World Academy of Science, Engineering and Technology
International Journal of Civil and Environmental Engineering
Vol:5, No:6, 2011

Combining Fuzzy Logic and Neural Networks in Modeling Landfill Gas Production

Mohamed Abdallah, Mostafa Warith, Roberto Narbaitz, Emil Petriu, and Kevin Kennedy

***Abstract*—**Heterogeneity of solid waste characteristics as well as the complex processes taking place within the landfill ecosystem motivated the implementation of soft computing methodologies such as artificial neural networks (ANN), fuzzy logic (FL), and their combination. The present work uses a hybrid ANN-FL model that employs knowledge-based FL to describe the process qualitatively and implements the learning algorithm of ANN to optimize model parameters. The model was developed to simulate and predict the landfill gas production at a given time based on operational parameters. The experimental data used were compiled from lab-scale experiment that involved various operating scenarios. The developed model was validated and statistically analyzed using F-test, linear regression between actual and predicted data, and mean squared error measures. Overall, the simulated landfill gas production rates demonstrated reasonable agreement with actual data. The discussion focused on the effect of the size of training datasets and number of training epochs.

***Keywords*—**Adaptive neural fuzzy inference system (ANFIS), gas production, landfill

I. Introduction

In the last few decades, biogas generation in landfills has been modeled via different methods that focused on describing the physical, chemical and biological processes inside landfills. An accurate landfill gas generation model can be used to design gas recovery systems, analyze the economic viability of gas recovery operations, and assess potential environmental impacts. Various mathematical models for simulating landfill gas are presented in the literature [1]-[3]. However, the uncertainties of solid waste characteristics, as well as the complex physical, chemical, and biological processes taking place within the landfill ecosystem, motivated advanced modeling techniques to be applied; stochastic modeling [4], [5], and fuzzy logic systems [6], [7].

Recently, soft computing methods such as the fuzzy logic (FL) and the artificial neural networks (ANN) have been shown to be powerful modeling tools with several advantages over traditional techniques. However, their application is limited due to certain weaknesses in their modeling capabilities. The integration of the FL and ANN, referred to as the Neuro-Fuzzy System, combines the merits of both systems, and is a more powerful modeling tool [8]. The basic idea of incorporating both systems is to design an architecture that uses a FL system to represent knowledge in an interpretable manner and implements the learning algorithms of ANN to optimize its parameters. Hence, the drawbacks of both systems, such as the black box behavior of ANN and the problems of tuning the membership values in FL systems, could be avoided.

The implementation of ANN as an adaptive learning technique in FL systems is known as the adaptive neuro-fuzzy inference system (ANFIS). The adaptive technique of ANN optimizes the parameters associated with the membership functions throughout a learning process. The computation of these parameters (or their adjustment) is facilitated by a gradient vector which provides a measure of the adequacy of the inference system in modeling the input/output data which can be optimized to tune the parameters so that the error between actual and simulated outputs is minimized [9].

The present work develops an ANFIS model to simulate biogas generation in lab-scale landfill cells. The model incorporates the effect of leachate recirculation and sludge addition rates as controllable input variables that can predict the output variable, biogas generation rate. The model is trained, verified, and validated using published experimental data of lab-scale landfill cells. Additionally, this paper focuses on assessing the effect of the size of the training data and the number of training epochs on the performance and stability of the ANFIS model.

II. Methodology

A. Experimental Data

Experimental data were compiled from the work by [10] which involved 8 liters lab-scale landfill cells containing a total mass and density of waste of 2.50 kg/reactor and 350 kg/m^3, respectively. Major components of the waste were paper (36.6%), food (36.2%), and yard trimmings (27.2%). The cells were recirculated with different rates of leachate and sludge that remained unchanged along the experiment.

Experimental measurements included temporal biogas generation rate as well as supplemental addition rates. The

M. Abdallah is with the Department of Civil Engineering, University of Ottawa, Ottawa, ON K1N6N5 CANADA (phone: 613-562-5800; fax: 613-562-5173; e-mail: mesabdallah@gmail.com).
M. Warith is with the Department of Civil Engineering, University of Ottawa, Ottawa, ON K1N6N5 CANADA (phone: 613-562-5800; fax: 613-562-5173; e-mail: mwarith@uottawa.ca).
R. Narbaitz is with the Department of Civil Engineering, University of Ottawa, Ottawa, ON K1N6N5 CANADA (phone: 613-562-5800; fax: 613-562-5173; e-mail: narbaitz@uottawa.ca).
E. Petriu is with the School of Information Technology and Engineering (SITE), University of Ottawa, Ottawa, ON K1N6N5 CANADA (phone: 613-562-5800; fax: 613-562-5173; e-mail: petriu@site.uottawa.ca).
K. Kennedy is with the Department of Civil Engineering, University of Ottawa, Ottawa, ON K1N6N5 CANADA (phone: 613-562-5800; fax: 613-562-5173; e-mail: kevin.kennedy@uottawa.ca).

experimental results resulted in 350 data vectors in the form of (time, leachate recirculation rate, sludge addition rate, biogas generation rate). The compiled data vectors were split into three independent groups for training, checking, and testing of the neuro-fuzzy inference system. It should be noted that the grouping procedures were carried out without discrimination using a simple resampling method [11]. Additionally, in order to assess the effect of the training data size on ANFIS, the random value generator in Microsoft Excel was used to build 6 training datasets of different sizes (50, 100, 150, 200, 250, and 300 data vectors) in addition to a single checking and testing datasets each of 50 vectors.

B. Fuzzy Logic Controller

The components of fuzzy logic controller structure include: (1) inputs, (2) fuzzifier unit, (3) data base, (4) rule base, (5) fuzzy inference engine, (6) defuzzifier unit, and (7) outputs.

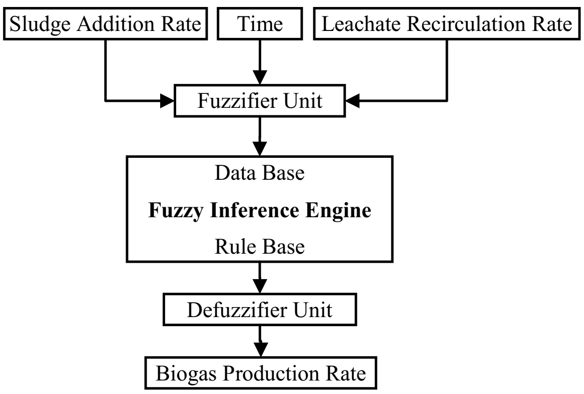


Fig. 1 Typical structure of a fuzzy logic controller

As shown in Fig. 1, the input variables included time (or operating phase), leachate recirculation, and sludge addition. Initially, the crisp values of the inputs are mapped into their corresponding universes of discourse in the fuzzifier unit. In other words, the numerical representation of the inputs is converted to suitable linguistic terms within the predefined membership functions. Any membership function (μ) is represented by a real number ranging between 0 and 1. Because of its smoothness and concise notation, the Gaussian membership function was used for specifying inputs' fuzzy sets. It is given by:

$$\mu_i(x) = exp[-(x-c)^2/2\sigma^2] \tag{1}$$

Where c is the mean and σ^2 is the variance of the i^{th} fuzzy set. Fig. 2 shows the membership functions defined for the three inputs of the ANFIS model. The linguistic labels used to describe the input rates were Low Rate (LR), Medium Rate (MR), and High Rate (HR). Time was defined by the five basic phases that typically characterize the landfill life span; Initial (I), Transition (T), Acid Formation (AF), Methane Fermentation (MF), and Final Maturation (FM). For more details on the basic characteristics of each phase, refer to [12].

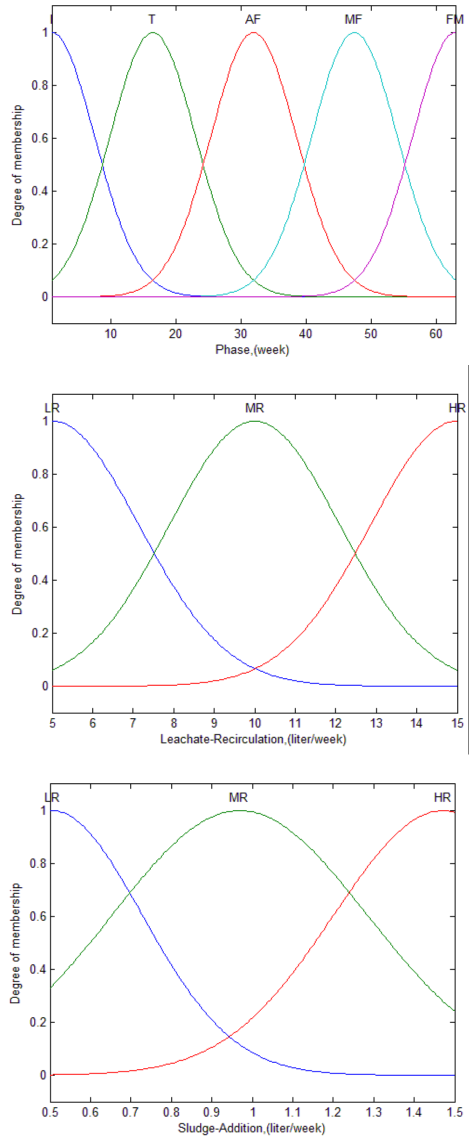


Fig. 2 Membership functions for the inputs of the ANFIS model: phase (time), leachate recirculation, and sludge addition.

The knowledge is introduced to the system in terms of data base and rule base. The data base is provided by defining the membership functions of the fuzzy sets used for each system variable. The rule base consists of fuzzy rules that describe the system behavior and replace the mathematical modeling of the system. These rules are built in the form of IF-THEN statements. In the present model, 45 statements were created to define the expected behavior at different operating scenarios. Based on that experienced knowledge, the fuzzy inference engine processes the fuzzy output. Finally, the defuzzifier unit is responsible for weighting and combining a number of fuzzy sets resulting from the fuzzy inference process in a calculation which gives a single crisp value for the output. In this work, defuzzification was processed according to the weighted average method in which the output is obtained by the weighing the average of each output of the set of rules stored in the knowledge base of system. The output is computed as;

$$O = \sum_{i=1}^{n} \mu^i W_i / \sum_{i=1}^{n} \mu^i \quad (2)$$

Where O is the defuzzified output, μ^i is the membership of the output of each rule, and w_i is the weight associated with each rule. This method is computationally faster and easier and gives fairly accurate result. After all the discussed steps and as a final point, the model gives the predicted output, biogas production rate, in its numerical (crisp) form.

C. Fuzzy Inference System

The present study applies the Takagi-Sugeno method for the fuzzy inference system. The Sugeno output membership functions are either linear or constant making them compact and computationally efficient and compatible to the use of adaptive techniques such as ANN. The learning algorithm of ANN is employed to tune the parameters of the membership functions along a training process.

D. ANFIS Architecture

The developed architecture included 11 input membership functions, 45 output membership functions, and 45 fuzzy rules. Due to the numerous interactions between system components in its complete structure, a simplified adaptive neuro-fuzzy inference system is shown in Fig. 3. The simplified ANFIS model includes two inputs T and P and one output f. The architecture of ANFIS, developed by [8], includes five layers;

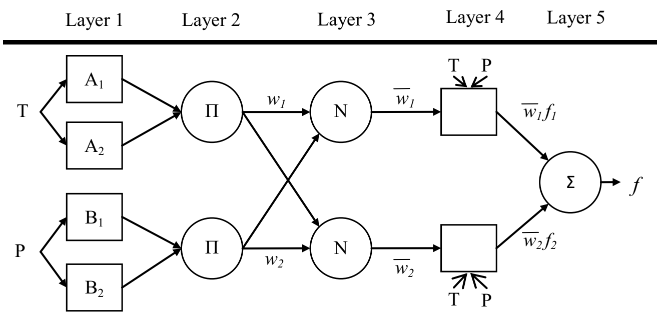


Fig. 3 Architecture of a minimized ANFIS model

Layer 1: Every node in this layer is an adaptive node with a node function that can be computed as:

$$O_i^1 = \mu_{Ai}(T) \quad (3)$$

Where T is the input to node i, and A_i is the linguistic label set associated with this node.

Layer 2: Every node in this layer is a fixed node labeled (Π) that represents the firing strength (W) of a fuzzy rule. The output of each node is the product of the incoming signals:

$$O_i^2 = W_1 = \mu_{Ai}(T)\mu_{Bi}(P) \quad (4)$$

Layer 3: Every node in this layer is a fixed node (N) which calculates the ratio of each rule's firing strength to the sum of all rules' firing strengths. The outputs of this layer are called normalized firing strengths and can be computed as:

$$O_i^3 = \overline{W_i} = \frac{W_1}{W_1 + W_2}, i = 1,2 \quad (5)$$

Layer 4: Every node in this layer is an adaptive node with a node function (i.e., linear combination of input variables). If p_i, q_i, r_i is the parameter set, then:

$$O_i^4 = \overline{W_i} f_i = \overline{W_i}(p_i T + q_i P + r_i) \quad (6)$$

Layer 5: The single node in this layer is a fixed node that computes the overall output as the summation of all incoming signals:

$$O_i^5 = \sum_i \overline{W_i} f_i = \frac{\sum_i W_i f_i}{\sum_i W_i} \quad (7)$$

E. Implementation of ANFIS Model

The ANFIS model was built and trained using the Fuzzy Logic Toolbox in MATLAB™. Practically, several conditions restrain the application of the ANFIS model. Firstly, the inference system must be Takagi-Sugeno type, and the system should have a single output obtained using weighted average defuzzification technique. Moreover, all output membership functions must be of the same type, either linear or constant, and there should be no rule sharing or weighing. Table I illustrates the main methods and characteristics of the developed ANFIS model based on these conditions.

TABLE I
MAIN METHODS AND CHARACTERISTICS OF THE ANFIS MODEL

Parameter	Type / Method
Fuzzy inference system type	Takagi-Sugeno
AND method	Min
OR method	Max
Defuzzification method	Weighted Average
Optimization method	Back propagation
Output membership function type	Linear

The flowchart of the ANFIS training algorithm in the Fuzzy Logic Toolbox is shown in Fig. 4. The process starts by loading the training and checking datasets which include input and output data vectors. Each input/output pair contains three inputs (time, sludge addition rate, and leachate recirculation rate) and one output (biogas generation rate). An error tolerance (ET) value is defined for the maximum acceptable difference between the actual and simulated output. The model starts the training process with the initial parameters of the membership functions, and the error for each data pair is calculated. If this error is larger than the ET value, the membership parameters are adjusted through an optimization step, otherwise, the process ends. Simultaneously, the error of the checking dataset is calculated. Typically, it decreases down to a certain point, and then increases. This overturn represents the point of model overfitting. The program chooses the model parameters associated with the minimum checking error. Finally, the model is validated against independent testing dataset.

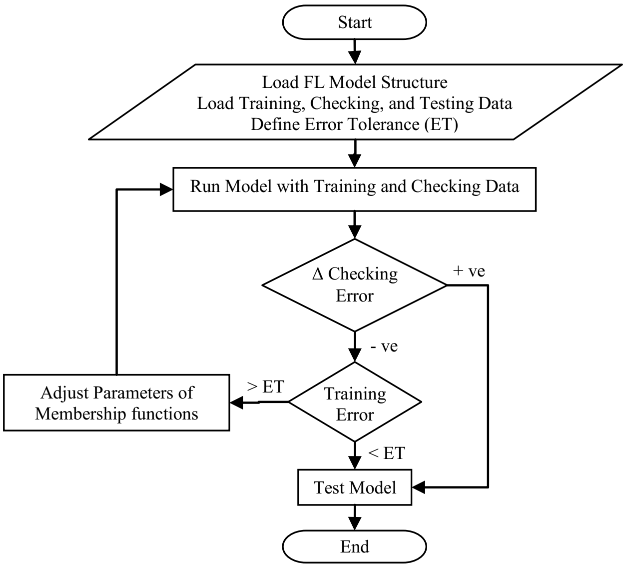


Fig. 4 Flowchart of the developed ANFIS model

The six training datasets, previously discussed, were used to train the ANFIS model separately. As a result, six sub-models were created with the same inputs, output, and basic structure. Table II shows the training data size, number of training epochs, and notation of the ANFIS sub-models.

TABLE II
NOTATION AND MAIN FEATURES OF THE ANFIS SUB-MODELS

ANFIS Model	Size of training dataset	Training Epochs
M-50	50	15
M-100	100	120
M-150	150	250
M-200	200	250
M-250	250	250
M-300	300	250

F. Model Evaluation Criteria

The model is evaluated statistically using a group of criteria that was established prior to the evaluation process and included *F*-test, linear regression between actual and predicted data, and mean squared error (MSE) measures. Firstly, in comparing two independent samples, the *F*-Test provides a measure for the probability that they have the same variance. The null hypothesis, H_0: $\sigma_1^2 = \sigma_2^2$, and the alternate hypothesis, H_A: $\sigma_1^2 \neq \sigma_2^2$. The "≠" sign indicates a 2-tailed test that is interested in both cases; $\sigma_1^2 > \sigma_2^2$ and $\sigma_1^2 < \sigma_2^2$. Secondly, the regression estimates of the intercept (a) and the slope (b) are good indicators of accuracy; the closer to zero and unity, respectively, the higher the accuracy. On the other hand, the correlation coefficient (R) is a good indicator of precision; the higher the R, the higher the precision [13].

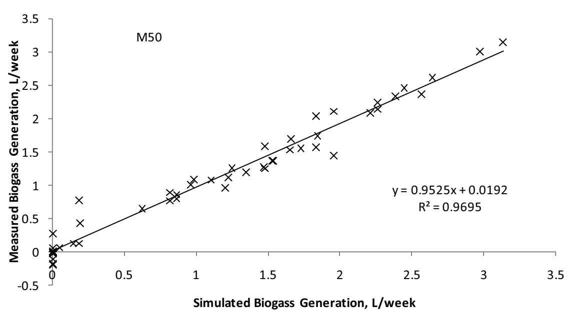
Finally, the MSE is the mean of the squared errors around the regression line in a plot of model simulation against measured values. Following the approach developed by [14], the MSE was partitioned into three components to achieve further understanding of model performance; square bias (SB), non-unity slope (NU), and lack of correlation (LC). These MSE components, which add up to give the MSE, have simple and distinct geometrical interpretation. SB, NU, and LC indicate the translation, rotation, and scattering around the regression line, respectively. Additionally, the root mean square error (RMSE) is calculated by square-rooting the MSE. The RMSE indicates the mean difference between observed and predicted values. In order to facilitate the assessment, the RMSE is normalized by dividing its value by the mean of the measured data.

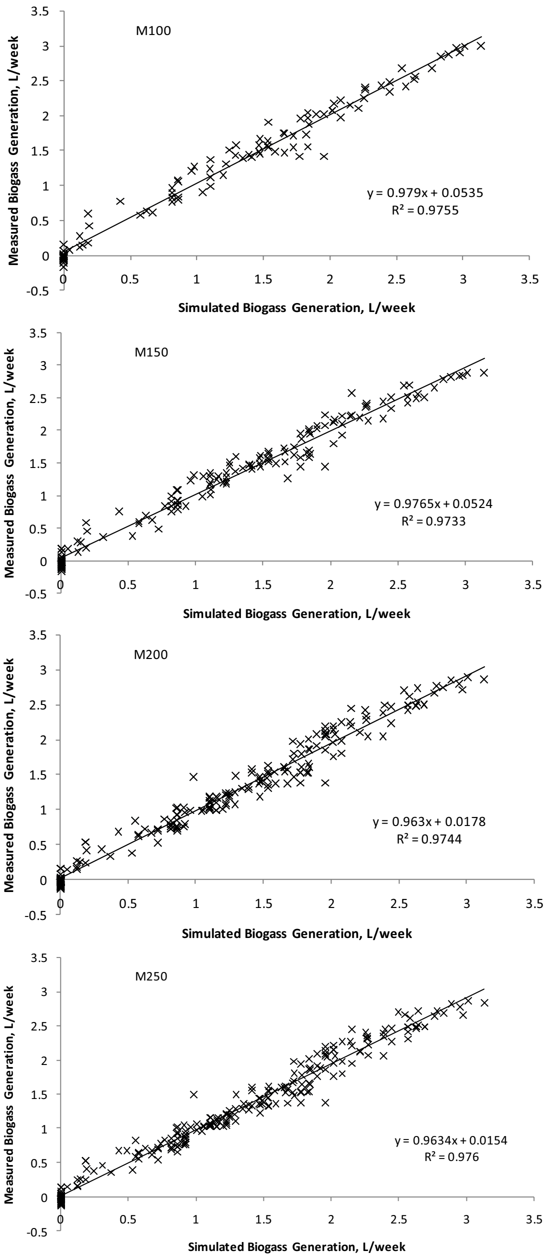
III. DISCUSSION

This work aims at evaluating the modeling capabilities of ANFIS model in simulating biogas production in landfills. The proposed model predicts the biogas generation rate in terms of time, leachate recirculation, and sludge addition. First, the model is verified by fitting adequacy of its predictions to the training datasets. Then, the model is validated using the testing dataset. Finally, two critical issues concerning the ANN learning algorithm are discussed; (1) the effect of the training data size on model efficiency, and (2) the effect of the number of training epochs on model stability and performance.

A. Model Verification

Fig. 5 shows the linear regression between model-based predictions and measured data for the six ANFIS sub-models. The values of slope, intercept, and coefficient of determination for the regression lines are shown on the same figures. It can be observed that, in all sub-models, the slope was close to unity and the intercept was close to zero. This could be a positive indication of the model accuracy. In addition, the slope did not exceed unity and the intercept was positive in all sub-models. This demonstrates that the ANFIS model underestimated the measured data in general. The model predictions were in excellent agreement with the training datasets (average correlation coefficient of 0.98). However, it should be clarified that this ideal correlation was achieved because the model was tuned on those datasets.





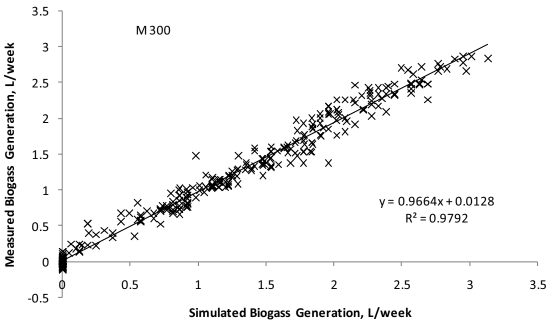


Fig. 5 Linear regression between modeled and actual data for ANFIS sub-models: M-50, M-100, M-150, M-200, M-250, and M-300

As shown in the statistical analyses in Table III, the simulation results were sufficiently agreeable to the experimental data based on paired *F*-test at a 95% confidence level ($F_{calculation} < F_{critical}$). Based on the MSE partitioning, LC was the main contributing component to MSE. The average percentages of SB, NU, and LC to MSE were 4%, 0%, and 96%, respectively. Therefore, deviations could be attributed mostly to scattering around the regression line. The difference between actual and simulated values was expressed in terms of the normalized RMSE. ANFIS model achieved an acceptable normalized RMSE that ranged between 11.24% and 14.42%.

B. Model Verification

Model validation is used to examine the potential of the ANFIS model under untrained range of operating conditions. Commonly, the main concern in model validation is selecting a data set that is both representative of the data via which the model was developed, yet sufficiently distinct from it otherwise the validation process becomes meaningless. The biogas generation predictions of the ANFIS sub-models are plotted together with actual experimental points in Fig. 6. Generally, the ANFIS model fairly reproduced the generation rates for most of the testing scenarios.

During the testing procedures, one critical observation was noticed regarding the model response to the noise and digressions which frequently occur due to the measuring accuracy and/or operating troubleshooting. As shown at week 21, the ANFIS model dealt with the sudden drop in biogas production nonsensically, yet efficiently. This proves that the ANFIS system works well only if the data used for training its membership function parameters is noiseless and fully representative of the features of the modeled system.

TABLE III
STATISTICAL TESTING OF THE ANFIS MODEL AT DIFFERENT TRAINING SIZES

Model	*F*-test		Mean Square Error						
	$F_{calculated}$	$F_{critical}$	P	SB	NU	LC	MSE	RMSE	N-RMSE
M-50	1.069	1.607	0.409	0.001	0.000	0.026	0.027	0.166	14.423
M-100	1.018	1.394	0.465	0.001	0.000	0.020	0.022	0.147	11.236
M-150	1.021	1.310	0.451	0.001	0.000	0.021	0.022	0.149	11.846
M-200	1.051	1.263	0.364	0.001	0.001	0.019	0.021	0.145	12.290
M-250	1.052	1.232	0.346	0.001	0.001	0.017	0.019	0.138	11.790
M-300	1.049	1.210	0.341	0.001	0.001	0.015	0.017	0.132	11.380

P, probability; SB, squared bias; NU, non-unity slope; LC, lack of correlation; MSE: mean square error; RMSE: root mean square error; N-RMSE: normalized root mean square error (in percent).

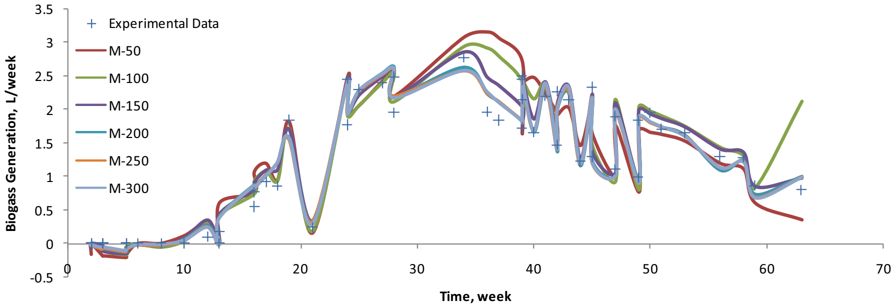


Fig. 6 Actual and simulated biogas generation rate for the ANFIS sub-models with time

The simulations of ANFIS are tested using three statistical measures; *F*-test, linear regression, and mean square error as illustrated in Table IV. The correlation coefficient ranged from 0.93 and 0.98 with an average value of 0.97. The simulation results were sufficiently agreeable to the actual data based on *F*-test at a 95% confidence level ($F_{calculation} < F_{critical}$). The sub-models achieved an acceptable average N-RMSE of 17.36%.

C. Size of Training Data

There was not much difference between the performances of ANFIS sub-models; all sub-models achieved high correlation with their training datasets. The best fit was achieved by M-300, the most trained model, whereas the least trained model, M-50, scored the lowest fit. During validation stage, the most distorted sub-models were M-50 and M-100. Statistically, the number of training vectors highly affected the performance of ANFIS to a certain point. Starting from M-150, a great enhancement was observed in terms of all statistical measures. However, beyond M-150, the model did not show significant improvement in response to a larger training dataset.

D. Number of Training Epochs

Fig. 7 shows the progress of the calculated error, during the learning algorithm, against the number of training epochs. The training dataset is indicated by the blue stars, whereas, the checking dataset is indicated by the blue circles. The main reason for using the checking dataset is to determine the optimum training effort needed for the system. Usually, the more epochs the system is trained with, the less the error will be until it reaches a certain point where overfitting starts. To avoid this, the program stops the training prior to the overfitting point. This point can be observed at iteration # 15 of M-50 and at iteration # 120 of M-100. It should be noted that the overfitting problem did not occur in other sub-models within 300 epochs.

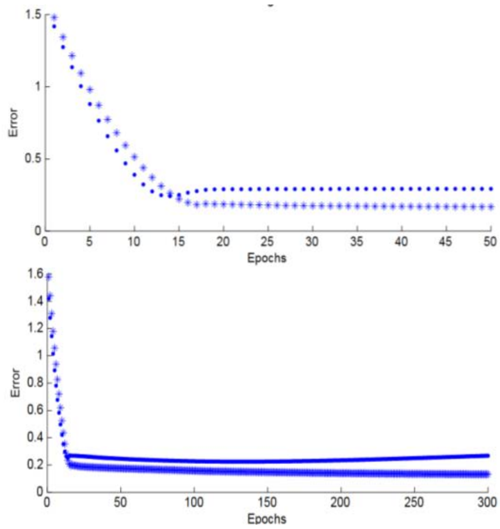


Fig. 7 Training and checking error with training epochs for M-50 (above) and M-100 (below)

IV. CONCLUSIONS

The complexity and heterogeneity of the landfill hinder conventional kinetic methods from modeling the system in an efficient and practical fashion. This work aimed at investigating the applicability of the advanced neural fuzzy technique in simulating and accommodating the uncertainties of such an ecosystem. The ANFIS model described the data that was used in its training stage ideally. Moreover,

TABLE IV
STATISTICAL TESTING OF THE ANFIS MODEL FOR THE TESTING DATASET

Model	F-test		Linear Regression			Mean Square Error		
	$F_{calculated}$	$F_{critical}$	a	b	R^2	MSE	RMSE	N-RMSE
M-50	1.249	1.607	0.016	1.040	0.865	0.141	0.375	29.193
M-100	1.169	1.607	0.083	1.024	0.898	0.109	0.33	25.661
M-150	1.072	1.607	0.048	1.020	0.970	0.032	0.178	13.829
M-200	1.012	1.607	0.033	0.979	0.971	0.023	0.152	11.852
M-250	1.005	1.607	0.030	0.983	0.972	0.023	0.151	11.75
M-300	1.001	1.607	0.030	0.986	0.972	0.023	0.152	11.848

a, intercept of regression line; b, slope of regression line; R^2, coefficient of determination.

throughout the validation step, the model was in good agreement with actual data. This attested its capabilities in predicting biogas generation over the presented universe of discourse for the inputs. In general, the model achieved acceptable statistical measures in terms of linear regression, *F*-test, and mean square error. On the other hand, the discussion revealed some limitations in the neural fuzzy model. First, its performance is highly dependent on the quality and quantity of training data as well as the number of training epochs. Besides, the overfitting problem may cause unexpected model distortion. On condition of proper training process, this technique can present a prospective alternative methodology in the modeling of landfill processes.

REFERENCES

[1] R. L., Peer, D. L., Darcy, D. L., Campbell, and P. V., Brook, "Development of an empirical model of methane emissions from landfills," *U.S. EPA, Air Energy Eng. Res. Lab.*, NC, no. EPA/600/SR-92/037, 1992.

[2] A. N., Findikakis, C., Papelis, C. P., Halvadakis, and J. O., Leckie, "Modeling gas production in managed sanitary landfills," *Waste Manage. Res.*, vol. 6, pp. 115-123, 1988.

[3] K. R., Gurijala, P., Sa, and J. A., Robinson, "Statistical modeling of methane production from landfill samples," *Appl. Environ. Micro.*, vol. 63, no. 10, pp. 3797-3803, 1997.

[4] N. K., Copty, D., Ergene, and T. T., Onay, "Stochastic model for landfill gas transport and energy recovery," *J. Environ. Eng.*, vol. 130, no. 9, pp. 1042-1049, 2004.

[5] A. I., Zacharof, and A. P., Butler, "Stochastic modeling of landfill leachate and biogas production incorporating waste heterogeneity: model formulation and uncertainty analysis," *Waste Manage.*, vol. 24, pp. 453-462, 2004.

[6] S., Rendra, L., Fernandes, and M., Warith, "Fuzzy logic simulation of biodegradation of municipal solid waste in simulated aerobic and anaerobic bioreactors landfill," *in Proc. 22nd Int. Conf. Solid Waste Technol. Manage.*, Philadelphia, 2007.

[7] A., Garg, G., Achari, and R. C., Joshi, "A model to estimate the methane generation rate constant in sanitary landfills using fuzzy synthetic evaluation," *Waste Manage. Res.*, vol. 24, pp. 363-375, 2006.

[8] J. S. R. Jang, "ANFIS: adaptive-network-based fuzzy inference system," *IEEE T. Syst.*, vol. 23, no. 3, pp. 665-685, 1993.

[9] J. S. R., Jang, and N., Gulley, *Fuzzy Logic Toolbox for Use with MATLAB*. *Massachusetts*, USA: The MathWork Inc., 1995.

[10] S., Rendra, *Comparative Study of Biodegradation of Municipal Solid Waste in Simulated Aerobic and Anaerobic Bioreactor Landfills*, PhD Thesis, University of Ottawa, Canada, 2007.

[11] P. I. Good, *Introduction to Statistics through Resampling Methods and Microsoft Office Excel*. New Jersy, USA: John Wiley & Sons, Ltd, 2005.

[12] F. G., Pohland, and B., Al-Yousfi, "Design and operation of landfills for optimum stabilization and biogas production," *Water Sci. Technol.*, vol. 30, no. 12, pp. 117-124, 1994.

[13] L. O., Tedeschi, "Assessment of the adequacy of mathematical models," *Agr. Syst.*, vol. 89, pp. 225-247, 2006.

[14] H. G., Gauch, J. T. G., Hwang, and G. W., Fick, "Model evaluation by comparison of model-based predictions and measured values," *Agron. J.*, vol. 95, no. 6, pp. 1442-1446, 2003.