

This item was submitted to Loughborough's Institutional Repository (<https://dspace.lboro.ac.uk/>) by the author and is made available under the following Creative Commons Licence conditions.



For the full text of this licence, please go to:  
<http://creativecommons.org/licenses/by-nc-nd/2.5/>

# Evolving Rule-based Models: A Tool for *Intelligent* Adaptation

**Plamen Angelov\***

Dept of Civil and Building Engineering  
Loughborough University  
Loughborough, Leicestershire, LE11 3TU, UK  
e-mail: [P.P.Angelov@Lboro.ac.UK](mailto:P.P.Angelov@Lboro.ac.UK)

**Richard Buswell**

Dept of Civil and Building Engineering  
Loughborough University  
Loughborough, Leicestershire, LE11 3TU, UK  
e-mail: [R.A.Buswell@Lboro.ac.UK](mailto:R.A.Buswell@Lboro.ac.UK)

## Abstract

An *on-line* approach for rule-base evolution by recursive adaptation of rule structure and parameters is described in the paper. An integral part of the procedure is to maximise the model transparency by simplifying the fuzzy linguistic descriptions of the input variables. The rule base evolves over time and utilising direct calculation approaches and hence minimising the reliance on the use of computationally expensive techniques, such as genetic algorithms. An *on-line* version of subtractive clustering recently introduced by the authors [1] is used for determination of the antecedent part of the fuzzy rules. Recursive least squares estimation [2]-[3] is employed to determine the parameters of the consequent part of each rule. The use of these efficient non-iterative techniques is the key to the low computational demands of the algorithm. The application of similarity measures improves the interpretability and compactness of the resulting *eR* model, with no significant detriment to the model precision. A time series prediction problem on data from a real indoor climate control (ICC) system has been considered to test and validate the proposed model simplification method.

## 1. Introduction

Fuzzy rule-based (FRB) models, and especially Takagi-Sugeno (TS) models [4], are widely used to represent complex non-linear systems. These models are also (relatively) easy to identify and their structure can be readily analysed [5]. Effective identification techniques treating the antecedent and consequent parts of the model [6]-[8] and methods for analysis of the stability of controllers based on these models [9] has been developed.

Alternative techniques for identification of both model structure and parameters that are, in principle, non-linear optimisation problems, include direct use of genetic algorithms [10]-[12] or gradient-based back-propagation [13]. The advantage of the latter is the higher precision that is gained by solving the parameter and structure identification simultaneously. These approaches include identification of the antecedent and the consequent part of fuzzy rules and their parameters [10]-[11]. The former approach ([1], [6] and [8]) is superior in terms of computational requirement. This is particularly evident when non-iterative clustering approaches (mountain [8] or subtractive [6] clustering instead of fuzzy-C-means [14]) are used. Both approaches are sometimes combined [1], [7] for this reason.

All these methods could be classified as data-driven rule/knowledge extraction. Expert knowledge plays a minor, if any role. This tendency in fuzzy model identification is typical in recent research, particularly over the past few years. One reason for the growing interest in these techniques could be due to the ease by which data can be gathered and distributed. At present, the real issue in many industries and organisations is how to effectively cope with the information in exponentially growing data-bases. This is especially so where the information is qualitative and imprecise.

One important aspect of real problems is the necessity to adapt models and systems in accordance with the changing environmental conditions. Current techniques do not accommodate this requirement [15]. Linear models and linear control theory have been developed to the point of effective solutions for these problems [16]. Complex, non-linear and linguistic models and systems have not. In [1] authors originally introduced an effective approach for recursive *on-line* identification of TS models. In this paper this approach

---

\* corresponding author

is developed further by the application of a model simplification methodology.

A brief summary of existing approaches for (*off-line*) identification of TS models is given with emphasis on non-iterative combination of clustering and linear least squares. *evolving* Rule-based (eR) models are considered as a tool for intelligent adaptation of complex systems description. The basic mechanism for rule-base evolution is presented followed by the model simplification procedure. In section 4 a time-series prediction problem is considered for testing and validation of the simplification methodology. After the analysis of the results conclusions are given.

### 1. (*Off-line*) identification of TS models

Takagi-Sugeno model [4] could be represented as:

$$R_i: \text{IF } (x_1 \text{ is } LV_{i1}) \dots \text{AND } (x_n \text{ is } LV_{in}) \\ \text{THEN } (y_i = p_{i1}x_1 + \dots + p_{in}x_n + q_i); \quad i=1, \dots, NR, \quad (1)$$

where  $R_i$  denotes the  $i^{th}$  fuzzy rule, NR is the number of rules,  $x$  is the input vector  $x=[x_1, x_2, \dots, x_n]^T$  and  $LV_{ij}$  denotes  $j^{th}$  linguistic variable of the antecedent part for the  $i^{th}$  fuzzy rule ( $j=1, 2, \dots, n$ ).  $y_i$  is the output of the  $i^{th}$  rule and  $p_{ij}$  and  $q_i$  are parameters of the consequence. The model output is calculated by weighted averaging of the individual rule contributions using the centre of area de-fuzzification operator.

TS models are quasi-linear in nature [17]; they result in smooth transition between linear sub-models, which are responsible for separate sub-space of states. This property allows separating the identification problem into two sub-problems:

- appropriate partitioning of the state space of interest by clustering;
- parameter identification of the consequent part.

As the output functions  $y_i$  are normally linear or singletons (constants), the second sub-problem is easy solvable by applying least square technique [2]-[3].

The first sub-problem uses clustering since it is more efficient than grid partitioning. Intuitively grid partitioning is closer to the linguistic concept of fuzzy variables, but it is impractical for larger dimensions, due to the so-called *curse of dimensionality* [10]-[11]. Fuzzy C-means have been used [7], but requires iterations. *Mountain* clustering [8] and its

modification, *subtractive* clustering [6] are therefore, preferred [1].

Subtractive clustering is based on the notion of potential of a data point to be a centre of a cluster. The following formula is used to express the potential as a sum of contributions of Euclidean distances between a given point and all other data points [6]:

$$P_i = \sum_{j=1}^N D_{ij}, \quad (2a)$$

$$D_{ij} = e^{-\alpha \|z_i - z_j\|^2}, \quad i=1, 2, \dots, N, \quad (2b)$$

where  $P_i$  is the potential of the data point  $z_i=[x_i, y_i]$  to be a cluster centre,  $D_{ij}$  denotes the contribution of every single distance,  $N$  is the number of training data samples and  $\alpha = 4/r^2$ ;  $r$  is the cluster radii.

Inspection of Equations 2a and 2b show that the potential of a data point to be a cluster centre is higher when more data points are closer to a specific candidate. The highest potential is called *reference* potential [1].

The procedure called *subtractive clustering* [6] is based on the successive process of determination of the point with highest potential. Potential of all other points are then reduced with an amount proportional to the potential of the chosen point and to the distance to this point:

$$P_i^{new} = P_i^{old} - P_k^* D_{ik}^*, \quad i=1, 2, \dots, N, \quad (3)$$

where  $P_k^*$  denote the potential of the  $k^{th}$  centre and  $D_{ik}^*$  is the modified contribution, which differs from  $D_{ik}$  by the parameter  $\alpha$  [1], [6].

When a data point is selected as a new cluster centre and its indices become the centres of new membership functions. The point is accepted as a centre if its potential is higher than certain threshold which is determined as a function of the reference. If the potential is less than a lower threshold (also a function of the reference potential) the point is rejected. If the potential falls between these limits and is sufficiently far away from the current centres, the point is also rejected. The distance criterion is based on the shortest of the distances ( $d_{min}$ ) between the new candidate to be a cluster centre ( $x_m^*$ ) and all previously found cluster centres. The following inequality, expresses the trade-

off between the potential value and the closeness to the previous centres [1], [6]:

$$\frac{P_m^*}{P_1^*} r \geq r - d_{\min} \quad (4)$$

Second sub-problem (this of parameters of the consequent part estimation) is easily solved by applying linear least square technique [1]-[3],[6]. It should be mentioned that parameters of the antecedent part can be further simplified and optimised, but only through the application of iterative non-linear approaches like GA [7], [10]-[12] or gradient-based techniques [13]. It is possible to improve precision up to 2 times and the model structure could be further simplified and optimised. The disadvantage, however, is computational expense.

## 2. Intelligent adaptation of rule-based models

In real-life problems a non-linear model which adapts to the changes in the environment and adapts to the object of modelling/control could be the basis for building *intelligent* systems that are able to learn more effectively. *eR* models (rule-based TS models evolving in structure and parameters) are seen as a promising candidate for this purpose [1]. A procedure for *on-line* recursive identification of TS models has been developed in [1]. Basically it consists of:

- calculation of the potential of new data points:

$$P^{k+1} = \frac{1}{N+k+1} \sum_{j=1}^{N+K+1} D_{j(N=k+1)} \quad k=0,1,2,\dots \quad (5)$$

where  $k$  denotes the *on-line* time sampling

- *on-line* recursive up-date of the potential of existing cluster (membership functions) centres:

$$[P_l^*]^{k+1} = A[P_l^*]^{k+1} + [\delta P_l^*]^{k+1}; \quad l=1,2,\dots,R \quad (6)$$

where  $A = \frac{N+k}{N+k+1}$ ;  $[\delta P_l^*]^{k+1} = \frac{D_{j(N=k+1)}}{N+k+1}$

- *on-line* recursive up-date of the reference potential

$$P_{N+k+1}^{\text{reference}} = \max(P_l^*, P_{N+k+1}) \quad (7)$$

- *on-line* recursive estimation of parameters of the consequent part.

In order to avoid overloading of memory a moving window has been introduced [1]. This is critical only

for calculation of the potential of the new data point (5).

The fuzzy rule-based model depicted in Equation 1 is generated automatically, *on-line*. It is used as a good initial estimate of the non-linear mapping between inputs and the output(s). Its optimality could be guaranteed by using a non-linear numerical optimisation algorithm, such as a GA [7], [10]-[11].

### FRB model simplification

It is desirable, especially in fault diagnosis, to have transparent models that are as simple as possible while maintaining a desired level of precision. In order to maximise the transparency, which also minimises the memory cost, it is necessary to minimise the number of membership functions describing each input variable. This procedure is depicted in Figure 1.

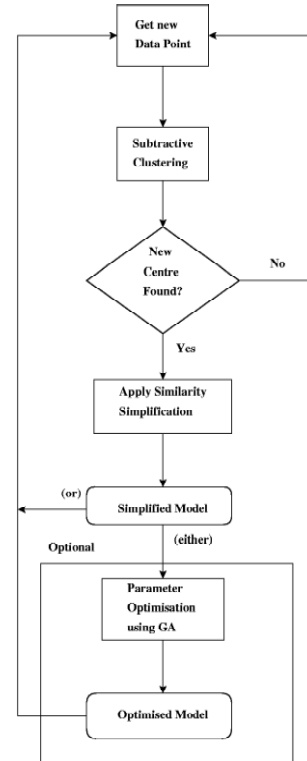


Figure 1.

With the *on-line* TS model identification, both the rules and parameters are considered [1]. The number of rules influences the precision of the model and is determined by the potential of data samples. The model simplification process seeks to minimise the number of membership functions associated with each input variable.

The approach allows for the optimisation of the membership function parameters, if the process is considered to be beneficial to the model. Often, however, this process will lead to over-fitting, with no real gain in the model representation of the process.

After the on-line search has yielded a new centre, the similarity of the membership functions of the new data point is compared to the existing model. Since the spreads of the membership functions are the same, the similarity can be judged on the values of the centre parameters alone. The centres are deemed to be similar if the distance is less than a given threshold (which is a predetermined percentage of the variable range; 10%→15% of the whole range seem to be reasonable values to use). Each membership function centre parameter in each input variable is sequentially checked against the new membership function. If the new membership function is similar to one that is existing, the new rule is rewritten to reflect the existing, similar membership function. If no similar membership function exists, it is added to the model. It should be noted that the selection of the distance threshold in the simplification criterion should not be too stringent, or the model precision will suffer. Conversely, it should not be too low, otherwise there will not be sufficient simplification of the model.

### 3. Results and Discussion

To demonstrate the reduction in the number of model parameters, through similarity, the approach was applied to a time series prediction problem. Figure 2 shows the training data used in the problem. The data was collected from a real system.

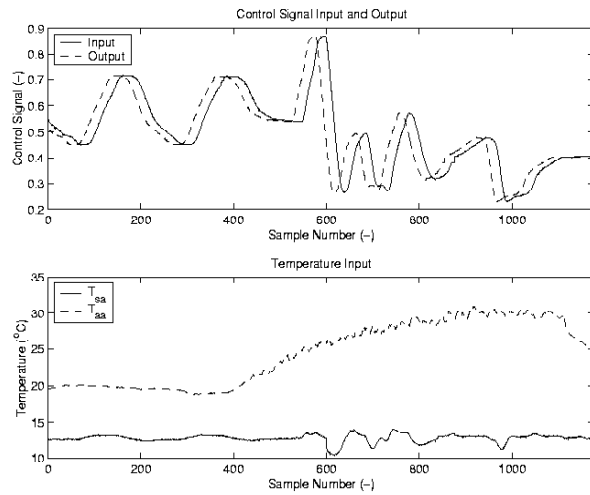


Figure 2

The plot shows the control signal (top plot) to a valve that controls the mass flow rate of water through a heat exchanger. The heat exchanger cools the warm air that flows on to the coil. The cool air is used to maintain comfortable conditions in an occupied space. One of the principle loads on the coil is generated due to the supply of ambient air; required to maintain a minimum standard of indoor air quality. The test system is shown in Figure 3.

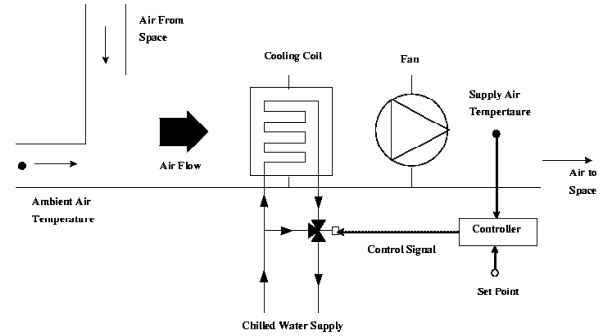


Figure 3

The ambient air ( $T_{aa}$ ) and supply air ( $T_{sa}$ ) temperatures (shown in the bottom plot) are sampled and the current and previous time intervals, as is the control signal. These are the model inputs. The model is then used to predict the control signal 20 samples ahead of the current sample. The sample interval is 1 minute. Data from the same system, but from a different day was used to validate the models.

Using a batch estimation approach on the training data, the subtractive clustering generated a model with 7 rules and 7 membership functions describing each input variable. Figure 4 shows these functions for each input. The number of membership functions in the test case is  $7 \times 6 = 42$ . Applying the similarity simplification (with a threshold value of 10%), this number is reduced to 12, as shown in Figure 5.

Figure 6 shows the correlation between the model predictions and the data for the training and validation cases. The correlation coefficients are noted on each plot. Figure 7 demonstrates the predictions compared to the measured data. The loss of precision of the model that this simplification results in is negligible. The root mean squared prediction error of the initial and simplified models on the training data was 0.030 and 0.031 respectively (no units for control signal). Application to the validation data set yields errors of

0.095 for both models. The correlation coefficients reflect similar results.

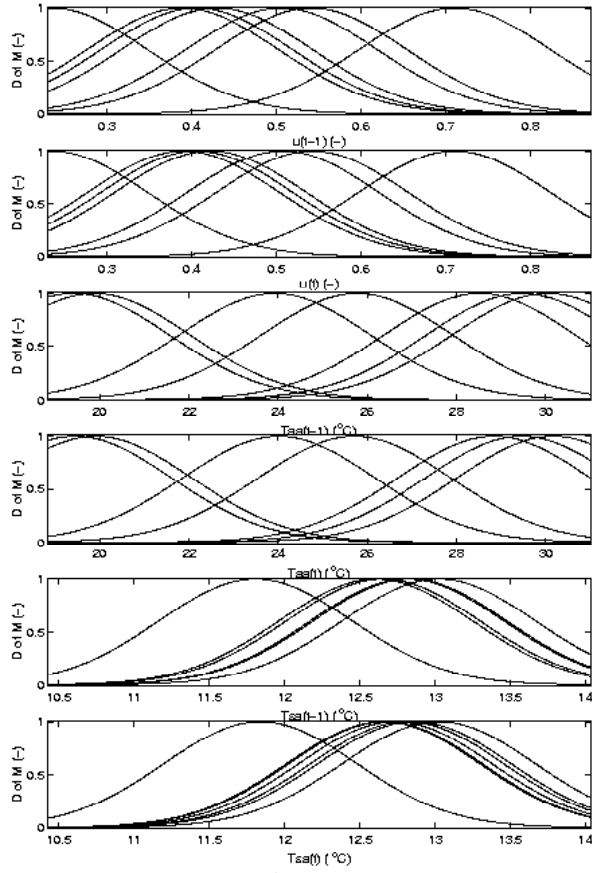


Figure 4

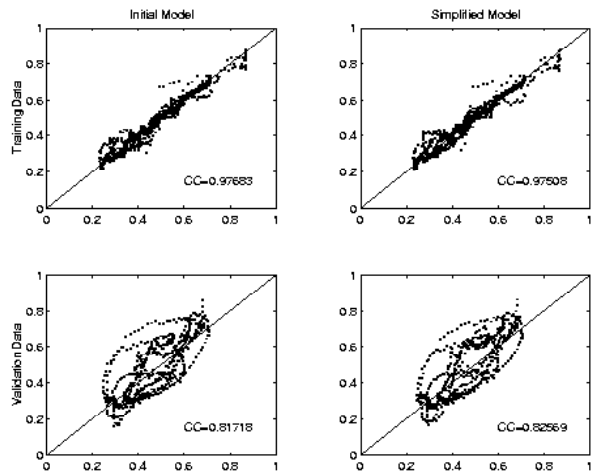


Figure 6

A further comment on the results is the similarity in the pairs of inputs. Each input pair consists of the variable value at the past ( $t-1$ ) and current ( $t$ ) sample, hence the

data is very similar, and the results in Figures 4 and 5 are to be expected. The simplified model has a number

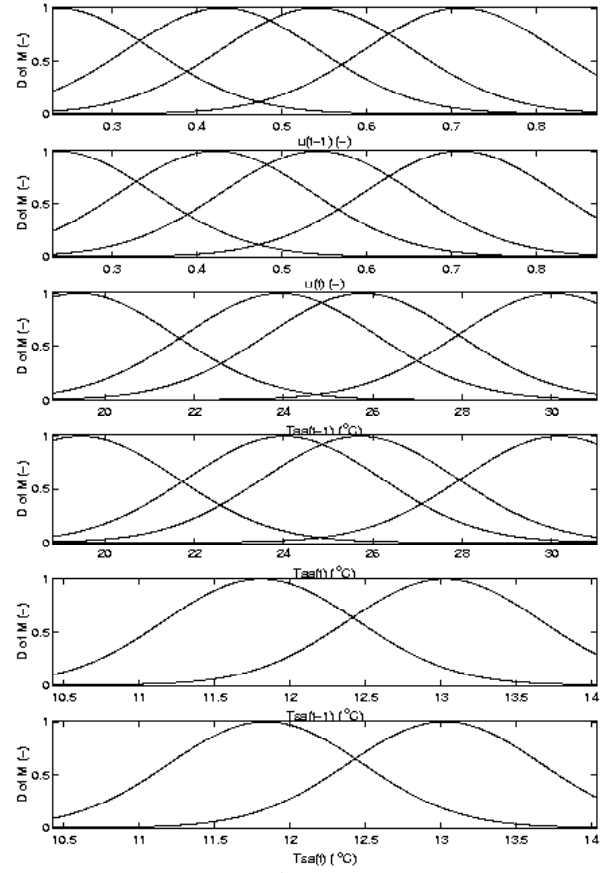


Figure 5

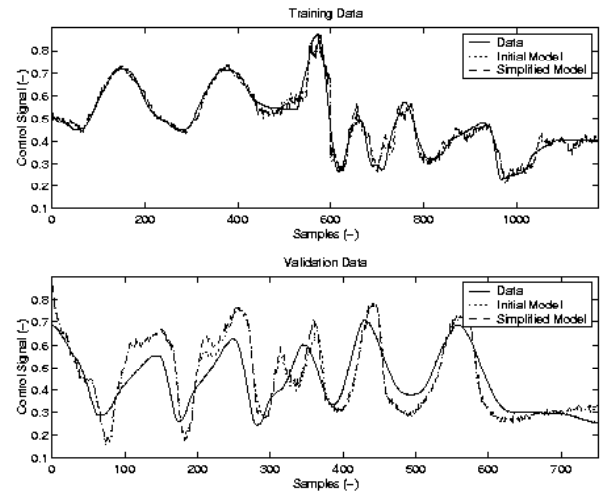


Figure 7

of advantages;

- it is more transparent,
- the linguistic terms are simpler,

- lower memory requirements result in lower computational demands and hence a faster execution time.

The application of the GA to optimise the parameters associated with this problem lead to over training with no improvement of captured process logic.

#### 4. Conclusions

An approach for rule-base evolution by recursive adaptation of rule structure and parameters and simultaneous model simplification in *on-line* was presented. An *on-line* version of the subtractive clustering recently introduced by the authors [1] is demonstrated. Recursive linear least squares [2]-[3] was applied for each rule separately to determine in *on-line* mode parameters of the consequent part of the rules. Application of similarity measures additionally improves the transparency and simplicity of the model, without significant degradation in the model precision. Additional benefits are a reduction in execution time and memory requirements.

Engineering applications to a time series prediction problem based on data from a real indoor climate control system has been considered to test and validate the model simplification method.

#### 5. Acknowledgements

The authors acknowledge support of the UK EPSRC through the grant GR/M27299 and use of the data courtesy of ASHRAE; generated from the ASHRAE funded research project RP1020.

#### 6. References

- [1] P.P. Angelov, and R.A. Buswell, "Recursive On-line Identification of Takagi-Sugeno Models by Rules and Parameters *Innovation*", *IEEE Transactions on Fuzzy Systems*, submitted paper
- [2] K. Astrom, and B. Wittenmark, "*Computer Controlled Systems: Theory and Design*", Prentice Hall, Englewood Cliffs, NJ, USA, 1984.
- [3] P. Strobach, "Linear Prediction Theory: A Mathematical Basis for Adaptive Systems", Springer Verlag, New York, USA, 1990.
- [4] T. Takagi, and M. Sugeno, "Fuzzy Identification of Systems and its Application to Modeling and Control", *IEEE Transactions on Systems, Man and Cybernetics*, Vol.15, 1985, pp 116-132.
- [5] R. R. Yager, and D. P. Filev, "*Essentials of Fuzzy Modeling and Control*", John Wiley and Sons, New York, USA, 1994.
- [6] S. L. Chiu, "Fuzzy Model Identification based on Cluster Estimation", *Journal of Intelligent and Fuzzy Systems*, Vol.2, 1994, pp 267-278.
- [7] M. Setnes, J. A. Roubos, "GA-Fuzzy Modeling and Classification: Complexity and Performance", *IEEE Transactions on Fuzzy Systems*, Vol.8 (5), 2000, pp 509-522.
- [8] R.R. Yager, D. P. Filev, "Learning of Fuzzy Rules by Mountain Clustering", *Proc. of SPIE Conference on Application of Fuzzy Logic Technology*, Boston, MA, USA, 1993, pp 246-254.
- [9] K. Tanaka, and M. Sugeno, "Stability Analysis and Design of Fuzzy Control Systems", *Fuzzy Sets and Systems*, Vol. 45 (2), pp 135-156.
- [10] P. Angelov, "Evolving Fuzzy Rule-based Models", *Journal of CIIE, special issue on Soft Computing Applications to Industrial Engineering*, Vol. 17, 2000, pp 459-468.
- [11] P.P. Angelov, V. I. Hanby, R. A. Buswell, J. A. Wright, "Automatic Generation of Fuzzy Rule-based Models from Data by Genetic Algorithms", *Proc. of the Conference on Advances in Soft Computing*, Leicester, UK, June 29-30, 2000 (R. John and R. Birkenhead Eds.), Springer Verlag.
- [12] K. Shimojima, T. Fukuda, Y. Hasegawa, "Self-tuning Fuzzy Modeling with Adaptive Membership Function, Rules, and Hierarchical Structure based on Genetic Algorithm", *Fuzzy Sets and Systems*, Vol. 71, 1995, pp 295-309.
- [13] J.S.R. Jang, "ANFIS: Adaptive Network-based Fuzzy Inference Systems", *IEEE Transactions on Systems, Man & Cybernetics*, Vol. 23 (3), 1993, pp 665-685.
- [14] J. Bezdek, "Cluster Validity with Fuzzy Sets", *Journal of Cybernetics*, Vol. 3 (3), 1974, pp 58-71.
- [15] EUNITE: *EUropean Network on Intelligent Technologies for Smart Adaptive Systems*, Contract No IST-2000-29207, Project Summary, p.4
- [16] L. Ljung, "*System Identification: Theory for the User*", New Jersey, USA, 1987, Prentice-Hall.
- [17] D. P. Filev, "Fuzzy Modeling of Complex Systems", *International Journal of Approximate Reasoning*, Vol. 5, 1991, pp 281-290.