

## ON-LINE CONSTRUCTION AND RULE BASE SIMPLIFICATION BY REPLACEMENT IN FUZZY SYSTEMS APPLIED TO A WASTEWATER TREATMENT PLANT

José Victor<sup>1,2</sup>, António Dourado<sup>2</sup> and Plamen Angelov<sup>3</sup>

<sup>1</sup>*School of Technology and Management  
Polytechnic Institute of Leiria, Portugal  
zevictor@estg.ipleiria.pt*

<sup>2</sup>*Adaptive Computation Group - CISUC  
University of Coimbra, Portugal  
{zevictor, dourado}@dei.uc.pt*

<sup>3</sup>*Department of Communications Systems  
Lancaster University, UK  
p.angelov@lancaster.ac.uk*

**Abstract:** Evolving Takagi-Sugeno (eTS) fuzzy models are used to build a computational model for the WasteWater Treatment Plant (WWTP) in a paper mill. The fuzzy rule base is constructed on-line from data using a recursive fuzzy clustering algorithm that develops the model structure and parameters. In order to avoid some redundancy in the fuzzy rule base mechanisms for merging membership functions and simplifying fuzzy rules are introduced. The rule base simplification is done by replacement allowing the preservation of the rule (cluster) centres as data points belonging to the original data set. Results for the WWTP show that it is possible to build less complex models and preserve a good balance between accuracy and transparency. *Copyright © 2005 IFAC*

**Keywords:** On-line learning, eTS fuzzy models, recursive fuzzy clustering, rule base simplification, interpretability, transparency.

### 1. INTRODUCTION

The transformation of data into useful information is probably the most important challenge for information engineers nowadays (Kruse, *et al.*, 1999). This information can be used for a better decision by humans or for better automatic control by (intelligent) controllers.

Computational Intelligence techniques, particularly fuzzy systems and neural networks, provide methods and tools for data analysis. Fuzzy systems can be used to create accurate predictive models from data. Interpretability of the models is decisive for the human decision process (Setnes, 2001).

The WasteWater Treatment Plant is a nonlinear process, time-varying, with a large number of variables. It is practically impossible to build a first-principles model. Fuzzy modelling, and particularly Takagi-Sugeno (TS) fuzzy models, provide a

powerful tool for modelling complex nonlinear systems (Yen, *et al.*, 1998) and by this way they seem to be the most appropriate approach because of its representational power and the possibility of using learning algorithms to identify its parameters; moreover they allow to gain insights into the local behaviour of the sub-models and they may have good interpretability properties.

The determination of the structure and parameters of TS fuzzy models (i.e., the learning problem) has been and is the subject of extensive research. Fuzzy systems have evolved from a seminal knowledge-driven approach, where expert knowledge plays an essential role, towards a data-driven approach, where expert knowledge plays a minor role (Angelov, 2002). The huge amount of data available in all sectors of human activities (industry, services, finance, etc.) demands a new kind of on-line learning algorithms in order to extract knowledge in a quick and efficient way.

In order to simplify the fuzzy model of the WasteWater Treatment Plant, and particularly of the effluent neutralisation process, a hierarchical structure is proposed to take advantage of the lower dimensions of the subsystems (Setnes, 2001). Since the process is time varying, on-line learning algorithms are needed. Evolving Takagi-Sugeno (eTS) models are used in for the effluent neutralisation process. The input-output space is partitioned by recursive fuzzy clustering for learning the antecedent parameters of the fuzzy model. The Evolving Clustering method (Kasabov and Song, 2002), based on an objective function, and the recursive version of the subtractive clustering (Angelov and Filev, 2004), based on the notion of the informative potential, are the most known at present times. For learning the consequent parameters in eTS fuzzy models the modified Recursive Least Squares (w)RLS estimation algorithm is used because it is assumed that the rule base will gradually evolve in number of rules (and by consequence in number of parameters).

The paper is organised in five sections. In section 2 the effluent neutralization process is described and several aspects related with its modelling are described, namely the hierarchical structure proposed for complexity reduction and the input selection mechanism. Section 3 briefly introduces the basic notions of the applied fuzzy modelling technique, eTS fuzzy models for on-line learning. Experimental results obtained for the second stage of the process are shown in section 4 and section 5 presents the conclusions.

## 2. EFFLUENT NEUTRALIZATION PROCESS

Three different stages compose the mill effluent treatment plant: primary, secondary and tertiary stage. The primary treatment consists in the effluent neutralisation and suspended solids removal. It is carried on the foam tower, the neutralisation basin and the primary clarifiers. The secondary or biological treatment uses activated sludge for the degradation of dissolved organic matter. It is carried out on the cooling tower, the selector, the aerator basins and the secondary clarifiers. The tertiary treatment consists in colour removal and effluent equalization, regarding to quality and temperature. It is carried on the equalization lagoon.

In this work we focus on the primary treatment, more particularly on the study of the effluent neutralisation process. The obtained models will be very important for control, fault diagnosis and process supervision.

### 2.1 Description of the neutralization process.

The acid effluent, the north alkaline effluent and the south alkaline effluent compose the plant effluent. Before the acid effluent enters in the WWTP it is partially neutralised in two stages. The first stage consists in the addition of dust collected by the electro filter of the limekiln. The second stage consists in the addition of hydrated lime. During the

normal functioning of the mill the industrial effluent is mainly acid but in certain situations, such as during maintenance, the effluent can be alkaline and in this case the neutralisation is made with the dosing of sulphuric acid at the end of the neutralisation basin. Fig. 1 gives a perspective of the effluent neutralisation process plant.

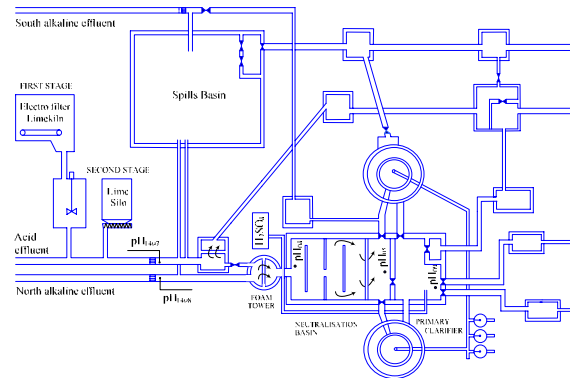


Fig. 1. Plant of the effluent neutralisation process.

The second stage has a much higher neutralisation capacity than the first stage, though with associated cost, since hydrated lime has to be bought, and the addition of the limekiln dust has no cost. While the effect of the first stage in the pH value of the acid effluent is very small, the effect of the second stage can be considerable. The correct neutralisation of the industrial effluent is vital for the functionality of the secondary or biological treatment and the quality of the final effluent. The goal is to achieve at the end of the primary treatment a pH in the interval [6.0; 8.0] and simultaneously minimizing the consumption of neutralisation materials, which are quite expensive.

### 2.2 Process modelling.

*Hierarchical structure.* One of the difficulties of the effluent neutralisation process modelling is the large number of variables that influence the values of pH. In order to deal with this problem a hierarchical structure is proposed. The initial model gives origin to five sub models: first stage, second stage, foam tower, neutralisation basin and primary clarifier. The hierarchical structure was created based on a priori knowledge obtained from process engineers and operators. With this knowledge the variables choice for each sub model was easier, since process engineers have a good perception about the importance of each process variable. Some variables were not considered relevant in this initial phase, since in normal conditions they are not significant. The two first sub models are related with the neutralisation of the acid effluent. In the foam tower the acid and alkaline effluents are combined resulting in the raw effluent. The uniformisation of the raw effluent starts on the neutralisation basin where the filtrates are added. The chemical reaction between the effluent and the neutralization materials continues on the primary clarifier in order to guarantee an efficient pH effluent uniformisation. Fig. 2 represents the input and output variables for the different models.

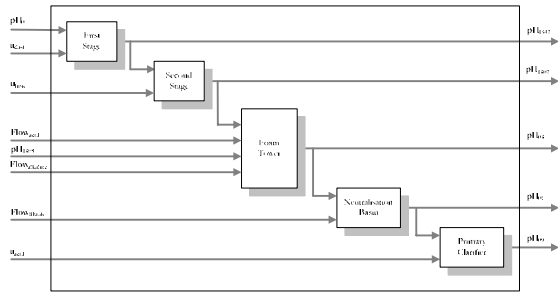


Fig.2. Hierarchical structure of the effluent neutralisation process.

*Structure of the models.* In the definition of the models structure one of the requisites is the selection of the variables and respective regressions to be included in the model. As stated before, one of the goals of this work is to construct fuzzy models that are interpretable, simple and applicable. In order to achieve this goal, each sub model will be a first-order model with time delay, according to Bröida definition, i.e., the output  $y_k$  will depend only on one past value of each of the inputs,  $u_{k-T_{d_p}}^p$ , and one past value of the output,  $y_{k-1}$ , as (1)

$$y_k = f\left(u_{k-T_{d_1}}^1, u_{k-T_{d_2}}^2, \dots, u_{k-T_{d_p}}^p, y_{k-1}\right) \quad (1)$$

*Input selection.* In order to obtain dynamic models it is necessary to determine the order and time transport delay for each one of the input variables. The order is already defined. To determine the time transport delay for each variable, information given by process engineers was precious, however this information must be completed by experimental data since some of the process inputs present a time-varying transport delay. ANFIS (Adaptive Network-based Fuzzy Inference System) was used as a straightforward way to determine the time transport delay (Jang, 1993; Jang, 1996). It employs an efficient hybrid learning method that combines gradient descent and least-squares method. The performance measure used was the MSE (Mean Square Error).

In input selection only one epoch of training is used for each ANFIS model. The number of possibilities tested and the amount of time spent in the determination for the time transport delay was considerably reduced, since the available a priori knowledge was used in the search.

### 3. eTS FUZZY MODELS

eTS fuzzy models have been recently introduced by Angelov and Filev and the approach is based on an on-line learning algorithm that recursively develops the model structure and parameters (Angelov and Filev, 2004). Structure identification includes estimation of the antecedent parameters by recursive fuzzy clustering and the consequent parameters are obtained by applying the modified (w)RLS estimation

algorithm. In the following the basics of the algorithm are briefly recalled, for more details see references (Angelov and Filev, 2004; Angelov, *et al.*, 2004).

#### 3.1 On-line estimation of the antecedent parameters.

The recursive fuzzy clustering algorithm uses the notion of informative potential (accumulated spatial proximity measure), introduced in the mountain clustering algorithm (Yager and Filev, 1993) and then refined in the subtractive clustering algorithm (Chiu, 1994), to compute the potential of each new data point in order to select the cluster centres (focals). The algorithm starts with the first data point established as the focal point of the first cluster. Its coordinates are used to form the antecedent part of the fuzzy rule using Gaussian membership functions and its potential is assumed equal to 1. Starting from the next data point onwards the potential of the new data points is calculated recursively. In contrast with the subtractive clustering algorithm there is not a specific amount subtracted from the highest potential, but update of all the potentials after a new data point is available on-line. The potential of the new data sample  $z_k$  is recursively calculated as follows (2),

$$P_k(z_k) = \frac{k-1}{(k-1)(g_k+1) + \sigma_k - 2v_k} \quad (2)$$

where  $g_k = \sum_{j=1}^{n+1} (z_k^j)^2$ ,  $\sigma_k = \sum_{l=1}^{k-1} \sum_{j=1}^{n+1} (z_l^j)^2$ ,  $v_k = \sum_{j=1}^{n+1} z_k^j \beta_k^j$  and  $\beta_k^j = \sum_{l=1}^{k-1} z_l^j$ .

Parameters  $g_k$  and  $v_k$  are calculated from the current data point  $z_k$ , while  $\beta_k^j$  and  $\sigma_k$  are recursively updated by (3).

$$\sigma_k = \sigma_{k-1} + \sum_{j=1}^{n+1} (z_{k-1}^j)^2; \beta_k^j = \beta_{k-1}^j + z_{k-1}^j \quad (3)$$

After the new data are available they influence also the potentials of the centres of the clusters ( $z_l^*$ ,  $l=1, \dots, R$ ), which are respective to the focal points of the existing clusters ( $x_l^*$ ,  $l=1, \dots, R$ ). The reason is that by definition the potential depends on the distance to all data points, including the new ones. The potential of the focal points of the existing clusters is recursively updated by (4),

$$P_k(z_l^*) = \frac{(k-1)P_{k-1}(z_l^*)}{k-2 + P_{k-1}(z_l^*) + P_{k-1}(z_l^*) \sum_{j=1}^{n+1} (d_{k(k-1)}^j)^2} \quad (4)$$

where  $P_k(z_l^*)$  is the potential of the cluster  $z_l^*$  at time  $k$ , which is a prototype of the  $l^{\text{th}}$  rule,

$d_{k(k-1)}^j = z_k^j - z_{k-1}^j$ , denotes the projection of the distance between two data points,  $z_k^j$  and  $z_{k-1}^j$ , on the axis  $z^j$  ( $x^j$  for  $j=1,2,\dots,n$  and on the axis  $y$  for  $j=n+1$ ).

If the potential of the new data point is higher than the maximum potential of the existing centres or lower than the minimum then the new data point is accepted as a new centre (Angelov et al., 2004) (5),

$$\begin{aligned} &\text{If } P_k(z_k) > \bar{P}_k \text{ or } P_k(z_k) < \underline{P}_k \text{ then} \\ &R = R + 1; x_R^* = x_k; P_k(x_R^*) = P_k(x_k) \end{aligned} \quad (5)$$

where  $\bar{P}_k = \max_{l=1}^R P_k(z_l^*)$  and  $\underline{P}_k = \min_{l=1}^R P_k(z_l^*)$ .

If in addition to the previous condition the new data point is close to an old centre then the new data point replaces this centre, instead of being accepted as a new one.

$$\begin{aligned} &\text{If } (P_k(z_k) > \bar{P}_k \text{ or } P_k(z_k) < \underline{P}_k) \text{ and} \\ &\delta_{\min} < \frac{r}{2} \text{ then } z_j^* = z_k; P_k(z_j^*) = P_k(z_k) \end{aligned} \quad (6)$$

where  $\delta_{\min} = \min_{l=1}^R \|z_k - z_l^*\|$ ,  $r$  is the constant radii and  $j$  is the index of the replaced centre (closest cluster centre to the new data point).

### 3.2 On-line estimation of the consequent parameters.

The problem of increasing size of the training data is handled by the RLS algorithm for the globally optimal case and by the weighted RLS algorithm for the locally optimal case.

In eTS fuzzy models, however, the rule base is assumed to be gradually evolving. Therefore, the number of rules as well as the parameters of the antecedent part will vary. Because of this evolution, the normalized firing strengths of the rules will change. Since this affects all the data (including the data collected before time of the change) the straightforward application of the RLS or wRLS algorithm is not correct. A proper resetting of the covariance matrices and parameters initialization of the algorithms is needed at each time a rule is added to and/or removed from the rule base. The modified RLS algorithm for global and local parameter estimation is described in reference (Angelov and Filev, 2004).

## 4. EXPERIMENTS

This section describes the experiments done in the second stage of the effluent neutralisation process. In the second stage, the input variables are the pH of the acid effluent (pH1347) and the addition of hydrated lime ( $u_{lime}$ ). The output variable is the pH of the acid effluent (pH1407).

The modelling approach described in the previous section is applied to predict the pH values for the second stage of the process.

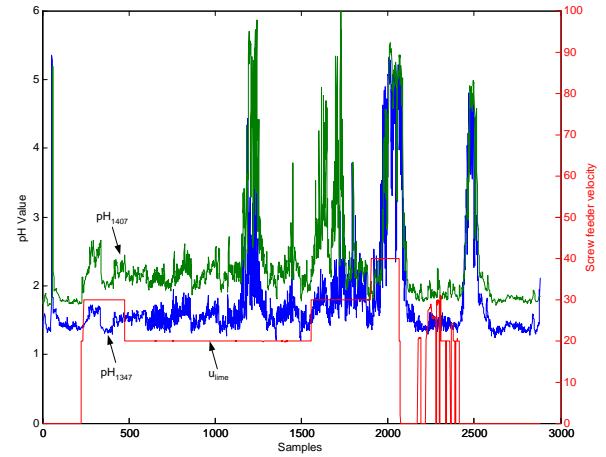


Fig. 3. Training data set for the second stage of the effluent neutralisation process.

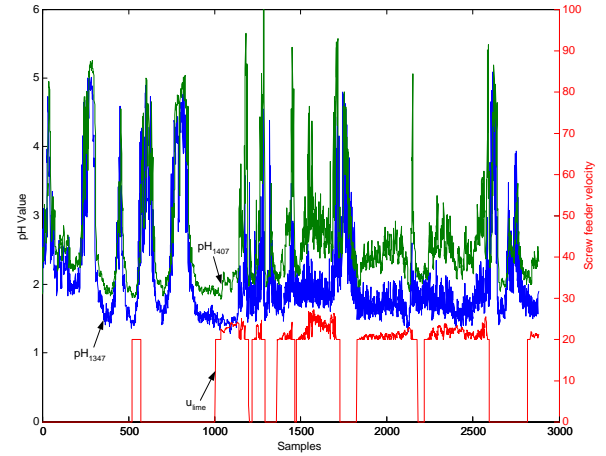


Fig. 4. Validation data set for the second stage of the effluent neutralisation process.

In the experiments, two plant data sets from consecutive days are used, the first one, Fig. 3 for training, and the second one, Fig. 4, to determine the generalisation capability of the fuzzy models. Each data set has 2880 samples, sampling interval was  $T=30$  seconds, and the data is normalized into de domain  $[0, 1]$ , to avoid negative influence in clustering results from variations in the numerical ranges of the different features (Babuska, 1998).

### 4.1 On-line learning.

The eTS fuzzy model approach is applied for on-line learning. The value for the constant radii was  $r = 0.5$  and the initialization parameter for RLS was  $\Omega = 750$ . Fig. 6 shows the evolution of the process output and model output for the validation data set. Table 1 presents the parameters of the eTS fuzzy model at the end of the learning process.

For validation the structure and parameters of the eTS fuzzy model are initialised with those obtained at the end of the training.

The number of rules at the end of the training was 8, and it remains the same after validation, but some fuzzy rules were replaced and the parameters of the model were constantly updated. In practice what happens is that at every instant, with the arriving of a new data point, a new model is constructed and even though the structure does not change the model parameters are varying.

**Table 1 eTS fuzzy model parameters for on-line learning.**

Rule	Antecedent Parameters	Consequent Parameters
1	[0.0795 0.0 0.0264]	[-0.0049 0.2919 0.0167 0.4528]
2	[0.5232 0.0 0.0433]	[0.0671 0.1509 1.2825 0.5386]
3	[0.8962 0.0 0.1145]	[-0.1382 0.0567 -2.9062 3.3540]
4	[0.9545 0.0 0.3133]	[-1.2997 1.1059 1.3470 1.7750]
5	[0.9809 0.0 0.8124]	[-0.2618 0.1350 0.0157 1.1675]
6	[0.0635 0.7486 0.1169]	[0.1195 0.2846 -0.1058 0.7601]
7	[0.0922 0.4961 0.1261]	[0.1316 0.9858 -0.3922 0.6218]
8	[0.9933 0.9956 0.8806]	[0.5024 0.1006 -0.2010 0.5228]

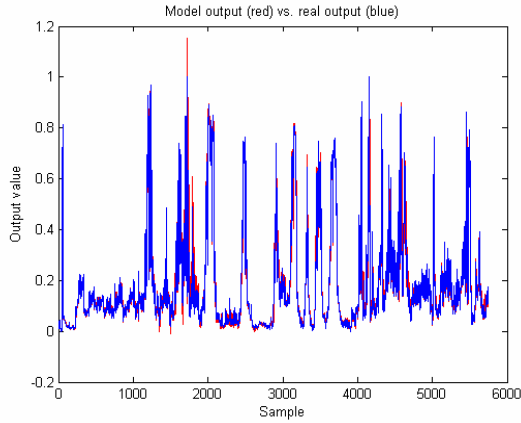


Fig. 6. Process and model output for on-line learning.

With on-line learning even if the training data is not representative of all the operating points of the process it is possible to achieve satisfactory results since the approach can evolve the structure and parameters of the fuzzy model according with the current dynamics of the process.

The performance of the eTS fuzzy model is satisfactory but the complexity and interpretability is poor since there are a lot of similar membership functions, particularly for the first and third input variables, Fig. 7 and Table 1.

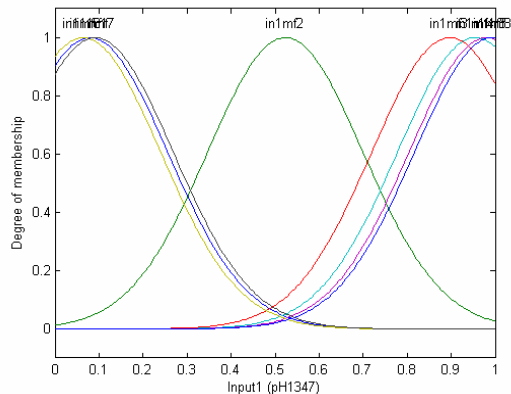


Fig. 7. Membership functions of the first input variable.

#### 4.2 On-line learning with rule-base simplification.

In order to reduce the redundancy in fuzzy models obtained from data rule base simplification and reduction methods must be combined with the data-driven modelling tools, resulting in a more transparent fuzzy modelling scheme (Setnes, *et al.*, 1988). The concept of on-line transparent modelling was recently introduced and results from the integration of rule base simplification and reduction procedures in the on-line learning process (Victor Ramos and Dourado, 2004).

First similarity between fuzzy sets, along each input dimension, is performed by the application of the similarity measure (7),

$$S(A, B) = \frac{1}{1 + d(A_1, A_2)}; S(\cdot) \in (0, 1] \quad (7)$$

$$d(A_1, A_2) = \sqrt{(c_1 - c_2)^2 + (\sigma_1 - \sigma_2)^2} \quad (8)$$

where  $c_i$  is the centre and  $\sigma_i$  is the width of the Gaussian membership functions. If the similarity between two fuzzy sets is greater than a predefined threshold  $\lambda \in (0, 1]$  the fuzzy set of the rule with higher frequency (measure of importance of a rule) is kept. This avoids the creation of artificial centres preserving as cluster centres only original data points. After merging of fuzzy sets the rule base may contain irrelevant fuzzy sets and merging of fuzzy rules is likely to happen. If the premise parts of two or more rules become equal these rules can be removed and replaced by one general rule with the same antecedents, as the rules which it replaces, and the consequents computed as the average of the consequents of all the replaced rules, (9).

$$\theta_{new} = \frac{1}{k} \sum_{i=1}^k \theta_i \quad (9)$$

Fig. 8 shows the evolution of the process output and model output for the validation data set with rule-base simplification. Table 2 presents the parameters of the eTS fuzzy model at the end of the learning process.

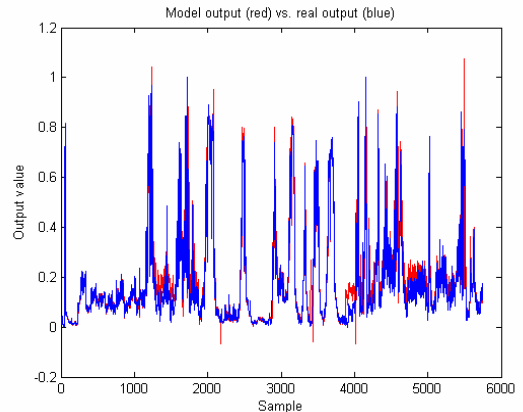


Fig. 8. Process and model output for on-line learning with rule-base simplification.

**Table 2 eTS fuzzy model parameters with rule-base simplification.**

Rule	Antecedent Parameters	Consequent Parameters
1	[0.1495 0.0366 0.1260]	[0.0049 0.2368 0.0821 0.8332]
2	[0.5210 0.0366 0.1260]	[-0.2649 0.9459 -0.2676 0.5395]
3	[0.9351 0.0366 0.1260]	[-1.1623 1.4106 1.1576 1.2917]
4	[0.9351 0.0366 0.7700]	[-0.9432 0.8751 0.5706 0.8614]
5	[0.1495 0.5124 0.1260]	[-0.1254 0.2742 0.1541 0.8202]

The interpretability of the model improved considerably since now only three linguistic terms (Low, Medium and High) are necessary to describe the first input variable and two linguistic terms (Low and High) to describe the second and third input variables.

Fig. 9 presents the membership functions for the first input variable (pH1347) at the end of the on-line learning process when rule-base simplification is applied.

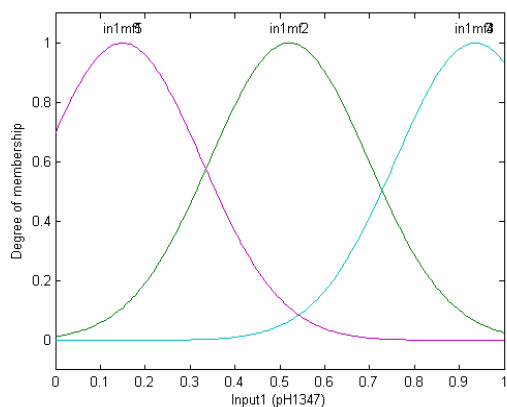


Fig. 9. Membership functions of the first input variable with rule-base simplification.

## 5. CONCLUSIONS

The on-line learning (eTS) and the rule-base simplification techniques presented in this paper have a great potential for building computational intelligence systems for monitoring and controlling industrial plants. This work presented the application of these techniques to a wastewater plant of a pulp and paper mill allowing the construction of partially interpretable models with reasonable accuracy. The conditions for new rule creation need tuning and further work to develop formal methods for rule simplification in the general case is under investigation.

## ACKNOWLEDGEMENTS

This work was supported by the Portuguese Ministry of Education, under the programme PRODEP III, and partially supported by POSI programme of Portuguese Foundation for Science and Technology (POSI/EEI/14155/2001). Soporcel Company gave the data for this work.

## REFERENCES

- Angelov, P. (2002). *Evolving Rule-Based Models: A Tool for Design of Flexible Adaptive Systems*. Studies in Fuzziness and Soft Computing, 92, Physica-Verlag.
- Angelov, P. and D. Filev (2004). An Approach to Online Identification of Takagi-Sugeno Fuzzy Models. *IEEE Transactions on Systems, Man, and Cybernetics – Part B: Cybernetics*, **34**(1), pp. 484-498.
- Angelov, P., J. Victor, A. Dourado and D. Filev (2004). On-line Evolution of Takagi-Sugeno Fuzzy Models. In: *Proceedings of the 2<sup>nd</sup> IFAC Workshop on Advanced Fuzzy/Neural Control (AFNC'04)*, Oulu, Finland.
- Babuska, R. (1988). *Fuzzy Modeling for Control*, International Series in Intelligent Technologies. Kluwer Academic Publishers.
- Chiu, S. (1994). Fuzzy Model Identification Based on Clustering Estimation. *Journal of Intelligent and Fuzzy Systems*, **2**, 267-278.
- Jang, J.-S. R. (1993). ANFIS: Adaptive-Network-based Fuzzy Inference Systems. *IEEE Transactions on Systems, Man, and Cybernetics*, **23**, pp. 665-685.
- Jang, J.-S. R. (1996). Input Selection for ANFIS Learning. In: *Proceedings of the IEEE International Conference on Fuzzy Systems*, New Orleans.
- Kasabov, N. and Q. Song (2002). DENFIS: Dynamic Evolving Neural-Fuzzy Inference System and Its Application for Time-Series Prediction. *IEEE Transactions on Fuzzy Systems*, **10**(2), pp. 144-154.
- Kruse, R., D. Nauck and C. Borgelt (1999). Data Mining with Fuzzy Methods: Status and Perspectives. In: *Proceedings of the EUFIT'99 Congress*, Aachen, Germany.
- Setnes, M. (2001). *Complexity Reduction in Fuzzy Systems*. PhD Thesis, Delft University of Technology.
- Setnes, M., R. Babuska, U. Kaymak and H.R. van Nauta Lemke (1988). Similarity Measures in Fuzzy Rule Base Simplification. *IEEE Transactions on Systems, Man, and Cybernetics – Part B: Cybernetics*, **28**(3), pp. 376-386.
- Takagi, T. and M. Sugeno (1985). Fuzzy Identification of Systems and Its Applications to Modeling and Control. *IEEE Transactions on Systems, Man, and Cybernetics*, **15**(1), 116-132.
- Victor Ramos, J. and A. Dourado (2004). On-Line Interpretability by Fuzzy Rule Base Simplification and Reduction. In: *Proceedings of the EUNITE'04 Conference*, Aachen, Germany.
- Yager, R. and D. Filev (1993). Learning of Fuzzy Rules by Mountain Clustering. In: *Proceedings of SPIE Conference on Applications of Fuzzy Logic Technology*, Boston.
- Yen, J., L. Wang and W. Gillespie (1998). Improving the Interpretability of TSK Fuzzy Models by Combining Global and Local Learning. *IEEE Transactions on Fuzzy Systems*, **6**(4), pp. 530-537.