

ORIGINAL ARTICLE

다중회귀분석을 활용한 하수처리시설 에너지 소비량 예측모델 개발

신원재 · 정용준¹⁾ · 김예진^{1)*}

(주)태성종합환경연구소, ¹⁾부산가톨릭대학교 환경공학과

Development of Energy Consumption Estimation Model Using Multiple Regression Analysis

Won-Jae Shin, Yong-Jun Jung¹⁾, Ye-Jin Kim^{1)*}

Tae Sung Environment Institute Co., Ltd., Ulsan 44992, Korea

¹⁾Department of Environmental Engineering, Catholic University of Pusan, Busan 46252, Korea

Abstract

Wastewater treatment plant(WWTP) has been recognized as a high energy consuming plant. Usually many WWTPs has been operated in the excessive operation conditions in order to maintain stable wastewater treatment. The energy required at WWTPs consists of various subparts such as pumping, aeration, and office maintenance. For management of energy comes from process operation, it can be useful to operators to provide some information about energy variations according to the adjustment of operational variables. In this study, multiple regression analysis was used to establish an energy estimation model. The independent variables for estimation energy were selected among operational variables. The R^2 value in the regression analysis appeared 0.68, and performance of the electric power prediction model had less than $\pm 5\%$ error.

Key words : Multiple regression analysis, Energy estimation, Wastewater treatment plant

1. 서 론

하수처리시설은 인간의 활동 및 산업 시스템의 구동 후 발생하는 하수가 주요한 유입 인자이기 때문에, 일반적으로 주민 혐오시설 및 다량의 에너지를 소비하는 시설물로 인식되어 왔다. 하수처리장에 사용되는 전력은 연간 총 전력 사용량의 0.5%를 차지하고 있으며, 이와 같이 다량의 에너지를 소비하는 공공하수처리시설의 전력비는 산업용 전력비가 적용됨에도 불구하고 공공하수처리시설의 20.0%로 인건비(23.9%) 다음으로 큰 비중을 차지하고 있다(2012년 기준, Ministry of Environment,

2013).

환경부는 현재 세계적으로 화두가 되고 있는 고유가 및 기후변화 협약에 대처하기 위해 지난 2010년 “에너지 자립화 기본계획”을 발표 하였다(Ministry of Environment, 2010). 하수처리 과정(소화가스 · 소수력 발전 · 하수열) 및 지역적(풍력 · 태양광 발전)특성을 활용한 신 · 재생에너지 시설의 확대를 통해 현재 공공하수처리 시설의 에너지 자립율(0.8%)을 2030년까지 단계적으로 향상시켜 최종적으로 50% 달성을 목표로 두고, 에너지 소비시설에서 에너지 재생산 시설로의 패러다임 전환을 시도했다. 이와 같은 신 · 재생에너지 기술들은 하수처리

Received 8 September, 2015; Revised 7 October, 2015;

Accepted 8 October, 2015

*Corresponding author : Ye-Jin Kim, Department of Environmental Engineering, Catholic University of Pusan, Busan 46252, Korea
Phone : +82-51-510-0621

E-mail : yjkim@cup.ac.kr

**본 논문은 2013년도 부산대학교의 석사 학위논문의 일부입니다.

© The Korean Environmental Sciences Society. All rights reserved.
© This is an Open-Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

장의 여유 부지 또는 처리 흐름의 수정으로도 충분히 에너지 생산이 가능한 기술들이며, 이미 하수처리장에 이들 신재생에너지 기술을 적용하여 에너지 자립화를 시도한 연구사례가 존재한다(Yoon et al., 2014). 또한 최근에는 고효율 장치를 도입함으로써, 동일한 전력으로도 더 높은 성능을 확보하여 에너지 사용량을 저감할 수 있는 방안도 개발되고 있고(Ovezza, 2009), 하수처리장의 에너지 효율성을 평가하기 위한 방법으로 비방사상 자료포락 분석(Data Envelopment Analysis)과 호기조 시스템의 산소전달 효율에 대한 다양한 방법론들(Hernández-Sancho et al., 2011; Kim et al., 2014; Liu et al., 2011)이 적용되고 있다. 장기적인 측면에서 이와 같은 신재생에너지 기술 및 고효율 장치 도입은 에너지 절감을 위한 효율적인 방법들 중 하나라고 판단할 수 있지만, 상기 언급된 방법들은 시설 설치 및 교체를 위한 초기 투자비가 발생하는 단점이 있으며 하수처리장의 운영과는 전혀 연계되지 않은 새로운 기술들을 도입한 것이

다. 따라서 하수처리장의 효율적인 운영을 통한 에너지 절감 가능성에 대해 고려해 볼 필요가 있다. 이에 본 연구에서는 운전자가 새로운 운전방안을 제시할 경우 그에 따른 에너지 절감량에 대해 제안할 수 있도록 다중 회귀 분석(Multiple regression analysis)을 통한 에너지 소비 상태를 예측할 수 있는 예측모델을 개발하였다.

2. 재료 및 방법

2.1. 대상 플랜트

본 연구는 처리용량 340,000 m³/day인 N 하수처리장 표준화성슬러지 공정을 대상으로 에너지 소비량을 진단 및 평가하기 위해 2008년부터 2010년까지, 20개 항목 528개의 data를 수집하였다. N 하수처리장의 처리공정도는 Fig. 1과 같다.

N 하수처리장의 에너지 소비패턴을 진단하기 위해 하수처리장에서 사용되는 주요 펌프류의 소비전력값을 바탕으로 총 소비되는 에너지 중 부문별 소비되는 에너지

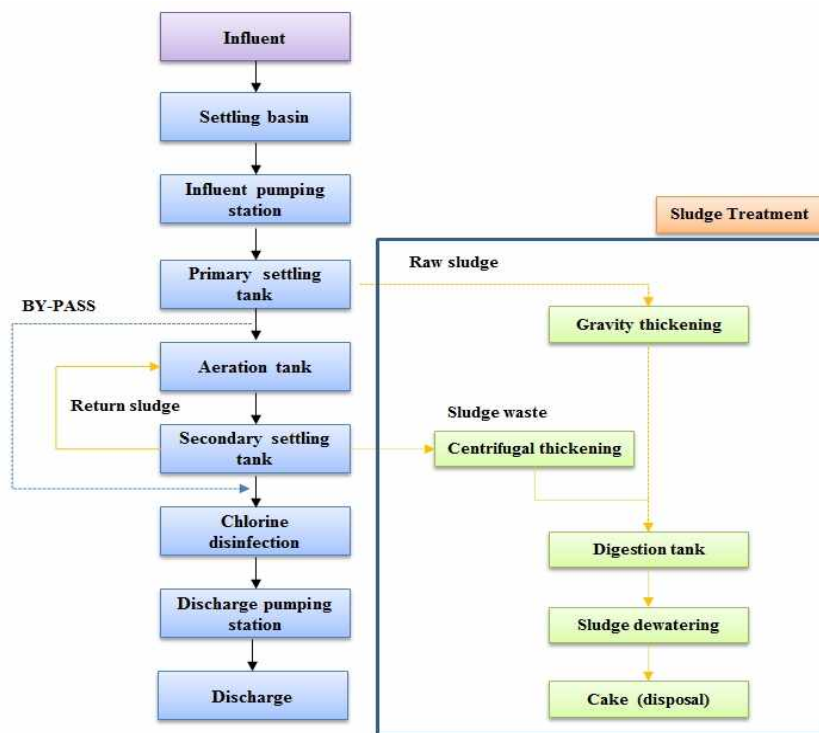


Fig. 1. The Process flow diagram of N WWTP.

양을 계산한 결과, Fig. 2와 같은 비율을 얻을 수 있었다. 일반적으로, 하수처리장의 운영비용 중 큰 비율을 차지하는 전력비는 호기조 송풍기가 가장 큰 요인인 것으로 선행 연구자들에 의해 조사된 바 있으며(KSWW, 2006), Fig. 2.1에서 하수처리장의 기기설비의 전력 소비율을 비교하였다. 이 중 송풍기의 전력 소비율은 37%를 차지하는 것으로 나타났으며, 나머지 부분은 주로 유입 및 방류를 위한 펌프류에 의한 전력 소비율이 높은 것을 알 수 있다. 이와 같이 큰 비율을 차지하는 송풍기에 의한 에너지 소비를 줄이기 위해, 송풍량 제어와 고효율 장비의 도입과 같은 연구들이 현재 활발히 진행되어 오고 있다(Åmand et al., 2012; Ekma et al., 2006; Fernández et al., 2011).

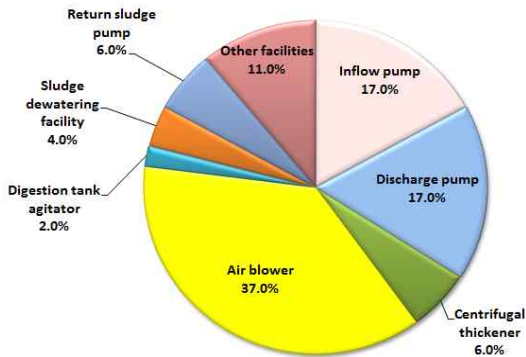


Fig. 2. The ratio of the power consumption of the mechanical equipments in N WWTP.

2.2. 다중회귀분석

회귀 분석(Regression Analysis)은 한 개 또는 그 이상의 독립변수들과 한 개의 종속변수들의 관계를 파악하기 위한 통계적 기법이다. 즉 종속변수에 영향을 미치는 여러 개의 독립변수들을 이용하여 종속변수의 변화를 예측하는 방법으로서 가장 대표적인 종속관계(dependence)에 관한 분석이다. 분석방법은 독립변수와 종속변수 사이에 존재할 법한 아래 식과 같이 표현되는 선형식을 구하여 독립변수들의 값이 주어졌을 때 종속변수의 값을 예측하고, 종속변수에 대한 독립변수의 예측력(영향력)을 분석한다. 공학 분야 뿐 아니라 사회과학 분야에서도 널리 쓰이는 방법론으로서, 환경공학 분야에서는 특정 오염물질의 원인이 되는 여러 가지 변수들을 분석하여 원하는 오염물질의 양을 정량적으로 예측하기 위한 연구 사례들이 다수 존재한다 (Çamdeviren et al., 2005).

$$Y = \beta_0 + \beta_1x_1 + \beta_2x_2 + \beta_3x_3 + \dots + \beta_nx_n + \epsilon_i \quad (1)$$

Y : dependent variable

β_i : partial regression coefficient

x_i : dependent variable

ϵ_i : error

종속변수를 표현하기 위한 독립변수가 하나면 단순회귀분석(Simple regression analysis)이라 하고 위 식과 같이 독립변수가 다수일 경우에는 다중회귀분석(Multiple regression analysis)이라고 한다(Sung, 2007). 다중회

Table 1. Operational conditions of N WWTP

Operational variables	Unit	Operational Conditions ('10.01 ~ '10.12)
Influent flowrate	m ³ /day	311,716
HRT	hr	4.7~8.8
SRT	day	3.9~15.4
BOD space loading	kg-BOD/m ³ · day	0.34~0.84
F/M ratio	kgBOD/kgMLSS · day	0.20~0.53
MLSS (suspended)	mg/L	1,018~1,858
MLVSS	%	764~1,310
Return sludge	Ratio	23.5
	TS	3,488~6,432

귀분석을 실시하기에 앞서 종속변수(Dependent variable)를 선택한 후 분석에 사용되는 독립변수(Independent variable)들을 선정하였다. 변수들의 선정 방법은 독립변수들 중 가장 중요하다고 판단되는 변수부터 모형에 추가 시키면서 더 이상 추가시킬만한 유의한 독립 변수가 없을 때 최종적으로 선택된 독립변수들만을 이용하여 모형을 적합 시키는 전진선택법(forward selection)과, 모든 독립변수들을 포함하는 회귀모형을 만든 후 기여도가 가장 적은 변수부터 차례로 제거시켜 나가는 후진 제거법(backward elimination) 중 후진제거법을 사용하였다.

회귀식의 유효성을 판단하기 위한 지표로서 기여율(결정계수)이 있다. 기여율은 R^2 로 표기하며 1에 가까울수록 우수한 회귀식임을 의미하나, 독립변수의 수가 증가될수록 값이 커지는 경향이 있어, 아래와 같은 식으로 인해 계산되는 수정된 R^2 값을 활용하여 회귀식의 유효성을 판단한다.

$$R^2 = \frac{SSR}{TSS} = 1 - \frac{SSE}{TSS} \quad (2)$$

TSS : Total Sum of Squares

SSR : Sum of Squares due to Regression

SSE : Sum of Squares due to Error

$$\text{수정된 } R^2 = 1 - \frac{n-1}{n-p-1}(1-R^2) \quad (3)$$

n : size of sample

p : number of dependent variables

회귀식이 개발된 후에는, 회귀모델의 유의성을 검정하고 다중공선성을 VIF 계수를 사용하여 확인하였다. 변수들 간에 다중공선성(Multicollinearity)이 존재하는 경우, 다른 독립변수들의 결합으로 영향을 받는 독립변수가 회귀모형 내에서 독자적인 정보를 제공하지 못하게 되므로 R^2 의 값은 큰 반면에 대부분의 회귀계수들이 유의하지 않은 경우가 발생하게 된다. 다중공선성은 독립변수들 사이의 선형종속(Linear Dependency)의 관계로 정의되며, 실질적으로 어느 독립변수가 다른 독립변수들로부터 밀접한 상관관계를 가지고 있는 상태를 말한다. 따라서 VIF 분산팽창인자(Variance inflation factor)의

유무를 확인하는 것이 바람직하며, VIF 값이 10 이상이면 다중공선성이 존재한다고 판단하여 독립변수로 사용하지 않는다(Ham, 2007).

$$VIF = \frac{1}{1-R^2} \quad (4)$$

3. 결과 및 고찰

3.1. 다중회귀식 개발

에너지 소비상태량 예측모델을 개발하기 위해 전력량(종속변수)을 예측하기 위해 데이터 항목들과의 상관관계를 고려한 후, 다중회귀분석을 실시하였다. 독립변수의 선정은 모든 독립변수들을 포함하는 회귀모형을 만든 후 기여도가 가장 적은 변수부터 차례로 제거시켜 나가는 후진제거법을 사용하여 수행되었다. 이에 따라 에너지 소비, 즉 전력량(kwh/d)를 모사하기 위한 다중회귀분석에 들어가는 변수로는 유입유량(Q_{in} , m^3/d), 수온(Temp, $^{\circ}C$), SRT(d), DO(mg/L), 반송슬러지유량(Q_{ras} , m^3/d), 1차 침전조 슬러지폐기유량(Q_{fir} , m^3/d), 2차침전조 슬러지폐기유량(Q_{sec} , m^3/d), 슬러지 원심탈수기 폐기유량(Q_{thic} , m^3/d), 소화조 유입 슬러지유량(Q_{din} , m^3/d)로 선택되었다(Table 2). 이는 단위공정별 전력량 및 단위공정에 설치되어 있는 펌프 전력사용량 등의 기록이 에너지 절감을 위한 핵심변수라는 기존의 연구(Cho et al., 2012)에 의해 선택된 것이다. 상용 통계 소프트웨어인 IBM SPSS 21v.을 사용하여 도출된 회귀식은 아래와 같다.

$$\begin{aligned} y(Elec) = & 52095.194 + (0.057 \times Q_{in}) \\ & + (-104.068 \times Temp) + (-198.114 \times SRT) \\ & + (167.094 \times DO) + (-0.017 \times Q_{ras}) \\ & + (-0.560 \times Q_{fir}) + (0.131 \times Q_{sec}) \\ & + (-0.256 \times Q_{thi}) + (0.483 \times Q_{din}) \end{aligned} \quad (5)$$

전력량을 종속변수로 한 전력량 예측모델의 적용 결과는 Fig. 3과 4에 도시하였다. 도출된 회귀모형에 따른 예측 전력량은 실제 측정값과 0.828의 높은 상관성을 보

Table 2. List of independent variables used for multiple regression analysis

No.	Independent variables	Mean	SD
1	Qin	312,396.30	41,911.22
2	Temp	20.45	4.73
3	SRT	5.33	1.59
4	DO	1.95	0.79
5	Qras	64,666.44	9,485.98
6	Qfir	1,206.11	353.34
7	Qsec	2,496.93	690.61
8	Qthic	275.10	140.38
9	Qdin	1,001.31	248.92

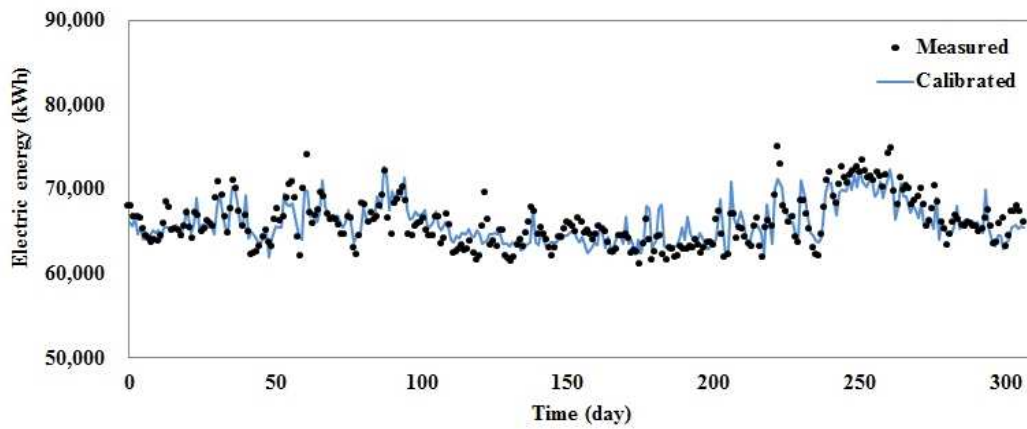


Fig. 3. Calibrated using prediction model of electric energy.

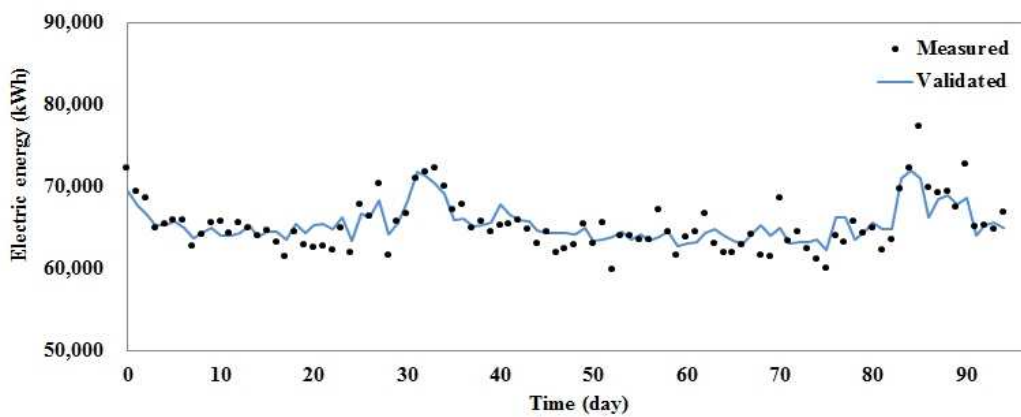


Fig. 4. Validated using prediction model of electric energy.

Table 3. Multiple regression analysis results

R	R ²	Adjusted R ²	Std. Error of the Estimate	Change Statistics				
				R ² Change	F Change	df1	df2	Sig.
.827	.684	.674	1632.45971	.684	71.377	9	2978	.000

Table 4. Result in the analysis of variance

Model	Sum of Squares	df	Mean Square	F	Sig.
Regression	1,711,930,763.889	9	190,214,526.321	71.37707	.000
Residual	791,482,640.671	297	2,664,924.718		
Total	2,503,413,404.560	306			

여주었고, 다중회귀분석에 사용되지 않은 95개의 데이터에 대한 검증 결과에서도 0.833의 높은 상관도를 보였다 (Fig. 4). 또한 예측모델의 정밀도(precision)를 보다 객관적으로 평가하기 위해 실제 전력 사용량과 모델이 예측한 값의 차이를 다룰 때 흔히 사용되는 RMSE(Root Mean Square Error)를 도출하였다. 예측 모델 개발에서 하수처리장에서 사용된 전력량과 모델 시뮬레이션 결과 사이의 RMSE는 약 1,603 kWh/day로 5% 이하의 낮은 오차로 확인되었다. 예측 모델의 검증결과에 대한 RMSE는 1,818 kWh/day로 모델 개발에서의 RMSE 값보다는 높지만 역시 5% 이하의 오차로써 예측 모델이 성공적으로 검증되었음을 확인하였다.

또한 회귀식의 유효성을 판단하기 위한 지표인 수정된 R² 값은 0.674로 계산되었다. 이것은 종속변수 y가 갖는 정보 중 67%를 x의 변동으로 설명할 수 있다는 것을 의미한다. F값은 독립변수가 종속변수 예측에 도움이 되는지 판단하는 값으로, F값이 2 이상이기 때문에 유효한 변수라 판단할 수 있다(Noh, 2007). 유의확률 또한 0.05 이내에 존재하기 때문에 매우 적합한 것으로 나타났다.

3.2. 회귀모형의 유의성 검증

N하수처리장의 에너지 소비량을 예측하기 위해 개발된 회귀모형의 유의성을 분산분석을 활용하여 검증한 결과, 아래 표와 같은 결과를 도출하였다. 검증될 가설은

Table 5. VIF values for determining the multicollinearity of the independent variables

Model	Unstandardized Coefficients		Colinearity Statistics	
	B	Std. Error	Tolerance	VIF
(Constant)	52,095.194	1,414.011		
Qin	.057	.003	.500	1.998
Temp	-104.068	32.269	.374	2.676
SRT	-198.114	77.338	.573	1.746
DO	167.094	150.689	.611	1.638
Qras	-.017	.012	.707	1.414
Qfir	-.560	.312	.716	1.397
Qsec	.131	.153	.779	1.283
Qthic	-.258	.832	.638	1.567
Qdin	.483	.464	.653	1.530

“회귀식은 의미가 없다”는 가설로 아래와 같이 표현할 수 있고 분산분석표의 “유의확률수치”에 주목하여 판정할 수 있다(Park, 2009). “회귀식에는 의미가 없다”는 귀무가설과 “회귀식에는 의미가 있다”라는 대립가설을 세우고, 유의수준 α 는 일반적으로 0.05로 설정된다($\alpha=0.05$). Table 4에서도 보이는 바와 같이, 개발된 회귀모형의 유의확률이 0.000 으로 귀무가설을 기각하므로 의미가 있다고 말할 수 있다.

3.3. 독립변수들의 다중공선성 검토

개발된 회귀식에 포함된 독립변수들의 VIF 값은 통계 소프트웨어 SPSS를 이용해 구하였으며 Table 5에 나타내었다. 본 연구에 사용된 독립변수의 VIF 값은 수식(6)과 같은 방법으로 계산되어 1.283~2.676의 값을 나타냈다. 일반적인 VIF 값의 범위인 10이내를 만족하였고, 공차한계(Tolerance)가 0.1 이상이므로 다중공선성의 영향은 거의 없다고 판단된다.

4. 결론

하수처리시설의 에너지 소비량을 예측하기 위한 모델을 다중회귀분석을 사용하여 개발하였다. 20개의 주요 운전변수 중 후진제거법을 사용하여 다중회귀모델에 사용된 변수는 폭기조의 폭기량과 각종 펌프류에 관한 변수들이었다. 개발된 회귀모델은 수정된 R^2 값이 0.68로서, 일반적으로 다중회귀식의 적합성 보증으로 거론되는 계수인 0.7보다는 낮으나 회귀식의 유의성이나 선택된 변수의 다중공선성의 검토 결과 무리가 없는 것으로 판단된다. 더군다나 회귀모델을 생성하기 위해 사용되었던 데이터셋에 관해서는 0.828의 상관계수로서 정확도를 보였고, 검증을 위한 데이터셋에 적용한 결과 0.833의 정확도를 보였다. RMSE의 값 또한 5%의 오차를 가짐을 보여주었다. 본 연구에서 개발된 다중회귀모델을 사용하여 하수처리장 운전자는 호기조 내 DO 농도의 조절이나 반송슬러지, 폐기슬러지의 유량이 조절될 경우 전력량은 어떻게 변동될 것인가를 추산해 볼 수 있다.

REFERENCE

Åmand, L., and Carlsson, B., 2012, Optimal aeration control in a nitrifying activated sludge process, *Water*

Research, 46, 2101-2110.

Cho, E. S., Han, D. H., Ha, J. S., 2012, Energy efficiency evaluation of publicly owned wastewater utilities, *Journal of Environmental Policy*, 11(4), 85-105.

Çamdeviren, H., Nilsun, D., Arzu, K., Keskin, S., 2005, Use of principal component scores in multiple linear regression models for prediction of Chlorophyll-a in reservoirs. *Ecological Modelling*, 181, 581 - 589.

Ekman, M., Bjorlenius, B., Andersson, M., 2006, Control of the aeration volume in an activated sludge process using supervisory control strategies. *Water Research*, 40, 1668-1676.

Fernández, F. J., Castro, M. C., Rodrigo, M. A., and Cañizares, P., 2011, Reduction of aeration costs by tuning a multi-set point on/off controller: A case study, *Control Engineering Practice*, 19(10), 1231-1237.

Ham, H. B., 2007, Data analysis and SAS programming, 328.

Hernández-Sancho, F., Molinos-Senante, M., and Salazar-Garrido, R., 2011, Energy efficiency in Spanish waste-water treatment plants: A non-radial DEA approach. *Science of the Total Environment*, 409(14), 2693-2699.

Kim, M. H., Ji, S. H., Chang, J. H., 2014, A study on energy saving effect from automatic control of air flowrate and estimation of optimal DO concentration in oxic reactor of wastewater treatment plant, *Journal of Energy Engineering*, 23(2), 49-56.

Korean Society of Water and Wastewater, 2006, Research on energy saving strategies at public sewerage system.

Liu, C., Li, S., and Zhang, F., 2011, The oxygen transfer efficiency and economic cost analysis of aeration system in municipal wastewater treatment plant. *Energy Procedia*. 5., 5, 2437-2443.

Ministry of Environment, 2010, Basic plan for energy saving at public wastewater treatment plants.

Ministry of Environment, 2013, Report on the management and operation of public wastewater treatment plants.

Noh, H., 2005, Theory and application of multivariate statistical analysis using Excel and SPSS, Hyungsul, 244-261.

Ovezza, A., 2009, Saving energy: Using fine bubble diffusers. Saving energy: Using fine bubble diffusers, *Filtration & Separation*, 46(1), 21-27.

Park, S. H., Cho, S. S., Kim, S. S., 2009, Understanding

- and application of SPSS 17.0, 261-270.
- Sung, T. J., 2007, Easy statistics using SPSS/AMOS, 263-265.
- Yoon, J. W., Kim, C. Y., Choi, C. K., 2014, Case studies of energy-saving method for renewable energy installation in sewage treatment plant, Journal of Korean institute of illuminating and electrical installation engineers, 28(4), 42-48.