Journal of Hazardous Materials 183 (2010) 441–447

Contents lists available at ScienceDirect

Journal of Hazardous Materials

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

# A new statistical framework for parameter subset selection and optimal parameter estimation in the activated sludge model

Y.S. Kim, M.H. Kim, C.K. Yoo*

*Center for Environmental Studies/Green Energy Center, Dept. of Environmental Science and Engineering, College of Engineering, Kyung Hee University, Seocheon-dong 1, Giheung-gu, Yongin-Si, Gyeonggi-Do 446-701, Republic of Korea*

ARTICLE INFO

*Article history:*
Received 26 January 2010
Received in revised form 7 June 2010
Accepted 10 July 2010
Available online 17 July 2010

*Keywords:*
Activated sludge model number 1 (ASM 1)
Sensitivity matrix
Similarity measure
Hierarchical clustering
Response surface method (RSM)

ABSTRACT

A new model-calibration method has been proposed to solve the problems associated with parameter subset selection and parameter estimation of the activated sludge model (ASM). We propose the use of a statistical methodology for reasonable parameter selection and parameter estimation that consists of sensitivity analysis, similarity measures, hierarchical clustering and response surface methods (RSM). The introduction of effluent quality index (EQI) can reduce all of the outputs of the ASM model into one factor. The EQI was used to calculate a sensitivity matrix. Then, the hierarchical clustering algorithm was used for parameter subset selection. This selection was based on a similarity measure using the sensitivity matrix and was used to reduce the number of model parameters by selecting only one parameter per cluster group (parameter subset selection step). Lastly, a RSM analysis was conducted in order to determine the optimal parameter values. This study was conducted in order to develop a new statistical framework that can greatly reduce the computational effort required to find the optimal solution by reducing the number of parameters. The experimental results indicated that the calibrated model can improve the prediction quality of the ASM model and the efficiency of the modeling.



## 1. Introduction

The activated sludge model number 1 (ASM 1) was developed by the International Association on Water Quality (IAWQ), and is known as one of the most effective models for the chemical oxygen demand (COD) oxidation, nitrification and denitrification processes. Dynamic simulation, process design, and control and process optimization are carried out for these processes in wastewater treatment plants (WWTPs) [1,2]. The modeling and calibration methods of the ASM model were developed to design and analyze the processes for biological wastewater treatment. The quality of the prediction, design and control performances of the ASMs depends on the methods which are used to calibrate their parameters. The default parameters must be calibrated for each individual WWTP, since the model parameters are not considered as universal for all of the activated sludge systems. Several factors contribute to the need for adjusting the model parameter calibration, including differences in plant operation types, biological kinetics, and the influent and sludge characteristics of the treatment plants [3]. The significant drawbacks of the ASM calibration include its large size and the complex optimization problem. In addition, these models are usually over-parameterized considering the limited availability of data for most WWTPs [4–6].

The use of advanced model calibration techniques, including the STOWA [7], BIOMATH [8], Water Environment Research Foundation (WERF) [9] and HSG [10] protocols, has been recommended in order to solve the problems associated with ASM calibration. These techniques are used to select important parameters and estimate those for the plants. The four calibration methods have their own advantages and disadvantages with regard to the applicability, usefulness, accuracy, cost, experimental work and chemical/biological aspects of the model calibration. Recently, Sin et al. [11,12] critically compared and reviewed the above four calibration protocols using the strengths, weakness, opportunities and threats (SWOT) analysis method. Four protocols may have similarities as well as differences, such as definition of modeling goal, data confidences, design of the measurement campaign, influent characterization, kinetic parameter estimation, and a selection of parameter subset. The differences between the four calibration methods used in Sin et al. [12] and the proposed method of this study is to focus on an automatic determination of representative parameter subset for an over-parameterization problem and then successive optimal parameter estimation by a statistical multiple optimization. An application point of view, the proposed method is more near to pragmatic calibration method than a scientific approach, which is useful in a ASM model calibration in a full-scale plant without an expert knowledge.

* Corresponding author. Tel.: +82 31 201 3824; fax: +82 31 202 8854.
*E-mail addresses:* ckyoo@khu.ac.kr, ChangKyoo.Yoo@biomath.ugent.be (C.K. Yoo).

0304-3894/$ – see front matter © 2010 Elsevier B.V. All rights reserved.
doi:10.1016/j.jhazmat.2010.07.044

There are several problems with the model calibration of the ASM families including identifiability, over-parameterization and sparsity. For example, if the parameters are not identifiable then even a low level of noise in the data will result in large variations in the parameters' estimated values, and therefore the parameters cannot be estimated accurately [13]. The identifiability of a set of parameters is determined by the effect of changes in the value of parameter on the value of output [13]. This effect is represented by sensitivity vectors. Several methods for parameter selection based on sensitivity vectors have been proposed in the literature [3,5,11]. Since the experiments are expensive to obtain the concentration of influent for the ASM model at the plants, it is limited the number of data sets available. To overcome this problem, a numerical method for parameter selection was used in this study. First, the number of parameters to be considered is reduced by grouping parameters that could be estimated together. Then, the model calibration is carried out by considering only one parameter per group for the parameter subset selection procedure.

The purpose of this study was to develop a statistical framework for ASM model calibration that can sort ASM model parameters into reasonable groups using parameter subset selection and then estimate the representative parameters using a multiple response surface method for optimal parameter estimation. One parameter per group is considered for the parameter set selection, which was reduced the size of the combinatorial problem resulting from a large number of parameters. The proposed new methodology will find a simultaneous solution to the parameter subset selection and optimal parameter estimation problems with a reduced computational load.

## 2. Materials and methods

Successful parameter estimation depends on parameter identifiability that can be determined either analytically or numerically. Analytical identifiability determines the uniqueness of the solution derived from parameter estimation while numerical identifiability focuses on the stability of the solution [13]. A sensitivity matrix for output can be used to find the likelihood of the parameter identifiability [6,13].

### 2.1. Sensitivity analysis

The optimization of the ASM is problematic due to the model complexity caused by many components as well as kinetic and stoichiometric parameters. As it is difficult to consider all of the parameters at the same time when predicting the results [14], sensitivity analysis is performed to select the key parameters which are influenced by the removal efficiency. This means helps to improve the prediction efficiency of the model.

The identifiability of a set of parameters is determined by sensitivity analysis that is the effect of change in the value of a parameter on the value of output [6]. The relative sensitivity of each parameter, $j$, to each model output, $y$, and at each time instant, $I(\mathbf{S}_{ij})$, can be calculated as:

$$\mathbf{S}_{ij} = \left(\frac{\theta_j}{y_j}\right)\left(\frac{\partial y_i}{\partial \theta_j}\right) \tag{1}$$

where $\partial y_i/\partial\theta_j$ is defined as the absolute sensitivity of the model output $y$ to the parameter $\theta_j$ at each time instant. The absolute sensitivity function can be approximated using a finite difference method. This method is valid only for small changes in the values of parameters, i.e. a small perturbation factor ($\Delta\theta$) [6]. Parameters with high relative sensitivity are considered to be the most likely parameters, to have a strong effect that is minimized the modeling error.

### 2.2. Similarity measures

Brun et al. [3] proposed a similarity measure for classifying a set of parameters which have at least one common property among the ASM parameters. However, when the sensitivity vectors of the output and that of the parameter are parallel to each other, the sensitivity measure and collinearity index proposed by Brun et al. [3] cannot be used. In this study, a similarity measure approach proposed by Chu and Hahn [13] was used to analyze the relationship between input and output of the ASM parameters. This approach does not require that the sensitivity matrix is to be nonparallel, but the angle between the sensitivity vectors should be small. The measure of the effect of two parameters on the output can be defined as:

$$\cos\phi_{ik} = \frac{|\mathbf{S}_i^T\mathbf{S}_k|}{||\mathbf{S}_i||_2||\mathbf{S}_k||_2} \tag{2}$$

where $\phi_{ik} \in [0,2\pi]$ is the angle between the relative sensitivity vectors $\mathbf{S}_i$ and $\mathbf{S}_k$. The term $|\mathbf{S}_i^T\mathbf{S}_k|$ is an inner-product of relative sensitivity vectors and $||\mathbf{S}_i||$ is the Euclidean norm of the $i$th column of relative sensitivity vector $\mathbf{S}$ [13]. The value of the similarity measure can range from 0 to 1. If the value is determined to be 0, then the parameters are orthogonal and they have a distinct effect on the output. The parameters can be clustered into groups based on the values of similarity measures.

### 2.3. Hierarchical clustering

There are two types of methods used for hierarchical clustering: the first method is to select the two objects which are all most similar and group them. The other method selects two objects which are all most dissimilar and divides them [15]. The methods available in the literature can be divided into many types based on the degree of similarity between the two clusters. These include the single linkage method, complete linkage method, average linkage method, centroid linkage method, and so on. In addition to the above methods, bottom-up and top-down methods are also used in hierarchical clustering. A hierarchical order of repeated division is found when all of data objects are grouped into the top level that consists of only one cluster [15].

### 2.4. Response surface method

The response surface method (RSM) is used to find the relationship between one or more response variables and a set of quantitative, experimental variables or factors. This method is often employed after a few controllable factors that are determined to be vital have been identified and the factor settings that optimize the response have been found. This method is usually chosen for an experiment when the response surface is expected to have some curvatures in it [16]. In generally, the RSM is defined as the independent variable or operating condition that optimizes any response variable. The optimal solution should be obtained experimentally. The relationship between the independent variables $(x_1, x_2, \ldots, x_k)$ and the response can be described using the following equation:

$$y = f(x_1, x_2, \ldots, x_k) + \varepsilon \tag{3}$$

where $y$ is the response of the system, $f$ is the unknown response function, $x_1, x_2, \ldots, x_k$ are the independent variables, $k$ is the number of variables and $\varepsilon$ is the modeling error. In most cases, after completion of the experimental design, the mathematical model can be written as a second-order regression of response surface model using Eq. (4):

$$y = \beta_0 + \sum \beta_i x_i + \sum \beta_{ii} x_i^2 + \sum \beta_{ij} x_i x_j + \varepsilon \tag{4}$$

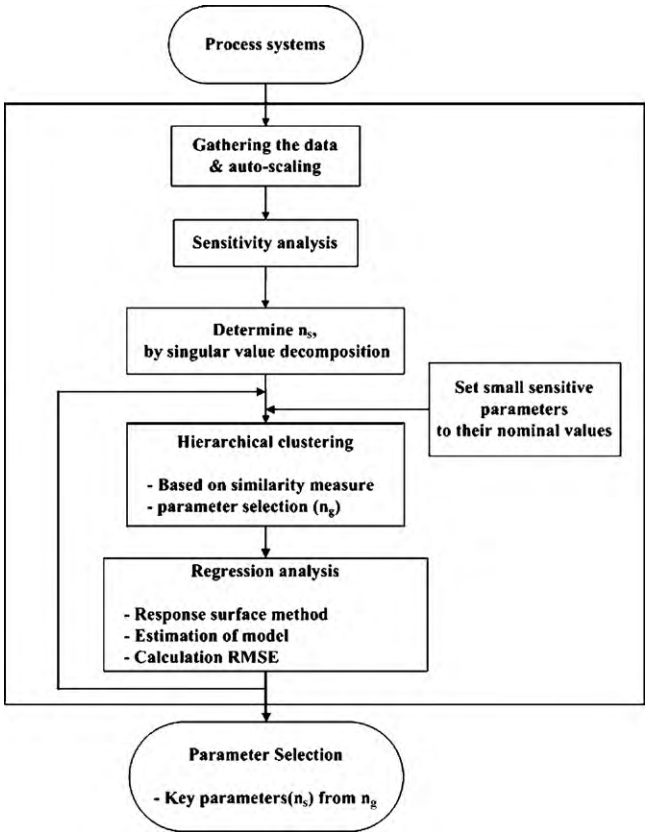


**Fig. 1.** A statistical framework for sensible parameter subset selection and optimal parameter estimation.

where $y$ is the response variable, $x_i$ and $x_j$ are the $i$th and $j$th independent variables and $\beta_0$ is a constant coefficient. $\beta_i$ is the regression coefficient for linear terms, $\beta_{ii}$ is the regression coefficient for quadratic terms and $\beta_{ij}$ is the coefficient for interaction terms of independent variables, respectively. In this study, the response surface methodology (RSM) is used to determine the optimal parameter estimation, because it can optimize the model parameters for the output variables [16].

### 2.5. A statistical framework for sensible parameter subset selection

Fig. 1 shows the statistical framework for parameter subset selection and optimal parameter estimation using hierarchical clustering and the response surface method. First, a data set was gathered from the plant and then normalized to solve the problem caused by different ranges and units. Second, the sensitivity vectors of the outputs were calculated with respect to the parameters and then singular value decomposition (SVD) of the sensitivity matrix was carried out. Third, the key parameters ($n_s$) are selected based on the results of the SVD, when the parameter's sensitivity vectors are shorter in length than their nominal values. $n_s$ is the number of parameters resulted from SVD. Then, the parameters are clustered into $n_g (n_g \geq n_s)$ groups by hierarchical clustering based on the similarity measure. $n_g$ is the number of parameters which is grouped by hierarchical clustering. The number of model parameters can be reduced by determining several groups of parameters for sensible parameter subset selection, where the parameters within a group are pair-wise indistinguishable. Next, the adequate number of $n_g$ parameters was determined using the root mean squares of error (RMSE) of the discrepancy between the model of the original model parameters and the model with a reduced number of parameters. For each group, the parameter that has the largest sensitivity vector was selected as the representative of the group. Then the key parameters within the $n_g$ parameters were selected. The RSM analysis proposed by Kim et al. [16] was conducted to find the optimal parameter values of the reduced parameter set with respect to an effluent quality index for the ASM calibration. Finally, the design for the plant model using the calibrated ASM model was finalized.

## 3. Results and discussion

Process data collected from a biological wastewater treatment plant is used in this study. The plant is located in N city, Korea, and is operated with an advanced nutrient removal process at the process capacity of 43,000 $m^3$/d. This plant utilizes the DNR process, which is one of the most advanced treatment processes available for the biological removal of nitrogen and phosphorus at the same time utilizing the characteristics of the microorganisms. This process requires four basins for denitrification and anaerobic, anoxic and oxic processes as well as a secondary clarifier shown in Fig. 2. Where the term of RAS is the amount of return activated sludge, WAS is the amount of wasted activate sludge, $Q_{inf}$ is a flowrate of influent, $Q_{eff}$ is a flowrate of effluent, NRCY is inner recycle, respectively. The daily mean values between March 9, 2007, and February 29, 2008, were used for the ASM model calibration.

In order to investigate the output effects with respect to various parameter conditions, a hundred data sets are used by choosing random combinations of the parameters. This process included 19 dynamic parameters of the ASM 1 model and three outputs of total suspended solids (TSS), chemical oxygen demand (COD) and total Kjeldahl nitrogen (TKN). In order to understand and analyze the effects of all of the parameters, the effluent quality index (EQI) is

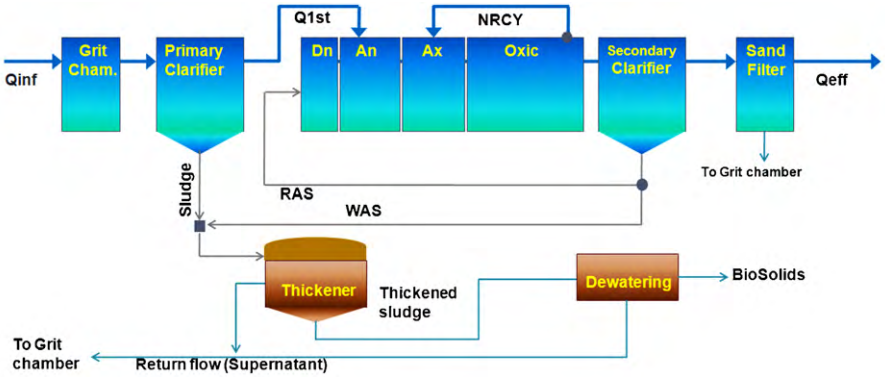


**Fig. 2.** A layout of the DNR process in the H-WWTP.

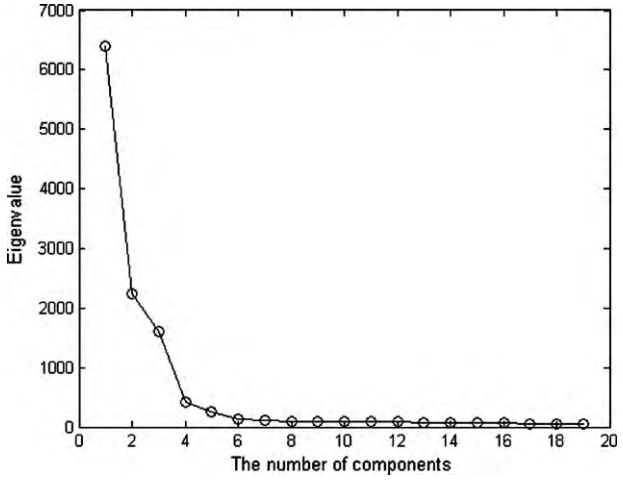


**Fig. 3.** The scree plot of a sensitivity matrix using singular value decomposition.

used. The EQI represents an index which combines many effluent components into only one index [17,18].

In this study, since the experimental values of biochemical oxygen demand (BOD) component of the effluent is not obtained, the EQI equation needed to be adjusted and it resulted the following Eq. (5):

$$\mathrm{EQI} = (\beta_{\mathrm{TSS}}\mathrm{TSS}_e + \beta_{\mathrm{COD}}\mathrm{COD}_e + \beta_{\mathrm{TN}}\mathrm{TN}_e) \times Q_e \quad (5)$$

where the $\beta$ factor represents the weights of the different types of effluent and the subscript $e$ indicates the type of effluent. We used the following values proposed by Copp [17] $\beta_{\mathrm{TSS}} = 2$, $\beta_{\mathrm{COD}} = 1$ and $\beta_{\mathrm{TN}} = 20$ in this study. The data set consisted of 19 model parameters and the EQI was normalized during a pre-process step. Then the sensitivity matrix is calculated using a relative sensitivity. This matrix was analyzed using SVD in order to determine the value $n_s$ and a scree plot is used as shown in Fig. 3. There are no specific rules or regulations for determining $n_s$. However, it is advised that if there is a gap of an order of magnitude or more between the singular values, then it is appropriate to choose the number of parameters to be estimated that is equal to the number of singular values that are larger than the predetermined cutoff value [13]. Following the above method, $n_s$ is determined to be 4 as it is clear from Fig. 3.

Next, the lengths of the sensitivity vectors were analyzed in order to select the meaningful parameters. The lengths of the sensitivity vectors were limited to the value less than 5% of the largest vector value in this study. Table 1 shows the parameter selection results based on the 5% limited lengths of sensitivity analysis, where the most sensitive parameter among all parameters is the autotrophic decay rate, $b_A$. The correction factor for the anoxic growth of heterotrophy, $\eta_g$, and the ammonia half-saturation coefficient (HSC) for the autotrophs, $K_{\mathrm{NH}}$, showed similar sensitivity in the output of the EQI. On the other hand, the HSC for heterotrophs, $K_S$, was very similar to the oxygen HSC for autotrophs, $K_{\mathrm{OA}}$. Only six parameters had a value higher than 5%. However, two parameters ($\mu_A$ and $K_X$) had values close to 5% of the largest sensitivity value. Therefore, a total of eight parameters were selected for the cluster analysis. These parameters are the mass N/mass COD of the products in biomass $i_{\mathrm{XP}}$, the correction factor for anoxic growth of heterotrophy, $\eta_g$, the ammonia HSC for autotrophs, $K_{\mathrm{NH}}$, the autotrophic decay rate, $b_A$, the autotrophic maximum specific growth rate, $\mu_A$, the HSC for the hydrolysis of the slowly biodegradable substrate, $K_X$, the correction factor for anoxic hydrolysis $\eta_H$ and the fraction of the biomass yielding particulate products $f_p$. In this study we were able to reduce the number of parameters to eight groups (eight parameters) out of a total of 19 parameters in the ASM model. This result reduced the size

**Table 1**
Parameter selection results when the lengths of sensitivity analysis are limited.

| Limited length | Maximum value of sensitivity | Parameters (unit) | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | $Y_H$ (g COD/g COD) | $i_{\mathrm{XB}}$ (g N/g COD) | $Y_A$ (g COD/g N) | $i_{\mathrm{XP}}$ (g N/g COD) | $K_S$ (g COD/m$^3$) | $K_{\mathrm{OH}}$ (g $O_2$/g COD) | $K_{\mathrm{NO}}$ (g $NO_3$-N/m$^3$) | $b_H$ (1/d) | $\mu_H$ (1/d) | $\eta_g$ |
| 5% | 77.4705 | 8.6 | 60.1 | 23.2 | 89.8 | 47.8 | 65.2 | 12 | 62.7 | 16.1 | 467.3 |
| Limited length | Maximum value of sensitivity | Parameters (unit) | | | | | | | | | |
| | | $K_{\mathrm{OH}}$ (g $O_2$/m$^3$) | $K_{\mathrm{NH}}$ (g $NH_3$-N/m$^3$) | $b_A$ (1/d) | $\mu_A$ (1/d) | $k_a$ (m$^3$/(g COD d)) | $K_X$ (g COD/g COD) | $k_h$ (g COD/(g COD d)) | $\eta_H$ | $f_p$ | |
| 5% | 77.4705 | 46.8 | 405.7 | 1549.41 | 73.2 | 50.9 | 75.8 | 57.3 | 85 | 111.4 | |

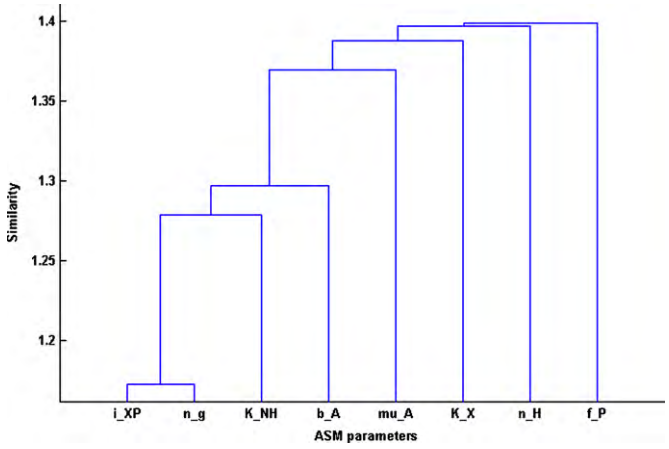


**Fig. 4.** The dendrogram of the hierarchical clustering of eight parameters in the ASM model.

**Table 2**
ANOVA results of the linear term of the RSM model in the effluent quality index.

| Source | DF | Seq SS | Adj SS | Adj MS | *F* | *P* |
|---|---|---|---|---|---|---|
| Regression | 8 | 346.76 | 346.76 | 43.3455 | 0.47 | 0.873 |
| Linear | 8 | 346.76 | 346.76 | 43.3455 | 0.47 | 0.873 |
| Residual error | 91 | 8348.76 | 8348.76 | 91.7446 | | |
| Total | 99 | 8695.52 | | | | |

of the combinatorial problem which involved a large number of parameters.

Fig. 4 is a dendrogram of the hierarchical clustering of the eight ASM parameters which are determined in the third step of the proposed method. It can be concluded that the similarity values of the eight parameters were very close, because their sensitivity vectors are almost parallel [13]. The diagram illustrates how the selection of the similarity value influenced the number of group, i.e. an increase in similarity values for the system led to an increase in the number of groups.

In order to determine the value of $n_g$, the response surface method (RSM) was carried out for the linear model with eight parameters. Table 2 shows the analysis results for the variance of the linear RSM model in the EQI, where DF is the degrees of freedom, Seq SS is the sequential sum of squares, Adj SS is the adjusted sum of the squares, Adj MS is the adjusted mean square, $F$ is the $F$-statistics and $P$ is the probability which is determined using the $F$-statistics. The linear model determined is not to be significant, since the $P$-value of the linear term is 0.873 over 0.05.

However, due to the characteristics of the microorganism, each parameter influenced the other parameters. Therefore, polynomial models that take into consideration the interaction and quadratic effects among the parameters are used instead of the linear model. Table 3 shows the analysis of variance (ANOVA) results of the polynomial RSM model in the EQI. In this model the $P$-value of the linear term is 0.965, the square term is 0.969 and the interaction term is 0.811.

**Table 3**
ANOVA results of the polynomial RSM model in the effluent quality index.

| Source | DF | Seq SS | Adj SS | Adj MS | *F* | *P* |
|---|---|---|---|---|---|---|
| Regression | 44 | 3056.05 | 3056.05 | 69.456 | 0.68 | 0.908 |
| Linear | 8 | 346.76 | 241.68 | 30.21 | 0.29 | 0.965 |
| Square | 8 | 599.47 | 232.36 | 293.045 | 0.28 | 0.969 |
| Interaction | 28 | 2109.81 | 2109.81 | 75.351 | 0.73 | 0.811 |
| Residual error | 55 | 5639.47 | 5639.47 | 102.536 | | |
| Total | 99 | 8695.52 | | | | |

**Table 4**
Modeling results of the calibrated ASM model with the different number of groups.

| No. of groups | 3 | 4 | 5 | 6 | 7 | 8 |
|---|---|---|---|---|---|---|
| Least similarity | 1.3791 | 1.3876 | 1.3697 | 1.2969 | 1.2788 | 1.1724 |
| RMSE | 180.2 | 171.8 | 229.8 | 246.5 | 192.8 | 7.5 |

In order to determine the modeling efficiency, the coefficient of each parameter was substituted for each parameter that represented groups 3, 4, 5, 6, 7 and 8. Then the residual mean square of the error (RMSE) was calculated, which is the difference between the actual value and the predicted value, and is represented by the following equation:

$$\mathrm{RMSE} = \sqrt{\frac{\sum_{i=1}^{n}(Y_i = \widehat{Y}_i)^2}{n-1}} \quad (6)$$

where $Y_i$ is a real value, $\widehat{Y}_i$ is a predicted value and $n$ is the number of experiments. For example, if 3 groups were selected, then the number of parameters in each group was 3: $i_{XP}$, $b_A$, and $\eta_g$. Considering the tradeoff between the model efficiency and the number of parameters, four groups ($n_g$) were selected in this study. These parameters are $i_{XP}$, $b_A$, $\eta_g$ and $K_{NH}$.

Table 4 shows the modeling results of the calibrated ASM model with the varied number of groups. The RMSE values in Table 4 changed in accordance with the varying number of groups, where the RMSE value was the smallest when there were eight groups. It was also noted from out results (not shown in the table) that although the number of groups was increased, the RMSE value by model was not decreased. However, it is a general trend that the RMSE values are usually smaller when there are larger groups.

As a final step, we selected $n_s$ parameters within the $n_g$ parameters. Since $n_s$ is same as $n_g$, four parameters were selected as key parameters: $i_{XP}$, $b_A$, $\eta_g$ and $K_{NH}$. In order to estimate a value for each key parameter, RSM was used for the final parameter estimation. The polynomial model containing linear, square and interaction terms for the four parameters is shown as follows:

$$\begin{aligned}\mathrm{EQI} = {} & (-1142.1 \times i_{XP}) + (32 \times \eta_g) + (-7265 \times b_A) + (-21.6 \times K_{NH}) \\ & + (8066.3 \times i_{XP}^2) + (-4.8 \times \eta_g^2) + (77377.4 \times b_A^2) \\ & + (-1.1 \times K_{NH}^2) + (-803.9 \times i_{XP} \times \eta_g) + (63387.9 \times i_{XP} \times b_A) \\ & + (62.2 \times i_{XP} \times K_{NH}) + (955 \times \eta_g \times b_A) + (10.7 \times \eta_g \times K_{NH}) \\ & + (1621.4 \times b_A \times K_{NH}) \end{aligned} \quad (7)$$

where effluent quality index (EQI) is an integrated index for evaluating the performance of WWTP and does not have any unit. The terms $i_{XP}$ is the mass N/mass COD of the products in biomass (g N/g COD), $\eta_g$ is the correction factor for anoxic growth of heterotrophy (no unit), $b_A$ is the autotrophic decay rate (1/d) and $K_{NH}$ is the ammonia HSC for autotrophs (g $NH_3$-N/$m^3$), respectively.

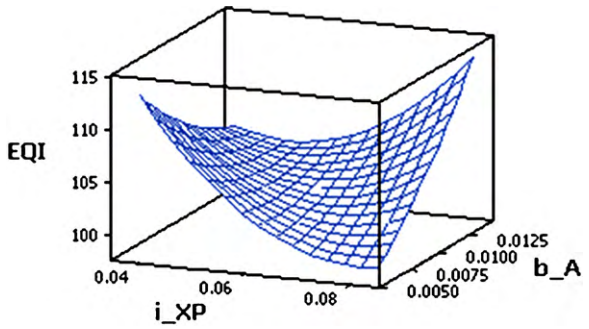


**Fig. 5.** The response surface plot of the effect of $i_{XP}$ and $b_A$ on the effluent quality index (EQI).

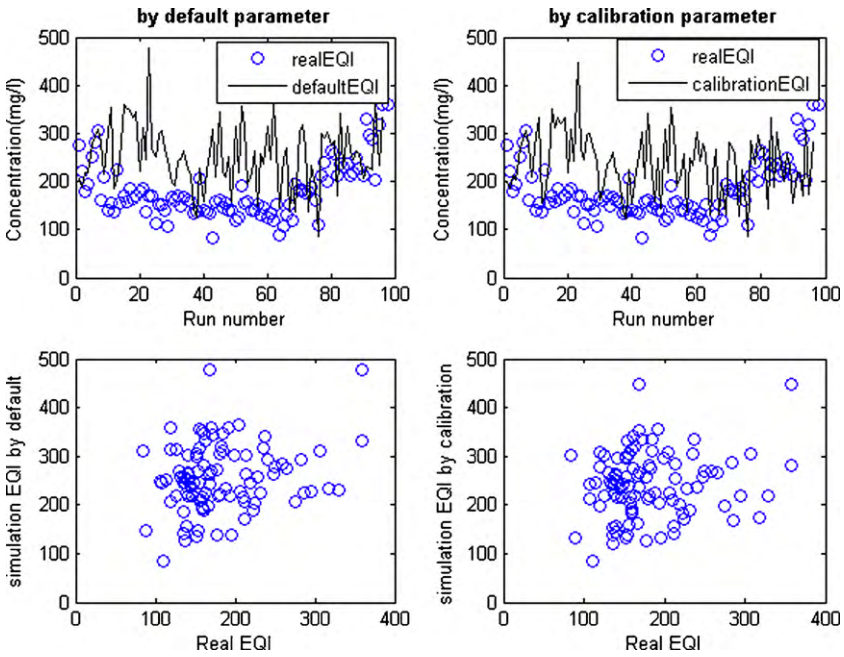


**Fig. 6.** Comparison of actual results to the simulated results using the default and calibrated parameters: (a) time series plot of the actual data and simulated data using the default model; (b) time series plot of the actual data and simulated data using the calibrated model; (c) actual data versus simulated data using the default model; and (d) actual data versus simulated data using the calibrated model.

The model confirms that the square terms of $i_{XP}$ and $b_A$ have the most sensible influence on the effluent quality index (EQI). Fig. 5 shows the effect of the two effective variables, $i_{XP}$ and $b_A$ (the other two parameters, $K_{NH}$ and $\eta_g$, were kept constant) on the response surface contours of the effluent quality index. The response surface plot can be used to determine the operation variables that optimize any of the response variables. This result is based on several objectives which determine the optimal solution by carrying out the minimum number of experiments [16]. Fig. 5 indicates that the EQI decreased when the value of $i_{XP}$ increased. However, the EQI increased when the value of $i_{XP}$ increased. The slope of the effect of $b_A$ was higher than that of $i_{XP}$. This result is reasonable when compared to Eq. (7). When the value of $i_{XP}$ was around 0.08, the EQI had the smallest value and when the value of $b_A$ was around 0.005, the EQI also had the smallest value. These results indicate that the optimum range for the two parameters of $i_{XP}$ and $b_A$ is approximately 0.08 and 0.005, respectively.

The optimal values for the four parameters are estimated based on the direction in which the EQI is minimized. With the minimum value of EQI close to the D-optimality of 92.08, the values of the four parameters were estimated to be: $i_{XP} = 0.08$, $b_A = 0.006$, $\eta_g = 0.95$ and $K_{NH} = 0.63$. These results are different from the default values of each of the parameters in the literature where $\eta_g = 0.8$ and $K_{NH} = 1.0$.

Then, we applied the calibration parameters to describe and simulate the ASM process, and then compared those results to the results when using default parameters. Fig. 6 shows the EQI results compared, when using default parameters and calibrated parameters in the simulation data. The black circles indicate the actual values and the black lines represent the simulation values by using the calibrated parameters. In Fig. 6(a) and (b), we can confirm that all of the simulated result has higher value than real measured values. The results shown in Fig. 6(c) and (d) indicate that the relationship between the actual values and simulation values can be confirmed directly. When the simulation value was closer to the real value, the linear relationship showed a 45° angle between them. In this plot, the linear relationship did not accurately display a 45° angle, but it represented a linear relationship between the actual value and simulation value.

For a more accurate comparison, we applied RMSE value to each simulation result. The RMSE value of EQI using calibrated parameters was 101.23 whereas the RMSE value using default parameters were 108.72, respectively. The errors between the real data and the calibration data can be less than those when the default values of parameters are used.

The reduction in the number of parameters by the parameter subset greatly reduced the search space in the ASM model parameter calibration. However, the RMSE values were slightly different between the eight and four parameters selected using the clustering and RSM. The results shown in Table 3 indicate that the calibrated model with the reduced parameter subset is still valid for model predictions and still it maintains its modeling efficiency. Table 4 indicates that the modeling results of the calibrated model are more accurate than those of the model using default parameters. If there are a large number of parameters in the model, ASM 2d or ADM for example, then the proposed method can be used as a very effective tool, since it can systematically select the parameter subset while maintaining the model prediction performance. This method reduces the size of the parameter search space.

## 4. Conclusion

This study proposes a new methodology for statistical model calibration that can sort the sensible parameter subsets and estimate the parameters using a multiple response surface of an effluent quality index. The selected parameters were clustered using the proposed method, because they had a similar effect on the model outputs. Only one parameter per group with the largest sensitivity vector was chosen as the representative parameter. The reduction in the number of parameters in the parameter subset greatly reduced the search space in the ASM model parameter calibration. A case study showed that the 19 total parameters of the ASM model were reduced to a set of eight parameters and then

a set of four parameters were identified from among the eight parameters by clustering. The optimal values of the four parameters were estimated using the multiple response method. The proposed statistical framework resulted in very little computational effort to find an optimal solution in comparison to the traditional calibration method which searches the full set of the total parameters.

## Acknowledgement

This work was supported by a Korea Science and Engineering Foundation (KOSEF) grant funded by the Korean government (MEST) (KRF-2009-0076129).

## References

[1] G. Olsson, B. Newell, Wastewater Treatment System—Modelling, Diagnosis and Control, IWA, London, 1999.
[2] M. Henze, W. Gujer, T. Mino, M.C.M. van, Loosdrecht, Activated Sludge Models: ASM1, ASM2, ASM2d and ASM3, IWA, London, 2000.
[3] R. Brun, M. Kuhni, H. Siegrist, W. Gujer, P. Reichert, Practical identifiability of ASM2d parameters–systematic selection and tuning of parameter subsets, Water Res. 36 (2002) 4113–4127.
[4] M.B. Beck, Water quality model: a review of the analysis of uncertainty, Water Resour. Res. 23 (1987) 1393–1442.
[5] S.R. Weijers, P.A. Vanrolleghem, A procedure for selecting best identifiable parameters in calibrating activated sludge model no. 1 to full-scale plant data, Water Sci. Technol. 36 (1997) 69–79.
[6] M.V. Ruano, J. Ribes, D.J.W. De Pauw, G. Sin, Parameter subset selection for the dynamic calibration of activated sludge models (ASMs): experience versus systems analysis, Water Sci. Technol. 56 (2007) 107–115.
[7] J.J.W. Hulsbeek, J. Kruit, P.J. Roeleveld, M.C.M. van, Loosdrecht, A practical protocol for dynamic modeling of activated sludge systems, Water Sci. Technol. 45 (2002) 127–136.
[8] P.A. Vanrolleghem, G. Insel, B. Petersen, G. Sin, D. De Pauw, I. Nopens, S. Weijers, K. Gernaey, A comprehensive model calibration procedure for activated sludge models, in: Proceedings: WEFTEC 2003, 76th Annual Technical Exhibition and Conference, USA, 2003.
[9] H. Melcer, P.L. Dold, R.M. Jones, C.M. Bye, I. Takacs, H.D. Stensel, A.W. Wilson, P. Sun, S. Bury, Methods for Wastewater Characterisation in Activated Sludge Modeling, Water Environment Research Foundation (WERF), Alexandria, VA, 2003.
[10] G. Langergraber, L. Rieger, S. Winkler, J. Alex, J. Wiese, C. Owerdieck, M. Ahnert, J. Simon, M. Maurer, A guideline for simulation studies of wastewater treatment plants, Water Sci. Technol. 50 (2004) 131–138.
[11] G. Sin, D.J.W. de Pauw, S. Weijers, P.A. Vanrolleghem, An efficient approach to automate the manual trial and error calibration of activated sludge models, Biotechnol. Bioeng. 100 (2008) 515–528.
[12] G. Sin, S.W.H. Van Hulle, D.J.W. De Pauw, G.A. Van Griensven, P.A. Vanrolleghem, A critical comparison of systematic calibration protocols for activated sludge models: a SWOT analysis, Water Res. 39 (2005) 2459–2474.
[13] Y. Chu, J. Hahn, Parameter set selection via clustering of parameters into pairwise indistinguishable groups of parameters, Ind. Eng. Chem. Res. 48 (2009) 6000–6009.
[14] M.H. Kim, W.Y. Lee, J.J. Lim, A.S. Rao, C.K. Yoo, A one-step model calibration methodology based on multiple error minimization method for the parameter estimation of activated sludge model no. 1, Korean J. Chem. Eng. (2010) in press.
[15] M.S. Kim, S.Y. Lee, A clustering method using neural networks for processing large data, J. Ind. Technol. 10 (2002) 1–5.
[16] M.H. Kim, A.S. Rao, C.K. Yoo, Dual optimization strategy for N and P removal in a biological wastewater treatment plant, Ind. Eng. Chem. Res. 48 (2009) 6363–6371.
[17] J.B. Copp, The COST-simulation benchmark: description and simulator manual, in: COST Simulation Benchmark Manual, IWA, London, 2001.
[18] M.H. Kim, C.K. Yoo, Design and environmental/economic performance evaluation of wastewater treatment plant using modeling methodology, Chem. Eng. Res. 46 (2008) 610–618.