

Application of Adaptive Network-Based Fuzzy Inference System to Predict the Fungal Growth in Food Mycology

何文獻
高雄醫學大學
醫療資訊管理學系
副教授
e-mail :

whho@kmu.edu.tw

方傑
高雄醫學大學
醫療資訊管理學系
學生
e-mail :

alex60510@hotmail.com

盧品君
高雄醫學大學
醫療資訊管理學系
學生
e-mail :

zikupp319@hotmail.com

卜怡云
高雄醫學大學
醫療資訊管理學系
學生
e-mail :

lemonedclaudia@hotmail.com

Abstract

In this paper, an adaptive network-based fuzzy inference system (ANFIS) with the genetic algorithm is used to predict the fungal growth in food mycology. For simultaneously finding the optimal premise and consequent parameters in the ANFIS, the improved genetic algorithm (i.e., hybrid Taguchi-genetic algorithm, HTGA) is applied to directly minimize the training root-mean-squared-error (RMSE) performance criterion. Experimental results show that the HTGA-based ANFIS approach outperforms the neural network (NN) methods reported recently in the literature in terms of prediction accuracy.

Keywords: prediction of fungal growth, adaptive network-based fuzzy inference system, genetic algorithm.

摘要

本文中，應用一基於基因演算法之適應性類神經模糊推論系統方法，來預測食品中真菌的成長。在適應性類神經模糊推論系統中，針對訓練資料的最小化均方根誤差，應用田口基因演算法，同時求解最佳化的前件部與後件部參數。運算的結果顯示，基於基因演算法之適應性類神經模糊推論系統方法，在預測食品中真菌成長的準確性上，優於現有文獻上所提出的類神經網路方法。

關鍵詞：真菌成長預測、適應性類神經模糊推論系統、基因演算法。

1. Introduction

Food spoilage by fungi is still a major concern for both food producers and regulatory agencies [1-3]. Therefore, improvement of food quality and safety demands the development of

appropriate tools allowing prediction of fungal growth. For this reason, Panagou and Kodogiannis [3] used the intelligent modelling methodology such as neural networks (NNs) [4] including a multilayer perceptron (MLP) network and a radial basis function (RBF) NN to model the relationship between the growth parameters (i.e., temperature, pH and water activity) and the desired output (maximum specific growth rate of *M. ruber* van Tieghem) [5] in the food mycology. In the work of Panagou and Kodogiannis [3], the predictive capability of NNs was tested on this fungus in comparison to conventional polynomial approach. The experimental results showed that the NNs outperform the conventional polynomial approach in terms of prediction accuracy, and the temperature was the most influential (sensitive) parameter affecting the maximum specific growth rate of *M. ruber* van Tieghem, followed by water activity and pH to a lesser extend. Despite the success of applying the NNs to predict the fungal growth in the food mycology given by Panagou and Kodogiannis [3], a more efficient intelligent modelling methodology shall be developed to improve the prediction accuracy for the fungal growth in the food mycology.

Recently, it is known that the adaptive network-based fuzzy inference system (ANFIS) is a successful approach for dealing with the nonlinear mapping [6]. By using a hybrid learning algorithm which combines the gradient method and the least squares estimate to identify parameters, the ANFIS can construct an input-output mapping based on both human knowledge (in the form of fuzzy if-then rules) and the stipulated input-output data pairs [6-8]. The hybrid learning algorithm may reach the local minima very fast and never find the global solution. Therefore, Ho et al. [9] have proposed the improved-genetic-algorithm-based ANFIS may obtain better results than the ANFIS with hybrid learning algorithm. In the work of Ho et al.

[9], the optimal antecedent and consequent parameters are simultaneously found with genetic algorithm by directly minimizing the root-mean-squared-error (RMSE) performance criterion. Therefore, same as the work of Ho et al. [9], the purpose of this paper is, by using an improved genetic algorithm in the ANFIS to find the optimal antecedent and consequent parameters, to investigate the promoted effectiveness of predicting the maximum specific growth rate of *M. ruber* van Tieghem in the food mycology, where the RMSE criterion is directly minimized. The improved genetic algorithm used in this paper is the hybrid Taguchi-genetic algorithm (HTGA). The reason why the HTGA is applied in this paper is that Tsai et al. [10] have shown that the HTGA may obtain the better results than those existing genetic algorithm reported in the literature. In this paper, some experimental results are also given to make the comparison of effectiveness with those obtained by Panagou and Kodogiannis [3] in terms of prediction accuracy.

2. HTGA-Based ANFIS Approach to Model the Fungal Growth

Three growth parameters, that have a major impact on the maximum specific growth rate of *M. ruber* van Tieghem [11], include temperature (x_1), pH (x_2) and water activity b_w (x_3) selected as the input variables of the ANFIS, and the maximum specific growth rate $\mu_{\max}$ (y) is selected as the output variable of the ANFIS, in which $x_i = [x_i(1), x_i(2), \dots, x_i(t)]^T$ ($i = 1, 2, 3$), $y = [y(1), y(2), \dots, y(t)]^T$ and t denotes the number of training data. The architecture of the ANFIS used in this paper is shown in Figure 1. A typical rule set with 27 fuzzy if-then rules can be expressed as

R^n : IF x_1 is M_{1j} AND x_2 is M_{2j} AND x_3 is M_{3j} ,

THEN $\hat{y}_n = a_{0n} + a_{1n}x_1 + a_{2n}x_2 + a_{3n}x_3$, (1)

where R^n ($n = 1, 2, \dots, 27$) denotes the n -th implication, M_{ij} ($i, j = 1, 2, 3$) are the linguistic terms of the precondition part with Gaussian membership functions $\mu_{M_{ij}}(x_i)$, $\hat{y}_n$ is the n -th output value of consequent part, and a_{kn} ($k = 0, 1, 2, 3$ and $n = 1, 2, \dots, 27$) are the consequent parameters.

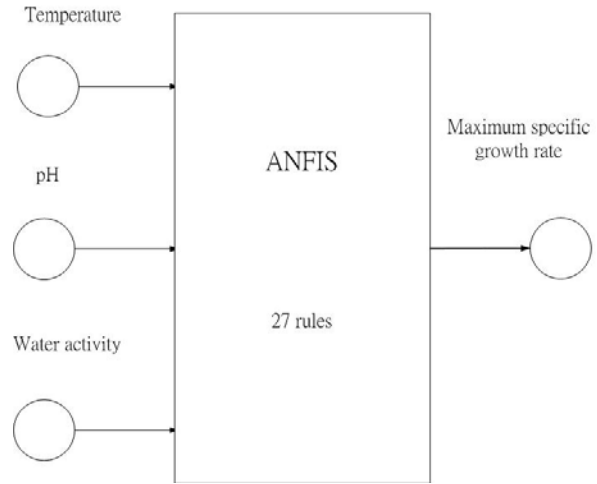


Figure 1. Three-inputs and one-output ANFIS architecture with 27 fuzzy rules.

The inferred ANFIS output from Eq. (1) is represented as

$$\hat{y} = \sum_{n=1}^{27} \bar{W}_n \hat{y}_n = \sum_{n=1}^{27} \bar{W}_n (a_{0n} + a_{1n}x_1 + a_{2n}x_2 + a_{3n}x_3), \quad (2)$$

where $\hat{y}$ denotes the inferred output (i.e., the predicted maximum specific growth rate) of the ANFIS, and $\bar{W}_n = W_n / \sum_{n=1}^{27} W_n$, in which $\hat{y} = [\hat{y}(1), \hat{y}(2), \dots, \hat{y}(t)]^T$ and $W_n = \mu_{M_{1p}}(x_1) \mu_{M_{2q}}(x_2) \mu_{M_{3r}}(x_3)$ ($n = 1, 2, \dots, 27$, and $p, q, r = 1, 2, 3$).

Let the antecedent parameters of Gaussian membership functions $\mu_{M_{1p}}(x_1)$, $\mu_{M_{2q}}(x_2)$ and $\mu_{M_{3r}}(x_3)$ in the precondition part be, respectively, $\{a_{M_{1p}}, b_{M_{1p}}\}$, $\{a_{M_{2q}}, b_{M_{2q}}\}$ and $\{a_{M_{3r}}, b_{M_{3r}}\}$, in which $a_{M_{1p}}$, $a_{M_{2q}}$ and $a_{M_{3r}}$ denote the center of the Gaussian membership function and $b_{M_{1p}}$, $b_{M_{2q}}$ and $b_{M_{3r}}$ are the width of the Gaussian membership function. It can be observed that the optimal values of antecedent and consequent parameters can be obtained after training by directly minimizing the following RMSE performance criterion:

$$J = \left[\sum_{m=1}^t \frac{(y(m) - \hat{y}(m))^2}{t} \right]^{\frac{1}{2}}, \quad (3)$$

where $y(m)$ is the value of actual experimental maximum specific growth rate, and $\hat{y}(m)$

denotes the value of predicted maximum specific growth rate.

From Eq. (3), it can be seen that the value of the performance criterion J is actually dependent on the set of $\{a_{M_{1p}}, b_{M_{1p}}, a_{M_{2q}}, b_{M_{2q}}, a_{M_{3r}}, b_{M_{3r}}, a_{kn}\}$ ($n = 1, 2, \dots, 27$, $p, q, r = 1, 2, 3$, and $k = 0, 1, 2, 3$), which means

$$J = g(a_{M_{1p}}, b_{M_{1p}}, a_{M_{2q}}, b_{M_{2q}}, a_{M_{3r}}, b_{M_{3r}}, a_{kn}) \\ \equiv g(f_1, f_2, \dots, f_{126}). \quad (4)$$

This is equivalent to the optimization problem in the following:

$$\text{minimize } J = g(f_1, f_2, \dots, f_{126}). \quad (5)$$

Then, the HTGA described-below can be employed to search for the optimal solution of the optimization problem in Eq. (5), where Eq. (5) is a nonlinear function with the continuous variables.

The HTGA combines the traditional genetic algorithm (TGA) [12] with the Taguchi method [13, 14]. In the HTGA, the Taguchi method is inserted between the crossover and mutation operations of a TGA. Then, by using two major tools (signal-to-noise ratio and orthogonal arrays) of the Taguchi method, the systematic reasoning ability of the Taguchi method is incorporated in the crossover operations to systematically select the better genes to achieve crossover operations, and consequently enhance the genetic algorithms. The detailed steps of the HTGA are described in the following. The detailed description of the Taguchi method can be found in the books presented by Taguchi et al. [13] and Wu [14]. Besides, the additional details regarding the HTGA can be found in the works proposed by Tsai et al. [10].

Detailed Steps: HTGA

Step 1: Parameter setting.

Input: population size M , crossover rate p_c , mutation rate p_m , and number of generations.

Output: the set of $V = \{f_1, f_2, \dots, f_{126}\}$ and the value of J in Eq. (3).

Step 2: Initialization. By using the J in Eq. (3), which is the fitness function defined for the HTGLA, the fitness values of the initial population are then calculated, where the initial population with the chromosomes of the form $V = \{f_1, f_2, \dots, f_{126}\}$ for the problem in Eq. (5) is randomly generated, where the

types of membership functions are specified in advance.

Step 3: Selection operation using the roulette wheel approach.

Step 4: Crossover operation. The probability of the crossover is determined by the crossover rate p_c .

Step 5: Select a suitable two-level orthogonal array $L_\gamma(2^{\gamma-1})$ for the matrix experiments, where γ denotes the number of experimental runs, and $\gamma - 1$ is the number of columns in the orthogonal array. The orthogonal array $L_{128}(2^{127})$ is used in the illustrative results given in the next section.

Step 6: Choose randomly two chromosomes at a time to execute the matrix experiments.

Step 7: Calculate the fitness values of the γ experiments in the orthogonal array $L_\gamma(2^{\gamma-1})$ by using the Eq. (3).

Step 8: Calculate the effects of the various factors.

Step 9: One optimal chromosome is generated based on the results from Step 8.

Step 10: Repeat Steps 6 through 9 until the expected number $M \times p_c$ has been met.

Step 11: The population via the Taguchi method is generated.

Step 12: Mutation operation. The probability of the mutation is determined by the mutation rate p_m .

Step 13: Offspring population is generated.

Step 14: Sort the fitness values in increasing order among the parents and offspring populations.

Step 15: Select the better M chromosomes feasible for the constraint as the parents of the next generation.

Step 16: Has the stopping criterion been met? If yes, go to Step 17. Otherwise, return to Step 3 and continue through Step 16.

Step 17: Compare the RMSE performance criteria and check if the specified stopping condition has been met. If yes, go to Step 18. Otherwise, return to Step 2 and continue through Step 17.

Step 18: Display the optimal chromosome and the optimal fitness value.

3. Results and Discussions

As the work of Panagou and Kodogiannis

[3], a total of the 60 sets data ($t = 60$) shown in Table 1 are selected from the total of 73 sets obtained in the experiments of maximum specific growth rate of *M. ruber* van Tieghem for the purpose of training in the ANFIS; the other 13 sets shown in Table 2 are then used for testing after the training is completed to verify the accuracy of the predicted values of maximum specific growth rate. The proposed HTGA-based ANFIS method directly compares with the NN approaches given by Panagou and Kodogiannis [3]. The evolutionary environments of the HTGA used in this paper are: the population size is 200, the crossover rate is 0.9, the mutation rate is 0.1, and the generation number is 200. Then, after training by using the HTGA-based ANFIS approach with $\underline{C}_i \leq f_i \leq \bar{C}_i$ ($i = 1, 2, \dots, 126$) in which the values of $\underline{C}_i$ and $\bar{C}_i$ are given in Table 3, we can obtain that the RMSE performance criterion J for the training, testing and full data are 0.062, 0.0151 and 0.085, respectively. Figure 2 shows the convergence result of the training RMSE performance criterion J in Eq. (3). Figures 3-5 show the final Gaussian membership functions of the three growth parameters scaled in the rank $[0, 1]$ after training process. From Figures 3 and 5, it can be seen that, for the parameters x_1 (temperature) and x_3 (water activity), the final membership functions after training experience considerable changes in the large and small areas, respectively. That is, both temperature (x_1) and water activity (x_3) have the most impacts on maximum specific growth rate of *M. ruber* van Tieghem, which corresponds with the conclusion presented by Panagou and Kodogiannis [3]. After the three growth parameters of any one of the 13 sets of testing data are entered in the HTGA-based ANFIS after training, an output value can be obtained from the calculated result. This output value is the predicted value of maximum specific growth rate. Figures 6 and 7 show that all predictions derived from the HTGA-based ANFIS were generally closer to the line, which was indicated better fitting performance of the HTGA-based ANFIS. In addition, Figures 8 and 9 show that the residuals were distributed symmetrically around 0, which was displayed no systematic tendency to the positive or negative side of the graph for the case of HTGA-based ANFIS. The spread of residual values was narrower for the HTGA-based ANFIS for all data

sets indicating better prediction performance.

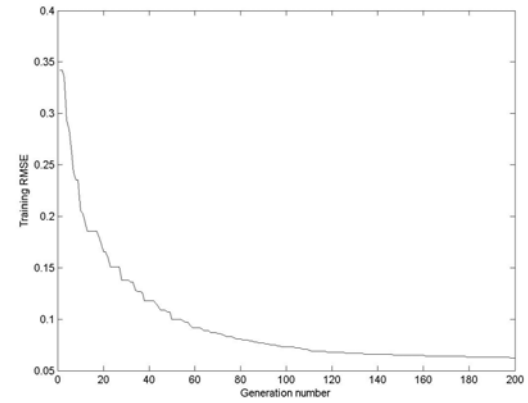


Figure 2. Convergence results of the training RMSE performance criterion J .

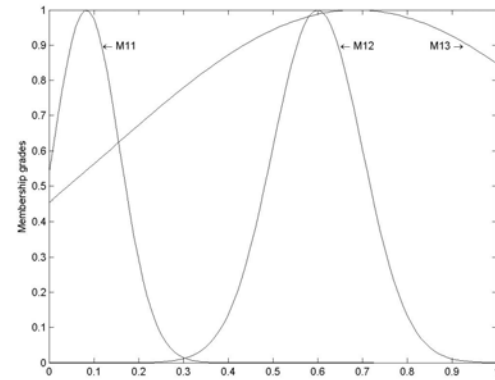


Figure 3. Final Gaussian membership functions of the temperature (x_1).

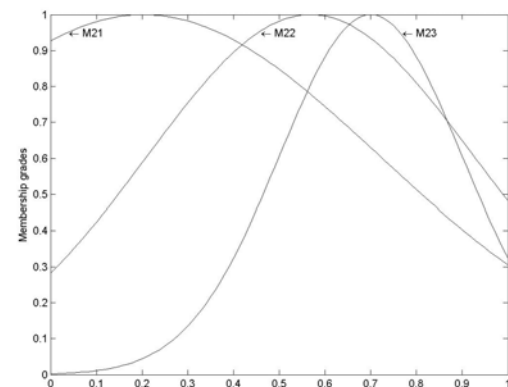


Figure 4. Final Gaussian membership functions of the pH (x_2).

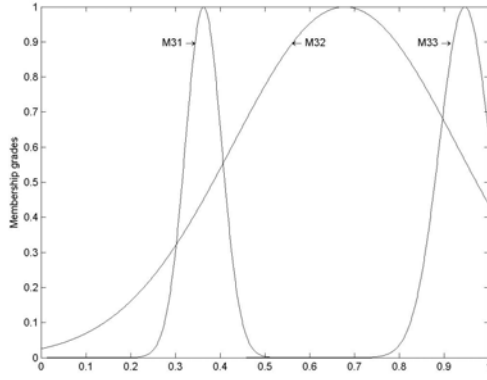


Figure 5. Final Gaussian membership functions of the water activity (x_3).

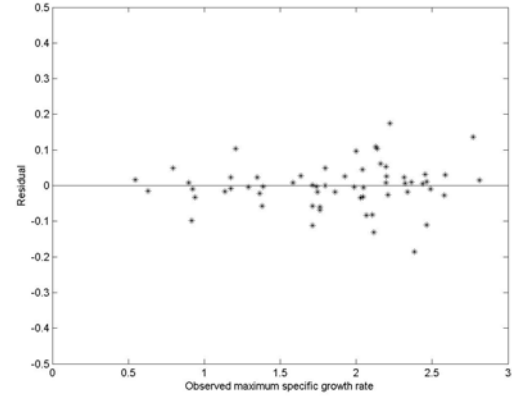


Figure 8. Comparison residual values (predicted-observed maximum specific growth rate of *M. ruber* van Tieghem) by the HTGA-based ANFIS for the training data set.

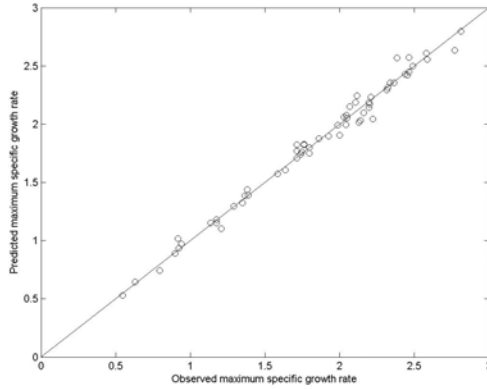


Figure 6. Comparison predicted the maximum specific growth rate of *M. ruber* van Tieghem and observed values by the HTGA-based ANFIS for the training data set.

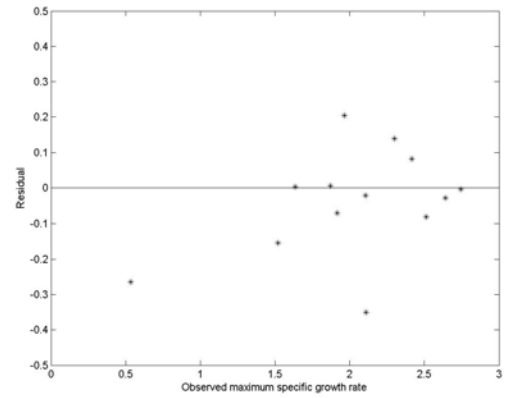


Figure 9. Comparison residual values (predicted-observed maximum specific growth rate of *M. ruber* van Tieghem) by the HTGA-based ANFIS for the testing data set.

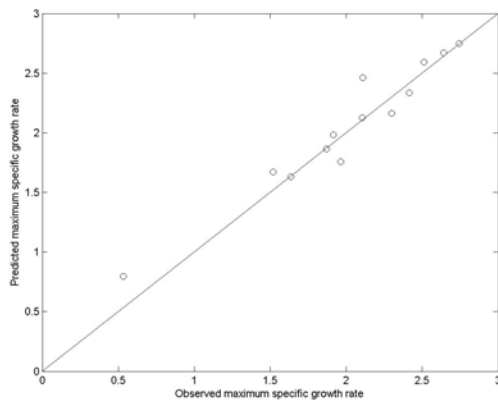


Figure 7. Comparison predicted the maximum specific growth rate of *M. ruber* van Tieghem and observed values by the HTGA-based ANFIS for the testing data set.

Table 4 shows the RMSE of the predictions of maximum specific growth rate obtained, respectively, from the proposed HTGA-based ANFIS approach (0.085 in full data), the NNs method including MLP (0.110 in full data) and RBF (0.092 in full data) models given by Panagou and Kodogiannis [3]. It can be seen that the prediction accuracy of the HTGA-based ANFIS approach is higher than the NNs methods presented by Panagou and Kodogiannis [3]. That is, the proposed HTGA-based ANFIS approach outperforms the NNs methods presented by Panagou and Kodogiannis [3] in terms of prediction accuracy.

4. Conclusions

An HTGA-based ANFIS approach has been proposed in this paper to effectively predict the maximum specific growth rate of *M. ruber* van

Tieghem in the food mycology by using three growth parameters (temperature, pH and water activity). The HTGA is applied in the ANFIS to simultaneously find the optimal antecedent and consequent parameters by directly minimizing the RMSE performance criterion. Experimental results have shown that the RMSE of the HTGA-based ANFIS approach is 0.085 in full data which outperforms the RMSEs 0.110 and 0.092 with MLP and RBF models in full data obtained, respectively, by Panagou and Kodogiannis [3].

Acknowledgements

This work was supported by the National Science Council, Taiwan, Republic of China, under grant number NSC 99-2320-B-037-026-MY2.

References

- [1] Gibson, A. M. and A. D. Hocking, "Advances in the Predictive Modelling of Fungal Growth in Food", *Trends in Food Science and Technology*, Vol. 8, pp. 353-358, 1997.
- [2] Murphy, P. A., S. Hendrich, C. Landgren and C. M. Bryant, "Food Mycotoxins: An Update", *J. of Food Science*, Vol. 71, pp. 51-65, 2006.
- [3] Panagou, E. Z. and V. S. Kodogiannis, "Application of Neural Networks as a Non-linear Modelling Technique in Food Mycology", *Expert Systems with Applications*, Vol. 36, pp. 121-131, 2009.
- [4] Tsai, J. T., J. H. Chou and T. K. Liu, "Tuning the Structure and Parameters of a Neural Network by Using Hybrid Taguchi-Genetic Algorithm", *IEEE Trans. on Neural Networks*, Vol. 17, pp. 69-80, 2006.
- [5] Rosso, L., J. R. Lobry, S. Bajard and J. P. Flandrois, "Convenient Model to Describe the Combined Effects of Temperature and pH on Microbial Growth", *Applied and Environmental Microbiology*, Vol. 61, pp. 610-616, 1995.
- [6] Jang, J. S. R., C. T. Sun and E. Mizutani, *Neuro-Fuzzy and Soft Computing*, Pearson Education, Taipei, 2004.
- [7] Lin, C. J. and Y. J. Xu, "A Self-Adaptive Neural Fuzzy Network with Group-Based Symbiotic Evolution and Its Prediction Applications", *Fuzzy Sets and Systems*, Vol. 157, pp. 1036-1056, 2006.
- [8] Lin, C. J., C. H. Chen and C. T. Lin, "Efficient Self-Evolving Evolutionary Learning for Neurofuzzy Inference Systems", *IEEE Trans. on Fuzzy Systems*, Vol. 16, 1467-1490, 2008.
- [9] Ho, W. H., J. T. Tsai, B. T. Lin and J. H. Chou, "Adaptive Network-Based Fuzzy Inference System for Prediction of Surface Roughness in End Milling Process Using Hybrid Taguchi-Genetic Learning Algorithm", *Expert Systems with Applications*, Vol. 36, pp. 3216-3222, 2009.
- [10] Tsai, J. T., T. K. Liu and J. H. Chou, "Hybrid Taguchi-Genetic Algorithm for Global Numerical Optimization", *IEEE Trans. on Evolutionary Computation*, Vol. 8, pp. 365-377, 2004.
- [11] Panagou, E. Z., P. N. Skandamis and G. J. E. Nychas, "Modelling the Combined Effect of Temperature, pH and a_w on the Growth Rate of *Monascus Ruber*, a Heat-Resistant Fungus Isolated from Green Table Olives", *J. of Applied Microbiology*, Vol. 94, pp. 146-156, 2003.
- [12] Gen, M. and R. Cheng, Genetic Algorithms and Engineering Design, *John Wiley and Sons, New York*, 1997.
- [13] Taguchi, G. S., Chowdhury and S. Taguchi, *Robust Engineering*, McGraw-Hill, New York, 2000.
- [14] Wu, Y., *Taguchi Methods for Robust Design*, The American Society of Mechanical Engineers, New York, 2000.

Table 1. Experimental results for training data set (60 data) [3].

Temperature (°C)	pH	b_w	$\ln(\mu_{\max})$
20	3.5	0.173	1.3483
20	4	0.173	1.1731
20	4.5	0.173	1.1346
20	5	0.173	0.9247
25	4	0.173	1.9864
25	4.5	0.173	2.1961
25	5	0.173	2.1372
30	3.5	0.173	2.4632
30	4	0.173	2.7704
30	5	0.173	2.581
35	3.5	0.173	2.5864
35	4.5	0.173	2.8113
40	3.5	0.173	2.3165
40	4	0.173	2.1162
40	4.5	0.173	2.3842
20	3.5	0.207	1.2921
20	4	0.207	1.1752
20	4.5	0.207	0.8983
20	5	0.207	0.6296
25	3.5	0.207	2.0487
25	4.5	0.207	1.7376
25	5	0.207	1.5829
30	3.5	0.207	2.4361
30	4	0.207	2.3366
30	5	0.207	2.0407
35	3.5	0.207	2.4896
35	4	0.207	2.4532
35	4.5	0.207	2.3217
35	5	0.207	2.0677
40	3.5	0.207	2.4623
40	4.5	0.207	2.3635
40	5	0.207	2.1958
20	3.5	0.238	0.9383
20	4	0.238	0.7944
20	4.5	0.238	0.5462
25	3.5	0.238	1.7595
25	4.5	0.238	1.3797
25	5	0.238	1.3886
30	3.5	0.238	2.1991
30	4	0.238	2.1593
30	4.5	0.238	1.9997
30	5	0.238	1.7117
35	3.5	0.238	2.2086
35	4	0.238	2.1058
35	4.5	0.238	2.0274
40	3.5	0.238	2.0452
40	4	0.238	2.2202
40	4.5	0.238	2.1272
25	3.5	0.25	1.2073
25	4	0.25	0.9179
30	3.5	0.25	1.7978
30	4	0.25	1.7615
30	4.5	0.25	1.7459
30	5	0.25	1.6355

35	4	0.25	1.9248
35	4.5	0.25	1.8596
35	5	0.25	1.7971
40	3.5	0.25	1.3653
40	4	0.25	1.7113
40	4.5	0.25	1.712

Table 2. Experimental results for testing data set (13 data) [3].

Temperature (°C)	pH	b_w	$\ln(\mu_{\max})$
25	3.5	0.173	2.1051
25	4	0.207	1.869
25	4	0.238	1.634
25	4.5	0.25	0.5328
30	4.5	0.173	2.7427
30	4.5	0.207	2.3003
35	4	0.173	2.5109
35	5	0.173	2.6405
35	3.5	0.25	1.9633
40	5	0.173	2.1102
40	4	0.207	2.4161
40	5	0.238	1.9147
40	5	0.25	1.5193

Table 3. Ranges of the antecedent and consequent parameters.

Parameters		Range
Premise parameters	$a_{M_{11}}, a_{M_{21}}, a_{M_{31}}$ (f_1, f_4, f_7)	0–0.4
	$a_{M_{12}}, a_{M_{22}}, a_{M_{32}}$ (f_2, f_5, f_8)	0.2–0.8
	$a_{M_{13}}, a_{M_{32}}, a_{M_{33}}$ (f_3, f_6, f_9)	0.6–1.0
	$b_{M_{1p}}, b_{M_{2q}}, b_{M_{3r}}$ ($p, q, r = 1, 2, 3$)	0–0.6
	($f_{10} - f_{18}$)	
Antecedent parameters	a_{kn} ($n = 1, 2, \dots, 27$, and $k = 0, 1, 2, 3$)	-1.0–1.0
	($f_{19} - f_{126}$)	

Table 4. Comparison of results.

Performance index	Model	Data set		
		Training	Testing	Full
RMSE	MLP	0.088	0.179	0.110
	RBF	0.063	0.167	0.092
	HTGA-based ANFIS	0.062	0.151	0.085