Chemical Engineering Journal 145 (2008) 78–85

Contents lists available at ScienceDirect

Chemical Engineering Journal

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

# Prediction of effluent quality of an anaerobic treatment plant under unsteady state through ANFIS modeling with on-line input variables

Altunay Perendeci [a,d], Sever Arslan [b], Serdar S. Çelebi [c], Abdurrahman Tanyolaç [c,*]

[a] Turkish Sugar Factories Corporation, Sugar Institute, Etimesgut 06790, Ankara, Turkey
[b] Turkish Sugar Factories Corporation, Electromechanic Instruments Factory, Etimesgut 06790, Ankara, Turkey
[c] Chemical Engineering Department, Hacettepe University, Beytepe 06800, Ankara, Turkey
[d] Environmental Engineering Department, Akdeniz University, 07058 Antalya, Turkey

ARTICLE INFO

*Article history:*
Received 13 December 2007
Received in revised form 1 March 2008
Accepted 12 March 2008

*Keywords:*
Anaerobic wastewater treatment
Adaptive network-based fuzzy inference system
Modeling
On-line variables

ABSTRACT

A neural fuzzy model based on adaptive network-based fuzzy inference system (ANFIS) was proposed in terms of on-line input variables $CH_4\%$, $Q_{gas}$, $Q_{anarecycle}$, $Q_{inf\text{-}bypass}$ and $Q_{inf}$ to estimate the effluent chemical oxygen demand, $COD_{eff}$, of a real scale unsteady anaerobic wastewater treatment plant of a sugar factory.

Two new variables were added into the input variables matrix of the model; phase vectors of the plant operation and the history of effluent *COD* values in order to increase the fitness of simulated results. ANFIS was able to estimate the water quality discharge parameter with success for the case when only limited on-line variables were available without requiring the measurement of inlet *COD*. Acceptable correlation coefficient (0.8354) and root mean square error (0.1247) were found between estimated and measured values of the system output variable, effluent *COD*, in the case of excluding inlet volumetric flow rate of the wastewater treatment plant from the on-line input variable matrix. The developed ANFIS model may be integrated into an advanced control system for the anaerobic treatment plant using different control strategies with further work.

© 2008 Elsevier B.V. All rights reserved.

## 1. Introduction

Discharge of inadequately treated wastewater either from industrial or municipal sources leads to serious ecological problems in receiving waterways. For high removal efficiency, wastewaters with a high chemical oxygen demand, COD, can be treated in an anaerobic digestion process, in which biodegradable organic materials are eventually decomposed to mainly methane in the absence of oxygen. Anaerobic digestion has several advantages compared to the traditional aerobic treatment; high capacity to degrade difficult substrates at high concentrations, very low sludge production, low energy requirements and energy recovery through methane combustion [1]. Anaerobic digestion is generally considered a non-linear, time varying three-stage process, which depends on a synergistic relationship of bacterial populations and is influenced by physico-chemical conditions and the process in the reactor [2]. Anaerobic wastewater treatment plants are normally designed with reference to nominal operating conditions where the loading rate is assumed constant. However, in practice, the steady state assumption is seldom met and in fact, the process is subject to wide fluctuations, both in flow and organic loading, which often results in corrupted performance or even plant failure. The successful management of these critical situations is of high relevance to industrial applications. This requires the solution of a complex control design, assuming that on-line information about the quality of the plant is available and that the plant is equipped with flow control devices to alter the flow patterns [3]. Moreover, the sensitivity of anaerobic reactors means that they need to be supervised by the plant personnel, especially when a process control system is run up in a new plant.

The control of anaerobic processes may be difficult due to two main reasons. First, there is no precise mathematical models for these complex systems, in which the interactions among the multiple microbial species are not understood exactly [4]. Second, there may be a significant delay before an alteration in the state of the system is stemmed from some significant changes in macroscopic process variables susceptible to easy measurements [4].

A common trend in modeling studies is to identify a model, suitable for on-line estimation and forecasting, providing accurate predictions of digester behavior, which will be sufficient for use in process control and controller design. There are numerous comprehensive mathematical models of anaerobic treatment systems in literature as reviewed by Harper and Suidan [5]. Of these, kinetic models are among the most widely used ones, however,

* Corresponding author. Tel.: +90 312 297 74 04; fax: +90 312 299 21 24.
*E-mail address:* tanyolac@hacettepe.edu.tr (A. Tanyolaç).

1385-8947/$ – see front matter © 2008 Elsevier B.V. All rights reserved.
doi:10.1016/j.cej.2008.03.008

many of them are very complex, often analytically insolvable, and not routinely useful in control applications. For instance, although the advanced structured models for anaerobic processes have some advantages, there are too many kinetic parameters to determine, which, in fact, are highly dependent on specific environmental conditions [6].

One of the most frequent and important challenges in the control of biochemical processes is to find adequate and reliable sensors to measure all important state variables of the plant. However, if a number of on-line sensors providing state information are today available at industrial scale, they are still very expensive and their maintenance is usually time consuming. In order to overcome these difficulties, the notation of software sensors has been introduced, which simply consists of using state estimation techniques to predict the values of unmeasured variables from the available on-line measurements [7,8]. Under the name of soft computing; theories, approaches and techniques are gathered to find solutions to a wide variety of problems such as pattern recognition, system control, prediction, optimization and others which share same characteristics of nonlinear nature, missing or disturbed data by noise, imprecision or uncertainty. Moreover, the sources of these data can be very heterogeneous, ranging from discrete to continuous variables, and include a spatial or temporal component. When tackling a real-world problem, it turns out that they are mostly partial and, sometimes, ill defined, difficult to model and the solutions are lost in huge search spaces. Now, precise models are impractical to use, costly, or simply non-existent. This makes soft computing approaches a flexible means to deal with such problems [9].

Over the last decade, significant advances have been made in two distinct technological areas; fuzzy logic and computational neural networks. The theory of fuzzy logic provides a mathematical framework to capture the uncertainties associated with human cognitive processes, such as thinking and reasoning. On the other hand, the computational neural network paradigms have evolved in the process of understanding the learning and adaptive features of neural mechanisms inherent in certain biological species. The integration of two fields has created an emerging technological field of the fuzzy neural networks. Since fuzzy set theory and neural networks were applied to anaerobic digestion in the 1980s, fuzzy and neural models have shown great advantages in the simulation, prediction and controlling of the anaerobic treatment systems. Adaptive network-based fuzzy inference system, ANFIS, proposed by Jang [10–13], is an integrated technique which was first applied to the anaerobic wastewater treatment system by Tay and Zhang [6,14].

In this study, a neural fuzzy model based on ANFIS was proposed for a real scale industrial anaerobic treatment plant (anaerobic methane production, ANAMET) operating at unsteady state to estimate the discharge variable (lamella clarifier effluent COD) by using only appropriate on-line input variables. Evaluation of ANFIS modeling with only on-line input variables for the operation of a real scale anaerobic wastewater treatment plant was studied for the first time in literature. In order to create a possibility for integrating model results into the decision making unit of the advanced process control and to reduce the computational time, only available on-line process variables in the process were directly used to develop a simple and realistic model which did not include any information about the inflow COD concentrations of the plant.

## 2. Materials and methods

### 2.1. Process configuration

The sugar beet processing factory (Ereğli, Turkey) has an actual capacity of 8000 tones of beet per day. The factory wastewater consists of two main streams; one is the mixture of flume (beet transportation water) and washing water and the second is the wastewater from miscellaneous use of water in the process. Water from soil settlement lagoons and miscellaneous use were balanced in the equalization basin from where pumped to the treatment plant for the removal of COD and nitrogenous compounds prior to reuse and discharge. Characteristics of wastewater fed to the wastewater treatment plant are presented in Table 1. All data in Table 1 are average values of wastewater parameters obtained from the daily composite samples. Operation period of wastewater treatment plant was divided into three phases based on COD loading. Period of first 33 days was called start-up, the period between the days 34 and 149 was defined as pseudo-steady state phase and last 42 days (150–192 days) was described as the end phase of the operation period. Furthermore, average values of the wastewater fed to the wastewater treatment plant during full operation period were presented in the last column of Table 1.

The treatment complex is a full-scale ANAMET type plant which consists of sequential anaerobic and aerobic biological treatment units. Anaerobic unit includes hydrolysis and anaerobic tanks, which are totally mixed reactors, and a lamella type sludge separation system. The ANAMET plant was designed for a wastewater flow of 4680 $m^3$/day and a COD load of 37,500 kg/day. Process scheme of anaerobic wastewater treatment plant is presented in Fig. 1. Average treatment efficiency of the anaerobic unit based on COD removal was realized as 97% during the operation period.

### 2.2. Model architecture and components

ANFIS is a technique for automatically tuning first-order Sugeno type inference systems based on training data [10]. It consists of five key components; inputs and outputs, database and preprocessor, a fuzzy system generator, a fuzzy inference system, and an adaptive neural network representing the fuzzy system [6]. Input and output variables are selected or generated from the variables com-

**Table 1**
Characteristics of wastewater fed to the wastewater treatment plant in different periods of operation

| Wastewater parameters | Average characteristics of wastewater at start-up period (1–33 days) | Average characteristics of wastewater in pseudo-steady state period (34–149 days) | Average characteristics of wastewater in the end period (150–192 days) | Overall average values of full operation period (1–192 days) |
|---|---|---|---|---|
| $Q_{inf}$ ($m^3$/day) | 1267.64 | 4009.53 | 2082.56 | 3107.00 |
| COD (mg/L) | 9069.85 | 4651.96 | 7506.35 | 6050.55 |
| pH | 4.62 | 6.01 | 6.43 | 5.89 |
| VFA (mg HAc/L) | 642.86 | 2318.24 | 4874.52 | 2643.26 |
| $N_{tot}$ (mg/L) | 68.36 | 50.26 | 22.71 | 44.94 |
| $NH_4$–N (mg/L) | 13.60 | 12.59 | 8.78 | 11.90 |
| $P_{tot}$ (mg/L) | 1.40 | 2.31 | 2.28 | 2.24 |
| $PO_4$–P (mg/L) | 0.67 | 1.49 | 2.00 | 1.52 |
| TSS (mg/L) | 84.57 | 392.62 | 1075.58 | 497.00 |
| VSS (mg/L) | 71.50 | 189.62 | 618.21 | 283.10 |

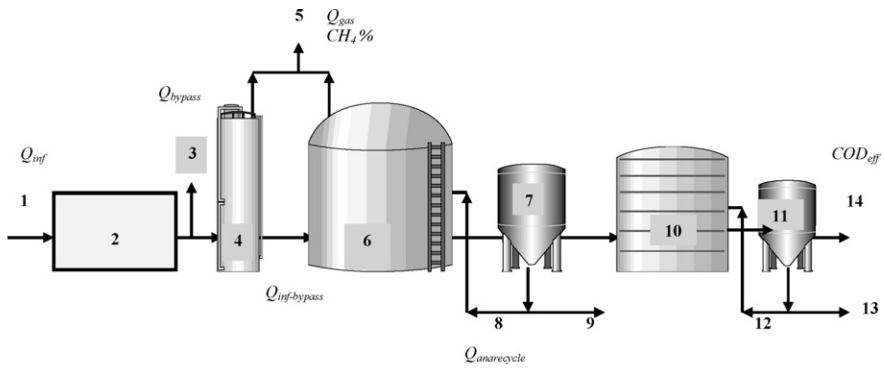


**Fig. 1.** Process scheme of anaerobic wastewater treatment plant, ANAMET: (1) wastewater from sugar factory, (2) equalization basin, (3) bypass to aerobic unit, (4) hydrolysis tank, (5) biogas to burner, (6) anaerobic tank, (7) lamella settler, (8) anaerobic sludge recycle line, (9) excess sludge line, (10) denitrification and nitrification tank, (11) final sedimentation, (12) aerobic sludge recycle line, (13) excess sludge line and (14) discharge.

monly used for system description. Database containing system performance information is a prerequisite for the model development. Generally, it is developed by collecting the data of regularly monitored variables. Since the ANFIS is usually started with a prototype fuzzy system, a fuzzy system generator is needed. The software MATLAB (Matworks Inc.) provides this function. Tay and Zhang [6] used MATLAB for model programming and they proved that the MATLAB language is suitable for the programming of the model.

### 2.3. Model implementation

Water quality data of 192 days from the treatment plant were used to derive the models. Daily composite samples were formed by collecting samples from lamella outlet every 2 h and raw data of COD analysis and on-line variables were normalized between 0 and 1 structure and formatted in Excel 7.0.

These normalized values were used to form the original database that serves the training and validation database for ANFIS. In the modeling study, the data of 129 days were used for training and the remaining data of 63 days were reserved for validation. As stated by El-Mansi and Bryce [15], selection of the data in the training and validation sets can be carried out either randomly or statistically depending on the design employed. In our study, training and validation data were selected randomly from the whole data set of 192 days using a macro written in Excel 7.0 program.

The software Matlab 6 (The Matworks Inc. Natick, MA) and the fuzzy logic toolbox (Version 2.1) were used to derive the model that assist the user in designing a Sugeno fuzzy system prototype for each trial from training data. The function “genfis 2” was applied to generate a first-order Sugeno fuzzy inference system (FIS) using subtractive clustering of the data set provided. Subtractive clustering, a technique for automatically generating FIS by detecting clusters in input–output training data, assumes each data point as a potential cluster center and calculates a measure of the likelihood that each data point would define the cluster center, based on the density of surrounding data points. Default values for range of influence (radius), squash factor, accept ratio and reject ratio in genfis 2 were selected as 0.5, 1.25, 0.5 and 0.15, respectively.

Once a fuzzy system prototype is available, the ANFIS algorithm tunes and optimizes the fuzzy system by learning from the training data and finally produces a Sugeno fuzzy system with the same structure as prototype. Schematic diagram of ANFIS models with all input variables and, input–output mapping structure with only on-line variables were presented in Fig. 2(a) and (b), respectively. Checking data option in ANFIS was used in all models in order to prevent the model from overfitting.

Operators used in automatically tuning Sugeno fuzzy system; AND method is product, OR method is probabilistic OR, defuzzification method is weighed average, implication method is minimum and aggregation method is maximum. The membership functions (MF) of the inputs were characterized as Gaussian MFs and MFs of the outputs were specified linear.

The resulting model was verified by using a validation database provided by the normalized database. If the resulting fuzzy system

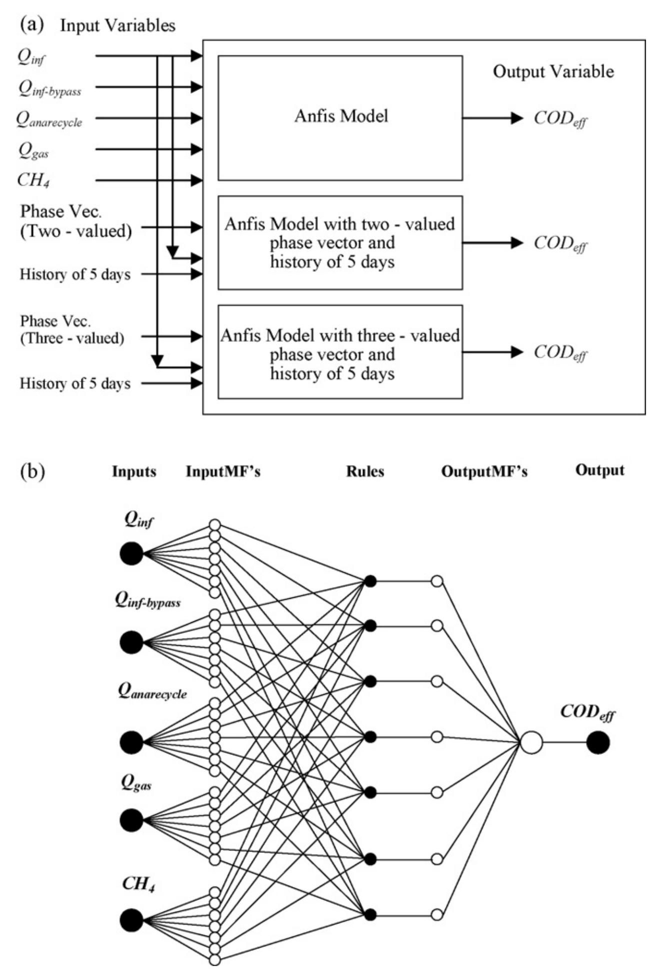


**Fig. 2.** Schematic diagram of (a) ANFIS models with all input variables and (b) input–output mapping structure of ANFIS model with only on-line input variables.

estimates satisfactorily the validation data, then the computation was terminated.

### 2.4. Determination of input–output variables

Lamella clarifier effluent COD was chosen as the output variable of the model in our study. Since the temperature in anaerobic tank was kept at approximately 35 °C during the whole operation period, temperature was not accounted as a variable in the model as previously stated by Tay and Zhang [6]. pH was also not regarded as a model variable as suggested by Weiland and Rozzi [16], because fluctuations in pH during operation is quite negligible. Volumetric flow rate of wastewater was measured by electromagnetic flow meters (Danfoss MagFlo). Gas flow rate was measured by a Bailey Fischer Porter vortex flow-meter. Gas composition of the biogas was determined by an on-line Varian Micro GC. Programmable logic control (PLC) and a TEOS 32 supervisory control and data acquisition (SCADA), systems were used to monitor and control the anaerobic wastewater treatment plant. Every 5 min, wastewater and gas flow rates and $CH_4$ content of biogas were read from TEOS 32 SCADA system to yield average daily values. COD analysis of the lamella outlet was carried out off-line according to a standard method [17]. Selected on-line input variables available in the wastewater treatment plant for model solutions along with operational ranges are presented in Table 2.

**Table 2**
Selected on-line input variables available in the wastewater treatment plant and operational ranges for model solutions

| Input variables | Unit | Determination | Minimum | Maximum | Overall average |
|---|---|---|---|---|---|
| $Q_{inf}$ | $m^3$/day | On-line | 218 | 4,856 | 3107 |
| $Q_{inf\text{-}bypass}$ | $m^3$/day | On-line | 218 | 4,768 | 3003 |
| $Q_{anarecycle}$ | $m^3$/day | On-line | 754 | 7,852 | 4631 |
| $Q_{gas}$ | $m^3$/day | On-line | 150 | 20,325 | 9458 |
| $CH_4$ | % (v/v) | On-line | 57.68 | 71.59 | 65.95 |

## 3. Results and discussions

For the treatment period, time profiles of inlet volumetric flow rate of the wastewater treatment plant, $Q_{inf}$, anaerobic sludge recycle flow rate, $Q_{anarecycle}$, 0inlet volumetric flow rate of the anaerobic tank, $Q_{inf\text{-}bypass}$, produced volumetric gas flow rate, $Q_{gas}$, COD in influent, $COD_{inf}$, and in effluent, $COD_{eff}$, are presented in Fig. 3(a)–(f), respectively. In the figures, the on-line variables as well as COD values fluctuated within wide ranges showing unsteady nature of

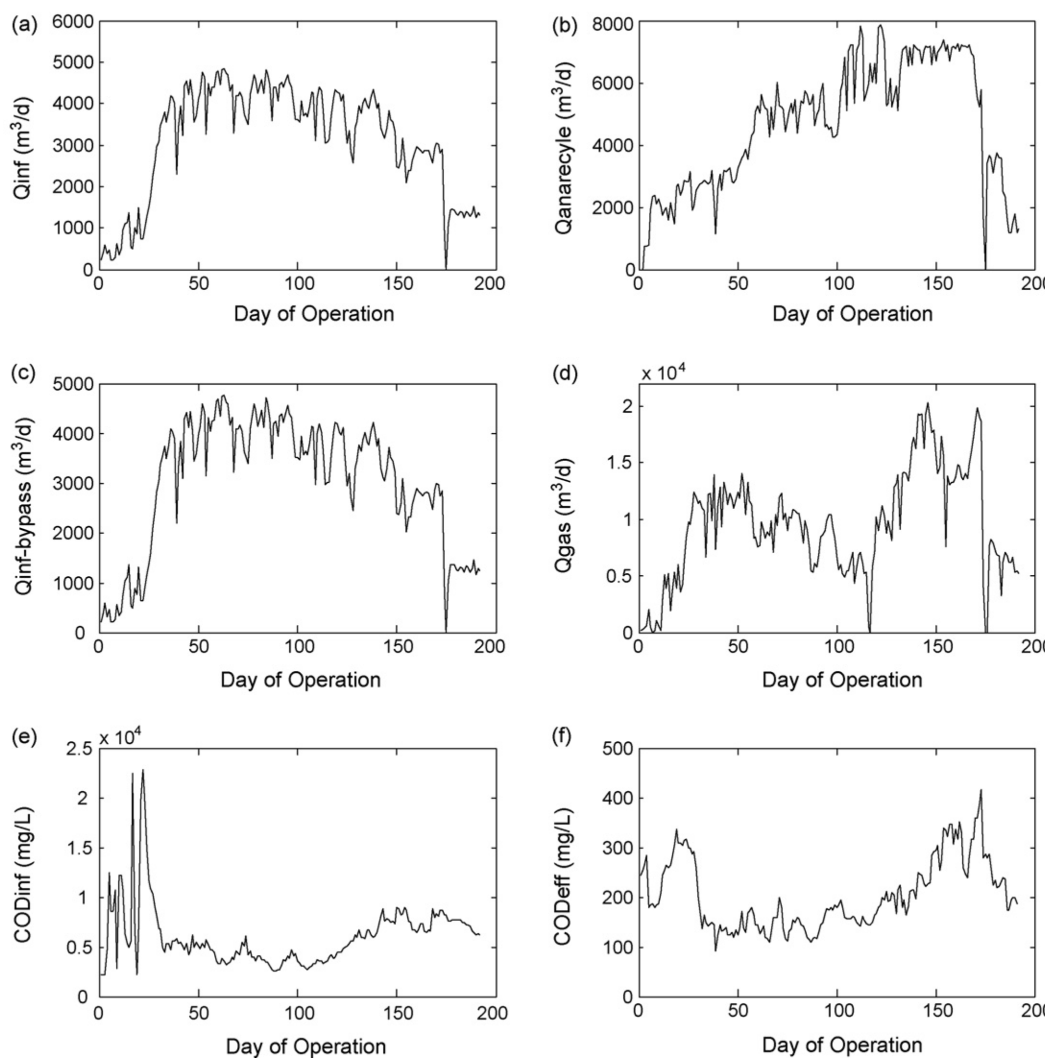


**Fig. 3.** Time profiles of (a) $Q_{inf}$, (b) $Q_{anarecycle}$, (c) $Q_{inf\text{-}bypass}$, (d) $Q_{gas}$, (e) $COD_{inf}$ and (f) $COD_{eff}$.

**Table 3**
Evaluation of on-line input variables with the inclusion of two-valued phase vector and history of last 5 days

| Excluded variable | Training (fuzout 2/dataout) | | Validation (valid fuzout 2/valid dataout) | |
|---|---|---|---|---|
| | RMSE | R | RMSE | R |
| None | 1.9476E−005 | 1.0 | 0.1522 | 0.8097 |
| History of 5 days | 0.0334 | 0.9874 | 0.1471 | 0.8011 |
| Phase vector (two-valued) | 6.5211E−005 | 1.0 | 0.1924 | 0.6824 |
| $CH_4$ | 3.3658E−004 | 1.0 | 0.2293 | 0.6897 |
| $Q_{gas}$ | 0.0055 | 0.9997 | 0.2844 | 0.6290 |
| $Q_{anarecycle}$ | 0.0404 | 0.9815 | 0.2049 | 0.6690 |
| $Q_{inf\text{-}bypass}$ | 1.0673E−004 | 1.0 | 0.1945 | 0.7514 |
| $Q_{inf}$ | 7.9758E−005 | 1.0 | 0.1651 | 0.7887 |

ANAMET plant. During the last 6 years of operation, plant did not experience any process failure or any imbalance occurrence. Therefore, in our work developed ANFIS models could not be trained with the data representing process failure occurrences such as organic or hydraulic overload, total shutdown or washout of the reactor.

Only available on-line process variables $CH_4$%, $Q_{gas}$, $Q_{anarecycle}$, $Q_{inf\text{-}bypass}$ and $Q_{inf}$ in Table 2 were used so as to create a simple and realistic model for the advanced process control of the treatment unit minimizing the complexity and the computation time. A stepwise approach was applied in this study to determine the individual effects of input variables on the estimation of output variable value. In the approach, separate networks were trained for each new case, which was formed by dropping one of the input variables sequentially while the others are kept in input variable matrix. Effect of each input variable on the estimation performance of resulting model is determined by evaluating the measured and estimated results via root mean squared error, RMSE, and correlation coefficient, R.

The features of phase vector of the plant operation and recent values of the output variable (history of last 5 days for effluent COD) were also included in model prediction/parameter evaluation studies mainly due to the success in our previous work of ANFIS modeling [18]. Two-valued phase vector differentiated start-up (0) and pseudo-steady state phases (1) and, the three-valued phase vector represented start-up (0), pseudo-steady state (0.5) and end phases (1) of the full operation period of the treatment plant. Recent values of the output variable covered the last 5 days since the hydraulic retention time of the wastewater stream within the treatment plant is approximately 5 days. The inclusion of recent values of the output variable within the matrix was expected to eliminate the dead band of the system, therefore increasing the accuracy of the model prediction.

In the first case, two-valued phase vector (0–1) and the history of last 5 days were integrated into on-line input variables matrix of the model and evaluation of the models are presented in Table 3. In this table, "dataout" and "valid dataout" are the measured output values (effluent COD values) of the system to be used for training and validation sets, respectively. The terms "fuzout 2" and "valid fuzout 2" are predicted ANFIS results as "dataout" and "valid dataout", respectively generated from the FIS after the training step with 200 iterations. General assessment of the model success was essentially based on validation results after training.

In Table 3, RMSE and R values were obtained as 0.1522 and 0.8097, respectively with the addition of two-valued phase vector and the history of last 5 days in on-line input variable matrix containing all on-line variables. When $Q_{anarecycle}$, two valued phase vector and $CH_4$% variables were excluded from the input variables matrix one at a time, correlation coefficients were calculated 0.6690, 0.6824, and 0.6897, respectively. Clearly, exclusion of these variables presented considerable drop in correlation coefficient. Also, individual absence of $Q_{gas}$ resulted in a very low correlation coefficient (0.6290) and a high error of RMSE (0.2844), denoting the high contribution of this parameter. Similarly, $Q_{inf\text{-}bypass}$ is more effective parameter than $Q_{inf}$ when compared the correlation coefficients 0.7887 and 0.7514, respectively, because anaerobic unit performance is directly related to feed rate to the reactor as indicated also by $Q_{gas}$.

Validation pattern of estimated and measured composite values of system output variable of normalized chemical oxygen demand ($COD_{norm}$) for the best model of Table 3 are shown as a function of time in Fig. 4 after the training step. Considering the fact that input variable matrix is constructed of only a few on-line variables, which in fact did not possess any information about the progress of the biological digestion except $Q_{gas}$, the prediction performance of the model was quite acceptable.

In the second case, three-valued phase (0–0.5–1) vector and recent values of output variable for last 5 days were again added to the on-line input variables matrix of the model and step-wise evaluation of all input variables is presented in Table 4.

In Table 4, correlation coefficient for the model including three-valued phase vector, the history of last 5 days and all on-line variables decreased to 0.7670 from previous 0.8097 attained in Table 3 and accordingly the RMSE increased. Interestingly when $Q_{inf}$ was omitted from the input variables matrix, correlation coefficient was found as 0.6321 showing the importance of this parameter in this case. When $Q_{inf}$, $Q_{gas}$, three valued phase vector and $Q_{anarecycle}$ variables were excluded from on-line input variables matrix, correlation coefficients were dropped down to 0.6321, 0.6821, 0.6824, and 0.6846, respectively, showing the effective weight of $Q_{inf}$. In the second case, best correlation coefficient was obtained as 0.7982,

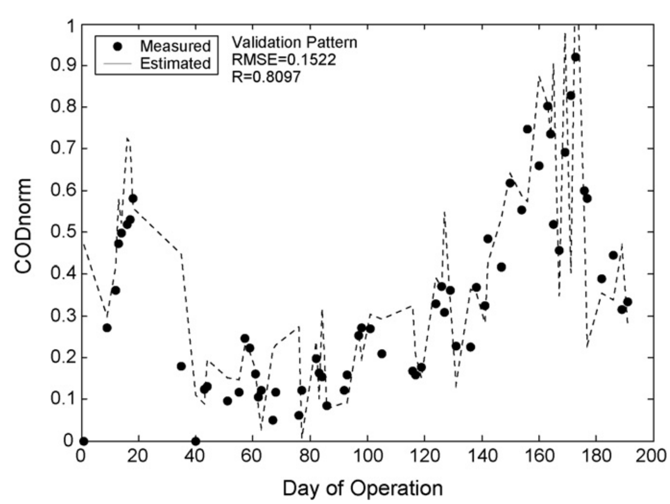


**Fig. 4.** Validation pattern of variations of estimated and measured values of system output variable $COD_{norm}$ with time after training step from the model including two-valued phase vector, history of last 5 days and all on-line variables.

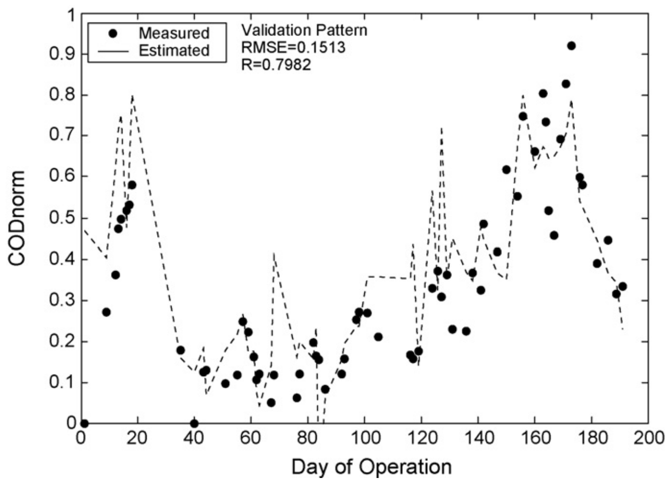


**Fig. 5.** Validation pattern of variations of estimated and measured values of system output variable $COD_{norm}$ with time after training from the model excluding history of last 5 days but including three-valued phase vector and all on-line variables.

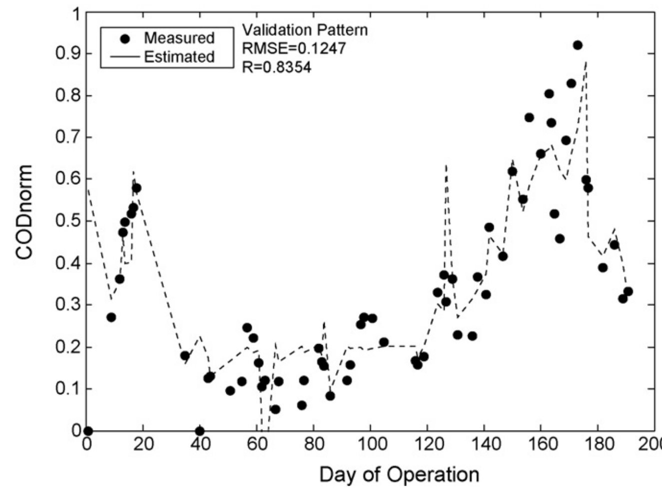


**Fig. 6.** Validation pattern of variations of estimated and measured values of system output variable $COD_{norm}$ with time after training step from the model including only on-line variables except $Q_{inf}$.

when history of last 5 days variable was expelled from the on-line input variable matrix.

In both cases, exclusion of two and three-valued phase vectors from the matrix yielded lower correlation coefficients than when history variable excluded, explaining the working nature of treatment plant in different phases of sugar processing operation period. For validation after training step, estimated and measured composite values of system output variable $COD_{norm}$ as a function of time is graphed in Fig. 5 for the case excluding of history of last 5 days variable but the rest. As a general trend in Fig. 5, peak values around unity cannot be represented by the model adequately, which is quite a common case for similar simulation studies [19].

For the last case, in order to assess the pure contribution of on-line input variables, only on-line variables were used in the variable matrix. For this case, model assessment of on-line input variables is presented in Table 5.

In Table 5 with all on-line input variables, RMSE and R between estimated and measured $COD_{norm}$ values were found 0.1431 and 0.8115, respectively, having a better fit than previous two cases. Furthermore, the exclusion of $Q_{anarecycle}$ yielded a correlation coefficient 0.8114, which was not significantly smaller than 0.8115 for the model containing all on-line variables, denoting that $Q_{anarecycle}$ is a weak parameter in the on-line input variable matrix. In the table, when $Q_{inf\text{-}bypass}$, $Q_{gas}$ and, $CH_4\%$ variables were omitted one at a time from the input variables matrix, correlation coefficients were dropped to 0.7779, 0.7952 and 0.7962, respectively. These variables are directly related to anaerobic reactor performance and therefore must be kept in the model.

Out of three cases evaluated in this study, the case with only on-line variables excluding $Q_{inf}$ yielded the highest R (0.8354) and smallest RMSE (0.1247). For validation, after training step of the third case excluding $Q_{inf}$ estimated and measured composite values of $COD_{norm}$ is depicted in Fig. 6.

Considering the fact that limited number of on-line input variables (4) was used in the variable matrix and the matrix did not contain any information about the wastewater inlet COD concentration, acquired RMSE of 0.1247 and R of 0.8354 in Fig. 6 were of acceptable level of fit. The addition of two- and three-valued

**Table 4**
Evaluation of on-line input variables with the inclusion of three-valued phase vector and history of last 5 days

| Excluded variable | Training (fuzout 2/dataout) | | Validation (valid fuzout 2/valid dataout) | |
|---|---|---|---|---|
| | RMSE | R | RMSE | R |
| None | 2.3691E−005 | 1.0 | 0.1969 | 0.7670 |
| History of 5 days | 0.0382 | 0.9835 | 0.1513 | 0.7982 |
| Phase vector (three-valued) | 6.5211E−005 | 1.0 | 0.1924 | 0.6824 |
| $CH_4$ | 3.5532E−004 | 1.0 | 0.2326 | 0.7465 |
| $Q_{gas}$ | 0.049 | 0.9997 | 0.2346 | 0.6821 |
| $Q_{anarecycle}$ | 0.0394 | 0.9824 | 0.2065 | 0.6846 |
| $Q_{inf\text{-}bypass}$ | 6.9030E−005 | 1.0 | 0.1613 | 0.7851 |
| $Q_{inf}$ | 5.2030E−004 | 1.0 | 0.2440 | 0.6321 |

**Table 5**
Evaluation of only on-line input variables

| Excluded variable | Training (fuzout 2/dataout) | | Validation (valid fuzout 2/valid dataout) | |
|---|---|---|---|---|
| | RMSE | R | RMSE | R |
| None | 0.0462 | 0.9757 | 0.1431 | 0.8115 |
| $CH_4$ | 0.0608 | 0.9576 | 0.1418 | 0.7962 |
| $Q_{gas}$ | 0.1010 | 0.8781 | 0.1372 | 0.7952 |
| $Q_{anarecycle}$ | 0.0955 | 0.8918 | 0.1316 | 0.8114 |
| $Q_{inf\text{-}bypass}$ | 0.0556 | 0.9646 | 0.1469 | 0.7779 |
| $Q_{inf}$ | 0.0816 | 0.9222 | 0.1247 | 0.8354 |

phase vector and history of last 5 days to the on-line input variable matrix apparently did not enhance the prediction performance, in contrast to our previous study where all available system variables were used in the model [18].

In literature, there has not been any study done yet for ANFIS modeling of a real scale anaerobic wastewater treatment with on-line variables to compare our results. Nevertheless, there have been some other neural network techniques for the prediction of performance in wastewater treatment plants. Tomida et al. [20], applied recursive fuzzy neural network to predict COD effluent for a real scale activated sludge treatment plant using on-line variables of operation time, influent COD, effluent flow rate, air-flow rate and temperature through a year, and they achieved an estimation of effluent COD concentration ranging from 0 to 20 mg/L with an average error of 0.40 mg/L. Kim et al. [21] applied polynomial neural network, PNN, and rule base compensator to estimate the $NO_x$–N and $NH_4$–N concentrations of a pilot scale sequencing batch reactor, SBR, by using only on-line values of oxidation reduction potential, dissolved oxygen and pH as on-line input variables, and they obtained estimations for $NO_x$–N (0–120 mg/L) and $NH_4$–N (0–80 mg/L) with RMSEs in the range 1.23–8.4 and 2.54–11.22, respectively, depending on loading rate. Grieu et al. [22] proposed a procedure based on a multi-layer perceptron neural network to obtain on-line estimation of influent and effluent concentrations of COD and $NH_4$ for a real scale activated sludge wastewater treatment plant. Using on-line variables of influent flow rate, air-flow rate and dissolved oxygen concentration together with off-line parameters, they estimated effluent concentration of COD and $NH_4$ with average relative errors in the range 6.6–15.2%. Holubar et al. [23,24] demonstrated that pH, volatile fatty acid, acetic acid, propionic acid, gas from the 20 L laboratory scale anaerobic CSTR process can be effectively modeled by the feed forward back propagation, FFBP, neural network. Three to nine input variables were used to estimate the parameters within a regression coefficient range of 0.80–0.90, but only pH, gas production and gas composition parameters were monitored on-line in Holubar's studies. In their work, they achieved to develop a decision support system that contains FFBP model of anaerobic system and a searching algorithm.

The results of our study cannot be compared with those of the literature due to the different number and type of the variables, size, type of treatment and highly fluctuated nature of the data, however, prediction power of our model is not lower than those of literature. Conclusively, acceptable correlation coefficient of 0.8354 and root mean square error of 0.1247 were acquired in our study for the estimation of effluent COD using only four on-line variables, $Q_{\text{inf-bypass}}$, $Q_{\text{anarecycle}}$, $Q_{\text{gas}}$ and *CH4*%, already measurable parameters of a real scale anaerobic wastewater treatment plant. This developed ANFIS model may be integrated into an advanced control system for the treatment plant using different control strategies with further work.

## 4. Conclusion

This paper presents a neural fuzzy model of ANFIS using only on-line input variables of *CH4*%, $Q_{\text{gas}}$, $Q_{\text{anarecycle}}$, $Q_{\text{inf-bypass}}$ and $Q_{\text{inf}}$ to predict effluent COD for the operation of a real scale industrial anaerobic wastewater treatment plant of a sugar factory, being first in literature. The model did not necessitate the measurement of COD, an off-line parameter to be determined experimentally, and made it possible to monitor the treatment performance of the unit as well as operating conditions for a reliable treatment process.

Two new input variables, phase vectors of the plant operation and the history of effluent COD, the output variable, were added into the input variables matrix with the anticipation of increasing the fitness of the model. However, these two descriptive variables did not enhance the prediction power of the model. Produced results from the developed ANFIS model was satisfactory for the estimation of effluent COD in the case of only limited on-line variables were available. Acceptable correlation coefficient of 0.8354, and root mean square error of 0.1247 were found for the assessment of the system output variable, effluent COD, in the case of excluding inlet volumetric flow rate of the wastewater treatment plant from the on-line input variable matrix. The information provided by the prediction procedure is sufficiently reliable for the plant monitoring and performance check. The developed ANFIS model may be integrated into an advanced control system of the anaerobic treatment plant using different control strategies with further work.

## Acknowledgments

The authors would like to express their gratitude to Turkish Sugar Factories Cooperation and Sugar Institute for the financial support and permission to use plant data in this study. This study has also been supported by the Project Management Unit of Akdeniz University, Turkey.

## Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.cej.2008.03.008.

## References

[1] J.P. Steyer, J.C. Bouvier, T. Conte, P. Gras, P. Sousbie, Evaluation of a four year experience with a fully instrumented anaerobic digestion, Water Sci. Technol. 45 (4/5) (2002) 495–502.
[2] G.C. Premier, R. Dinsdale, A.J. Guwy, F.R. Hawkes, D.L. Hawkes, S.J. Wicox, A comparison of the ability of black box and neural network models of arx structure to represent a fluidized bed anaerobic digestion process, Water Res. 33 (4) (1997) 1027–1037.
[3] A. Müller, S. Marsili-Libelli, A. Aivasidis, T. Lloyd, S. Kroner, C. Wandrey, Fuzzy control of disturbances in a wastewater treatment plant, Water Res. 31 (12) (1997) 3157–3167.
[4] J. Flores, B. Arcay, J. Arias, An intelligent system for distributed control of an anaerobic wastewater treatment process, Eng. Appl. Artif. Intell. 13 (2000) 485–494.
[5] S.R. Harper, M.T. Suidan, Anaerobic treatment kinetics: discussers' report, Water Sci. Technol. 24 (8) (1991) 61–78.
[6] J.H. Tay, X. Zhang, A fast predicting neural fuzzy model for high-rate anaerobic wastewater treatment systems, Water Res. 34 (11) (2000) 2849–2860.
[7] C. Aubrun, D. Theilliol, J. Harmand, J.P. Steyer, Software sensor design for COD estimation in an anaerobic fluidized bed reactor, Water Sci. Technol. 43 (7) (2001) 115–120.
[8] A. Karama, O. Bernard, A. Genovesi, D. Dochain, A. Benhammou, J.P. Steyer, Hybrid modeling of anerobic wastewater treatment processes, Water Sci. Technol. 43 (1) (2001) 43–50.
[9] L.A. Belanche, J.J. Valdes, J. Comas, I.R. Roda, M. Poch, Towards a model of input–output behavior of wastewater treatment plants using soft computing techniques, Environ. Model. Softw. 14 (1999) 409–419.
[10] R.J.S. Jang, ANFIS: adaptive-network-based fuzzy inference system, IEEE Trans. Syst. Man and Cybern. 23 (3) (1993) 665–685.
[11] R.J.S. Jang, C.T. Sun, Functional equivalence between radial basis function networks and fuzzy inference system, IEEE Trans. Neural Netw. 4 (1) (1993) 156–159.
[12] R.J.S. Jang, C.T. Sun, Neuro-fuzzy modeling and control, Proc. IEEE 83 (3) (1995) 378–405.
[13] R.J.S. Jang, N. Gulley, Fuzzy Logic Toolbox User's Guide, The Mathworks Inc., 24 Prime Park Way, Natick, MA, 1995, 01760–1500.
[14] J.H. Tay, X. Zhang, Neural fuzzy modeling of anaerobic biological wastewater treatment systems, J. Environ. Eng. 125 (12) (1999) 1149–1159.
[15] E.M.T. El-Mansi, C.F.A. Bryce, Fermentation, Microbiology and Biotechnology, Taylor & Francis Inc., 325, Chestnut Street, 8th Floor, Philadelphia, PA, USA, 1999, ISBN 0-7484-0733-2.
[16] P. Weiland, A. Rozzi, The startup, operation and monitoring of high rate anaerobic treatment systems: discusser's report, Water Sci. Technol. 24 (8) (1991) 257–277.

[17] APHA, AWWA, WEF, Standard Methods for the Examination of Water and Wastewater, 19th ed., APHA, Washington, 1995.
[18] A. Perendeci, S. Arslan, A. Tanyolaç, S.S. Çelebi, Evaluation of input variables in adaptive-network based fuzzy inference system (anfis) modeling for an anaerobic wastewater treatment plant under unsteady state, J. Environ. Eng. 133 (7) (2007) 765–771.
[19] D.S. Lee, C.O. Jeon, J.M. Park, K.S. Chang, Hybrid neural network modeling of a full-scale industrial wastewater treatment process, Biotechnol. Bioeng. 78 (6) (2002) 670–682.
[20] S. Tomida, T. Hanai, H. Honda, T. Kobayashi, Construction of COD simulation model for activated sludge process by recursive fuzzy neural network, J. Chem. Eng. Jpn. 34 (3) (2001) 369–375.
[21] Y. Kim, J. Ko, K.M. Boo, H.J. Woo, C. Kim, Software sensor using PNN model and knowledge base for sequencing batch reactor, in: Liong, Phoon, Babovic (Eds.), Proceeding of the 6th International Conference on Hydroinformatics, World Scientific Publishing Company, 2004, ISBN 981-238-787-0.
[22] S. Grieu, A. Traore, M. Polit, J. Colprim, Prediction of parameters characterizing the state of a pollution removal biologic process, Eng. Appl. Artif. Intell. 18 (2005) 559–573.
[23] P. Holubar, L. Zani, M. Hager, W. Froschl, Z. Radak, R. Braun, Advanced controlling of anaerobic digestion by means of hierarchical neural networks, Water Res. 36 (2002) 2582–2588.
[24] P. Holubar, L. Zani, M. Hager, W. Froschl, Z. Radak, R. Braun, Start up and recovery of a biogas reactor using a hierarchical neural network-based control tool, J. Chem. Technol. Biotechnol. 78 (2003) 847–854.