. Since effluent CBOD, TSS, TSP, and TDP reflect treatment quality, the objective can be transformed to minimize their concentrations in the effluent. Temperature and pH

Table 2
Description of the variables used in the analysis.

Variable	Description	Unit	Type
u_1	Influent flow rate (IFR)	MGD	uncontrollable
u_2	Returned sludge flow	MGD	Controllable
u_3	Temperature	°C	Uncontrollable
u_4	pH	–	Uncontrollable
y1	Dissolved oxygen (DO)	mg/L	Target
x	Airflow rate (AFR)	scfm	Controllable
y_2	Effluent CBOD (ECBOD)	mg/L	Target
y_3	Effluent TSS (ETSS)	mg/L	Target
y_4	Effluent TDP (ETDP)	mg/L	Target
y_5	Effluent TSP (ETSP)	mg/L	Target

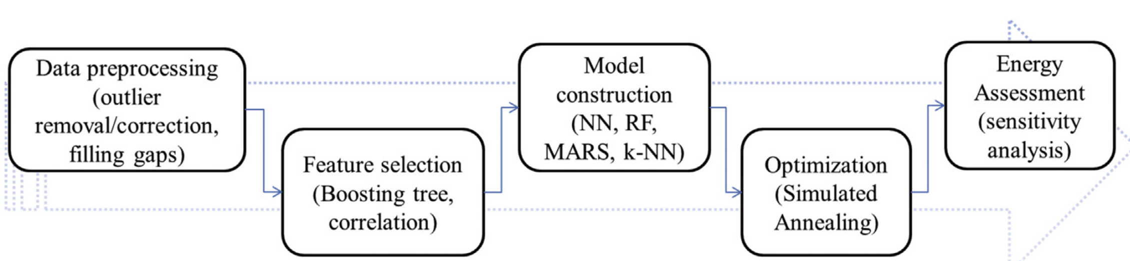

Fig. 2. Overall framework of the developed approach.

are uncontrollable variables which also affect the quality of the effluent.

3.2. Model construction and prediction of effluents

A multi-objective model that minimizes the dissolved oxygen (y_1), effluent COBD (y_2), effluent TSS (y_3), effluent TSS (y_4), effluent TSS (y_5) and is formulated in (1). Since the CBOD and TSS of the effluent were daily data, the hourly data of influent flow rate, returned sludge flow rate, and the concentration of DO was averaged to daily values. A generic multi-objective formulation of the model is given below. The description of the variables in provided in Table 2.

$$F = \min(y1, y2, ...y5) \quad (1)$$

Where:

$$y1, y2, ...y5 = f(u_1, u_2, u_3, u_4, u_{1avg}, x, x_{avg})$$

Due to lack of better physical models, the relationship between input, controllable (u1, u2, …, x) and output variables (y1,y2, …y5) is constructed by data mining algorithms. Four well-known algorithms namely multi-adaptive regression spline (MARS), Artificial Neural Networks (ANN), Random Forest (RF), K-nearest neighbor (k-NN) are employed in the model construction phase. The selected algorithms are known to map the highly non-linear relationship among the input and output variables, such as in wastewater treatment processes. MARS is a non-parametric approach that does not rely on the underlying data distribution and hence suitable for modeling highly non-linear processes such as wastewater treatment process. MARS approximate the functional relationship between input and output from a set of coefficient and basis functions, derived from the regression (Friedman, 1991; Kusiak et al., 2013). Artificial Neural Networks (ANN) is a simple emulation of brain. ANN's parameters are determined by maximum likelihood estimation and minimization of the error function over the training data. (Hsu et al., 1995; Kusiak and Wei, 2012). Random Forest (RF) is an ensemble learning method. RF relies on generation of many classification trees and concluding the result based on the result of all generated classifiers (Liaw and Wiener, 2002; Wei et al., 2012). *K-nearest* neighbor (k-NN) with only one input parameter, namely number of clusters (k) uses data directly to determine class member ship of each data point (Denoeux, 1995). Neighborhoods in this model can provide more information about the reasoning and helps understanding the underling reasoning better than many most of black box models. Most of ensemble learning methods lead to black box model predictions which despite high accuracy are not interpretable with regard to input features; hence k-NN has the advantage in terms of model interpretability comparing to most of black box models. On the other hand, defining the parameter k the hardest part of the modeling and usually is done by trial and error (Wei and Kusiak, 2015). Figs. 3–4 display ECBOD, DO, values predicted using these four algorithms.

As it can be seen from Figs. 3–4, MARS algorithm was able to approximate the trend very well when other algorithms tend to over (NN) or under predict (RF). The four selected data mining algorithms (described in section 3.2) developed predictive models on training data with high correlation coefficient, i.e. 0.7–0.9. To get better assessment of the models developed by four data mining algorithms, Mean Absolute Error (MAE) and Coefficient of determination (R^2) presented in Equations (2) and (3) are used. In Equation (2), $\hat{y}_i$ is the value of observation i predicted by the model, y_i is the actual observed value, and n is the number of total observations in the dataset. Whereas, coefficient of determination measures the proportion of the variance in the target variable that is predictable from input variables (Equation (3)).

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|\hat{y} - y_i| \quad (2)$$

$$R^2 \equiv 1 - \frac{SS_{res}}{SS_{tot}} \quad (3)$$

Table 3 displays the error values obtained by four data mining algorithms when tested on hold out data set for estimating effluent CBOD, TSS, DO, TSP, and TDP. The analyses indicate the performance of MARS is better than other algorithms. Predictions results indicate that most of the time MARS followed by k-NN algorithms were able to approximate the output values. Ranking of the algorithms based on two evaluation criteria is also provided in Table 3.

In general, the models developed by the selected data mining algorithms indicate a good approximation, but mostly fails to predict the peaks and valleys, resulting large error. The error could have been improved by including more high frequency data points. Overall, the models developed by MARS and k-NN found to be more accurate than RF and ANN. However, the models developed by MARS algorithm are used for optimization as they are easy to comprehend. A sample description of DO as a function of input variables is shown in Equation (4). The subscript at the end of variables explains tank 1.

$$\begin{aligned}
DO_1 = {} & 0.534 - 0.919{*}max(0, MLE_1 - 0.469) + 0.751{*}max(0, 0.469 - MLE_1) - 0.096{*}max(0, O2A_1 - 0.350) \\
& -0.153{*}max(0, 0.351 - O2A_1) + 0.143{*}max(0, SVIG_1) - 0.081{*}max(0, IFR_1 - 0.504) \\
& -0.161{*}max(0, 0.504 - IFR_1) - 0.496{*}max(0, 0.780 - RFR_1) - 0.134{*}max(0, CBODD.peas_1) \\
& -0.0738{*}max(0, TPA.peas) - 0.549{*}max(0, CBODD.Primary - 0.663) - 0.109 \\
& {*}max(0, 0.663 - CBODD.Primary) - 0.584{*}max(0, RTR_1 - 0.253) + 2.163{*}max(0, 0.253 - RTR_1) \\
& +0.164{*}max(0, TSSD.peas) - 0.073{*}max(0, TSSD.Primary) + 0.509{*}max(0, O2A_1 - 0.743) \\
& +0.049{*}max(0, MLSS_1) + 1.137{*}max(0, RFR_1 - 0.901) - 1.56{*}max(0, RFR_1 - 0.932) \\
& +0.632{*}max(0, MLE_1 - 0.769) - 0.5{*}max(0, RFR_1 - 0.556) + 1.869{*}max(0, RTR_1 - 0.429) \\
& -0.148{*}max(0, O2A_1 - 0.535) - 0.752{*}max(0, CBODD.Primary - 0.699) - 0.045 \\
& {*}max(0, Influent.PH.O - 0.338) - 0.112{*}max(0, 0.338 - Influent.PH.O) + 0.0473 \\
& {*}max(0, Influent.PH.NI) - 0.1384{*}max(0, DT_1) + 0.2394{*}max(0, RFR_1 - 0.72)
\end{aligned} \quad (4)$$

Based on comparing the algorithms on training dataset, MARS algorithm was selected to perform prediction. An hour (t+1 h) ahead time window is selected for prediction for model validation

Please cite this article in press as: Asadi, A., et al., Wastewater treatment aeration process optimization: A data mining approach, Journal of Environmental Management (2016), http://dx.doi.org/10.1016/j.jenvman.2016.07.047

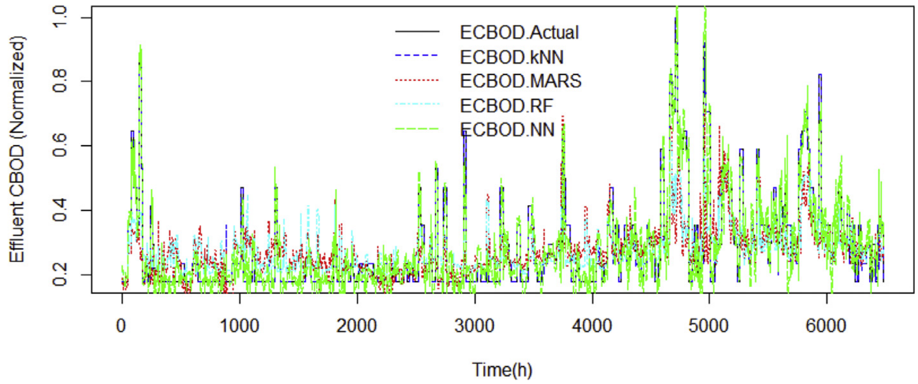

Fig. 3. Data-mining algorithms performance on training dataset (normalized ECBOD).

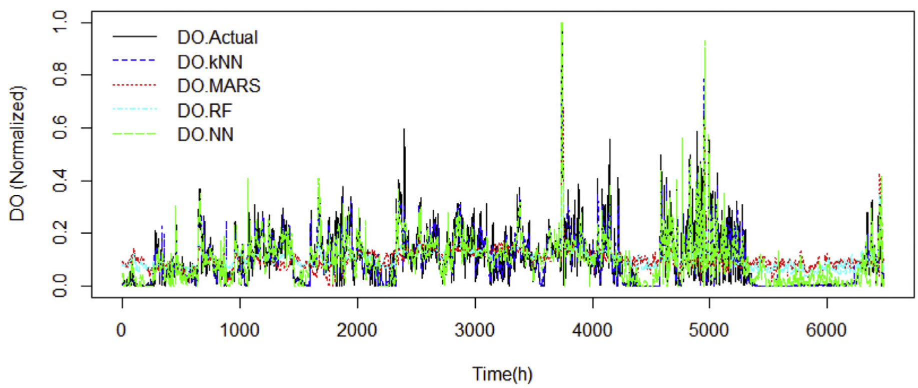

Fig. 4. Data-mining algorithms performance on training dataset (normalized DO).

Table 3
Performance of data-mining algorithms on testing dataset.

Variable	Criteria	MARS	RF	ANN	k-NN
ECBOD	MAE	2.12%	8.42%	7.55%	3.57%
	R^2	93.67%	57.27%	47.4%	94.36%
ETSS	MAE	2.05%	4.64%	4.21%	1.29%
	R^2	92.9%	59.9%	77.5%	98.6%
DO	MAE	2.29%	7.36%	6.51%	4.12%
	R^2	91.9%	46.0%	68.1%	82.8%
ETDP	MAE	2.03%	7.89%	6.50%	2.57%
	R^2	96.9%	29.9%	63.2%	93.3%
ETSP	MAE	2.08%	8.04%	6.54%	2.49%
	R^2	94.8%	62.2%	79.6%	97.5%
Overall Rank		**I**	**IV**	**III**	**II**

as it can provide time for maintenance operator to change control set point as needed. In next section, the underlying optimization model is formulated.

3.3. Aeration process optimization models

In the optimization models, the aim is to minimize the effluents (i.e. CBOD, TSS, TSP, and TDP) concentration to account for water quality, and to minimize the average dissolved oxygen (DO) as a measure to minimize the process energy consumption. The optimization models are subjected to constraints to ensure acceptable water quality, blower capacity, and enough oxygen for mixing. Two optimization scenarios are investigated, viz. (1) optimization w.r.t. permit, and (2) optimization w.r.t. plant current operations. While, the structure of the model stays same, the limits on the effluents vary.

3.3.1. Optimization model 1

Given below is the optimization model formulated with the individual effluents and DO models approximated from MARS algorithms. The constraint limits are obtained from the plant from where the process data was obtained.

$$\Xi = \min(w_1 \times y_1 + w_2 \times y2 + w_3 \times y_3 + w_4 \times y_4 + w_5 \times y_5) \tag{5}$$

Subjected to:

$$0 \le y_1 \le 6.5 \tag{6}$$

$$0 \le y_2 \le 25 \tag{7}$$

$$0 \le y_3 \le 30 \tag{8}$$

$$0.2 \le y_4 \le 1 \tag{9}$$

$$0.2 \le y_5 \le 1 \tag{10}$$

In the equation above y_1–y5 are the objective functions corresponding to dissolved oxygen (DO), effluent CBOD, effluent TSS, effluent TSP, and effluent TDP respectively, whereas, w_1-w5 being the weights associated with individual objectives, where $0 \le$ w1,w2, …w5 ≤ 1. The constraints mentioned in Equation (5)–(9) set limits on underlying objectives. As, the data-points are normalized for each objective, the corresponding constraints for normalized objectives would be:-

Please cite this article in press as: Asadi, A., et al., Wastewater treatment aeration process optimization: A data mining approach, Journal of Environmental Management (2016), http://dx.doi.org/10.1016/j.jenvman.2016.07.047

$$\Xi = \min(w_1 \times y_1 + w_2 \times y2 + w_3 \times y_3 + w_4 \times y_4 + w_5 \times y_5) \tag{11}$$

Subjected to:

$$0 \le y_1 \le 0.21 \tag{12}$$

$$0 \le y_2 \le 1 \tag{13}$$

$$0 \le y_3 \le 0.39 \tag{14}$$

$$0.2 \le y_4 \le 1 \tag{15}$$

$$0.2 \le y_5 \le 1 \tag{16}$$

3.3.2. Optimization model 2

This optimization model is intended to optimize the current operating strategies being followed at the plant. The current operating limits of the plants are obtained from 2 years of data that is used in overall analysis. To be precise, the established water quality limits are further narrowed down and replaced with the plant's recent limits. Mathematically, the optimization model is defined as

$$\Xi = \min(w_1 \times y_1 + w_2 \times y_2 + w_3 \times y_3 + w_4 \times y_4 + w_5 \times y_5) \tag{17}$$

Subjected to:

$$0 \le y_1 \le 6.5 \tag{18}$$

$$0 \le y_2 \le \delta * \sum_{i=1}^{N} y_{2i} \Big/ N \tag{19}$$

$$0 \le y_3 \le \delta * \sum_{i=1}^{N} y_{2i} \Big/ N \tag{20}$$

$$0.2 \le y_4 \le \delta * \sum_{i=1}^{N} y_{2i} \Big/ N \tag{21}$$

$$0.2 \le y_5 \le \delta * \sum_{i=1}^{N} y_{2i} \Big/ N \tag{22}$$

Equations (17)–(22) display the updated optimization model based on the new limits on water quality indicators. δ is a sensitivity parameter which is adjusted to show a tighter or loose bound on water quality indicators. It basically explains if plant can consider some deviation in their current delivery of water quality if savings in energy is possible and vice-versa. $\delta = 0.8$ (10% improvement in current setting, and $\delta = 1.2$ (20% deviation from current control setting) were analyzed along with $\delta = 1.0$ (current setting).

For both optimization models, i.e. model 1 and 2. Two optimization scenarios, namely (1) water quality oriented (referred as *WQ* in subsequent figures), and (2) energy (referred as *Energy* in subsequent figures) oriented were derived by adjusting the weights associated with three objectives. In scenario 1, higher weights are assigned to effluent CBOD, TSS, TDP, and TSP, while weight associated with DO is kept low, i.e. $w_1 = w_2 = 0.4$ and $w_3 = 0.2$. In scenario 2, lower weights are assigned to y_1 and y_2, i.e. $w_1 = w_2 = 0.1$, and $w_3 = 0.8$. The weight values shown here are arbitrary selected and reflect the preference of one optimization criteria over other. The objective functions developed in this work are highly nonlinear (one such mapping is provided in Equation (4)) and therefore exact programming approaches are not suitable. In this work, the developed multi-objective function is optimized using simulated annealing (SA) algorithms. SA optimization is a point based evolutionary local search technique that explores the search space iteratively based on a probability distribution proportion to the temperature. In the developed research, SA explores the search space until no change in the objective function is found for consecutive 300 iterations (Kirkpatrick et al., 1983; Verma et al., 2007). In next section, the results obtained on two scenarios are presented.

4. Optimization results

The objective function developed in previous section is solved for all three scenarios. Due to stochastic nature of simulated annealing algorithm, the optimization routine was run 5 times and the average optimized results are compared against the actual values. Figs. 5–10 display the comparison of scenario 1 and 2 with the observed values. Compared with the observed values, the airflow rate is significantly lower in energy oriented optimization than water quality oriented (Fig. 5). The water quality indicators namely effluent CBOD, effluent TSS, effluent TDP, and effluent TSP are not affected much while optimizing for energy (Figs. 6–9). Average DO was however least affected (Fig. 10). Overall, the reduction in airflow rate is clearly an indication of energy savings (Fig. 5).

Table 4 Compares the optimized output against actual water quality (ECBOD, ETSS, ETDP, and ETSP) and energy indicators (DO), tested against actuals. As optimized-energy scenario focused on minimizing energy consumption, it affects the water quality and vice-versa for optimized-water quality scenario. On an average, energy oriented optimization improves DO concentration (i.e. energy indicators) by 5.4% at the expense of 18% increase in effluent quantity (i.e. water quality indicators). On the other hand, optimized-water quality scenario improves effluents (i.e. water quality indicators by) 16% at the expense of 9.6% increase in DO concentration (i.e. energy indicators).

Results on Table 4 indicate a trade-off between water quality and energy consumption indicators, i.e. improving one indicator reversely affects other. However, when optimizing for the best water quality, it is possible to reduce air flow rate by more than 30% almost without sacrificing water quality which means none of the indicators exceed the permit as defined in Equations (6)–(9). To gain further insights into the operating procedure of the plant, their current control scheme is analyzed for improvement (from optimization model 2). To prove robustness of energy saving methodology as well as providing a guarantee that effluent quality would not be effected under influent fluctuation a sensitivity analysis with 20% wastewater influent increase and decrease has been carried out. The current operational limits obtained from 2 years of data for water quality metrics which are fed to the mathematical formulation and optimized. In this section, the current operating limit of the plants are obtained and improved upon via optimization. Figs. 11 and 12 compares the results obtained using optimization on different sensitivity settings.

Analysis above shows a 31.4% decrease in the energy consumption in terms of aeration oxygen reduction with keeping the effluent water quality almost the same and still better than the standard requirements. A comprehensive sensitivity analysis both reassure robustness of the solution under wastewater input parameter fluctuation and also helps to apportioned effect of

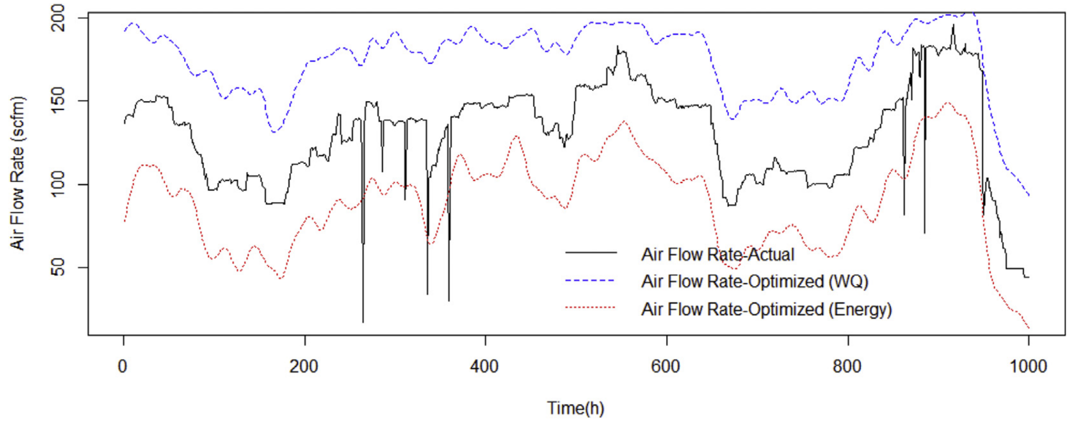

Fig. 5. Comparisons of observed and optimized air flow rate.

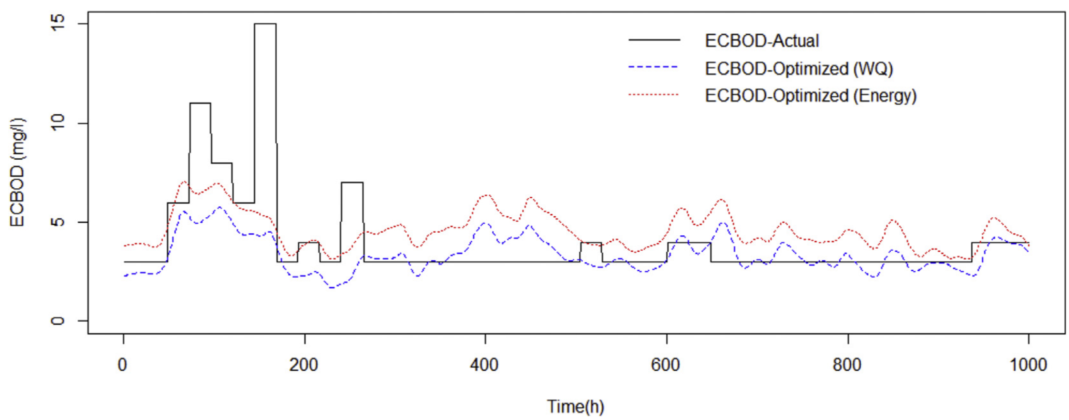

Fig. 6. Comparisons of observed and optimized effluent CBOD.

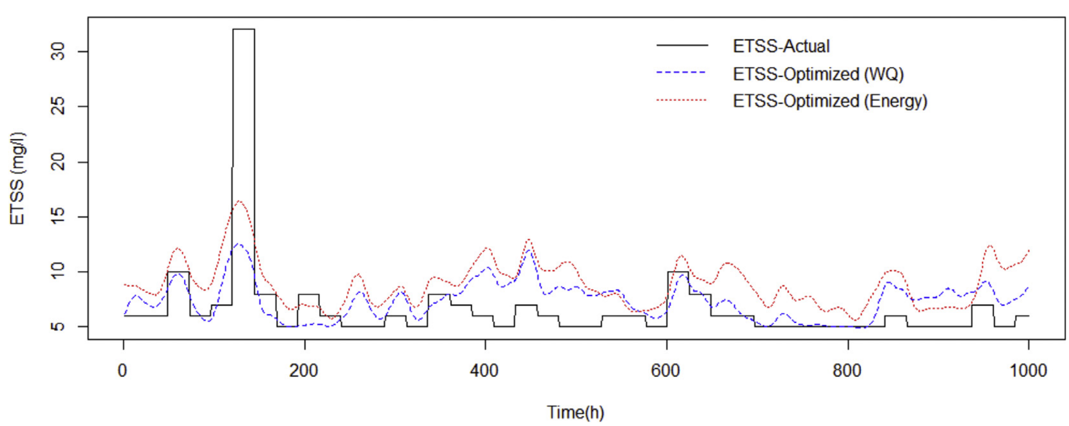

Fig. 7. Comparisons of observed and optimized effluent TSS.

different input variables. We have done sensitivity analysis for TSS, CBOD, DO and Air flow rate. For all four scenarios parameters under energy optimization procedure, and with 1.2 and 0.8 time energy consumption is studied. By bringing the effluent permit down to current operating strategy, a further 12.79% improvement in the TSS quality is possible. Overall, at max15.7% energy improvement could be achieved in 0.8 times energy consumption whereas still minimum 4.46% TSS increase in effluent concentration can save process energy consumption. Bringing the effluents permit down to current operating strategy improve effluent CBOD by 4% while a maximum of 27.5% energy improvement could be achieved in low energy consumption while it would have minimum 3.25% increase in effluent concentration can save process energy consumption. DO sensitivity analysis doesn't indicate big shifts in terms of effluent quality or energy. By bringing the effluents permit down to current operating strategy, only 1.6% improvement in the DO consumption is possible. Maximum 5.4% improvement with developed optimization, whereas, minimum 7.32% increase in DO concentration if

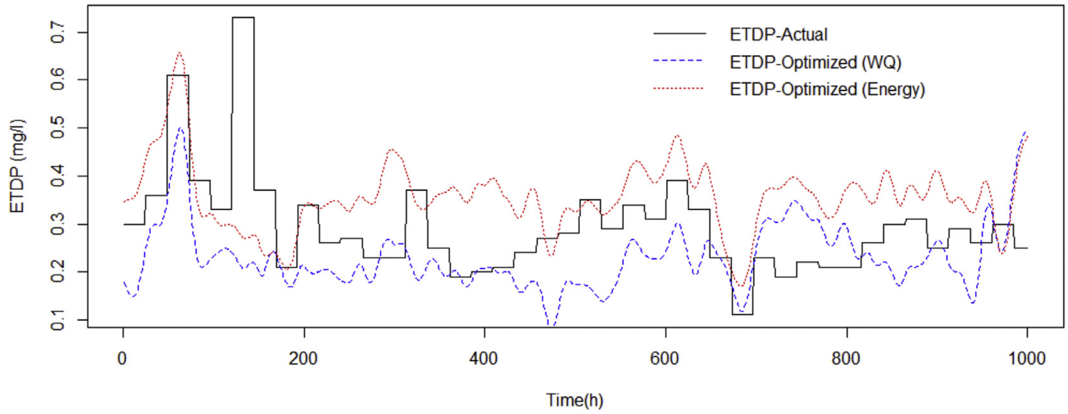

Fig. 8. Comparisons of observed and optimized TDP.

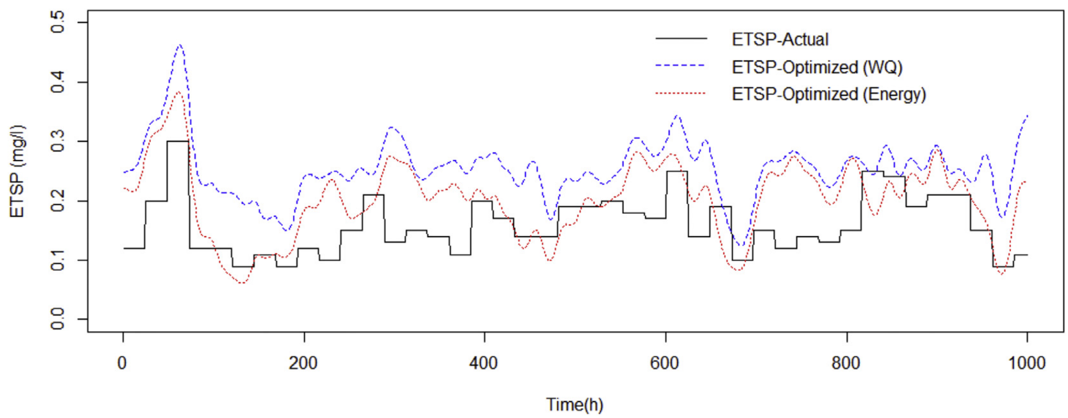

Fig. 9. Comparisons of observed and optimized TSP.

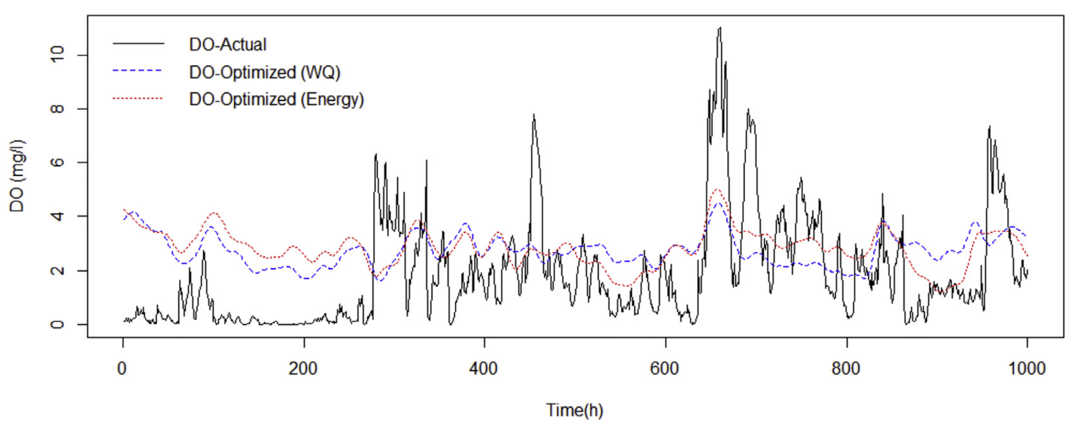

Fig. 10. Comparisons of observed and optimized DO.

Table 4
Percentage improvement in water quality and energy consumption.

Scenario	ETSS	ECBOD	ETDP	ETSP	DO	AFR
Optimized-Energy	−11.1%	−13.7%	−21.7%	−25.8%	5.4%	31.5%
Optimized-Water Quality	9.7%	15.7%	21.1%	−59.8%	−9.6%	−33.7%

water quality is given preference.

In nutshell, the analysis performs here provides an opportunity to plant owners and operators flexibility in selecting different methods to save energy while maintaining water quality. Moreover,

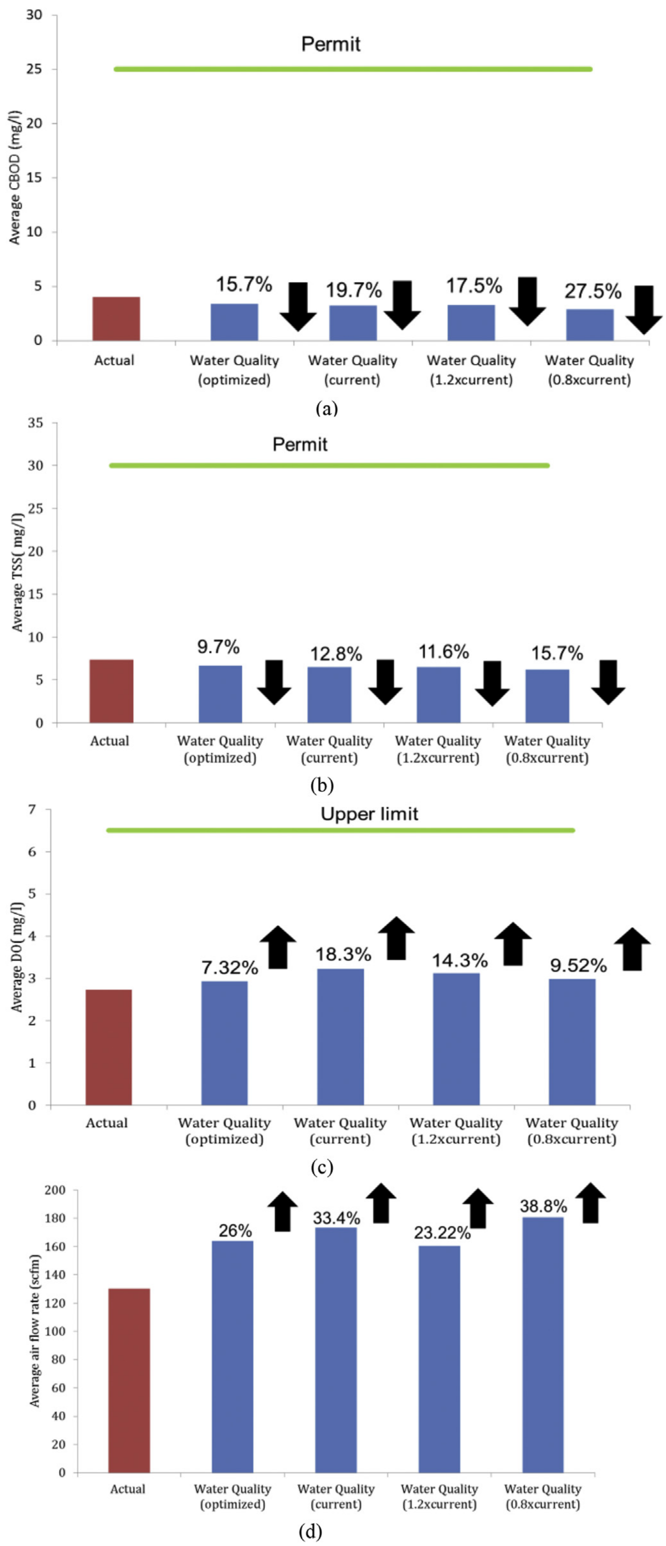

Fig. 11. Results of sensitivity analysis using water quality optimization scenario.

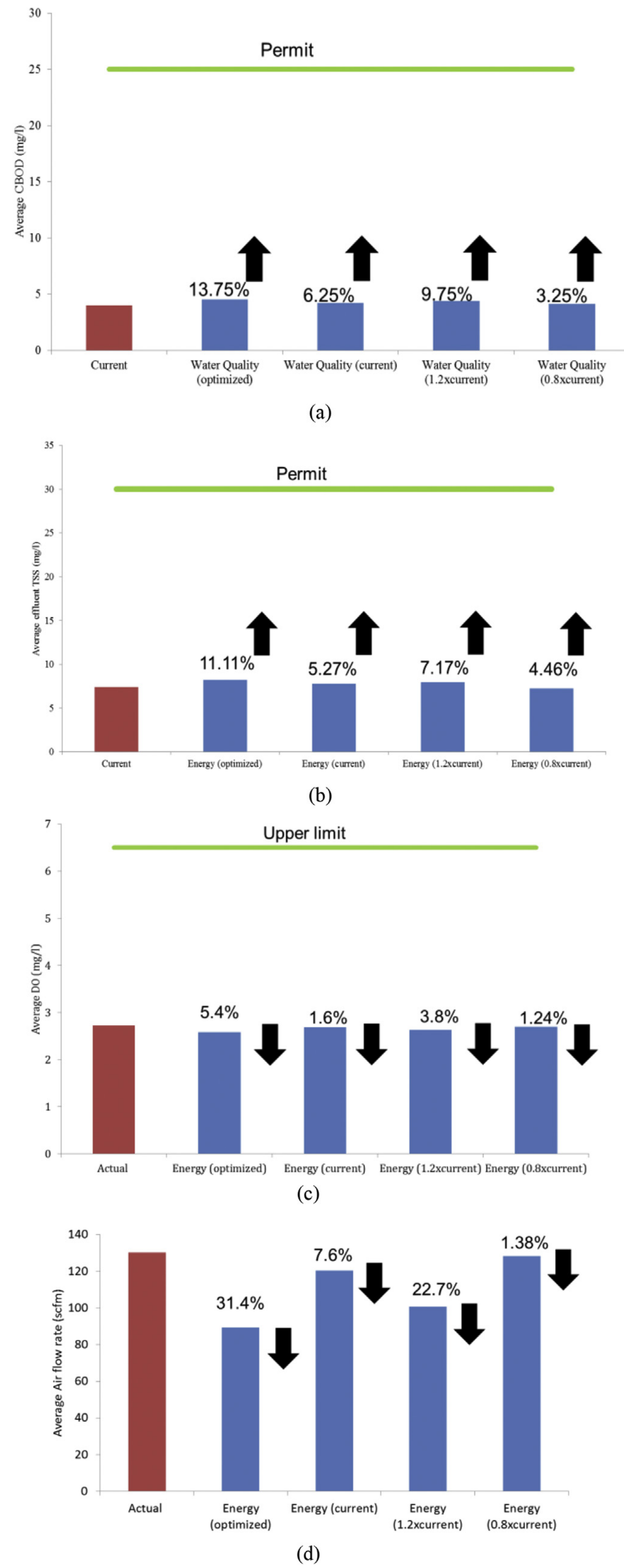

Fig. 12. Results of sensitivity analysis using energy optimization scenario.

the plant operations practice can be improved by optimizing the overall control process.

5. Conclusions

The research performed in the present work was based on 3 years of data collected form DWSD. Of four well known data mining models selected, the models developed by MARS provided the best estimation of effluent CBOD, effluent TSS, average DO, effluent TDP, and effluent TSP. Overall, two optimization models were formulated with different set of control limits. The models were optimized with due consideration of energy and water quality improvements. Results obtained in energy oriented scenario yielded more than 31% reduction in the airflow rate while keeping the water quality within acceptable range. Due to lack of high frequent data such as CBOD, TSS, the optimized results had more noise. For developing a better control system, more frequent sampling of those influent variables is needed.

Acknowledgment

Authors would also like to thank Detroit Water and Sewerage Department (DWSD), Detroit, MI for data access.

References

Bae, H., Kim, S., Kim, Y., Lee, M.H., Woo, K.B., 2003. E-prognosis and diagnosis for process management using data mining and artificial intelligence. In: Proc. Industrial Electronics Conf., Roanoke, VA, vol. 3, pp. 2537–2542.

Chen, W.C., Chang, N.B., Chen, J.C., 2003. Rough set-based hybrid fuzzy-neural controller design for industrial wastewater treatment. Water Res. 37 (1), 95–107.

Denoeux, T., 1995. A k-nearest neighbor classification rule based on Dempster-Shafer theory. Systems, Man and Cybernetics. IEEE Trans. 25 (5), 804–813.

Dixon, M., Gallop, J.R., Lambert, S.C., Healy, J.V., 2007. Experience with data mining for the anaerobic wastewater treatment process. Environ. Model. Softw. 22 (3), 315–322.

Electric Power Research Institute, Inc. (EPRI, 2002. Water & Sustainability (Volume 4): U.S. Electricity Consumption for Water Supply & Treatment - the Next Half Century.

Flores-Alsina, Xavier, et al., 2008. Multi-criteria evaluation of wastewater treatment plant control strategies under uncertainty. Water Res. 42.17, 4485–4497.

Friedman, J.H., 1991. Multivariate adaptive regression splines. Ann. Statistics 19 (1), 1–67.

Garcıa, H.L., González, I.M., 2004. Self-organizing map and clustering for wastewater treatment monitoring. Eng. Appl. Artif. Intell. 17 (3), 215–225.

Gernaey, K.V., van Loosdrecht, M.C., Henze, M., Lind, M., Jørgensen, S.B., 2004. Activated sludge wastewater treatment plant modelling and simulation: state of the art. Environ. Model. Softw. 19 (9), 763–783.

Hernández-del-Olmo, F., Llanes, F.H., Gaudioso, E., 2012. An emergent approach for the control of wastewater treatment plants by means of reinforcement learning techniques. Expert Syst. Appl. 39 (3), 2355–2360.

Hernandez-del-Olmo, F., Gaudioso, E., Nevado, A., 2012. Autonomous adaptive and active tuning up of the dissolved oxygen setpoint in a wastewater treatment plant using reinforcement learning. Systems, Man, and Cybernetics, Part C: applications and Reviews. IEEE Trans. 42 (5), 768–774.

Hsu, K.L., Gupta, H.V., Sorooshian, S., 1995. Artificial neural network modeling of the rainfall-runoff process. Water Resour. Res. 31 (10), 2517–2530.

http://water.epa.gov/infrastructure/sustain/waterefficiency.cfm.

Kirkpatrick, S., Gelatt, C.D., Vecchi, M.P., 1983. Optimization by simulated annealing. Sciences 220 (4598), 671–680.

Krioukov, A., 2011. Integrating renewable energy using data analytics systems: challenges and opportunities. IEEE Data Eng. Bull. 34.1, 3–11.

Kusiak, A., Oct. 2000. Decomposition in data mining: an industrial case study. IEEE Trans. Electron. Packag. Manuf. 23 (4), 345–353.

Kusiak, A., Wei, X., 2012. Optimization of the activated sludge process. J. Energy Eng. 139 (1), 12–17.

Kusiak, A., Verma, A., Wei, X., 2013. A data-mining approach to predict influent quality. Environ. Monit. Assess. 185 (No. 3), 2197–2210.

Liaw, A., Wiener, M., 2002. Classification and regression by random Forest. R. news 2 (3), 18–22.

Long, Y., Li, J., 2015. Application of data mining technology in an integrative intelligent transportation system. ICTE 2109–2114.

Lu, Xin, Dong, Zhao Yang, Li, Xue, 2005. Electricity market price spike forecast with data mining techniques. Electr. Power Syst. Res. 73, 19–29.

Mashayekhy, L., Nejad, M.M., Grosu, D., Vasilakos, A., 2015. An Online Mechanism for Resource Allocation and Pricing in Clouds.

Seo, H., Yang, J., Choi, J., 2001. Building intelligent systems for mining information extraction rules from web pages by using domain knowledge. In: Proceedings IEEE International Symposium Industrial Electronics, Pusan, Korea, 1, pp. 322–327.

Sin, G., Gernaey, K.V., Neumann, M.B., Loosdrecht, M.C.M.V., Gujer, W., 2009. Uncertainty analysis in WWTP model applications: a critical discussion using an example from design. Water Res. 43, 2894–2906.

Singh, P., Carliell-Marquet, Cynthia, Kansal, Arun, 2012. Energy pattern analysis of a wastewater treatment plant. Appl. Water Sci. 2.3, 221–226.

Tan, P.N., Steinbach, M., Kumar, V., 2006. Introduction to Data Mining. Pearson Education/Addison Wesley, Boston.

Tay, J.H., Zhang, X., 1999. Neural fuzzy modeling of anaerobic biological wastewater treatment systems. J. Environ. Eng. 125 (12), 1149–1159.

Verma, A., Shukla, N., Shankar, R., Tiwari, M.K., 2007. "Solving Machine Loading Problem of FMS: an Artificial Intelligence (AI) Based Random Search Optimization Approach," Handbook of Computational Intelligence in Manufacturing and Production Management. Idea Group Publishing, pp. 19–43.

Verma, A., Wei, X., Kusiak, A., 2013. Predicting the total suspended solids in wastewater: a data-mining approach. Eng. Appl. Artif. Intell. 26 (No. 4), 1366–1372.

Villez, K., Ruiz, M., Sin, G., Colomer, J., Rosen, C., Vanrolleghem, P.A., 2008. Combining multiway principal component analysis (MPCA) and clustering for efficient data mining of historical data sets of SBR processes. Water Sci. Technol. 57 (10), 1659–1666.

Wei, X., Kusiak, A., 2015. Short-term prediction of influent flow in wastewater treatment plant. Stoch. Environ. Res. risk Assess. 29 (1), 241–249.

Wei, X., Kusiak, A., Sadat, H.R., 2012. Prediction of influent flow rate: data-mining approach. J. Energy Eng. 139 (2), 118–123.

Zhang, Z., Kusiak, A., Zeng, Y., Wei,, X., 2016. Modeling and optimization of a wastewater pumping system with data-mining methods. Appl. Energy 164, 303–311.