SOFT SENSOR BASED ON FUZZY MODEL IDENTIFICATION

Elaine Yassue Nagai *,[1]
Lucia Valeria Ramos de Arruda *

* *Federal Center of Technological Education of Paraná*
Av. Sete de Setembro,3165,80.230-901,Curitiba, PR, Brazil
e-mails: (eynagai,arruda)@cpgei.cefetpr.br

Abstract: This paper presents an approach to build a soft sensor based on computational intelligence techniques. The goal is to identify fuzzy models from numerical data. First of all, the fuzzy model input variables are selected from the secondary variables set by applying Kohonen maps. Then, the Lipschitz quotients are used to select the lag structure of the fuzzy model. A fuzzy clustering algorithm is applied to find an initial rule base, and, to conclude the identification process, this initial rule base is simplified by merging the similar membership functions. The validity of the proposed identification method is demonstrated by the development of a soft sensor to infer the top composition of a distillation column.

Keywords: distillation columns, estimators, fuzzy modelling, neural networks, soft sensing.

1. INTRODUCTION

Soft sensor or inferential estimator has been an alternative approach for estimating process variables when hardware sensors are not available, or when their high cost or technical limitations prevent their on-line use. Typically, in the chemical industries, soft sensors can be used to estimate product compositions from temperature and other secondary variables. However, because chemical processes are quite complex to model, and are characterized by strong nonlinearities, theoretical modeling approaches based on Kalman filter or other classical estimation techniques are hard to implement in real world (Zambrogna *et al.*, 2005; Lant *et al.*, 1991).

For these reasons, alternative approaches based on computational intelligence techniques have been recently proposed (Qin, 1996; Rallo *et al.*, 2002; Fabro, 2003), in which successful applications of Artificial Neural Networks (ANN) and Partial Least Squares (PLS) to develop soft sensors for different processes have been reported. In this context, this work proposes a fuzzy modeling technique that can be used to build soft sensors in chemical industries.

This methodology is composed by three stages. First, the input variables to the soft sensor are selected by applying Kohonen maps, and the Lipschitz quotients are used to select the structure of a linguistic fuzzy model. Second, a fuzzy clustering algorithm is applied to find an initial rule base that can model the interaction between variables

[1] Financial support of the Brazilian National Agency of Petroleum -ANP- and of the Research and Projects Financing -FINEP- through of Program of Human Resources of ANP for the Petroleum and Gas sector - PRH-ANP/MCT (PRH10-CEFET-PR)

and process non-linearities. Finally, the model is simplified by merging the similar membership functions from each variable domain.

This paper is organized as follows. Section 2 details each procedure in a theoretical point of view, and section 3 synthesizes the proposed methodology. Section 4 presents simulation results obtained with the application of the developed soft sensor to infer the top composition of a distillation column. Section 5 outlines some concluding remarks.

2. IDENTIFICATION PROCEDURE

The fuzzy models used in this work are similar to those described in Nagai and Arruda (2002), in which the model rule base is composed by M linguistic rules, with n input variables (selected from process secondary variables) and one output variable (inferred variable), as shown in figure 1.

If x_1 is A_{11} and ... and x_n is A_{1n} Then y is B_1
If x_1 is A_{21} and ... and x_n is A_{2n} Then y is B_2
⋮
If x_1 is A_{M1} and ... and x_n is A_{Mn} Then y is B_M

Fig. 1. Linguistic fuzzy model.

The fuzzy inference engine uses the Zadeh Inference Method and the inferred output y^* is obtained by means of the Center of Area (COA) defuzzification method (Lee, 1990).

The first phase of the proposed fuzzy model identification method is concerned with the selection of a proper input set, and the determination of the input-output lag space. These two problems are not independent, but, in this methodology, they are solved in independent way, through two distinct procedures, as described below.

2.1 Input Variables Selection

In this work, the technique proposed by Rallo *et al.* (2002) is adapted to select the most adequate subset of secondary variables to compose the soft sensor input set. In this case, the most adequate subset is the smallest secondary variables set containing most of the relevant informations about the input space.

Initially, the secondary variables subsets are specified as follows. First, it is computed the correlation coefficient between each secondary variable and the output variable. The secondary variables are sorted by their correlation coefficient absolute value, creating the set S of ordered secondary variables. Let p the number of secondary variables. The p subsets are constructed according to the sequence of variables in S, i.e., the first subset is composed by the first variable in S and the output variable; the second subset is composed by the first and second variables in S and the output variable. Then, a Kohonen map is built for each subset, and the quality of each map is calculated by means of the dissimilarity measure between two distinct maps L_q and L_r (Kaski and Lagus, 1996), defined by equations (1) and (2). The dissimilarity measure informs about the relevance of each combination of secondary variables in relation to the output variable (Rallo et al., 2002).

$$D(L_q, L_r) = E\left[\frac{|d_{Lq}(x) - d_{Lr}(x)|}{d_{Lq}(x) + d_{Lr}(x)}\right], \qquad (1)$$

$$d(x) = \|x(t) - w_{bmu1}(t)\| + \sum_{i=p_{bmu1}}^{p_{bmu2}-1} \|w_i(t) - w_{i+1}(t)\|, \qquad (2)$$

where: E is the average expectation; p_{bmu1} is the first bmu (best map unit) position; p_{bmu2} is the second bmu position; and $d(x)$ is the distance from x to the second bmu, denoted by $w_{bmu2}(t)$, beginning at the first bmu, denoted by $w_{bmu1}(t)$.

The distance $d(x)$ combines an indication of the continuity of the mapping from the data set to the map structure with a measure of the accuracy of the map in representing the data set (Kaski and Lagus, 1996). The smallest average of dissimilarity value calculated through equation (1) for any given subset of secondary variables indicates the similarity in quality and quantity of the information represented by the maps. Figure 2 shows the procedure of input variables selection.

- Step 1: Let $U = \{u_1, \ldots, u_p\}$ the secondary variables set and y the output variable. Calculate the correlation coefficient between each secondary variable and the output variable;
- Step 2: Sort the secondary variables by their correlation coefficient absolute value, creating the set $S = \{s_1, \ldots, s_p\}$, in which $\|corr(s_i, y)\| \geq \|corr(s_{i+1}, y)\|$, and $\|.\|$ is the absolute value;
- Step 3: Define the number W of units in the self-organizing maps ($SOMs$);
- Step 4: Let p the number of secondary variables. Define the p subsets of secondary variables $(c_1, \ldots, c_p)$ from the set S, in which $c_1 = \{s1, y\}$, $c_2 = \{s1, s2, y\}$, ..., $c_p = \{s1, \ldots, s_p, y\}$;
- Step 5: Define a SOM with W units for each subset c_i, and train these $SOMs$ by using the training data set;
- Step 6: For each pair of $SOMs$ L_q and L_r, $q \neq r$, calculate the dissimilarity measure (equation (1));
- Step 7: Calculate the average of dissimilarity of each SOM L_q with relation to the other $SOMs$;
- Step 8: The map with the smallest value of average of dissimilarity contains the most adequate subset of secondary variables.

Fig. 2. The input variables selection algorithm.

2.2 Lag Space Selection

The method proposed in He and Asada (1993) is applied to select the lag structure of the fuzzy model. This method uses the Lipschitz quotients, that are calculated from the data set $\{\mathbf{u}(t), y(t)\}, t = 1, \ldots, N_d$, where $\mathbf{u}(t)$ is the vector containing all input variables of the fuzzy model (previously selected) and $y(t)$ is the fuzzy model output, and the regressors vector is defined by equations (3) and (4):

$$\varphi(t) = [y(t-1), \ldots, y(t-lag_y), \mathbf{u}(t-d-1), \ldots, \mathbf{u}(t-d-lag_u)]^T, \quad (3)$$

$$Z^{N_d} = \{[\varphi(t), y(t)], \quad t = 1, \ldots, N_d\}, \quad (4)$$

where lag_y is the output lag, d is the input delay, lag_u is the input lag, and Z^{N_d} is the set of N_d input-output pairs. Thus, for all combinations of input-output pairs, the Lipschitz quotients are calculated as follows:

$$q_{i,j} = \frac{|y(t_i) - y(t_j)|}{|\varphi(t_i) - \varphi(t_j)|}, \quad i \neq j. \quad (5)$$

The entire procedure is outlined in figure 3 (He and Asada, 1993):

- Step 1: For given choice of lag space (z), determine the *Lipschitz* quotients of all combinations of input-output pairs (equation (5));
- Step 2: Select the r largest quotients, $r = 0.01N_d$ $\approx 0.02\mathrm{N}_d$. The largest quotients typically occur when the differences $\varphi(t_i) - \varphi(t_j)$ are small;
- Step 3: Evaluate the criterion:

$$\bar{q}^{(z)} = \left(\prod_{t=1}^{r} \sqrt{z} q^{(z)}(t) \right)^{\frac{1}{r}}, \quad (6)$$

  where $q^{(z)}$ is the quotient of the lag space z, and $\bar{q}^{(z)}$ is the criterion value for the lag space z.
- Step 4: Repeat the calculations for a number of different lag structures.
- Step 5: Plot the criterion as a function of lag space and select the optimal number of regressors as the "knee-point" of the curve.

Fig. 3. The lag selection algorithm.

In the phase of fuzzy model parameters estimation, a fuzzy clustering algorithm is applied to the input-output space, and it generates an overestimated rule base, in which each cluster corresponds to a rule. A high value of the initial rule base size (ncr) is used as a way to cover all the important regions of the input-output space, thus the clustering results become less dependent on the initial partition of the variables domain (Setnes, 2000).

2.3 Fuzzy Clustering

Clustering is a tool that attempts to assess the relationships among patterns of the data set by organizing the patterns into groups or clusters such that patterns within a cluster are more similar to each other than are patterns belonging to different clusters (Xie and Beni, 1991). Fuzzy clustering methods provide fuzzy partitions such that each object or data point could belong to two or more clusters with different degrees of membership. So, fuzzy clustering methods provide an adequate tool for representing real-data structures. A fuzzy c-partition is defined as follows.

Definition 2.1. Fuzzy c-partition: Let a data set $X = \{\mathbf{x}_1, \ldots, \mathbf{x}_n\}$, where $\mathbf{x}_k \in \Re^p$. V_c is the set of all real $c \times n$ matrices $U = [u_{ik}]$. The set M_{fc} of fuzzy c-partitions matrices is:

$$M_{fc} = \{U \in V_c \mid u_{ik} \in [0,1], \ 1 \leq i \leq c, \ 1 \leq k \leq n; \ \textit{and} \ \sum_{i=1}^{c} u_{ik} = 1, \forall k \in \{1, 2, \ldots, n\}\}, \quad (7)$$

where u_{ik} is the membership value of the data sample $\mathbf{x}_k$ to the cluster c_i.

In the fuzzy c-means (FCM) clustering algorithm, the objective is to find $U[u_{ik}] \in M_{fc}$ and $V = (\mathbf{v}_i, \ldots, \mathbf{v}_c)$, $\mathbf{v}_i \in \Re^p$ such that $J_m(U, V)$ (equation (8)) is minimized.

$$J_m(U, V) = \sum_{k=1}^{n} \sum_{i=1}^{c} (u_{ik})^m \|\mathbf{x}_k - \mathbf{v}_i\|^2, \quad (8)$$

where $m \in (1, \infty)$ is the fuzziness index. The approximate optimization given by the FCM algorithm is described in figure 4 (Bezdek, 1987).

- Step 1: Let the data set X, set $c \in \{2, 3, \ldots, n-1\}$, $m \in (1, \infty)$, and initialize $U^{(0)} \in M_{fc}$.
- Step 2: In the iteration q, $q = 0, 1, 2, \ldots$, compute the vectors of V, from the new center matrix C_i:

$$\mathbf{v}_i^{(q)} = \frac{\sum_{k=1}^{n} \mathbf{x}_k (u_{ik}^{(q)})^m}{\sum_{k=1}^{n} (u_{ik}^{(q)})^m}. \quad (9)$$

- Step 3: Refresh $U^{(q)} = [u_{ik}^{(q)}]$ to $U^{(q+1)} = [u_{ik}^{(q+1)}]$ such that:

$$u_{ik}^{(q+1)} = \frac{1}{\sum_{j=1}^{c} \left(\frac{\|\mathbf{x}_k - \mathbf{v}_i^{(q)}\|}{\|\mathbf{x}_k - \mathbf{v}_j^{(q)}\|} \right)^{\frac{2}{m-1}}}, \quad (10)$$

  where: $1 \leq i \leq c, \ 1 \leq k \leq n$
- Step 4: If $\|U^{(q+1)} - U^{(q)}\| < \epsilon$, then stop the process; otherwise, $q = q + 1$ and go back to the step 2.

Fig. 4. The fuzzy c-means algorithm.

2.4 Initial rule base generation

In this phase, the samples from the data set are classified into ncr clusters. Each one of these clusters corresponds to a rule, and each cluster center coordinate corresponds to a point in the input and output domains (the rule antecedent and consequent parts), as shown in figure 5.

The corresponding membership functions of the fuzzy sets in the rules are obtained from the partition matrix U. One dimensional fuzzy sets A_{ij} and B_i (from the ith cluster) are obtained from the multidimensional fuzzy sets defined point-wise in the ith row of the partition matrix $U = [u_{ik}]$ by fuzzy projections (Lee, 1990) onto the input

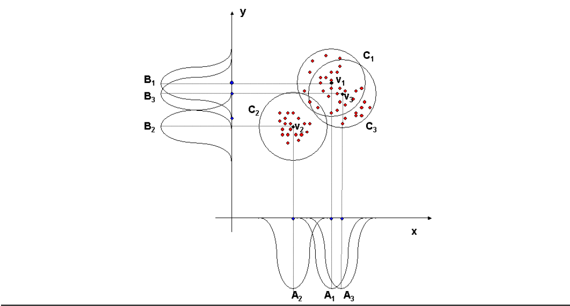

Fig. 5. The clusters projections.

and output domains. This procedure generates the fuzzy sets discrete points, which can be approximated by parametric membership functions. The gaussian function has been selected to represent the fuzzy sets. In this case, the parameters of the membership functions can be easily found from the discrete fuzzy sets. For each fuzzy set, the element with the highest grade of membership function is the center of the gaussian function, and the standard deviation is computed from the point-wise values. The figure 6 shows the rule base obtained from the clusters projection.

C_1: If x is A_1 Then y is B_1
C_2: If x is A_2 Then y is B_2
C_3: If x is A_3 Then y is B_3

Fig. 6. Rule base from clusters projection.

2.5 Initial rule base simplification

Fuzzy rule-based models obtained from data often contain redundancy present in the form of similar fuzzy sets representing compatible concepts. This hampers the transparency and interpretability of the model (Babuska et al., 1998). In order to look for the existence of similar fuzzy sets, the parameters of the membership functions are verified as follows.

Considering each domain, the membership functions whose center values are too close are considered similar. In this way, these membership functions are merged, and the new center value is the average of the similar centers, and the new standard deviation value is the highest value among the respective standard deviation values. After this process, the rules are refreshed with the new membership functions.

This simplification procedure is concerned with the merging of the membership functions that describe almost the same region on each domain. Reduction of the rule base is only a consequence of this merging. The structure improvement performed in this phase plays an important role in the fuzzy model identification because it increases the model interpretability and transparency, two essential features of the fuzzy systems (Guillaume, 2001).

2.6 Conflicting rules treatment

The union of similar fuzzy sets can generate conflicting rules, i.e., rules with the same antecedent part, and distinct consequents. In order to overcome this problem, a variant of the well-known confidence factor (CF) (Quinlan, 1987) measure is used. Let $A \to C$ denote a rule, in which A is the rule antecedent (a conjunction of conditions) and C is the rule consequent (the value predicted for the goal attribute). The CF is denoted by the equation (11) (Nagai and Arruda, 2002):

$$CF(R_i) = \frac{|A \wedge C| - 1/2}{|A|}, \tag{11}$$

where $|.|$ denotes the cardinality.

In a group of conflicting rules, the rule with the highest value of CF is kept, and the others are removed from the rule base.

3. PROPOSED SOFT SENSOR

This section summarizes the fuzzy model identification method:

- Step 1: Select the fuzzy model input set from the secondary variables set by means of the algorithm in figure 2;
- Step 2: Select the input-output lag space by means of the algorithm in figure 3;
- Step 3: Establish the training and test data sets, the initial rule base size ncr, the FCM algorithm parameter m, and the acceptable error rate ξ;
- Step 4: Apply the FCM algorithm to the training data set, in order to obtain ncr clusters, it means, the initial rule base;
- Step 5: Realize the clusters projection into each variable domain to get the point-wise defined membership functions, and calculate the standard deviation from these values;
- Step 6: Simplify the initial rule base by merging the similar membership functions, and refresh the rule base with the new membership functions;
- Step 7: Verify the conflicting rules existence in the rule base, calculate the confidence factor for each of them, and the rule with the highest confidence factor value is kept, and the others are removed;
- Step 8: Calculate the test error by applying the obtained fuzzy model. If the error is bigger than ξ, then return to the Step 4. Finish the procedure, otherwise.

4. RESULTS

This section presents the simulation results obtained from the application of the proposed fuzzy model identification methodology. The proposed methodology has been applied for the development of a soft sensor (inferential model) to infer the top composition of a distillation column.

The simulation data has been obtained from (Fabro, 2003; Neves-Jr and Aguilar-Martin, 1999), in which a binary distillation column that separates water from methanol has been modeled by means of $HYSYS^{©}$ [2]. The column has 20 trays (figure 7) and it presents the following behaviour.

- feed composition: 50% water and 50% methanol;
- feed temperature: 75.9^oC;
- feed molar flow: 236 kgmol/h (inserted at tray 14);
- reboiler level: 50% (liquid volume);
- bottom product flow and composition: 152 kgmol/h, with methanol concentration of less than 0.01%;
- condenser level: 20% (liquid volume);
- distillate flow and composition: 85 kgmol/h, with methanol concentration more than 99.9%;
- dry initial conditions for the reboiler, condenser and trays;
- reflux flow: 247 kgmol/h;
- column pressure: 1 atm (101.3 kPa);
- temperature profile:
  - · tray 9: 67.54^oC;
  - · tray 13: 86.41^oC;
  - · tray 16: 101.8^oC;
  - · tray 18: 102.3^oC.

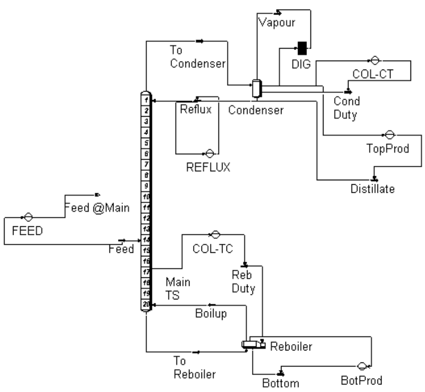

Fig. 7. The distillation column.

The temperature sensors are positioned at the trays 9, 13, 16, and 18. As the temperature at these trays approaches the above-specified values, the column is approaching its steady-state conditions. When the temperature profile is reached, the column is at its nominal production, and the startup procedure can be considered ended. At this point, the composition of the top and bottom flows are better than 99.9% pure.

In this case, the simulations provide the compositions information, but in the real processes, these composition values are often obtained from laboratory analysis (Qin, 1996). In Fabro (2003) neural estimators (neural soft sensors) have been developed to infer the top and bottom compositions by using the temperature readings. In this paper a soft sensor based on fuzzy model is developed to infer the top product composition during the startup procedure.

At the beginning of the startup, when the column is still being warmed, it is not critic to know the compositions. As the startup advances, it is more important to obtain good predictions, especially when the compositions approach the value of 98%. So, the developed soft sensor must infer the compositions of the top product (distillate) from 90% to 99% of purity.

The available secondary variables are described in table 1. The table 2 presents the selected input variables ($u_1, \ldots, u_7$), and the number of membership functions for each model variable (including the output variable) obtained after the fuzzy model simplification procedure (section 2.5).

Table 1. Secondary variables description.

#	variable name	description
1	$feed1\_pv$	feed molar flow
2	$reflux\_pv$	condenser reflux
3	pv_press	top column pressure
4	$bottom\_pv$	reboiler liquid level
5	top_pv	condenser liquid level
6	$tray9$	temperature of tray 9
7	$tray13$	temperature of tray 13
8	$tray16$	temperature of tray 16
9	$tray18$	temperature of tray 18
10	$cond\_flow$	flow to condenser
11	top_flow	top product flow
12	$bottom\_flow$	bottom product flow

Table 2. Selected variables.

#	variable name	# membership functions
u_1	$feed1\_pv$	8
u_2	top_pv	8
u_3	$tray9$	9
u_4	$tray13$	10
u_5	$tray16$	8
u_6	$tray18$	7
u_7	$cond\_flow$	8
u_8	methanol composition	9
y	methanol composition	11

As described in table 2, seven input variables have been chosen, and by considering a feedforward predictive model with the selected lag, the fuzzy model parameters are shown in table 3. In the present implementation, in order to avoid a very time-consuming computation, the number of past inputs and output is increased simultaneously. Thereby, as presented in table 3, $lag_y = lag_u = 2$.

The model performance is shown in table 4, in which the employed performance indexes have been the mean-square error (mse) and the normalized root mean-square error ($nrmse$) (Nagai and Arruda, 2002).

[2] HYSYS 3.0, HYPROTECH Ldt.

Table 3. The fuzzy model parameters.

description	value
initial rule base size	100
simplified rule base size	45
number of input variables	15
number of output variables	1
selected lag	2

Table 4. The performance indexes.

	train	test	total
mse	0.0482	0.0505	0.0494
$nrmse$ (%)	0.2287	0.2348	0.2318

Figure 8 shows the methanol composition inference generated by the obtained soft sensor (doted line) and the expected composition (solid line).

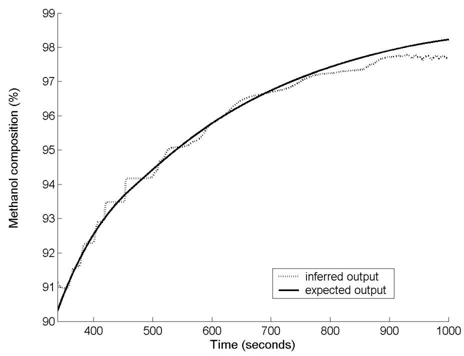

Fig. 8. Methanol composition inference.

5. CONCLUDING REMARKS

This work presents a fuzzy model identification approach to build soft sensors. This approach is done in three phases. First of all, the input variables to the soft sensor are selected by applying Kohonen maps and the Lipschitz quotients are used to select the structure of a linguistic fuzzy model. Second, the FCM algorithm is applied in order to find the model variables membership functions, and also the rules that describe the process behaviour. In the third phase, the similar membership functions are merged of such way to eliminate the redundancy on each variable domain description. This simplification can also reduce the rule base size, increasing the model interpretability and transparency.

The proposed methodology has been evaluated by the construction of a soft sensor for the inference of the distillation column top composition. Some of the obtained results in section 4 indicate that the methodology has been capable to generate a composition estimator whose performance could be improved through fuzzy model refinement techniques, such as a method for tuning of the membership functions parameters, and a method for the evaluation of each rule relevance in order to reduce the rule base size.

REFERENCES

Babuska, R., U. Kaymak, H. R. van Nauta and M. Setnes (1998). Similarity measures in fuzzy rule base simplification. IEEE Trans. on Sys., Man, and Cyb. - Part B **28**(03), 376–386.

Bezdek, J. C. (1987). Pattern Recognition with Fuzzy Objective Function Algorithms. Plenum Press. New York, NY.

Fabro, J. A. (2003). A neuro-fuzzy approach to predictive control of multiple stages process. PhD thesis. CPGEI-CEFET-PR. Curitiba-PR-Brazil. In Portuguese.

Guillaume, S. (2001). Designing fuzzy inference systems from data: an interpretability-oriented review. IEEE Trans. on Fuzzy Systems **09**(03), 426–443.

He, X. and H. Asada (1993). A new method for identifying orders of input-output models for nonlinear dynamical systems. In: Proc.of ACC'93. SF, California. pp. 2520–2523.

Kaski, S. and K. Lagus (1996). Comparing self-organizing maps. In: Proc. of ICANN'96.

Lant, P. A., G. A. Montague, A. J. Morris and M. T. Tham (1991). Soft-sensors for process estimation and inferential control. J. Proc. Cont. **01**, 3–14.

Lee, C. C. (1990). Fuzzy logic in control systems: Fuzzy logic controller-part i and ii. IEEE Trans. on Sys., Man and Cyb. **20**, 404–418.

Nagai, E. Y. and L. V. R. Arruda (2002). Automatic generation of fuzzy models by genetic algorithms. In: 15th IFAC. Barcelona, Spain.

Neves-Jr, F. and J. Aguilar-Martin (1999). Qualitative event-based expert supervision for transient condition monitoring. Applied Intelligence **10**, 197–210.

Qin, S. J. (1996). Neural networks for intelligent sensor and control — Practical issues and some solutions. Chap. 8, pp. 215–236. Academic Press.

Quinlan, J. R. (1987). Generating production rules from decision trees. In: Proc. of IJCAI-87. pp. 304–307.

Rallo, R., J. Ferre-Gine, A. Arenas and F. Giralt (2002). Neural virtual sensor for the inferential prediction of product quality from process variables. Comput. and Chem. Eng. **26**, 1735–1754.

Setnes, M. (2000). Supervised fuzzy clustering for rule extraction. IEEE Trans. on Fuzzy Systems **08**(04), 416–424.

Xie, X. L. and G. Beni (1991). A validity measure for fuzzy clustering. IEEE Trans. on Pattern Analysis and Machine Intel. **13**, 841–847.

Zambrogna, E., M. Barolo and D. E. Seborg (2005). Optimal selection of soft sensor inputs for batch distillation columns using principal component analysis. J. of Proc. Cont. **15**, 39–52.