AIChE℠

SENSOR TECHNOLOGY

Conference Proceedings

AIChE Annual Meeting
November 3-8, 2002
Indianapolis, Indiana

Conference Chairs

Chung-Chiun Liu
Wallace R. Persons Professor of Sensor Technology and Control, Director, Electronics Design Ctr.
Case Western Reserve University
Bingham Building
10900 Euclid Ave.
Cleveland, OH 44106 USA
cxl9@po.cwru.edu

Heidi B. Martin
Nord Assistant Professor of Engineering
Chemical Engineering
Case Western Reserve University
A.W. Smith Building
10900 Euclid Ave.
Cleveland, OH 44106
hbm@po.cwru.edu

John C. Angus
Kent H. Smith Professor of Engineering
Chemical Engineering
Case Western Reserve University
A.W. Smith Building
10900 Euclid Ave.
Cleveland, OH 44106
jca3@po.cwru.edu

ISBN 0-8169-0891-5

AIChE Pub. No. P-172

FORECASTING PRODUCT QUALITY IN INDUSTRIAL PROCESSES WITH VIRTUAL SENSORS

R. Rallo, J. Ferre-Giné, A. Arenas and Francesc Giralt[#]

Departament d'Enginyeria Informàtica i Matemàtiques. Escola Tècnica Superior d'Enginyeria.
[#]Departament d'Enginyeria Química. Escola Tècnica Superior d'Enginyeria Química (ETSEQ).
Universitat Rovira i Virgili. Av. dels Països Catalans, 26, 43007 Tarragona, Catalunya, SPAIN.

Prepared for presentation at the 2002 Annual Meeting, Indianapolis, IN, Nov. 3-8
Unpublished

Abstract.

Several models for the design and implementation of virtual sensor systems based on advanced neural architectures that are capable to infer on-line the properties of a manufactured product from real process variables were developed and tested. The architectures considered were a modified Fuzzy ARTMAP network, and two hybrid networks each combining a dynamic unsupervised classifier with a different kind of supervised predictor. A new method to construct dynamically the unsupervised layer was developed. The values of the melt index or quality of different LDPE grades produced in a tubular reactor were inferred at the beginning of the process cycle for different operation conditions and feed compositions. The most relevant process variables to build the sensor were selected using self-organizing maps and dissimilarity measures. All neural sensors outperformed predictors based on linear correlation techniques when appropriate information was provided for training. The results obtained indicate that the virtual sensors developed are capable of learning the relationships between process variables measured at the beginning of the production cycle and the quality parameters of the final product.

1. Introduction and Motivation.

Real time quality control requires robust product quality measurement techniques, which often are either unavailable on-line or entail long time delays. Virtual or software sensors are an alternative to the use of hardware sensors or laboratory procedures. There are tree kinds of approaches available to build models that can be used as soft sensors in manufacturing processes: mechanistic models, statistical regression models and artificial intelligence based models.

A specific area of intrinsic interest to chemical process industries is the estimation of the quality of final products in polymerization processes. Due to the high nonlinearity, complexity and uncertainty of these processes, it is often difficult to obtain mechanistic models. Furthermore it is very complex to implement reliable and fast on-line analyzers to measure the properties of these products and to establish appropriate control strategies. In these situations data-driven approaches such are statistical methods or artificial neural networks (ANN) provide a useful alternative. Using these techniques, on-line estimators of product quality can be developed based on available process information. One of the most powerful and increasingly used methodologies is the inferential measurement (Martin, 1997). This method consists in the inference of product quality or difficult to measure process indicators from other more reliable or easily performed plant measurements, such are pressures, flow rates, concentrations or temperatures.

The purpose of the current study is to develop a virtual sensor to infer product quality from other more easily measured process variables using several adaptive neural network architectures. Different techniques for the selection of relevant variables and the construction of appropriate training and test sets are also proposed. The networks that have been considered are a modified Fuzzy ARTMAP network (Ferre-Giné et al. 1996), and a hybrid network that combines the construction of a dynamic unsupervised classifier with a supervised predictor (Rallo et al., 2002). The neural virtual sensors developed following this approach are applied to infer quality indicators of different low-density polyethylene (LDPE) grades measured on-line in operating plants. The quality of the polymer is determined essentially by the Melt Index (MI), which is measured by the flow rate of polymer through a die. The on-line measurement of this quantity is difficult and requires close human intervention because the extrusion die often fool and blocks. As a result, in most plants the MI is evaluated off-line with an analytical procedure that takes between 2 to 4 hours to complete in the laboratory. The paper is organized as follows. In the next section the architectures of virtual sensors are described, outlining the learning algorithms and operation procedures. Section 3 focuses on the preprocessing techniques used to perform the selection of relevant variables and relevant examples. Finally the results obtained are presented and discussed in section 4, together with some concluding remarks about the design and implementation of virtual sensors systems.

2. Architecture of the Virtual Sensor.

The proposed architecture for the virtual sensor consists of two main components: The preprocessing module, that performs the selection of the most relevant information to train the system, and a core module that contains the inferential predictor. Two neural models have been developed and evaluated as core modules: (i) A predictive Fuzzy ARTMAP architecture that has been capable of learning the dynamics of large-scale structures in a turbulent flow (Giralt et al., 2000) and (ii) an unsupervised classifier layer that is connected to an output layer, where outputs are either the average target property value of all the input patterns belonging to the cluster that is activated or the target property value that results from the incorporation of a kernel function in the cluster centers.

2.1. Fuzzy ARTMAP.

The Fuzzy ARTMAP neural network is formed by a pair of fuzzy ART modules linked by an associative memory and an internal controller (Carpenter et al., 1992). The Fuzzy ART architecture was designed by Carpenter et al. (1991) as a classifier for multidimensional data clustering based on a set of features. The classification procedure of fuzzy ART is based on Fuzzy Set Theory (Zadeh, 1965) and clusters the data having a value for its grade of fuzzy membership greater than a certain *vigilance parameter* into groups or classes. This parameter controls the granularity of the classes and allows the specification of the desired accuracy criterion in the classification procedure. The mechanisms to speed up the process and to conduct the classification properly can be found elsewhere (Carpenter et al., 1991). The Fuzzy ARTMAP architecture, which has been successfully applied to educe the different classes of large-scale events present in free turbulence (Ferre-Giné et al., 1996), was designed to classify data and, thus, cannot generate an output pattern after the training stage. To implement predictive capabilities the categories educed by the system from the learned information are linked to the desired outputs (Giralt et al., 2000). The accuracy of the procedure increases asymptotically towards a constant value with the number of examples used for training, i.e., when the space of outputs is accurately mapped.

2.2 Dynamic Unsupervised Layer.

The construction of the unsupervised layer is performed in three steps: (i) Set-up an initial configuration with the center of the first cluster formed by one pattern chosen randomly from the training dataset; (ii) define the minimal mean distance between training patterns as the maximum *attention radius* to control the generation of new nodes; (iii) present a new input pattern to the network and compute the Euclidean distance between the pattern and all nodes. The structure of the network is adapted according to the following two rules. If the input pattern is located inside the region of influence of any node, the pattern is classified and the center of this node is adapted using a winner takes-all approach based on the Kohonen's learning rule (Kohonen, 1982). Otherwise, a new node is created with the center located at the point that defines the input pattern. The procedure is repeated until the number of nodes stabilizes and either the classification or the number of iterations reaches a predetermined minimum or maximum value, respectively. The current algorithm creates an appropriate number of clusters since it determines the *attention radius* based on the distribution of the training patterns.

2.3. Output Models for the Dynamic Unsupervised Layer.

Two techniques, based on a hybrid approach that combines the current dynamic unsupervised classifier with a supervised learning engine, are used to obtain an output from the unsupervised layer. The first is a *clustering average* that labels the unsupervised layer using the values of the target variable. One of the most common labeling processes consists in averaging the target value for each of the training patterns belonging to a given cluster, like in the k-means algorithm (MacQueen, 1967). This averaged value is subsequently the output of the network. Once the dynamic unsupervised layer is labeled, the network is ready to infer the target property values, i.e., act as a virtual sensor. The second technique consists in the placement of a set of Radial Basis Functions (RBF) (Moody and Darken, 1989) over the cluster centers using a supervised training procedure to adjust the output. This neural network is herein after identified as *Dynamic Radial Basis Function network* (Rallo et al. 2002). The idea is to pave the input space (or the part of it where the input vectors lie) with the receptive field of these Gaussian functions. A map between the RBF outputs and the desired process outputs is then constructed in a second supervised training stage. First, the unsupervised layer that defines the number of radial functions in the hidden layer as well as their position and width in the input space is constructed. The activation of the RBF layer for a given input pattern is related to the desired output in a second supervised learning stage.

3. Preprocessing of data.

In chemical processing plants, the number of process variables that can be measured is very large and the sampling rates used for these measurements are usually high. This implies the generation of large datasets containing lots of features. In those situations it is very useful to have an "intelligent system" capable of selecting the most relevant features needed to build an accurate and reliable model for the process. The problem of feature selection is a common challenge in most fields of engineering, mainly due to problems related to noise during the data acquisition process, and the presence of contradictory information due to the inclusion of irrelevant or redundant variables. The selection of a good set of process variables to learn a given concept during the training of the neural network is a key issue in data pre-processing. A reduction in the dimension of the input space would also simplify the input layer of the neural architecture and reduce the time needed for training. Additionally, if the class to which a certain input pattern belongs is known it should be possible to figure out the features that best discriminate between the different values of the target properties. Thus, the problem of feature selection is an optimization problem that involves searching the space of all possible feature subsets to identify one that is optimal or near optimal with respect to a certain performance measure.

3.1. Selection of relevant Variables.

Variable selection methods can be represented as heuristic search problems and, thus, be classified in terms of four basic parameters that drive its operation mode. (i) The *starting point* (or points) in the search space, which determines the components of the initial variable configuration to perform the selection process. This leads to *forward selection*, in which the process is started without variables or with only a small subset of them. Then it proceeds adding new variables after each iteration, *backward elimination*, in which the process is started with the complete set of variables and then deleting some of them after each iteration. (ii) The *organization of the search* process in the space of variables. An exhaustive search is usually impractical, because for a given problem there are 2^n possible subsets of n attributes and this high dimension makes this approach infeasible in the majority of real applications. Another more realistic approach is based on a greedy method to traverse the variable space. In this situation, only a subset of all the possible configurations is explored. (iii) The *strategy to evaluate the subsets* of attributes. A metric to evaluate the goodness of an attribute is needed once the search over the space of variables starts. Commonly used metrics are the attribute's ability to discriminate among classes that are present in the training data set. (iv) The *criterion to stop* the search procedure. One may stop adding or removing attributes when none of

the alternatives improves the estimated accuracy of the previous set or continue generating candidate sets until reaching the other end of the search space and then selecting the best candidate. A more robust alternative is to perform an ordering of the variables according to some relevance measure and then using some threshold value to determine the break point.

The method currently used (Espinosa et al., 2001) is based on the projection of selected subsets of the input space onto the space generated by a topology-preserving clustering process such as the Self Organizing Map (Kohonen, 1990). The comparison of the resulting maps using some dissimilarity measure gives an indicator of the relevance of each combination of variables. In the present work we use the dissimilarity measure proposed by Espinosa et al. (2001) to compare the positions of the reference vectors in different map structures. This measure is based on a *map goodness* measure (Kaski and Lagus, 1997) that combines and index of the continuity of the mapping from the dataset to the map grid with a measure of accuracy of the map. The dissimilarity of two maps L and M is defined as the average difference of its goodness

$$D(L,M) = E\left[\frac{|d_L(x) - d_M(x)|}{d_L(x) + d_M(x)}\right] \qquad [1]$$

In this equation E is the average expectation, and $d(x)$ the distance over the map from the winner neuron or best matching unit (BMU), denoted by $m_{bmu(x)}$, to the second best cluster or BMU, denoted by $m_{bmu'(x)}$. Of all possible paths between $m_{bmu(x)}$ and $m_{bmu'(x)}$ the shortest path passing continuously between neighbor units is selected,

$$d(x) = \|x - m_{bmu}(x)\| + \min_i \sum_{k=0}^{K_{bmu'(x)}-1} \|m_{I_i(k)} - m_{I_i(k+1)}\| \qquad [2]$$

To confirm that equation [1] measures dissimilarity between pairs of maps the dissimilarity between an initial map and other maps trained using the same dataset and learning parameters but progressively contaminated by Gaussian noise was computed. The dissimilarity between maps increased with noise level, indicating that the maps were becoming progressively more different, as should be the case for a function of this type.

3.2. Methodology.

The identification of redundant variables has been carried out with a redundancy index based on correlation properties of the input data and their projection over the SOM. Once the SOM is properly trained the weight vectors are used to build a "sliced" representation of the map showing the distribution of each component (variable). This representation is named *component-plane*. Also, the *unified distance matrix* (u-matrix) (Ultsch and Siemon, 1990) can be computed as the distance between the weight vectors of adjacent units in the map. From the u-matrix, the different component planes can be obtained as well. Figure 1a shows the representation of the component planes of the original SOM while figure 1b depicts the component planes of its associated u-matrix. Components planes show the distribution of the data over the SOM space while the u-matrix planes represent the clustering of each variable over the map space. The redundancy index (RI) that takes into account not only the correlation between variables but also the correlation of their distribution and its clustering properties is given by,

$$RI_{i,j} = |corr(i,j)| + |corr(cplane_i, cplane_j)| + |corr(umat_i, umat_j)| \qquad [3]$$

If $RI_{i,j}$ is greater than a certain threshold the variable j is discarded from the dataset because the information that it provides is redundant with that of variable i.

The next step is the definition of the *starting point* in the variable space to start the search for the best set of variables. This process is based on the clustering of the SOM and in the assumption that variables located in the same cluster contribute with similar information to the model. Thus, the starting point for the search is determined by choosing a representative variable for each cluster. The cluster prototypes are the variables with the highest correlation with the target variable. To avoid the inclusion in the initial set of irrelevant variables only those variables with a correlation value with the target property higher than the average correlation for the whole set of variables are considered. The criterion of a minimum Davies-Bouldin index (Davies and Bouldin, 1979) has been applied to determine the optimal number of clusters. This index is a function of the ratio between the sum of cluster compactness and of cluster separations.

The organization of the search proceeds from the starting point by building subsets of variables with the addition of the rest of variables, one by one, ranked by its correlation with the target property. For each of these subsets the dissimilarity with the rest of possible subsets is computed using equation [1]. The process stops when the dissimilarity between all possible configurations has been computed. This procedure reduces the complexity of the search algorithm from $O(2^n)$ to $O(n^2)$, being n the number of input variables. The smallest average dissimilarity value for any given set of input variables indicates the similarity in quality and quantity of the information presented by the maps and, thus, the process of including variables to

form the best set of input variables can be stopped when the dissimilarity measure stabilizes, i.e., the maps for these different variables are very similar. Note that any increase in dissimilarity with the inclusion of additional input variables to the previous subset indicates that they do not provide any additional relevant information.

4. Results and Discussion.

In the current study the 25 process variables (pressures, flow rates, temperatures of the cooling/heating streams of the reactor, etc.) listed in Table 1 have been chosen to characterize and to predict the time-variation of MI with the virtual sensors. The data presented in the following analysis correspond to time intervals of 10 minutes. The mean residence time of the materials in the reactor was $\tau \approx 30$ minutes.

Variable Name	Units	|R| with MI_{plane}	Selected	Cluster
compressor throughput	tm/h	0.041		I
concentration 1	%	0.529	√	V
concentration 2	%	0.248	√	II
concentration 3	%	0.169	√	II
concentration 4	%	0.022		I
density	g/cm^3	0.180	√	I
extruder power	A	0.590	√	I
extruder speed	rpm	0.052		V
flow rate 1	kg/h	0.021		I
flow rate 2	kg/h	0.043		I
flow rate 3	kg/h	0.330		IV
flow rate 4	kg/h	0.526	√	V
level	%	0.125		IV

Variable Name	Units	|R| with MI_{plane}	Selected	Cluster
Melt Index	g/10min	-	(target variable)	III
Pressure	kg/cm^2	0.054		I
temperature 1	°C	0.313	√	I
temperature 2	°C	0.026		I
temperature 3	°C	0.522	√	I
temperature 4	°C	0.114		I
temperature 5	°C	0.122		I
temperature 6	°C	0.138	√	IV
temperature 7	°C	0.429	√	I
temperature 8	°C	0.447	(redundant≈temp7)	-
temperature 9	°C	0.111		I
vol. flow rate 1	ℓ/h	0.600	√	III
vol. flow rate 2	ℓ/h	0.630	√	V

Table 1. Variable identification and correlation with the Melt Index for the ensemble of all LDPE grades produced. The columns on the right show the most relevant variables detected using the Self Organizing Map and its cluster assignment. Temperature 8 has been detected as a redundant variable.

The process of detection of redundant variables yields the component planes shown in figure 1.

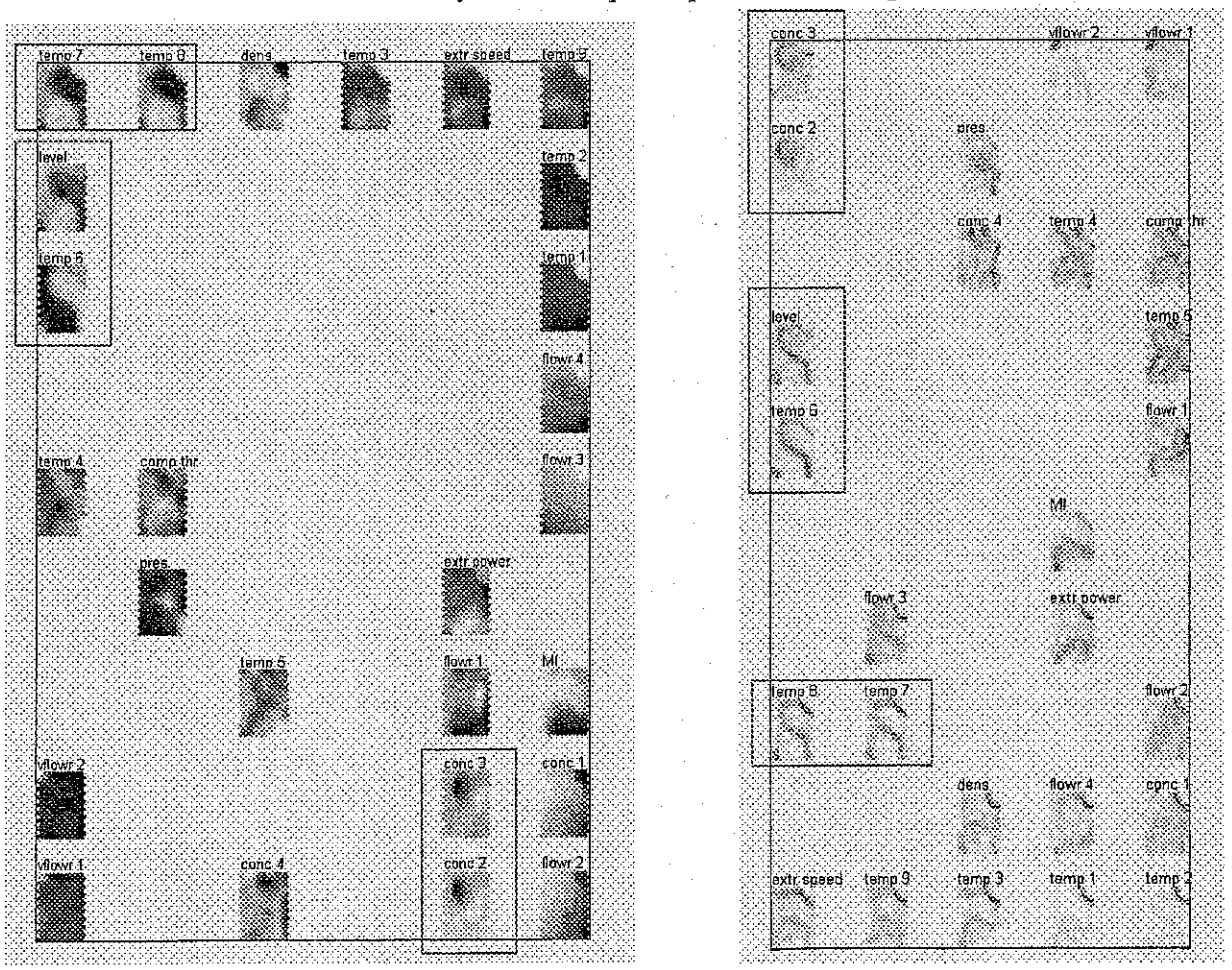

Figure 1. Topologically ordered component planes of a Self organizing Map trained using all the available variables. (a) Component planes of the SOM, showing the distribution of values of input variables; (b) component planes of the u-matrix showing cluster borders in dark gray.

Based on these component planes and the redundancy index [3], the pair of variables labeled as [*temp8,temp7*] is detected as redundant. Also, this figure shows that other pairs of variables contribute with similar information, e.g., [*level,temp6*] and [*conc2, conc3*], but their redundancy index is below the threshold value and consequently are not discarded. After removing the redundant variables from the dataset a map of 17x12 units was trained and the component planes of the u-matrix extracted. Figure 2 shows the resulting U-matrix of these components after being clustered using a 10x10 SOM. In this representation each map node has been labeled with the name of the classified variables. Since clusters are difficult to detect unambiguously by visual inspection a cluster validity index has been used for this purpose. Figure 3 shows the five clusters obtained after the minimization of the Davies-Bouldin index for the k-means clustering of these component planes.

The cluster assignment for each variable is presented in Table 1. In cluster I, which is the largest, the variables located in the upper side are variables with low correlation values with the target and the variables located in the lower part of the cluster correspond to variables with higher correlations. The representative variable or prototype of this cluster is the *extruder power consumption*. Cluster II is formed by the pair of variables [*conc2,conc3*], which are not included in the initial set since none of them has a correlation value higher than the average. For the rest of the clusters three flow rates are selected as representative variables.

After computing the dissimilarity for each of the possible subsets of variables, nine additional variables were chosen as members of the set of most relevant process information. Table 1 identifies the reduced set of 13 selected variables. The initial dimension of the problem is thus reduced by nearly to the 50% by using the current method.

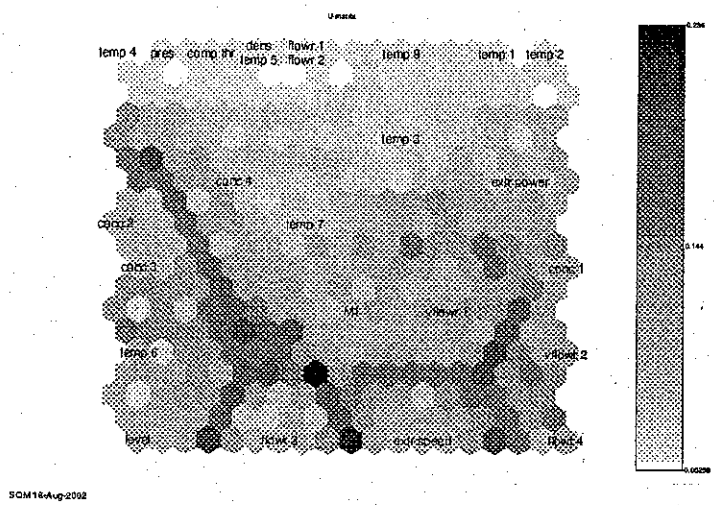

Figure 2. U-Matrix of the SOM trained using the component planes corresponding to the set of variables without redundant information. High values of distance (dark gray) represent cluster borders. Low values represent compact clusters with similar variables.

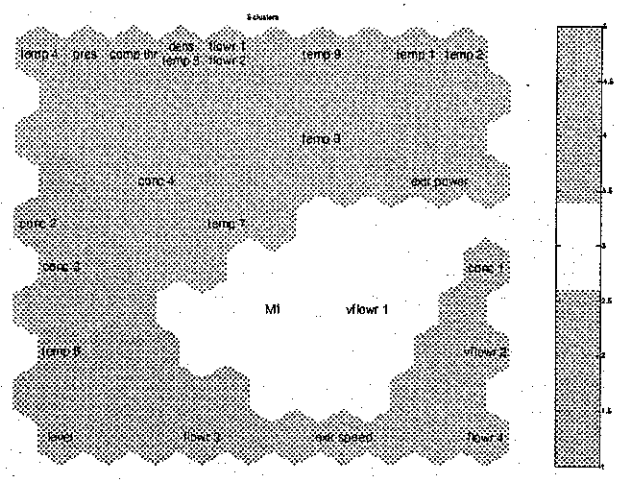

Figure 3. Detection of the optimal set of clusters by minimization of the the Davies-Bouldin index. Gray levels identify the different clusters.

The three neural sensors were trained using a data set of 14,111 patterns corresponding to the operating condition of the production of six different product grades of LDPE. The validation of the trained sensors was performed using a test dataset formed by 589 patterns never seen before by the neural network. Table 2 summarizes the results obtained for each sensor using both the complete and the reduced sets of process variables.

	Modif. Fuzzy ARTMAP	Clustering Average	DYNARBF
Complete set of 24 non-redundant variables	0.207	0.130	0.135
Reduced set of 13 relevant variables	0.156	0.122	0.123

Table 2. Absolute mean errors of the inferred MI obtained using the three virtual sensors with the complete set of data and with the reduced set.

It can be observed in Table 2 that despite the important reduction in the dimension of the input space from 24 to 13 variables and, thus, in the complexity of the virtual sensor, the performances of all the three virtual sensors slightly increase since only relevant (not noisy or contradictory) information is used. Figure 4 depicts the results obtained for the clustering average virtual sensor trained using the complete and the reduced sets of process variables. The use of the reduced set of 13 variables enhances the capacity of the virtual sensor to properly classify the process states and adequately predict MI values. As a

consequence, a reduction on erroneous sensor responses is observed in the time series depicted in figure 5. This effect is more evident for LDPE grades having high values of MI.

The proposed virtual sensor architecture combined with the automated procedure for the selection of relevant information provides a reliable and accurate framework for the design and implementation of virtual sensor systems. The results obtained indicate that virtual sensors are capable of learning the relationships between process variables measured at the beginning of the production cycle and the quality parameter of the final product.

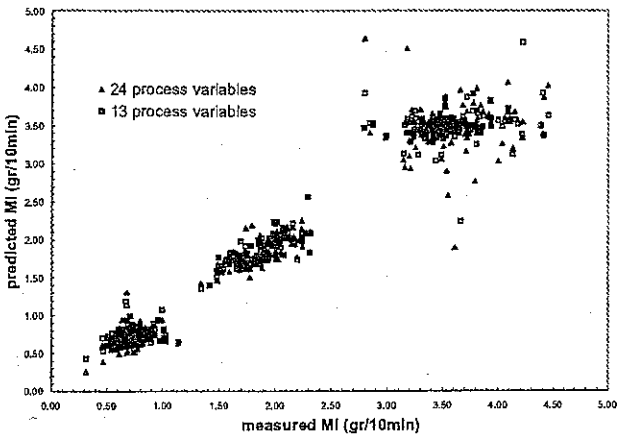

Figure 4. Measured and predicted MI using the Clustering average model with the complete and the reduced sets of input variables.

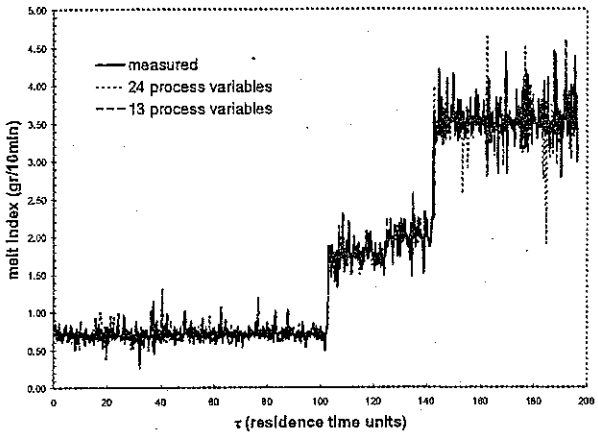

Fig. 5. Measured and predicted time-records using the Clustering average model with both sets of input process variables for 3 LDPE grades.

Acknowledgement.

This research was supported by DGICYT (Spain), projects PB96-1011 and PPQ2000-1339, and Comissionat per a Universitats i Recerca (Catalunya), projects 1998SGR00102 and 2000SGR00103.

References.

Carpenter, G.A., and Grossberg, S. and Rosen, D. "Fuzzy ART: Fast stable learning and categorization of analog patterns by an adaptative resonnace system" Neural Networks, 4, 759 (1991).

Carpenter, G.A., Grossberg, S., Marcuzon, N., Reinolds, J.H., and Rosen D.B. "Fuzzy ARTMAP: A neural network architecture for incremental supervised learning of analog multidimensional maps" IEEE Transactions on Neural Networks, 3, 698 (1992).

Davies, D.L., Bouldin, D.W. "A cluster separation measure" IEEE Transactions on Pattern Analysis and Machine Intelligence, 1, 224 (1979).

Espinosa, G., Yaffe, D., Arenas, A., Cohen, Y., and Giralt, F. "A fuzzy ARTMAP based quantitative structure-property relationships (QSPRs) for predicting physical properties of organic compounds" Industrial and Engineering Chemistry Research, 40(12), 2757 (2001).

Ferre-Giné, J., Rallo, R. , Arenas, A. and Giralt, F. "Identification of coherent structures in turbulent shear flows with a fuzzy ARTMAP neural network" International Journal of Neural Systems, 7, 559(1996).

Giralt, F., Arenas, A., Ferre-Giné, J., Rallo, R, Kopp, G.A. "The simulation and Interpretation of free turbulence with a cognitive neural system" Physics of Fluids, 12, 1826 (2000).

Giralt, F., Espinosa, G., and Arenas, A. "An Integrated Cognitive based System for the evaluation of Physicochemical Properties and Biological Activities", V Girona Seminar on Molecular Similarity, Girona, July, (2001).

Kaski, S. and Lagus, K. "Comparing Self-organizing Maps" Proceedings of ICANN'96, International Conference on Artificial Neural Networks, 809 (1997).

Kohonen, T. "Self-organized formation of topologically correct feature maps" Biological Cybernetics, 43, 59-69 (1982).

Kohonen, T. "The Self-Organizing Map" Proceeding of IEEE, 78(9), 1464 (1990).

Martin, G. "Consider Soft Sensors" Chemical Engineering Progress, july 66 (1997).

MacQueen, J. "Some methods of classification and analysis of multivariate observations" Proceeding of 5th Berkeley Symposium on Mathematics, Statistics and Probability (I), 281 (1967).

Moody, J., and Darken, C. J. "Fast learning in networks of locally-tuned processing units" Neural computation, 1, 281 (1989).

Rallo, R., Ferre-Giné, J., Arenas, A., Giralt, F. "Neural Virtual Sensor for the inferential prediction of product quality from process variables" Computers & Chemical Engineering , 2002 (in press).

Ultsch, A., Siemon, H. "Kohonen's Self Organizing feature maps for exploratory data analysis" Proceedings of INNC'90. International Neural Networks Conference, 305 (1990).

Zadeh, L. "Fuzzy sets" Information and Control, 8, 338 (1965).