Towards a Model of Input-Output Behaviour of Wastewater Treatment Plants using Soft Computing Techniques

Lluís A. Belanche[1] Julio J. Valdés[1] Joaquim Comas[2]
Ignasi R.-Roda[2] Manel Poch[2]

[1] Secció d'Intel·ligència Artificial.
Dept. de Llenguatges i Sistemes Informàtics.
Universitat Politècnica de Catalunya.
c/Jordi Girona 1-3. 08034 Barcelona, Spain.
{belanche, valdes}lsi.upc.es
[2] Laboratori d'Enginyeria Química i Ambiental.
Facultat de Ciències.
Universitat de Girona.
Campus de Montilivi s/n 17071 Girona, Spain.
{quim, irroda, mpoch}@lequia1.udg.es

Abstract. Wastewater Treatment Plants (WWTPs) control and prediction under a wide range of operating conditions is an important goal in order to avoid breaking of environmental balance, keep the system in stable operating conditions and suitable decision-making. In this respect, the availability of models characterizing WWTP behaviour as a dynamic system is a necessary first step. However, due to the high complexity of the WWTP processes and the heterogeneity, incompleteness and imprecision of WWTP data, finding suitable models poses substantial problems. In this work, an approach via soft computing techniques is sought, in particular, by experimenting with fuzzy heterogeneous time-delay neural networks to characterize the time variation of outgoing variables. Experimental results show that these networks are able to characterize WWTP behaviour in a statistically satisfactory sense and also that they perform better than other well-established neural network models.

Keywords: Wastewater Treatment Plants; System Identification; Heterogeneous Neural Networks; Environmental Modeling.

SOFTWARE AVAILABILITY

Name of the software:	HeterogNeu - Heterogeneous Neural Networks
Developer:	Lluís A. Belanche, Julio J. Valdés and R. García
Contact address:	Secció d'Intel·ligència Artificial. Dept. de Llenguatges i Sistemes Informàtics. Universitat Politècnica de Catalunya. c/Jordi Girona 1-3. 08034 Barcelona, Spain.
Telephone/Fax:	+34 93 401 56 44/+34 93 401 70 14
E-mail:	{belanche, valdes}lsi.upc.es
Year first available:	1999
Program language:	C++
Availability:	It is a prototype. Available only for personal use.

1 INTRODUCTION

Nowadays, proper management of wastewaters in modern industrialized societies is not only an option, but a necessity. The main objective is to maintain natural water systems at as high a quality level as possible, and to ensure equilibrium between supply and demand through a rational use and management of water resources. Moreover, the wastewater treatment would help to reach the attainment of rivers as biological corridors, which means to ensure a good quality of life for animals and vegetals living in the water. Wastewater coming from different municipal uses contains a wide variety of contaminants. Among them, the most commonly found in municipal wastewater are total suspended solids (TSS), organic matter —measured as biochemical oxygen demand (BOD) and chemical oxygen demand (COD)— pathogens, and nutrients. The basis of wastewater treatment processes lies in oxidizing biodegradable organics from raw water into stabilized, low-energy compounds, maintaining a mixture of microorganisms and supplying oxygen by aerators (WEF, 1996).

The autonomous Government of Catalonia, according to the European directive of the Council 91/271/CEE, is developing its *Pla de Sanejament* (Government, 1996), which foresees wastewater treatment for populations greater than $2,000$ inhabitants-equivalents before year 2005. To achieve these purposes, more than 200 Wastewater Treatment Plants (WWTP) have already been built in Catalonia, treating an average daily wastewater flow of about $2,000,000\ m^3$.

Although it is very important to ensure the quality of the treated wastewater prior to discharge, the correct control and operation of the process carried out in the WWTP is not a well established task. Some of the factors which affect the real-time control of the process are:

– the biological nature of the process, involving the presence of a true trophic web,
– the great complexity and variability of the influent composition,
– the lack of on-line sensors and signals,
– the delay of the analytical results from the laboratory (WEF, 1992): minutes, hours or even days according to the different TSS (30 minutes), COD (2 hours) or BOD (5 days) determinations, and
– the dynamic state of the process.

Different, classical control methods (among which we can find feed-forward, feedback, adaptative, predictive, etc.), have been used to improve and optimize WWTP operation (Dochain, 1991; Moreno, 1992; Isaacs, 1992; Heinzle, 1993; Nejjari, 1997; Queinnec, 1998).

It seems then interesting to predict the behaviour of the plant under a wide range of operating conditions. The objective is to improve the control of the process, avoiding poor treated discharges that break the environmental balance. This is why our goal is focused on the development of a prediction model, through the applicability of fuzzy modeling, which could contribute to a better management of the process. This method is concerned with extracting useful patterns or relationships between different variables by means of analysing the historical WWTP database provided by years of operation. This analysis will allow to build a fuzzy model that would help to understand dynamics of the system and could support in making decisions in WWTP management. The developed model characterizes the effluent quality as a function of the influent characteristics and control actions, by means of developing a model for each variable. The aim of this work (much of which is presented in (Belanche, 1998b) has been to find, as a first step, models able to characterize the time variation of outgoing WWTP variables using soft computing techniques; in particular, rough sets and time-delay neural networks of two kinds: fuzzy heterogeneous and classical.

The paper is organized as follows. Section 2 describes the problem at hand, the particular WWTP under study. Section 3 briefly introduces the reader to soft computing methods, while Section 4 reviews the concept of fuzzy heterogeneous neurons and their use in configuring hybrid neural networks, which will be then used to find input-output models of the plant. Rough set theory, used in the second part of the experiments, is outlined in section 5. The experiment setup itself and the obtained results are collectively presented in Section 6. Finally, Section 7 presents the conclusions.

2 A WWTP CASE STUDY

The database utilized to build the characterization model corresponds to a WWTP of a touristic resort situated in Costa Brava (Catalonia). This plant provides primary and secondary treatment using the activated sludge process to remove organic load and suspension solids contained in the raw water of about $30,000$ inhabitants-equivalents in winter and about $150,000$ in summer. An schematic of this WWTP is illustrated in figure 1.

The available historical data comprises a large amount of information corresponding to an exhaustive characterization of the plant. This information is also being used in other approaches to improve WWTP operation (Comas, 1998). It includes analytical results of water and sludge quality, together with on-line signals coming from sensors (wastewater, recycle, purge sludge and aeration flow rates, pH, temperature and dissolved oxygen concentration at the biological reactor).

The first work was focused on selecting an homogeneous amount of days, to cover a representative period of time. Then, it was necessary to select the most relevant variables of the process, corresponding to the analysis of water quality and flow-rates at different points of the plant. These variables are presented below, distinguishing between the on-line and the off-line values, and specifying the sample point (AB or influent, OP1 or primary settler effluent, and AT or effluent). Global

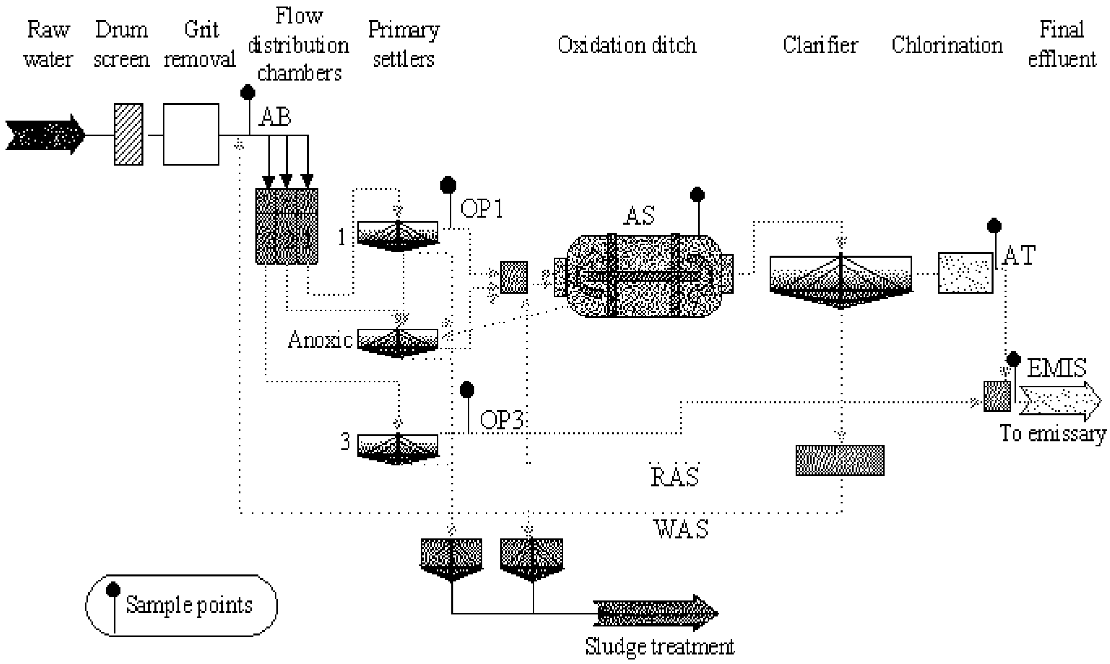

Fig. 1. Schematic of this WWTP that provides primary and secondary treatment.

process variables are related to the three control actions that the plant's manager can modify when removal efficiency decreases, in order to reconduct the process to normal performance: purge (Q-P), recycle (Q-R) and biological aeration (Q-A) flow rates.

Sample Point	On-line Data	Analytical Data
AB (influent)	Q-AB (influent flow)	COD, BOD, TSS
OP1 (primary treatment effluent)	Q-OP1 (primary effluent flow)	COD, BOD, TSS
AT (final treated effluent)	-	COD, BOD, TSS
Global process	Q-R (biological recycle flow) Q-P (biological waste flow) Q-A (biological aeration flow)	-

Table 1. Basic variables for characterizing the behaviour of the studied WWTP.

The final database studied in this paper covers a period of 609 consecutive days, considering each day as a new sample. Note that the period studied has an important amount of missing data due to the frequency of analysis, so the data utilized to build the model includes only those days with an actual value in the effluent (see table 1). Finally, to simplify the description of the influent characteristics, the set of internal variables (Q-OP1, COD-OP1, BOD-OP1 and TSS-OP1) has been excluded from the preliminary analysis.

The second work has comprised a statistical analysis of the studied database variables. Basic statistical descriptors are shown in table 2. In it, the extremely high incidence of missing values for most variables is the relevant feature. Specially in the case of target variables from the point of view of developing prediction models (COD-AT, BOD-AT and TSS-AT) and variables characterizing the physical-chemical state of incoming waters (COD-AB, BOD-AB and TSS-AB), the proportion of missing values is very severe (between 60-80%, approximately), i.e., there are much more missing data than actual information. Clearly, this situation makes considerably hard the search for models to characterize WWTP behaviour and must always be taken into account in evaluating the quality of the learned model.

Variable	Unit	Number of non-Missing	Number of Missing	Mean	StDev	Min	Max
Q-AB	m^3/d	591	18	10707	3634	0.0	23681
COD-AB	mg/l	229	380	795.8	198.0	150.0	1644.0
BOD-AB	mg/l	129	480	390.70	95.70	70.00	620.00
TSS-AB	mg/l	229	380	315.85	91.35	69.00	647.00
Q-R	m^3/d	608	1	5597.7	2287.1	0.0	12086.0
Q-P	Kg TSS/d	598	11	771.6	756.6	0.0	6523.0
Q-A	Kg O_2/d	548	61	4138.6	1878.4	0.0	8643.0
COD-AT	mg/l	229	380	55.80	18.52	20.00	134.00
BOD-AT	mg/l	129	480	8.959	4.876	2.300	32.000
TSS-AT	mg/l	233	376	9.562	5.750	2.000	42.000

Table 2. Basic statistical descriptors for selected WWTP variables.

The linear intercorrelation structure among variables is shown in figure 2 as an average clustering of the (absolute) correlation matrix of variables. With the exception of incoming water discharge (Q-AB), the actuation (Q-), output (-AT) and input (-AB) variables are clustered into three –not too homogeneous– groups. The fact that the highest intercorrelations are observed in output variables (0.736-0.764) indicates that once a reasonable model is found for one of them, similar ones should be also found for the rest.

The complexity of the WWTP behavior problem is reflected in the frequency distribution of their values. For example, Kolmogorov-Smirnov tests applied to the incoming TSS-AB and outgoing TSS-AT variables confirms what direct inspection suggests, in the sense that, whilst the first variable distributes normally, the second does not. Actually, it has a right-skewed distribution, reflecting strong non-linear distortions introduced by the WWTP dynamics (see figure 3).

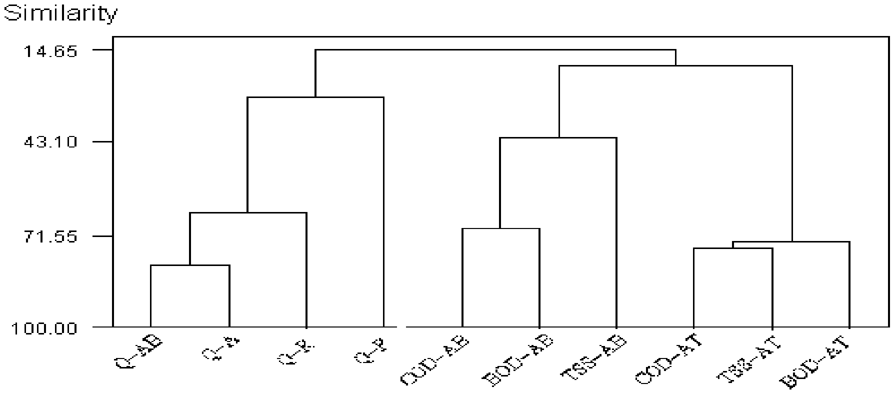

Fig. 2. Average clustering of the absolute correlation for the studied WWTP variables.

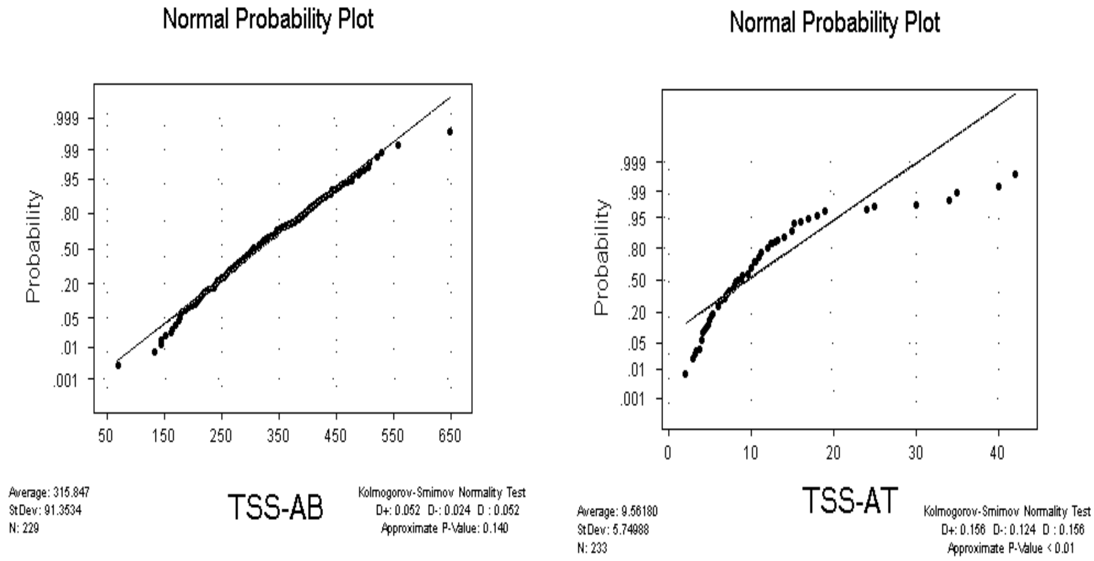

Fig. 3. Kolmogorov-Smirnov test for Total Suspended Solids (TSS). Left: incoming. Right: outgoing.

3 A SOFT COMPUTING APPROACH

Under the name of *soft computing* several theories, approaches and techniques are gathered together with a common purpose: to find solutions (usually in the form of models) to a wide variety of problems (such as pattern recognition, systems control, prediction, optimization and others) which share some characteristics: the nature of the problem is usually non-linear, and data is disturbed by noise, imprecision or uncertainty, and is often missing. Moreover, the sources of these data can be very heterogeneous, ranging from discrete to continuous variables, which can also be scalar, vectorial, etc, and include a spatial or temporal component. The most common theories and methods employed make use of fuzzy logic (Klir, 1988), genetic algorithms (Goldberg, 1989), artificial neural networks (Hertz, 1991), probabilistic reasoning (Pearl, 1987) and rough set theory (Pawlak, 1982; Pawlak, 1991).

Fuzzy Logic brings a formalism with its own syntax and semantics capable to express qualitative knowledge about the problem under study. Its excellence relies specially in the strength of its interpolative reasoning mechanism. *Genetic Algorithms* are general adaptive search methods based on the main ideas of Darwinian evolution. They maintain a population of individuals –each of which represents a possible solution to a given problem– that evolves from generation to generation through two main processes: a) selection of the fittest and b) application of genetic operators to recombine and somehow alter surviving individuals in the hope of finding better ones. These two mechanisms together form a powerful domain-independent search method. *Neural Networks* are structures capable of universal computation where knowledge and function are distributed among nodes or units each performing some simple (usually non-linear) computation. They can be given a training set, based on examples of some unknown input-output relation or system for which one is interested in finding a functional expression. *Probabilistic Reasoning* offers a means to evaluate the output of systems affected by randomness or other types of probabilistic uncertainty. In essence, it provides ways of updating the expected results in light of new acquired knowledge. Finally, in *Rough Set* theory structural set relationships in qualitative or imprecise data are explored from the point of view of their ability to approximate concepts –expressed as subsets of a universe of data objects– in terms of subsets of variables describing these objects. One of the main problems addressed by this technique is the discovery, representation and analysis of data regularities, aimed at discovering cause-effect relationships, identifying dependencies among group of variables and evaluating their relative importance.

Despite their obvious (and beneficial) differences, the common denominator of these approaches is that they leave behind non-flexible concepts such as binary logic, analytic models, rigid classifications and deterministic search. Ideally, the perfect system to be modeled or controlled can be described in a precise and complete way. In such cases, it is possible to use formal reasoning systems to associate boolean truth values to descriptions of the state or behaviour of this ideal physical system. However, when tackling a real-world problem, it turns out that they are mostly partly (and, sometimes, ill) defined, difficult to model –if one wishes to understand the nature of the process– and the solutions are immerse in huge search spaces. Now, precise models are impractical to use, costly, or simply non-existent. This makes soft computing approaches a flexible means to deal with such problems.

4 HETEROGENEOUS NEURAL NETWORKS

A fuzzy heterogeneous neuron is defined as a mapping $h : \hat{\mathcal{H}}^n \to \mathcal{R}_{out} \subseteq \mathbb{R}$, satisfying $h(\emptyset) = 0$ ($\emptyset$ is the empty set). Here $\mathbb{R}$ denotes the reals and $\hat{\mathcal{H}}^n$ is a cartesian product of an arbitrary number of *source sets*. Source sets may be families of extended reals $\hat{\mathcal{R}} = \mathbb{R} \cup \{\mathcal{X}\}$, extended fuzzy sets $\hat{\mathcal{F}}_i = \mathcal{F}_i \cup \{\mathcal{X}\}$, and extended finite sets of the form $\hat{\mathcal{O}}_i = \mathcal{O}_i \cup \{\mathcal{X}\}$, $\hat{\mathcal{M}}_i = \mathcal{M}_i \cup \{\mathcal{X}\}$, where each of the $\mathcal{O}_i$ has a full order relation, while the $\mathcal{M}_i$ have not. In all cases, the special symbol $\mathcal{X}$ denotes the unknown element (missing information) and it behaves as an incomparable element w.r.t. any ordering relation. According to this definition, neuron inputs are possibly empty arbitrary tuples, composed by n elements among which there might be reals, fuzzy sets, ordinals, nominals and missing data. Heterogeneous neurons are classified according to the nature of its image set (which does not have to be necessaryly restricted to a subset of the reals). In the present study, since the image set is given by $\mathcal{R}_{out}$ the model is of the *real kind*, which is easily coupled with other, classical neuron models (i.e. accepting only real inputs), thus leading to hybrid networks in a straightforward way. These networks have been used successfully in classification problems reported elsewhere (Valdés, 1997; Valdés, 1998; Belanche, 1998c), but their potential of application in other fields was not yet assessed experimentally. The purpose of this paper is to explore the performance of fuzzy heterogeneous networks (in hybrid architectures) for the identification of valid input-output models of a wastewater treatment plant.

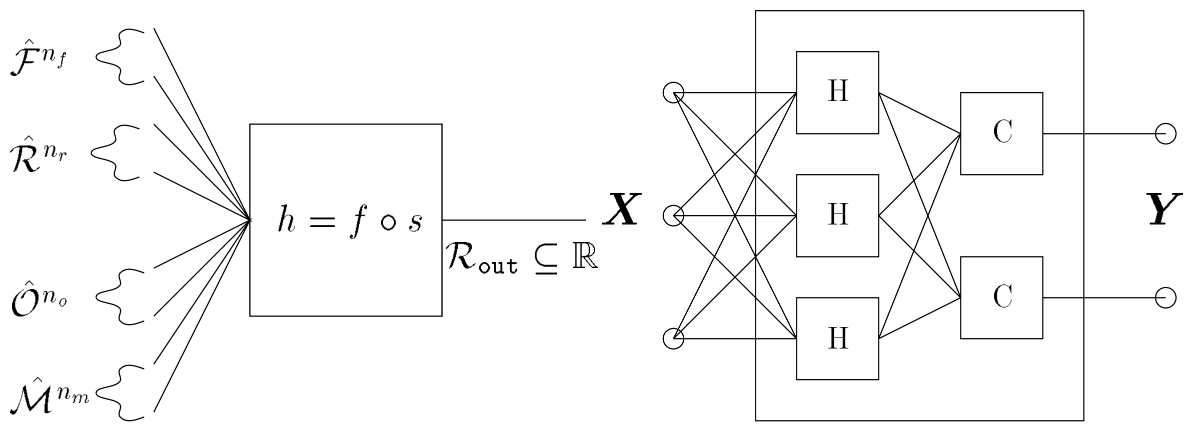

Fig. 4. Left: The fuzzy heterogeneous neuron model. Right: An example of a hybrid neural network composed by a hidden layer of heterogeneous neurons (H) and an output layer of classical neurons (C).

The use of the resulting heterogeneous neuron (shown in Fig. 4 (left)) to configure network architectures is thus straightforward, and a layered structure having a hidden layer composed of heterogeneous units and an output layer consisting of classical neurons is an immediate hybrid feed-forward choice (Fig. 4 (right)).

A particular class of heterogeneous networks (HNNs) is constructed by considering h as the composition of two mappings $h = f \circ s$, such that $s : \hat{\mathcal{H}}^n \to \mathcal{R}' \subseteq \mathbb{R}$ and $f : \mathcal{R}' \to \mathcal{R}_{out} \subseteq \mathbb{R}$. The mapping h can be considered as a n-ary function, parameterized by a n-ary tuple $\hat{\boldsymbol{w}} \in \hat{\mathcal{H}}^n$ representing neuron's weights, that is, $h(\hat{\boldsymbol{x}}, \hat{\boldsymbol{w}}) = f(s(\hat{\boldsymbol{x}}, \hat{\boldsymbol{w}}))$. In particular, the function s represents a *similarity* and f a squashing non-linear function with its image in $[0, 1]$. Accordingly, the neuron is sensitive to the degree of similarity between its inputs —composed in general by a mixture of continuous and discrete quantities possibly with missing data— and its weights. More precisely, s is understood as a *similarity index*, or proximity relation (transitivity considerations are put aside). That is, a binary, reflexive and symmetric function $s(x, y)$ with image on $[0, 1]$ such that $s(x, x) = 1$ (strong reflexivity).

The concrete instance of the model under study in the present paper uses as aggregation function a *Gower-like* similarity index in which the computation for heterogeneous entities is constructed as a weighted combination of partial similarities over subsets of variables. This coefficient has its values in the real interval $[0, 1]$ and for any two objects i, j given by tuples of cardinality n, is given by the expression

$$s_{ij} = \frac{\sum_{k=1}^{n} g_{ijk}\, \delta_{ijk}}{\sum_{k=1}^{n} \delta_{ijk}}$$

where:

- g_{ijk} is a similarity *score* for objects i, j according to their value for variable k. These scores are in the interval $[0, 1]$ and are computed according to different schemes for numeric and qualitative variables.
- The factor δ_{ijk} is a binary function expressing whether objects i, j are comparable or not according to their values w.r.t. variable k. It is 1 if and only if both objects have values different from $\mathcal{X}$ for variable k, and 0 otherwise.

Gower's original definitions for real-valued and discrete variables are kept (see (Gower, 1971) for details), but other similarity functions are possible. For variables representing fuzzy sets, similarity relations from the point of view of fuzzy theory have been defined elsewhere (Dubois, 1997) and different choices are possible. In our case, if $\mathcal{F}_i$ is an arbitrary family of fuzzy sets from the source set, and $\tilde{A}$, $\tilde{B}$ are two fuzzy sets such that $\tilde{A}$, $\tilde{B} \in \mathcal{F}_i$, the following similarity relation is used

$$g(\tilde{A}, \tilde{B}) = \sup_x \, (\mu_{\tilde{A} \cap \tilde{B}}(x))$$

where

$$\mu_{\tilde{A}\cap\tilde{B}}(x) = min(\mu_{\tilde{A}}(x), \, \mu_{\tilde{B}}(x))$$

For the activation function, a modified version of the classical logistic is used, which is an automorphism of the real interval $[0, 1]$.

$$f(x,p) = \begin{cases} \frac{-p}{(x-0.5)-a(p)} - a(p) & \text{if } x \leq 0.5 \\ \frac{-p}{(x-0.5)+a(p)} + a(p) + 1 & \text{otherwise} \end{cases}$$

where $a(p)$ is an auxiliary function given by $a(p) = \frac{-0.5+\sqrt{0.5^2+4*p}}{2}$ and p is a real-valued parameter controlling the curvature, set in the experiments to 0.1. The general training procedure for the HNN is based on genetic algorithms, since the heterogeneity of the variables involved and the non-differentiability of the similarity function prevent the use of gradient-based techniques.

5 ROUGH SETS

From the methodological point of view, lowering the precision with which a given system is observed usually makes easier the characterization of that system's interdependencies and makes data regularities more visible. This is the case, for example, of processes described by continuous variables which undergo a discretization process by converting these variables into categorical ones, introducing qualitative ranges. Obviously, this implies a loss of precision in the representation of objects, and a reduction in the ability to discern or differentiate among distinct objects. However, with a suitable compromise, this loss is compensated with the increase in the ability to reveal interesting data regularities.

Let U be a finite set of objects called the universe. Any subset $X \subseteq U$ is called a concept or category in U and the main interest is when concepts form a partition (classification) of the universe. In general, one deals with families of partitions over U which are called a *knowledge base* over U. A knowledge base is a relational system $K = (U, \mathbf{R})$ where $U \neq \emptyset$ is the universe (finite) and $\mathbf{R}$ is a family of equivalence relations over U. This is indeed the case, for example, in a collection of observed objects described by a set of categorical variables, where each defines an equivalence relation on U. Given $\mathbf{P} \subset \mathbf{R}$ and $\mathbf{P} \neq \emptyset$, the intersection of all these equivalence relations is also an equivalence relation denoted by $IND(\mathbf{P})$ and called an *indiscernibility relation* over $\mathbf{P}$. It expresses the knowledge associated with the family of equivalence relations $\mathbf{P}$, also called $\mathbf{P}$-*basic knowledge* about U in $\mathbf{K}$. If $X \subseteq U$, X is said to be *R-definible* if X is the union of some R-basic knowledge; otherwise X is called R-undefinible. The R-definible sets are those subsets of U which can be exactly defined in terms of the knowledge base K whereas the R-undefinible sets are called *rough sets*. This leads to the idea of approximation of a set by other sets.

With each subset $X \subseteq U$ and an equivalence relation $R \in IND(K)$ two subsets can be associated, called the *lower* and *upper* approximation, respectively, as follows:

$$R_L = \{\bigcup Y : Y \in U/R : Y \subseteq X\}$$
$$R_U = \{\bigcup Y : Y \in U/R : Y \cap X \neq \emptyset\}$$

where U/R is the equivalence class (partition) induced by R. The lower approximation (also called the *positive region* $POS_R(X)$) is the set of elements which can be certainly classified as elements of X whereas the upper approximation is the set of elements which can be possibly classified as elements of X.

An important issue in the analysis of dependencies among variables is the identification of information-preserving reduction of redundant variables. In particular, to find a minimal subset of interacting variables having the same discriminatory power as the original ones, which would lead to the elimination of irrelevant or noisy variables without any loss of essential information. A set of variables P is independent w.r.t. the set of variables Q if for every proper subset R of P, $POS_P(Q) \neq POS_R(Q)$; otherwise P is said to be dependent w.r.t Q. Moreover, the set of variables R is a minimal subset or *reduct* of P, if R is an independent subset of P w.r.t. Q, such that

$$POS_R(Q) = POS_P(Q)$$

A variable $a \in P$ is superfluous if $POS_P(Q) = POS_{P-\{a\}}(Q)$; otherwise a is said to be *indispensable* in P. The set of all indispensable relations is the *core*. An important property of the core is that it is equal to the intersection of all reducts.

6 EXPERIMENT SETUP

If some fixed-length segment of the most recent input values is considered enough to perform the task successfully, then a temporal sequence can be turned into a set of spatial patterns on the input layer of a multi-layer feedforward net trained with an appropriate algorithm such as backpropagation. These architectures are called *Time-delay neural networks* (TDNNs), since several values from an external signal are presented simultaneously at the network input using a moving window (shift register or tapped delay line) (Hertz, 1991). A main advantage of TDNNs in front of recurrent neural networks is their lower cost of training, which is very important in the case of long training sequences. TDNNs have been applied extensively in recent years to different tasks, in particular to prediction and system modeling (Lapedes, 1987).

In the present study, two different TDNN approaches that differ in the training method have been tested: a hybrid procedure composed by repeated cycles of simulated annealing coupled with conjugate gradient algorithm (TDNN-AC) (Ackley, 1987) and the HNN model presented. In the former case the hidden layer uses the hyperbolic tangent as neuron output function whereas the output layer was composed by a linear neuron. It should be noted that the HNN model as used here (TDNN-HG) is viewed as a TDNN that incorporates heterogeneous neurons and is trained by means of genetic algorithms. The TDNN-HG and the TDNN-AC architectures were fixed to include 1 output unit, 8 hidden units, and 13 input units, corresponding to the model $y(t+1) = F < x(t), x(t-1), x(t-2), y(t-1) >$, where $x(t)$ denotes the current value of the input variable and $y(t)$ denotes the value of the output. Selected inputs were Q-AB, Q-A, Q-P and Q-R, that is, the incoming flow rate and the three actuation variables.

In the testing process, the normalized mean square error (in percentage) between the predicted output value, $\hat{y}(t)$, and the controller output, $y(t)$, is used to determine the quality of each of the inferred models. This error is given by $MSE = \frac{E[(y(t)-\hat{y}(t))^2]}{s_y^2} \cdot 100\%$ where s_y^2 denotes the variance of $y(t)$. For each studied output variable, the TDNN-HG was trained using a standard genetic algorithm with the following characteristics: binary-coded values, probability of crossover: 0.6, probability of mutation: 0.01, number of individuals: 150, linear rank scaling with factor: 1.5, selection mechanism: stochastic universal, replace procedure: worst. The algorithm stopped when no improvement was found for the last $1,000$ generations (typical values were about $7,000$). The TDNN-AC was trained in only one run and the process was stopped when a reasonable error was attained. In both cases, the training set chosen was the first half of available data (about 300 days).

6.1 Results of the experiment

The WWTP characterization produced via neural networks trained with a hybrid simulated annealing-conjugate gradient procedure was worse than the corresponding obtained by using a fuzzy heterogeneous neural network model, as illustrated by normalized squared errors shown in table 3 for BOD-AT and COD-AT output variables. In both cases the same neuron architecture was used but the errors obtained are appreciably lower for the heterogeneous model w.r.t the classical neural one, although it uses a very sophisticated and robust training procedure.

	Classical Neural Model TDNN-AC	Fuzzy Heterogeneous Model TDNN-HG
BOD-AT	45.55%	20.74%
COD-AT	30.76%	11.64%

Table 3. Normalized MSE errors of the two neural network models used for characterizing some WWTP outgoing variables.

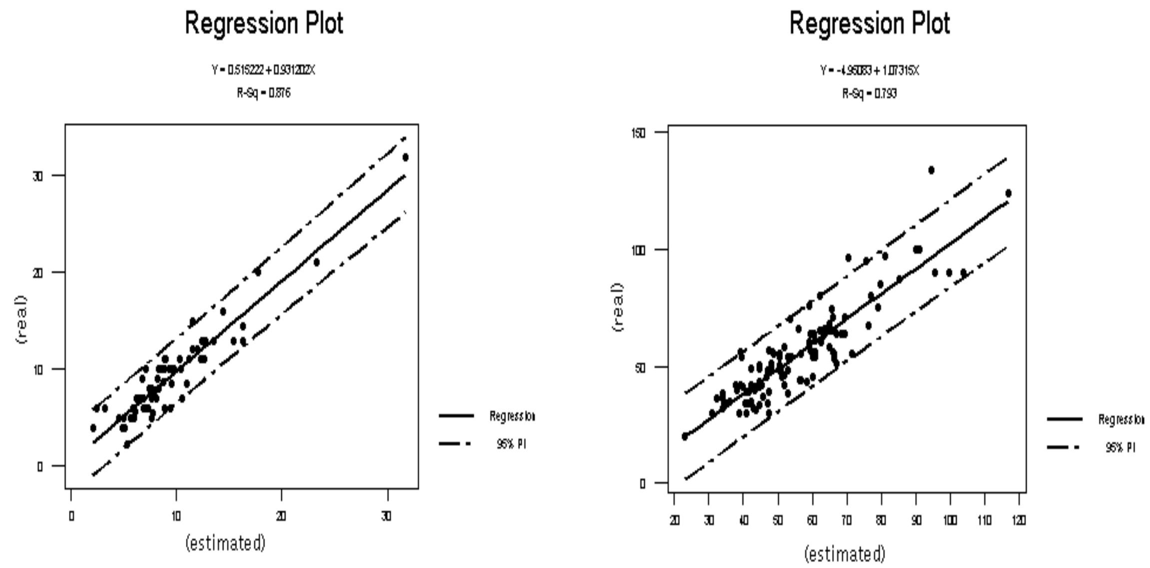

Fig. 5. Relation between estimated vs. real BOD-AT (left) and estimated vs. real COD-AT (right).

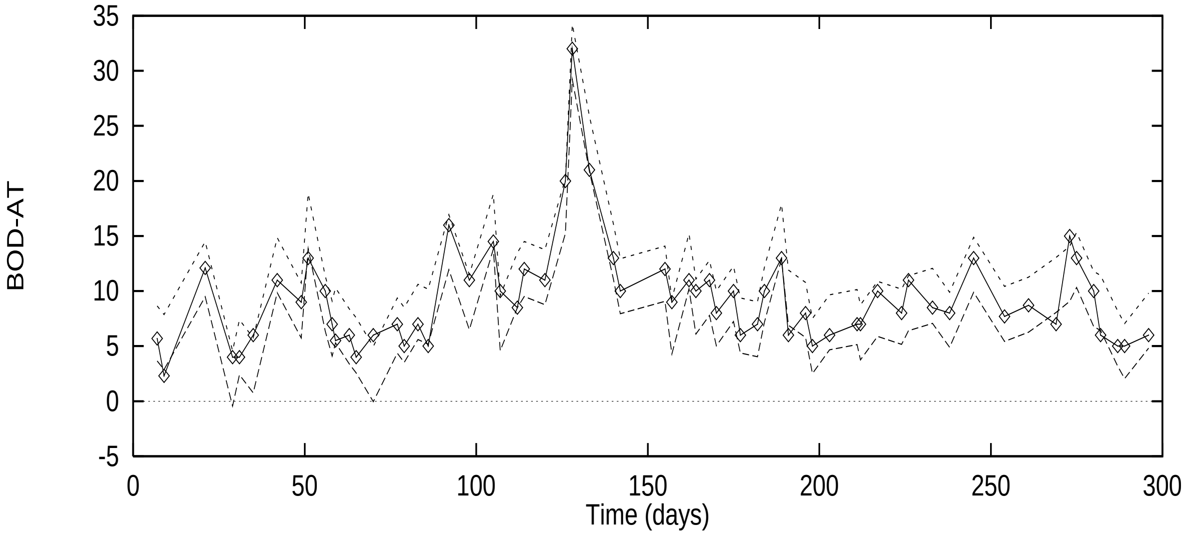

Fig. 6. Time behaviour of BOD-AT during the first 300 days (solid line) with observed points. Upper and lower dashed lines indicate the 95% confidence estimation interval (according to the TDNN-HG model).

The relation between the BOD-AT output variable as estimated by the heterogeneous neural network and the corresponding observed values is shown in fig 5 (left). There is a significant linear correlation between them and model adequacy is revealed by the fact that almost all points are enclosed by the 95% confidence band. The corresponding time behavior is illustrated in fig 6, where the observed BOD-AT values are displayed together with the 95% confidence band given by the neural network model (upper and lower dashed curves). In spite of the fact that

78.7% of the data, corresponding to the 300 day period chosen for the characterization were missing, almost all observed values are within the confidence band with only very slight exceptions. A similar behavior is exhibited by the COD-AT variable (figs. 5 (right) and 7). The fact that the HNN model outperforms the classical one has been observed in other application contexts (Valdés, 1997; Belanche, 1998c; Belanche, 1998a) and therefore deserves further attention as it seems to indicate a more general property of these recently introduced hybrid models.

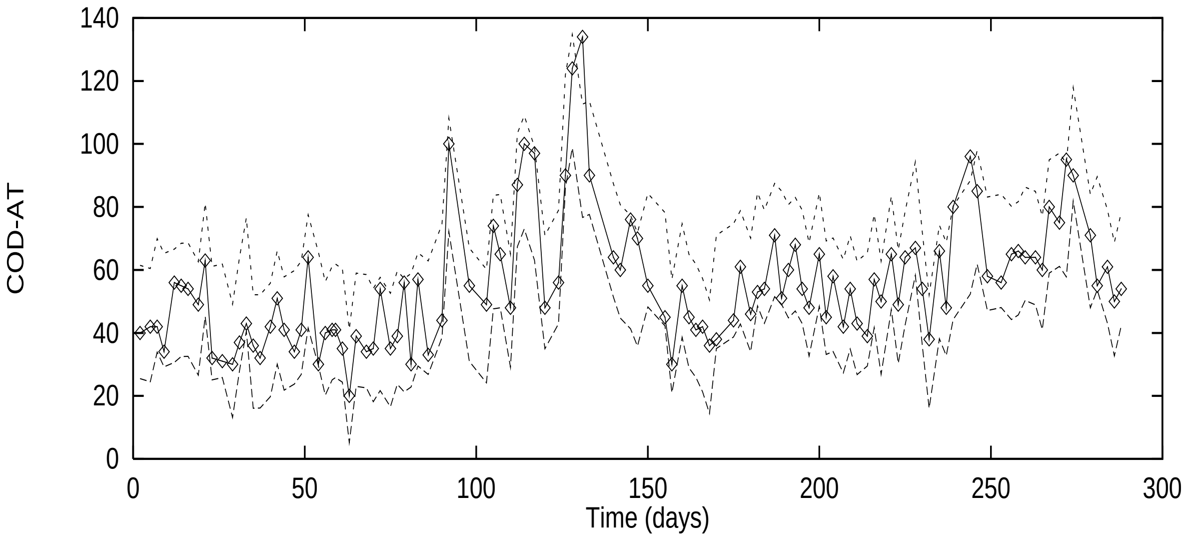

Fig. 7. Time behaviour of COD-AT during the first 300 days (solid line) with observed points. Upper and lower dashed lines indicate the 95% confidence estimation interval (according to the TDNN-HG model).

6.2 Predicting WWTP behaviour

The previous experiment showed how a model characterizing WWTP behaviour can be found. The next step to be taken is to develop a model able to predict future WWTP output in situations never seen before (that is, not used in the formation of the model) again in light of available past values of its variables. This is a very difficult issue, again complicated by the presence of missing data in a set of characterizing variables that is already very heterogeneous in nature and plays non-linear interaction roles in the overall process. A further complication that arises is the different time scales of the variables. While some of them are available almost at will, others may take days (as, for example, the BOD-AT, that takes 5 days). For this reason, the COD-AT variable has been the one chosen as a prediction target. This variable is available in about 2 hours and, thus, previous values w.r.t. the present day are always known. To this end, a still more careful selection and a clever treatment of the data to be used as training is a means to pave the way. Specifically, three treatments have been performed:

– First, the correlation structure reflected in figure 2 shows that variables (Q-AB, Q-A), (COD-AB, BOD-AB) and (COD-AT, TSS-AT, BOD-AT) are reasonably similar. This suggests the use of Q-AB, Q-R, Q-P, COD-AB and TSS-AB as input variables when considering the construction of prediction models. The choice of COD-AB is favored by the fact that it is a much simpler and faster analytical procedure than BOD-AB from the chemical point of view.
– Second, we observed that part of the errors of the models inferred in the previous experiments were due to the high peaks present in both studied variables (BOD-AT and COD-AT). For this reason, COD-AT was log_{10} transformed.
– Third, the delays used in such models were proposed intuitively, but without any regard to actual underlying significance. This is where rough set theory comes to play.

No doubt that one of the most important tasks when finding useful dependencies from the point of view of constructing prediction models for WWTP behavior is the discovery of those time delays in the input variables and in the predicted variable itself which carries essential functional relationship. In the present study an experiment was made by forming a data matrix containing the information concerning the behavior for each day of the last 10 days for variables Q-AB, Q-R, Q-P, COD-AB, TSS-AB and the target variable COD-AT itself. This makes a total of 60 new variables potentially related with the value of the COD-AT for each day, with a dependency coefficient found to be 0.9699 (a value of 1 means that the selected variables convey *all* the information present in the whole data available).

The continuous process represented by these data was transformed into a discrete one by analysing the empirical probability distribution of all variables involved and defining suitable categories introducing corresponding cut-point values. In particular, the following were set: Q-AB (8500, 13000, 16500), COD-AB (650, 950), TSS-AB (250, 400), Q-R (5000, 7000), Q-P(1000) and log_{10}(COD-AT) (1.65, 1.85).

The core and reducts were obtained for the discrete process obtained via categorization of the original data and it was found that, from the original 60 potential predictor variables, only 13 were really indispensable, whereas adding another 7 makes them an optimal reduct. That is to say, from the point of view of relative size of the positive region defined by these 20 variables and w.r.t. the positive region defined by the whole set. The core itself was composed of the following variables: Q-AB (delay 1), TSS-AB (delay 7), Q-R (delays 1, 2, 3, 4, 5, 7, 9, 10) and log_{10} (COD-AT) (delays 5, 7). The optimal reduct is completed by variables Q-AB (delays 5, 10), COD-AB (delays 2, 4), Q-P (delay 5) and log_{10}(COD-AT) (delays 2, 3). It is interesting to observe that almost all information coming from the recirculation flow was considered essential (a variable controlled by the WWTP human operator). The information given by the optimal reduct was used to set up a prediction model based on a fuzzy heterogeneous neural network (but using the original continuous variables) with log_{10}(COD-AT) with time delay 0 as target.

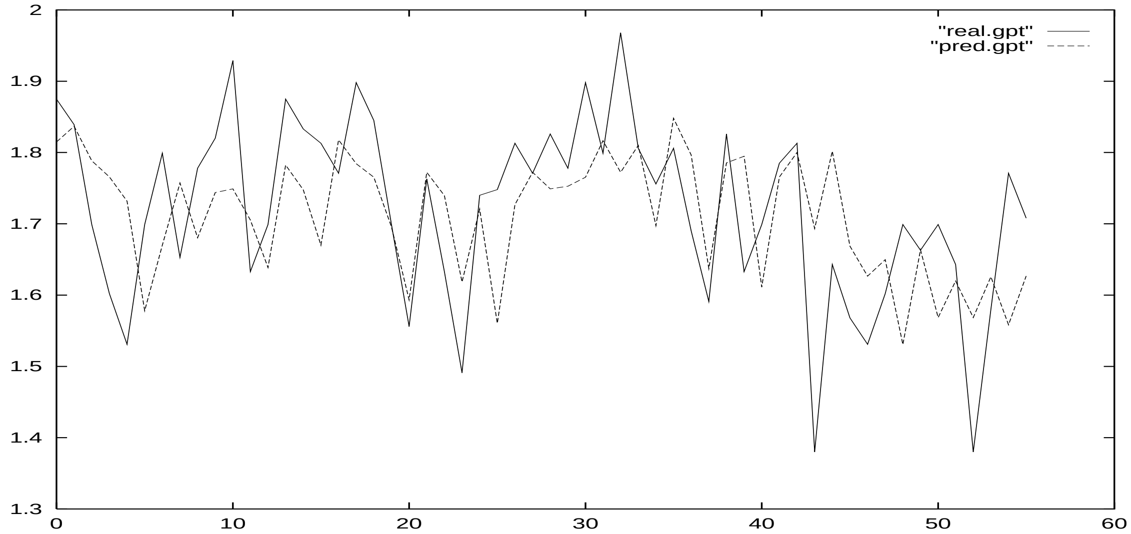

Fig. 8. Actual time behaviour of COD-AT during the last 56 days (solid line) with observed points against prediction according to the TDNN-HG model. The normalized MSE for prediction is 60.0%.

A very simple HNN architecture consisting of just 2 neurons was utilized (20 inputs, 1 output), with the same training set used for the previous experiments (50% of the total available). The last 25% (56 days) as data to be predicted. Also, very small GA settings were used (26 individuals, 500 generations) to avoid excessive data overfitting.

The behavior of the predicted COD-AT values w.r.t. the real observed ones is shown in Fig. 8. In spite of the fact that the fit is not as accurate as before, the relation between the two is highly significant, both the linear correlation coefficient and the linear regression points of view, as tested with the corresponding *t-test* for the correlation coefficient and the *F-test* for the analysis of variance (for 95% confidence in both cases). Actual numbers for the *t-test* are: $R = 0.504$ with 54 degrees of freedom ($t = 4.288$). The result for the *F-test* is 18.39 for one degree of freedom in the numerator and 54 in the denominator (Fig. 9). All this shows that the model, although far from perfect, does capture prediction information and is able to prognose outputs within a 95% confidence band. This is also particularly important having into account the WWTP complexity and the big quantity of missing information spread in the available data set.

7 CONCLUSIONS

For the WWTP under study, three main aspects have been found that deeply characterize the processes that are taking place. First, with the exception of incoming water discharge, actuation, outgoing and incoming variables are clearly distinguished from one another, reflecting an internal structure that must be taken into account

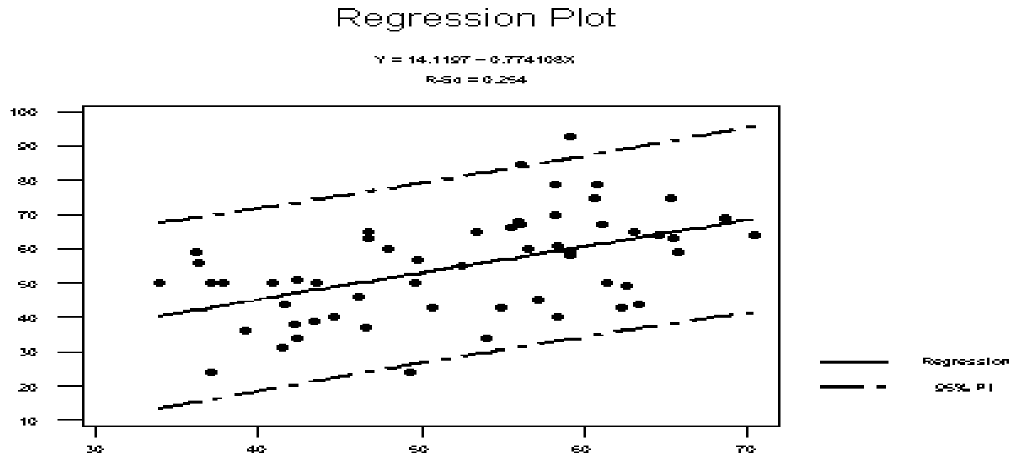

Fig. 9. Relation between predicted vs. real COD-AT (solid line). Upper and lower dashed lines indicate the 95% confidence estimation interval (according to the TDNN-HG model).

during the search for accurate models of the process. Second, the process dynamics introduces strong non-linear distortion between incoming and outgoing variables. Third, these outgoing variables are significantly related and, therefore, could be described by similar models. Soft computing techniques –in particular, fuzzy heterogeneous neural networks– have shown to be capable to describe and predict the behaviour of some of these processes in a statistically significant sense, despite the imprecision associated to raw real-world information and the high degree of incompleteness and fragmentation, due to the number of missing values and their time distribution in many small chunks. This, together with the fact that the TDNN-HG model outperformed the classical TDNN-AC, suggests that it fits better the especial requirements posed by the WWTP problematic. Further experiments tackling the difficult prediction task are a significant part of the future work.

Acknowledgements:

The authors wish to thank the *Consorci de la Costa Brava* for the data and information provided. This work has been supported by CICYT Projects AMB-97/889 and TIC96-0878.

References

Ackley, D. (1987) A connectionist machine for genetic hillclimbing. Kluwer Acad. Press.

Belanche, Ll., Valdés, J.J. Alquézar, R. (1998a) Fuzzy Heterogeneous Neural Networks for Signal Forecasting. In Procs. of *ICANN'98, Intl. Congress on Natural and Artificial Neural Networks*. Skövde, Sweden.

Belanche, Ll., Valdés, J.J., Comas, J., Roda, I., Poch, M. (1998b) Modeling the Input-Output Behaviour of Wastewater Treatment Plants using Soft Computing Techniques. In Procs. of *BESAI'98. Binding Environmental Sciences and AI.* Workshop held as part of *ECAI'98: European Conference on Artificial Intelligence.* Brighton, UK.

Belanche, Ll., Valdés, J.J., (1998c) Using Fuzzy Heterogeneous Neural Networks to Learn a Model of the Central Nervous System Control. In Procs. of *EUFIT'98, 6th European Congress on Intelligent Techniques and Soft Computing.* Aachen, Germany.

Comas J., R.-Roda I., Poch M., Sánchez M., Gimeno J.M., Cortés, U. (1998) An integrated intelligent system to improve wastewater treatment plant operation - part II: a real application. Waste-Decision 98, International Workshop on Decision and Control on Wastes Bio-Processing, Narbonne, France.

Dochain D. (1991) Design of adaptative controllers for non-linear stirred tank bioreactors: extension to the MIMO situation, *Journal of Process Control*, 1, pp 41-48.

Dubois D., Prade H., Esteva F., García P., Godo L., López de Mántaras, R., , (1997) Fuzzy set modeling in case-based reasoning. Intl. Journal of Intelligent Systems (to appear).

Government (1996) Junta de Sanejament. Pla de Sanejament (Autonomous Government of Catalonia official document). Generalitat de Catalunya, Departament de Medi Ambient. Barcelona.

Goldberg, D.E. (1989) Genetic Algorithms for Search, Optimization & Machine Learning. Addison-Wesley.

Gower, J.C., (1971) A general coefficient of similarity and some of its properties. Biometrics, 27, pp. 857-871.

Heinzle E., Dunn I., Ryhiner G. (1993) Modelling and control for anaerobic wastewater treatment. Adv. in Biochemical Engineering, 48, pp. 79-114.

Hertz, J., Krogh A., Palmer R.G., (1991) Introduction to the Theory of Neural Computation. Addison-Wesley, Redwood City.

Isaacs S.H., Soeberg H., Kummel M. (1992) Monitoring and control of a biological nutrient removal process: Rate data as a source of information. IFAC Modelling and Control of Biotechnological Processes, Colorado, USA.

Klir G.J., Folger T.A. (1988) Fuzzy Sets, Uncertainty and Information. Prentice Hall Int. Editions.

Lapedes A., Farber, R. (1987) Nonlinear signal processing using neural networks: prediction and system modelling, Tech. Rep. LA-UR-87-2662, Los Alamos National Laboratory, Los Alamos, NM.

Moreno R., de Prada C., Lafuente J., M. Poch M., Montague G. (1992) Non-linear predictive control of dissolved oxygen in the activated sludge process. *ICCAFT 5 / IFAC-BIO 2 Conference*, Keystone, pp 289-293.

Nejjari F., Benhammou A., Dahhou B. and Roux G. (1997) Nonlinear multivariable control of a biological wastewater treatment process, ECC 97, Brussels, Belgium.

Pawlak, Z. (1982) Rough Sets. Intl. Journal of Computer and Information Sciences 11(5): 341-356.

Pawlak, Z. (1991) Rough Sets: Theoretical aspects of reasoning about data. Kluwer Academic Publ. 229 pp.

Pearl, J. (1987) Probabilistic reasoning in intelligent systems: networks of plausible inference. Morgan-Kaufmann Inc.

Queinnec, I. (1998) Robust control of a denitrification biofilter. Proc. of Waste Decision 98, International Workshop on Decision and Control in Waste Bio-Processing, Narbonne, France.

Valdés, J.J., García, R., (1997) A model for heterogeneous neurons and its use in configuring neural networks for classification problems, *Procs. of IWANN'97, Intl. World Conf. on Artificial and Natural Neural Networks.* Lecture Notes in Computer Science 1240, Springer-Verlag, pp. 237-246.

Valdés J.J, Belanche, Ll. (1998) Fuzzy Heterogeneous neurons for Imprecise Classification Problems. LSI Research Report LSI-98-33-R. Universitat Politècnica de Catalunya. Dept. of Languages and Information Systems.

WEF (1992) Standard Methods for the Examination of Water and Wastewater, 16th Edition, Water Environment Federation, Washington APHA.

WEF (1996) Operation of Municipal Wastewater Treatment Plants. Manual of Practice No. 11, 5th Edition, Water Environmental Federation, Alexandria.