KSCE Journal of Civil Engineering (2015) 19(4):880-886

DOI 10.1007/s12205-014-0059-4

Environmental Engineering

pISSN 1226-7988, eISSN 1976-3808
www.springer.com/12205

Examination of Possible Energy Conservation in a Biological Water Treatment Process using a Multiple Regression Model

Seung-Pil Lee*, Man-Soo Kim, Jin-Sik Kim***, and Ihn-Sup Han******

Received February 3, 2014/Revised April 16, 2014/Accepted May 27, 2014/Published Online December 19, 2014

Abstract

This study examined various controls and operating parameters of a sewage treatment plant by performing multiple regression analyses on the operating parameters and the effluent quality of the sewage treatment plant. For the examination, operating data from April to November, 2012, was collected from a sewage treatment plant using the media anaerobic-anoxic-oxic method. The chemical oxygen demand (COD_{Mn}) and total nitrogen (T-N) forecasting models for the secondary sedimentation basin effluent were built through multiple regression analysis and showed Mean Absolute Percentage Error (MAPE) 6.6-9.3 and 16.0-23.7, respectively. All models showed similar results to real observational data. When controlling the operating parameters of COD_{Mn} and T-N, by using the regression model and the standardization regression coefficients without violating legal water quality criteria, operating conditions can be set that save an average of 23% of power consumption. Using this result, an operating guide for low energy consumption can be provided to the operators of sewage treatment plants.

Keywords: *biological water treatment, optimal control, multiple regression analysis, energy conservation, process diagnosis*

1. Introduction

Sewage treatment is principally a biological process. It is not easy to understand the process behavior because of its complicated reactions, and numerous analyses are needed to know the process state and behavior. Accordingly, at present, most sewage treatment plant operations rely on the empirical knowledge of skilled operators. This is not desirable in terms of the automation of water treatment and the reduction of labor costs. Moreover, although operators' empirical knowledge becomes more reliable as the amount of accumulated knowledge increases, the continuity, objectivity, and available amount of information are not guaranteed. In addition, when operations of the sewage treatment plant rely solely on the operators' empirical knowledge, the objective of energy conservation in the treatment process will be difficult to accomplish.

The Activated Sludge Model No.1 (ASM1) was presented in the IAWPRC (International Association on Water Pollution Research and Control) in 1987 as a method to help examine the sewage treatment process through mathematical calculations without relying on operators' empirical knowledge.(Henze *et al.*, 1987) Since the first ASM1 model was presented, in addition, the ASM2, ASM2d, and ASM3 have been published. These models show realistic and outstanding simulation capabilities. Commercial programs utilizing these models have been developed and are in use; however, because these models are cumbersome in analyses for substance classification and fail to measure some parameters, real time predictions and control of sewage treatment processes are not possible. To solve this problem, various studies have been conducted using the ASM models for model simplifications that help operators easily understand and eliminate the items that are not measurable. Furthermore, to avoid ASMs, some studies are being conducted to develop process controls by using various mathematical and statistical analysis methods that use the field data of sewage treatment plants. Such methods include the Benchmark Simulation Model (Benedetii *et al.*, 2010), ARIMA Model (Dellana and West, 2009), Neural Network Model (Machon *et al.*, 2007), Multi Objective Model (Fu *et al.*, 2008; Hakanen *et al.*, 2013), and Multi Regression Model (Min *et al.*, 2010).

Accordingly, this study built a model that predicts and controls the operating state of the sewage treatment process through multiple regression analysis using operating data that were analyzed daily in a real sewage treatment plant. After analyzing the influence of each operating parameter with this model, a method to determine the order of priority of operating parameters

*Member, Researcher, Environmental Technology Institute, SamchullyEnbio, Seoul 150-093 Korea; Ph.D. Student, School of Environmental Engineering, University of Seoul, Seoul 130-743, Korea (E-mail: ysp8230@empal.com)
**Principal Researcher, Environmental Technology Institute, SamchullyEnbio, Seoul 150-093 Korea (E-mail: lbuddha@hanmail.net)
***Researcher, Environmental Technology Institute, SamchullyEnbio, Seoul 150-093 Korea (E-mail: j1405j@hanmail.net)
****Member, Professor, School of Environmental Engineering, University of Seoul, Seoul 130-743 Korea (Corresponding Author, E-mail: ishan@uos.ac.kr)

related to the discharged water quality for control of the treatment process was examined. Furthermore, by examining the energy consumption levels per major facility, we suggested an economical operating model selection method that both satisfies legal water quality standards and minimizes energy consumption.

2. Study Materials and Methods

2.1 Target Treatment Plant

The target plant of this study was a sewage treatment plant operating the media anaerobic-anoxic-oxic method with a processing capacity of 47,000 m^3/day. The target process of the model was the bioreactor process, and the inflow into the bioreactor to the outflow from the secondary sedimentation basin was examined. Daily data analyzed in the subject plant was used as research data. Data from April to November, 2012, was collected for analysis.

2.2 Multiple Regression Analysis

Regression analysis is a statistical technique to analyze the relationship between a dependent variable (Y) and independent variables (X_1, X_2, X_3, ...). A common form of multiple regression models is expressed in Eq. (1) (Kim, 2008):

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \ldots + \beta_n x_n + \varepsilon_i \quad (1)$$

Here, Y is the Dependent variable, β_i is the Regression

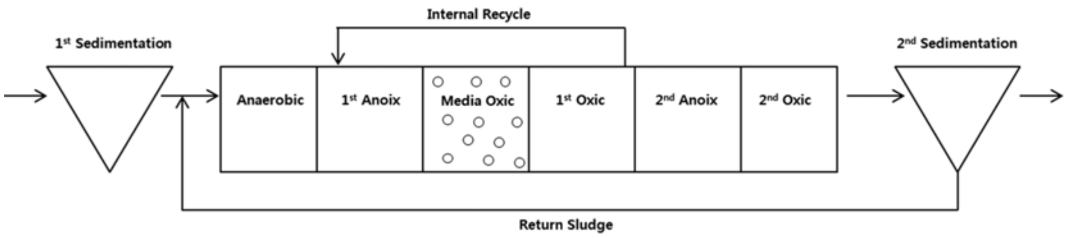

Fig. 1. Process Chart of Sewage Treatment Plant

Table 1. Multiple Regression Variables; IV = Independent Variable, DV = Dependent Variable

Category	Variable		Category	Variable	
Primary Sedimentation basin effluent temperature	R_I_TEMP	IV	1st aerobic tank -DO	AE_T1_DO	IV
Primary Sedimentation basin effluent pH	R_I_PH	IV	1st aerobic tank -MLSS	AE_T1_MLSS	IV
Primary Sedimentation basin effluent BOD_5	R_I_BOD	IV	1st aerobic tank -SVI	AE_T1_SVI	IV
Primary Sedimentation basin effluent COD_{Mn}	R_I_COD	IV	2nd anaerobic tank -temperature	AX_T2_TEMP	IV
Primary Sedimentation basin discharge SS	R_I_SS	IV	2nd anaerobic tank -pH	AX_T2_PH	IV
Primary Sedimentation basin effluent T-N	R_I_TN	IV	2nd anaerobic tank -ORP	AX_T2_ORP	IV
Primary Sedimentation basin effluent T-P	R_I_TP	IV	2nd anaerobic tank -DO	AX_T2_DO	IV
Anaerobic Tank temperature	AN_TEMP	IV	2nd anaerobic tank -MLSS	AX_T2_MLSS	IV
Anaerobic Tank -pH	AN_PH	IV	2nd anaerobic tank -SVI	AX_T2_SVI	IV
Anaerobic Tank -ORP	AN_ORP	IV	2nd aerobic tank –temperature	AE_T2_TEMP	IV
Anaerobic Tank -DO	AN_DO	IV	2nd aerobic tank -pH	AE_T2_PH	IV
Anaerobic Tank -MLSS	AN_MLSS	IV	2nd aerobic tank -ORP	AE_T2_ORP	IV
Anaerobic Tank -SVI	AN_SVI	IV	2nd aerobic tank -DO	AE_T2_DO	IV
1st Anaerobic Tank -Temperature	AX_T1_TEMP	IV	2nd aerobic tank -MLSS	AE_T2_MLSS	IV
1st Anaerobic Tank -pH	AX_T1_PH	IV	2nd aerobic tank -SVI	AE_T2_SVI	IV
1st Anaerobic Tank -ORP	AX_T1_ORP	IV	Hydraulic Retention Time	HRT	IV
1st Anaerobic Tank -DO	AX_T1_DO	IV	Sludge Retention Time	SRT	IV
1st Anaerobic Tank -MLSS	AX_T1_MLSS	IV	F/M ratio	FM	IV
1st Anaerobic Tank -SVI	AX_T1_SVI	IV	C/N ratio	CN	IV
Media aerobic tank-temperature	MO_TEMP	IV	C/P ratio	CP	IV
Media aerobic tank -pH	MO_PH	IV	BOD space loading	BOD_VL	IV
Media aerobic -ORP	MO_ORP	IV	Return sludge quantity	ORSF	IV
Media aerobic -DO	MO_DO	IV	Excess sludge quantity	ESF	IV
Media aerobic -MLSS	MO_MLSS	IV	Raw sludge quantity	EXS	IV
Media aerobic -SVI	MO_SVI	IV	Internal return flow rate	ICF	IV
1st aerobic tank-temperature	AE_T1_TEMP	IV	Secondary sedimentation basin COD_{Mn}	SL_COD	DV
1st aerobic tank -pH	AE_T1_PH	IV	Secondary sedimentation basin T-N	SL_TN	DV
1st aerobic tank -ORP	AE_T1_ORP	IV			

Table 2. Regression Analysis Method for each Case

Category	Variable selection method	Variables	Outliers and influential observations elimination
case 1	Stepwise	All variables	X
case 2			O
case 3	Stepwise	All variables & Essential choice variables must be included	X
case 4			O
case 5	Method not applied	Essential choice variables	X
case 6			O

coefficient, x_i is the Independent variable, and ε_i is the Error term.

In the regression analysis, the variables for which multi-collinearity occurred were eliminated from the model, and outliers and influential observations were examined and considered through residual analysis.

The SAS 9.2 program was used for the multiple regression analysis, and the stepwise method was used for selecting variables. For reviewing models and in the variable-selecting method, the significant level was set as 0.05.

For the multiple regression analysis, 55 variables were set; and among these variables, SL_COD (secondary sedimentation basin effluent, COD_{Mn}) and SL_TN (secondary sedimentation basin effluent, T-N) were selected as dependent variables (Table 1).

2.3 Analytical Method

2.3.1 Regression Analysis

Among the collected data, the data of April-October, 2012, was used to build the model, and the accuracy of the model was examined by using the data of November, 2012.

For the examination of outliers and influential observations, the Studentized Residual and the Difference of Fits (DFFITS) proposed by Belsley *et al.* (1980) were used. When meeting the following conditions, the Studentized Residual and DFFITS were considered to be the outliers and influential observations, respectively (Min *et al.*, 2010; Kim, 2008; Park, 2003; Belsley *et al.,* 1980).

The criteria of DFFITS influential observation :

$$|DFFITS(i)| \geq 2[(k+1)/n]^{1/2} \tag{2}$$

The criteria of Studentized Residual outlier :

$$|\gamma_i^*| \geq t(n-k-2_{;\alpha}/2) \tag{3}$$

Here, k is the number of independent variables, and n is the number of samples.

In the examination of multicollinearity, the variables were selected using the stepwise method, and then the Variance Inflation Factor (VIF) was calculated. Subsequently, in case VIF value was higher than 10, the variables with multicollinearity were eliminated from the model, and then variables were selected again using the stepwise method.

The controllable variables of sewage treatment operations (Sludge Retention Time, return sludge quantity, excess sludge quantity, raw sludge capacity, internal return flow rate, Media aerobic-DO, 1st aerobic tank -DO, and 2nd aerobic tank -DO) and the inflow load of each dependent variable were selected and set as essential choice variables for the regression analysis. Regression analysis was performed by classifying six cases as shown in Table 2.

The RMSE (Root Mean Square Error) and MAPE (Mean Absolute Percentage Error) are measures commonly used to deal with the difference between estimated or predicted values by a model and values observed in the real environment. These measures are suitable for expressing accuracy:

$$RMSE = \sqrt{\sum(\hat{y}_l - y_i)^2/n} \tag{4}$$

Here, y_i is the observation value, and $\hat{y}_l$ is the ith prediction value of the prediction data, and n is the number of data.

Each value of difference is called “residual”. The RMSE is used to synthesize the residuals to one measure. The smaller the RMSE is, the more accurate it is.

$$MAPE = 1/n[\textstyle\sum_{i=1}^{l}(|y_i - \hat{y}_l|)/y_i] \tag{5}$$

Here, y_i is the ith observation value of the prediction data, $\hat{y}_l$ is the ith prediction value of the prediction data, and l is the number of prediction data. Previous studies have interpreted that if MAPE < 10, the prediction is highly accurate; if 10 MAPE < 20, the prediction is good; if 20 ≤ MAPE < 50, the prediction is reasonable; and if MAPE > 50, the prediction is inaccurate (Lewis, 1982).

Standardized regression coefficients are used by Eq. (6):

$$B = b(SD_X/SD_Y) \tag{6}$$

Here, SD_X is the standard deviation of an independent variable, SD_Y is the standard deviation of a dependent variable and b is the non-standardized regression coefficient. Therefore, each standardized regression coefficient represents the unit change of the dependent variable standard deviation against the unit change of the independent variable standard deviation. In other words, through the size of standardization coefficients, which are irrelevant to the measurement unit of variables, the “relative influence” of dependent values on independent values can be identified, thus the relatively more important variables can be selected (Freund and Littell, 2000).

2.3.2 Power Consumption

Important equipments (return sludge pump, excess sludge pump, raw sludge pump, and internal recycle pump) are installed on electricity meter for collecting operation time, operation rate data, and the relevant power usage data. The electricity meter (KEN-900, KDPOWER Co. Ltd.) was used for the measurements; the measurement items included phase voltage, current, active power, reactive power, apparent power, and power factor. The unit power consumption ratio of daily power consumption and daily processing flow rate of the equipment was calculated.

3. Results and Discussion

The multiple regression model for the two dependent variables, SL_COD and SL_TN, showed that in all cases, the result of the F test (significance test of the model) was within 0.05, which is statistically significant.

3.1 Regression Analysis Results

The multiple regression analysis, which was carried out for the six classified cases on each dependent variable, showed that in all cases, the change behaviors of the observation values and the predicted values were similar. Figs. 2 and 3 show the graphs that compare the observation values and the predicted values of each case for each dependent variable.

Evaluation results of the MAPE and RMSE, which can evaluate the explanation ability of the measurement values of the multiple regression model, and selected variables are shown in Table 3.

The cases of each dependent variable showed that most of the MAPE and RMSE values were similar; however, as case 1 and 2 only have 1-4 controllable variables, these were determined difficult to apply in process management and, thus, were excluded from the regression equation selection when progressing.

The MAPE of cases 3, 4, 5, and 6 were reviewed based on the study of Lewis (1982). In the case of SL_COD, the MAPE values were less than 10, which indicate that the prediction was highly accurate. In the case of SL-TN, the MAPE values ranged between 10 and 20, which indicate that the prediction was good. In addition, the RMSE showed similar values in each case for each dependent variable.

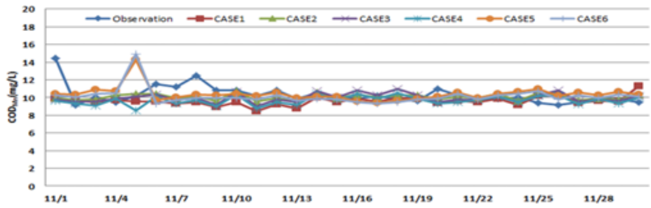

Fig. 2. Predictions of COD for Cases 1-6

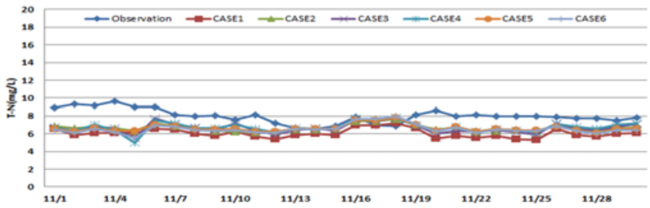

Fig. 3. Predictions of TN for Cases 1-6

3.2 Selection of Multiple Regression Model

As discussed in the previous section, except for cases 1 and 2, which had few controllable variables for each dependent variable, the MAPE and RMSE of cases 3, 4, 5, and 6 had very similar ranges. Standardization regression coefficients were checked to know the influence of operating parameters. The results indicate that the higher the absolute value is, the bigger impact the standardization regression value has on the predicted value (Tables 4 and 5).

The standardization regression coefficient expresses the degree of influence on the predicted value; it means that the larger the ratio of the standardization regression coefficient of variables measured for energy consumption levels among the controllable variables (ICF, EXS, ORSF, and ESF), the greater is the influence of the operator's operation on legal water standards and operating costs. Therefore, the higher the ratio of the standardization regression coefficient of the controllable

Table 3. Regression Analysis Results

Dependent Variable	case	MAPE	RMSE	Number of selected variables (Controllable variable)
SL_COD	case 1	9.2538	1.4047	6(1)
	case 2	6.6151	1.1272	7(2)
	case 3	7.5073	1.2404	15(8)
	case 4	8.5581	1.3446	16(8)
	case 5	8.6781	1.3486	9(8)
	case 6	8.8540	1.4490	9(8)
SL_TN	case 1	23.6863	2.1461	8(4)
	case 2	16.8263	1.5851	5(3)
	case 3	17.1858	1.6829	12(8)
	case 4	15.9980	1.6329	11(8)
	case 5	17.5044	1.6721	9(8)
	case 6	18.7879	1.7948	9(8)

Table 4. Standardization Regression Coefficient of each Dependent Variable (COD)

Rank	case 3		case 4		case 5		case 6	
	Variable	coefficient	Variable	coefficient	Variable	coefficient	Variable	coefficient
1	ICF	0.59188	ICF	0.64717	ICF	0.5286	ICF	0.57749
2	AE_T2_PH	-0.36115	AE_T2_PH	-0.29675	ESF	0.15255	ESF	0.16659
3	HRT	-0.23158	HRT	-0.29385	EXS	-0.13881	R_I_COD	0.13289
4	R_I_TN	0.21336	R_I_TN	0.24951	AE_T1_DO	-0.11601	SRT	0.12809
5	AE_T1_DO	-0.20201	AX_T1_PH	-0.23578	SRT	0.0905	EXS	-0.12747
6	AE_T2_DO	0.19513	AX_T2_DO	-0.18698	R_I_COD	0.0726	AE_T2_DO	-0.12666
7	AX_T2_DO	-0.19428	AE_T1_ORP	-0.18629	MO_DO	0.06236	AE_T1_DO	-0.10604
8	AX_T1_PH	-0.1474	ESF	-0.15882	ORSF	0.00619	MO_DO	0.08915
9	AE_T1_ORP	-0.14319	AE_T1_DO	-0.15683	AE_T2_DO	-0.00131	ORSF	0.02679
10	ESF	-0.0759	AE_T2_DO	0.06261				
11	ORSF	-0.06969	R_I_COD	0.06105				
12	R_I_COD	0.0475	SRT	-0.04622				
13	EXS	0.04319	EXS	0.0424				
14	SRT	-0.00087	ORSF	-0.02086				
15	MO_DO	0.000795	MO_DO	-0.00243				
16			SRT	-0.00699				

Table 5. Standardization Regression Coefficient of each Dependent Variable (TN)

Rank	case 3		case 4		case 5		case 6	
	Variable	coefficient	Variable	coefficient	Variable	coefficient	Variable	coefficient
1	ICF	0.83949	ICF	0.7714	ICF	0.68376	ICF	0.65614
2	ORSF	-0.3151	ORSF	-0.32555	ORSF	-0.23822	EXS	0.27079
3	AX_T2_SVI	0.23845	MO_SVI	0.28365	EXS	0.19692	ORSF	-0.22411
4	EXS	0.23081	EXS	0.25101	AE_T1_DO	-0.13763	R_I_TN	0.13528
5	ESF	-0.21405	ESF	-0.20864	R_I_TN	0.10078	AE_T1_DO	-0.12317
6	R_I_TN	0.19608	R_I_TN	0.15323	MO_DO	0.09656	MO_DO	0.08737
7	R_I_COD	-0.15843	AE_T2_DO	0.13451	ESF	-0.06898	AE_T2_DO	0.08119
8	AE_T2_PH	-0.15649	MO_DO	0.13349	AE_T2_DO	0.04834	ESF	-0.07991
9	AE_T1_DO	-0.15579	AE_T1_DO	-0.13123	SRT	-0.00994	SRT	-0.01879
10	AE_T2_DO	0.11648	R_I_COD	-0.12854				
11	MO_DO	0.05027	SRT	-0.03956				
12	SRT	-0.0365						

Table 6. Among Standardization Regression Coefficients, the Percentage of Variables for which Power Consumption was Measured

Dependent variable	case 3	case 4	case 5	case 6
SL_COD	31.0%	32.7%	70.7%	60.7%
SL_TN	59.1%	60.8%	75.1%	73.4%

variable, the more effective the model is in controlling energy consumption levels (Table 6).

Case 5 was selected for examination of whether the MAPE, RMSE, and controllable variables of each model were to be included or not and how big percentage the variables with measured power consumption take among the standardization regression coefficients. The regression model could be established using the regression coefficients of each variable (Table 7).

One can build multiple regression equations for each dependent variable using the regression coefficients shown in Table 6, and can make up, for each dependent variable, the water quality prediction models according to the changes of independent

Table 7. Regression Coefficient for each Dependent Variable of the Selected Model

Variable	SL_COD (case 5)	SL_TN (case 5)
Intercept	114,662.7	40,389.4
OREF	0.088123	-2.93883
ESF	43.84149	-17.1717
EXS	-89.0715	109.4559
ICF	2.532108	2.837251
MO_DO	3,047.09	4,087.313
AE_T1_DO	-7,035.28	-7,230.39
AE_T2_DO	-134.088	4,289.412
SRT	209.5281	-19.9405
R_I_COD	0.008684	
R_I_TN		0.019002

Table 8. Set Value of Control Parameters for which Power Consumption can be Measured

Internal return flow rate (ICF)		Excess sludge quantity (ESF)		Raw sludge quantity (EXS)		Return sludge quantity (ORSF)	
Condition (Q)	Applied value (m^3/d)	Condition (Q)	Applied value (m^3/d)	Condition (Q)	Applied value (m^3/d)	Condition (Q)	Applied value (m^3/d)
0.2	5,503	0.02	550	0.004	110	0.04	1,101
0.5	13,758	0.1	2,752	0.010	275	0.5	13,758
1.0	27,516	0.15	4,127	0.015	413	1.0	27,516
1.5	41,274	0.2	5,503	0.020	550	1.5	41,274
		0.3	8,255			2.0	55,032

variables, using the regression model.

3.3 Evaluation of the Prediction of and Possibility of Reduction in Power Consumption

In the regression model selected for examining the power consumption and water quality according to changes of current operating conditions for each dependent variable, the independent variables for which power consumption were measured are ORSF, ESF, EXS, and ICF(Table 8).

The actual power consumption of the dependent variable control parameters (ORSF, ESF, EXS, and ICF) was about 1,000 kWh/d; the power consumption of each control parameter ORSF, ESF, EXS, and ICF was 0.017, 0.341, 1.559, and 0.018 kWh/m^3, respectively.

Figure 4 shows the change in water quality against power consumption, which was illustrated by applying the selected case 5 model. The results show that for both dependent variables, the distribution of power consumption was broad even in the same range of water quality. The results also show that even in the case of using lower power than actual consumption levels (about 1,000 kWh/d), water treatment without exceeding legal discharge water quality standards is possible. It is concluded that energy can be conserved if the multiple regression model selected in this study is used in operations of the biological process of the sewage treatment plant from which we gathered the data.

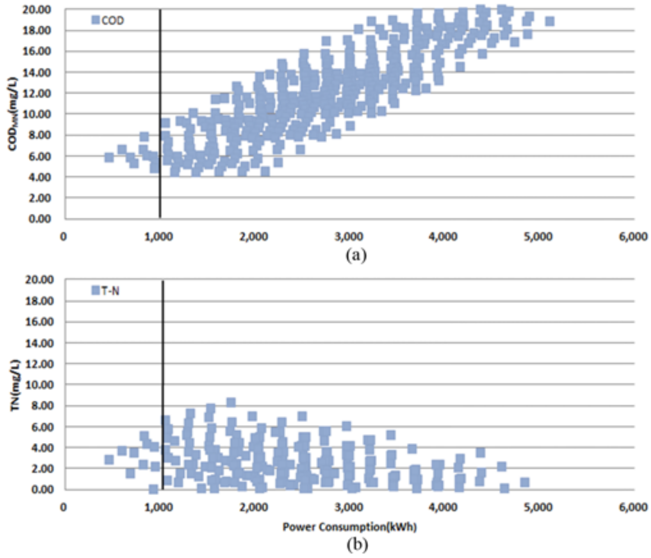

Fig. 4. Water Quality Change Related to Power Consumption: (a) COD_{Mn} Change related to Power Consumption, (b) T-N Change related to Power Consumption

Table 9. Power Consumption and Predicted Water Quality according to Operational Conditions

ICF (Q)	ESF (Q)	EXS (Q)	ORSF (Q)	Energy consumption level (kWh)	SL_COD (mg/L)	SL_TN (mg/L)
0.2	0.02	0.004	0.04	470.33	5.78	2.83
0.5	0.02	0.004	0.04	607.53	5.82	1.48
0.2	0.02	0.010	0.04	727.56	5.25	3.48
0.5	0.02	0.004	0.50	833.14	5.29	2.13
1.0	0.02	0.004	0.04	836.19	4.80	4.03
0.5	0.02	0.010	0.04	864.76	6.54	3.68
0.2	0.02	0.004	1.00	941.16	6.58	2.33
0.2	0.02	0.015	0.04	942.70	6.01	4.33
0.2	0.02	0.001	0.50	953.17	7.81	5.10

Table 10. Comparison of Predicted and Actual Power Consumption (kWh)

Category	Expected energy consumption			Actual energy consumption (average)			Conservation rate (%)
	1 day	1 month	1 year	1day	1 month	1 year	
Average consumption	770.1	23,104.4	281,103.5	1,000	30,000	365,000	23.0
Minimum consumption	470.3	14,110.0	171,672.3				53.0
Maximum consumption	953.2	28,595.0	347,905.3				-

An average energy conservation of 23% was confirmed when computing and estimating energy consumption levels assuming the plant was operated under control parameters lower than the actual energy consumption level evaluated in the selected model (1,000 kWh) (Table 10). Table 9 shows the control parameter conditions and estimated water quality. Table 10 shows the actual and the estimated energy consumption levels.

4. Conclusions

Using multiple regression analysis, a COD_{Mn} and T-N estimation model for the secondary sedimentation basin outflow water were established by considering MAPE, RMSE, the predictability of the behavior change aspect and the inclusion of controllable variables. The result of this attempt showed that all models gave similar results to actual measurements.

Case 5 was selected as the application model after comprehensively evaluating the ratio of variables measured for energy consumption among the standardization regression coefficients. The actual energy consumption level of the four control parameters (ORSF, EFS, EXS, and ICF) measured using electricity meters recorded approximately 1,000 kWh/d, while estimation of the water quality and energy consumption level of each dependent variable using the established multiple regression model showed a wide range between 450-5,000 kWh for similar processed water qualities. Furthermore, when controlling power at a smaller range than the actual energy consumption level, the estimated average energy conservation was 23%.

The selected regression model enables the prediction of the effluent quality of COD_{Mn} and T-N of the secondary sedimentation basin discharge according to operational conditions. If information on the power (energy) consumed by more operational parameters in addition to the four parameters identified in this study can be collected, by using the regression model, the operational power consumption (energy) can be saved through the control of parameters without violating legal water quality criteria of COD and T-N. Using such results, we believe that a low energy consumption operating guide can be provided to the operators of sewage treatment plants.

Acknowledgements

This study was supported by the Korea Ministry of Environment as a “Global Top Project” (Project No. : GT-11-B-02-014-2).

References

Belsley, D. A., Kuh, E., and Welsch, R. E. (1980). *Regression diagnostics: Identifying influential data and sources of collinearity,* John Wiley and Sons, New York.

Benedetti, L., De Baets, B., Nopens, I., and Vanrollenghem, P. A. (2010). “Multi-criteria analysis of wastewater treatment plant design and control scenarios under uncertainty.” *Environmental Modeling and Software*, Vol. 25, No. 5, pp. 616-621, DOI: 10.1016/j.envsoft.2009.03.003.

Dellana, S. A. and West, D. (2009). “Predictive modeling for wastewater applications: Linear and nonlinear approaches.” *Environmental Modeling and Software,* Vol. 24, No. 1, pp. 96-106.

Freund, R. J. and Littell, R. C. (2000). *SAS System for regression (3rd ed.)*, Wiley Inter-Science, SAS Institute.

Fu, G., Butler, D., and Khu, S.-T. (2008). “Multiple objective optimal control of integrated urban wastewater systems.” *Environmental Modeling and Software*, Vol. 23, No. 2, pp. 225-234, DOI: 10.1016/j.envsoft.2007.06.003.

Hakanen, J., Sahlstedt, K., and Miettinen, K. (2013). “Wastewater treatment plant design and operation under multiple conflicting objective functions.” *Environmental Modeling and Software*, Vol. 46, pp. 240-249, DOI: 10.1016/j.envsoft.2013.03.016.

Henze, M., Grady, C. P. L. Jr., Gujer, W., Marais, G. V. R., and Matsuo, T. (1987). *Activated Sludge Model No. 1*, IAWQ Scientific and Technology Report No. 1, London, UK.

Jung, K. M. and Kim, M. G.(2007). *Multivariate analysis*, Kyo Woo Sa, Seoul.

Kim, J. D. (2008). *Linear regression analysis using SAS*, Free Academy, Seoul.

Lewis, C. D. (1982). *Industrial and business forecasting methods: A practical guide to exponential smoothing and curve fitting*, Butterworth Scientific, London.

Machón, I., López, H., Rodriguez-Iglesias, J., Marañón, E., and Vázquez, I. (2007). “Simulation of a coke wastewater nitrification process using a feed-forward neuronal net.” *Environmental Modeling and Software*, Vol. 22, No. 9, pp. 1382-1387, DOI: 10.1016/j.envsoft.2006.10.001.

Min, S.-Y. Lee, S.-P., Kim, J.-S., Park, J.-U., and Kim, M.-S. (2012). “Development and validation of multiple regression models for the prediction of effluent concentration in a sewage treatment process.” *Environ. Eng. Res*., Vol. 34, No. 5, pp. 312-315, DOI: 10.4991.

Park, B. J. (2006). *Theory and application of modern statistics*, Sigma Press, Seoul.