# Neural Networks complemented with Genetic Algorithms and Fuzzy Systems for Predicting Nitrogenous Effluent Variables in Wastewater Treatment Plants

NARCIS CLARA
Universitat de Girona
Departament d'Informàtica i Matemàtica Aplicada
Campus de Montilivi s/n, 17071 Girona
SPAIN
narcis.clara@udg.edu

*Abstract:* This work focuses on the prediction of the two main nitrogenous variables that describe the water quality at the effluent of a Wastewater Treatment Plant. We have developed two kind of Neural Networks architectures based on considering only one output or, in the other hand, the usual five effluent variables that define the water quality: suspended solids, biochemical organic matter, chemical organic matter, total nitrogen and total Kjedhal nitrogen. Two learning techniques based on a classical adaptative gradient and a Kalman filter have been implemented. In order to try to improve generalization and performance we have selected variables by means genetic algorithms and fuzzy systems. The training, testing and validation sets show that the final networks are able to learn enough well the simulated available data specially for the total nitrogen.

*Key–Words:* Environmental modeling, Wastewater treatment plant, Total nitrogen, Total kjedhal nitrogen, Neural networks, Genetic algorithms, Fuzzy systems

## 1 Introduction

Neural networks and, globally, all the techniques belonging to Soft Computing, have become a successful tool to deal with a great set of industrial and environmental applications, like modeling, diagnose, object recognition or control ([3, 15, 17, 20]). This paper is focused in a highly-complex environmental application, represented as a numerical input and output vector (as chemical concentrations) whose mappings are considered differentiable. A neural network with feed-forward back-propagation architecture, which approximates the outputs variables with any predetermined error, can always be built for this kind of functions ([19]). Therefore, the set of patterns for testing become fundamental to analyze the degree of confidence of the network. However, the real conditions can suddenly vary, making very complex the whole process.

General studies on this fields have been carried out by other authors ([1, 5, 6, 8, 21, 23]), studying in depth about the possibility of giving some qualitative indicators in order to design an early warning system alerting on the incoming of critical working phenomena such as heavy period of rain ([16]) or using real data with missing and fuzzy information ([2]).

A mathematical model is suitable for model the data incoming in a Wastewater Treatment Plant (WWTP) due to the great complexity and variability of the influent composition, the lack of on-line sensors and signals, the delay of some analytical results, and, specially, the great number of missing values. The set of data used for this study were obtained by simulation of the processes involved the most extensively system used for wastewater treatment. This system involves two main components: biochemical reactors, where biological reactions take place to remove pollutants, and settling units to clarify the treated wastewater. The Activated Sludge Model ♯1 ([12]) is a mathematical model having the capability of realistically modeling the performance of simple-sludge systems carrying out carbon oxidation, nitrification and denitrification. This model is probably still the most widely used for describing wastewater treatment processes over the world and it can still considered a 'state-of-the-art' model. The double-exponential settling velocity function is chosen as a fair representation of the settling process ([22]). Three key operational parameters -controlled by the plant's operators- have been considered: aeration energy, pumping energy for activated sludge recycling and sludge purge. An schematic WWTP design is described in Figure 1.

The simulated database is formed by records cho-

sen each 15 minutes during two weeks thus 1344 records. A period of rain (or storm) makes more complex and realistic the whole set of data. The developed model characterizes the effluent quality as a function of the influent variables and control actions, by means of developing a model for each nitrogenous variable. Thirteen are the number of inputs; ten state variables in the affluent: $t$ (time), quickly biodegradable substrate ($S_S$), heterotrophic biomass ($X_{BH}$), Slowly biodegradable substrate ($X_S$), non-biodegradable particulate organic matter ($X_I$), nitrogen NH3+NH4 ($S_{NH}$), non-biodegradable soluble organic matter ($S_I$), soluble biodegradable organic nitrogen ($S_{ND}$), particulate biodegradable organic nitrogen ($X_{ND}$), inflow rate (Q); and three operational parameters: aeration energy (AirE), pumping energy (PumpE), and sludge purge (WAS). Two main state variables describe the water quality in the effluent with respect the nitrogen: total Kjedhal nitrogen (TKN) and total nitrogen (Ntotal). These nitrogenous variables are defined by

$$\text{TKN} = S_{NH} + S_{ND} + X_{ND} + 0.08(X_{B,H} + X_{B,A}) + \\ + 0.06(X_P + X_I)$$

$$\text{Ntotal} = \text{TKN} + \text{SON}$$

Based on measurements of Total Kjeldahl Nitrogen, the nitrogen is divided into free and saline ammonia ($S_{NH}$), organically bound nitrogen and active mass nitrogen, that is, a fraction of the biomass which is assumed to be nitrogen. The organically bound nitrogen is divided into soluble and particulate fractions, which in turn maybe biodegradable or non-biodegradable. It should be noted that only particulate biodegradable organic nitrogen ($X_{ND}$) and soluble biodegradable organic nitrogen ($S_{ND}$) are explicitly included in the model. The active mass nitrogen is included in the model only in the sense that decay of biomass will lead to a production of particulate biodegradable organic nitrogen. Variables $X_{B,A}$ and $X_P$ represent atrophic biomass and particulate products resulting of the biomass death respectively. No inert soluble nitrogen is modeled. Finally, $NO_2$ and $NO_3$ are put into one variable (SON), as a way to simplifying the model. All the units are measured on mg/l except for AirE (Kwh/d), PumpE (Kwh/d), WAS (Kg/d) and Q ($m^3$/day). The main statistical parameters for inputs and outputs are showed in Table 1. and Table 2.

For each nitrogenous variable we have built two kind of neural networks architectures. The first one (NN1) with only one output, obviously, the considered nitrogenous variable, namely, total nitrogen or total Kjeldahl nitrogen. The second one (NN2) with five outputs: total nitrogen, total Kjeldahl nitrogen, suspended solids (TSS),and the two main oxygen state variables that describe the water quality at the effluent: biochemical organic matter (BOD) and chemical organic matter (COD). These new three variables are defined by

$$\text{TSS} = 0.75(X_S + X_{BH} + X_{BA} + X_P + X_I)$$

$$\text{BOD} = 0.25(S_S + X_S + 0.92(X_{BH} + X_{BA}))$$

$$\text{COD} = S_S + S_I + X_S + X_{BH} + X_{BA} + X_P + X_I$$

Table 1: Basic statistical descriptors for selected input variables

| | Min | Max | Mean | StDev |
|---|---|---|---|---|
| $S_S$ | 13.828 | 120.011 | 60.533 | 21.178 |
| $X_{B,H}$ | 4.927 | 42.745 | 24.446 | 8.573 |
| $X_S$ | 38.900 | 293.814 | 177.965 | 57.481 |
| $X_I$ | 5.448 | 109.831 | 42.049 | 21.610 |
| $S_{NH}$ | 7.033 | 50.000 | 27.945 | 8.632 |
| $S_I$ | 10 | 30 | 27.76 | 5.51 |
| $S_{ND}$ | 1.383 | 12.001 | 6.053 | 2.118 |
| $X_{ND}$ | 1.853 | 16.071 | 9.190 | 3.223 |
| Q | 10000.0 | 52126.4 | 21329.8 | 8997.45 |

Table 2: Basic statistical descriptors for operational parameters and nitrogenous variables

| | Min | Max | Mean | StDev |
|---|---|---|---|---|
| AirE | 6020.8 | 8449.2 | 7205.2 | 708.08 |
| PumpE | 753.4 | 3847.1 | 1709.3 | 701.7 |
| WAS | 1930.5 | 3251.2 | 2491.8 | 304.45 |
| Ntotal | 8.14 | 21.3 | 16.2 | 2.245 |
| TKN | 1.88 | 12.7 | 4.79 | 2.379 |

Different studies have been published in this subject based on some techniques of soft computing ([2, 9, 10]). We will follow this general strategy but implementing other procedures intending to improve the adaptability to the theoretical model and to do your best for reducing errors.

## 2 Methodology

The purpose of developing a neural model ([3],[13],[19]) is to produce a formula that captures essential relationships between inputs and outputs ($y_i$). Once developed, this formula is used to

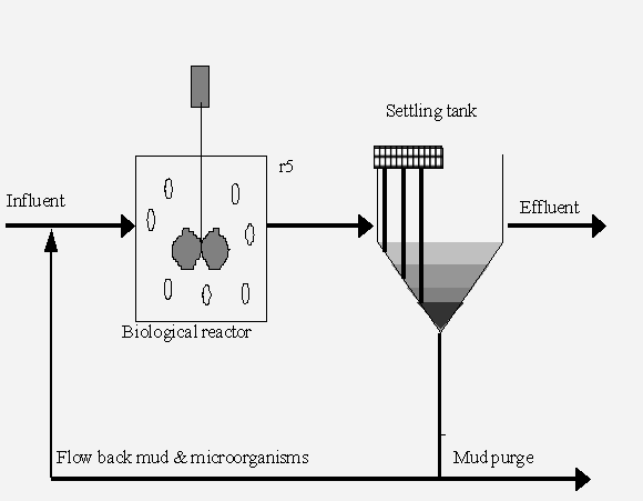


Figure 1: Schematic WWTP design

interpolate from a new set of inputs to corresponding predicted outputs ($\hat{y}_i$). In neural nets this process is called generalization. Several studies based on this approach have showed the capacity of learning knowledge when have been applied to Wastewater Treatment Plants ([4, 16]).

The neural network is constructed incrementally by adding hidden nodes, usually just one or two at a time. Each hidden node or pair of hidden nodes has its weights trained from several initializations. The best initialization is established in the network, and the all the weights to the output nodes of the net are retrained.

## 2.1 Partition of the data set

The train set is a subset of the input data which is used to train (updated the weights of) the neural network. During training, the score on the test set determines when to stop building the network and it is also used to choose between hidden node candidates. Moreover, the test set is used as part of the model building process to prevent overfitting (the opposite of generalization). Although the explicit training of the model uses only the training set, heuristics involving the performance on the test set are used to guide choices during the construction of the model thus, it is common to use the validation data set -a subset of the input data that is different from de data used to build the network- to finally estimate model performance in a deployed environment. The last 20 records are used to prove well generalization out of the train-test-validation interval. The percentages of train, test and validation data are, respectively, the $50\%$, $30\%$ and $20\%$ of the remaining input data, all chosen randomly.

## 2.2 Variable selection

Effective variable selection can substantially improve model performance and generalization. The power of variable selection is the ability to find small synergistic subsets of variables which solve the problem as well as or better than the full set of measured variables.

Genetic algorithms ([11],[14]) are loosely based on some of the processes involved in biological evolution. Any genetic algorithm has a population of individuals which change from one generation to the next, usually by combining characteristics of two *parent* individuals to create a *child* individual. Every individual is assigned a fitness and the concept of *survival of the fittest* is implemented by selecting fitter parents more frequently than less fit parents. In our case, the individuals are sets of inputs variables. A set of inputs variables derives its fitness from how successful a model can be built based on just those variables. As the algorithm depends on the initialization values we have begin with different initializations and the variables which have consistently fail to be included in the final population are omitted.

We have chosen fuzzy systems ([24]) as a reliable tool to compare genetic algorithms with other techniques of soft computing. The implemented procedure of fuzzy variable selection is based on a fuzzy ranking (FR) which has showed its well performance in other subjects ([18]). Summarizing, the basic idea behind this method is to assess the flatness of a fuzzy curve -defined by gaussian membership functions- characterizing a given input variables, since the output variable is scarcely influenced by the input variable if the related fuzzy curve is nearly flat. The subset of most outstanding variables, which define the subset of selected variables, depends on a threshold $\vartheta$ defined by an expert.

## 2.3 Learning rules

The adaptative learning rule uses back-propagated gradient information to guide an iterative line search algorithm ([7]). The search direction in the weight space is modified by the previous search direction, a decay term -to avoid overfitting- is obtained by and heuristic algorithm which constructs a two dimensional grid with one axis associated with the parameter *weight decay* for the hidden layer and the other associated to the weight decay for the output layer. Finally, a random vector is added to the line search direction vector. During training, the learning rule modifies the weights in response to the training data. If left unchecked, the weights for a processing element can latch onto spurious information in the training data, such as data that does not represent a general trend in

the input data. By slowly decaying the weights during the course of the training, only the general trends remain encoded in the weights.

Kalman filter learning rule considers the desired outputs to be the observations within a discrete state space transition framework. Standard non-linear Kalman filter theory is used to obtain the best estimate of the weights based on the stream of training data. Overfitting can be avoid by increasing the observation noise of the filter.

# 3 Results

Five methods has been implemented: adaptative gradient without variable selection (AG), adaptative gradient with genetic algorithm variable selection (AGG), adaptative gradient with fuzzy variable selection (AGF), Kalman filter with genetic algorithm variable selection (KG) and a Kalman filter with fuzzy variable selection (KF). For each kind of methodology several analysis have been made in different conditions of noise. We only present the most indicative and performing. Acronyms: $R$=linear correlation between real and model outputs and $RMS$=*root mean square error*. *Mean absolute percentage errors*=$100n^{-1}\sum_{i=1}^{n}|y_i-\hat{y}_i|/y_i$ is calculated for the whole ($E$), training ($E1$), test ($E2$), validation ($E3$), out-interval ($EG$) and rain ($Erain$) sets of records. We present the results for all the methods with NN1. We do not present the results for all the methods for NN2 in order to simplify the whole set of tables and figures. Due to the fact that in NN2 we seek for a model for the whole water quality we can not expect that the best methods for NN1 will be the best for NN2.

## 3.1 Total Nitrogen

Table 3 shows the results of the different methods applied for NN1 being, in general terms, the adaptative gradient with fuzzy variable selection and the Kalman filter with genetic algorithm variable selection the most performing. Taking as reference the adaptative gradient we observe that RMS and E diminish $20.9\%$ and $21.6\%$ when the fuzzy selection is applied. In the same way RMS and E are reduced $19.6\%$ and $20.6\%$ if we apply a Kalman Filter with genetic algorithm variable selection. More important are still the diminutions during the period of rain. In effect, Erain reduces by $32.2\%$ and $33\%$ applying AGF and KG. This fact permit to assert that a good election of variables can significantly improve the forecasting process. Table 4 shows the results for NN2 being $\%\Delta$ the increment of errors between the results of NN1 and NN2.

In Figure 2 and Figure 3 is plotted the real data versus the predicted data for the Kalman filter with genetic algorithm variable selection for NN1 and NN2 respectively. It is clear that is more difficult to fit in the peak zones especially for NN2.

Table 3: Main analytical results for Ntotal with NN1

| | AG | AGG | AGF | KG | KF |
|---|---|---|---|---|---|
| R | 0.926 | 0.921 | 0.955 | 0.953 | 0.918 |
| RMS | 0.854 | 0.883 | 0.675 | 0.687 | 0.902 |
| E | 4.16 | 4.32 | 3.26 | 3.30 | 4.33 |
| E1 | 4.17 | 4.37 | 3.21 | 3.26 | 4.44 |
| E2 | 3.94 | 4.01 | 3.37 | 3.35 | 4.22 |
| E3 | 4.40 | 4.56 | 3.26 | 3.30 | 4.15 |
| EG | 2.33 | 2.55 | 2.79 | 1.08 | 2.42 |
| Erain | 7.76 | 7.56 | 5.26 | 5.20 | 7.93 |

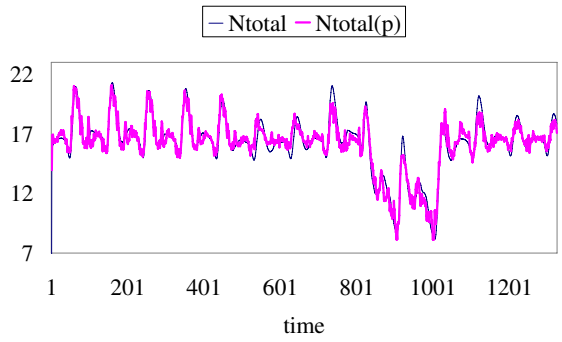


Figure 2: Ntotal real data versus predicted data based on a Kalman filter with genetic algorithm variable selection for NN1

Due to the long time interval it is difficult to appreciate the fitting of the predictions versus the real data. We can evaluate more accurately this fact taken in account only the data in the period of rain which is more complex to forecast. Figure 4 and Figure 5 describe this behaviour applying the same methodology that in Figure 2 and Figure 3. It is clear the difficulty of forecast with a vector composed by five components.

In order to check the behaviour of predicting new data we present the Figure 6 and Figure 7 based on

Table 4: Main analytical results for Ntotal with NN2

| | AG | %Δ | AGG | %Δ | KG | %Δ |
|---|---|---|---|---|---|---|
| R | 0.886 | -4.3 | 0.865 | -6.1 | 0.875 | -8.2 |
| RMS | 1.257 | 47.2 | 1.29 | 46.1 | 1.09 | 58.6 |
| E | 6.30 | 51.6 | 6.97 | 61.3 | 5.13 | 55.5 |
| E1 | 6.72 | 61.2 | 7.37 | 68.7 | 5.36 | 64.4 |
| E2 | 5.47 | 38.9 | 6.14 | 53.2 | 4.80 | 43.3 |
| E3 | 6.22 | 41.3 | 7.03 | 54.1 | 4.87 | 47.6 |
| EG | 4.00 | 71.5 | 4.97 | 95.1 | 2.81 | 160 |
| Erain | 11.9 | 53.7 | 13.2 | 75.3 | 11.4 | 119 |

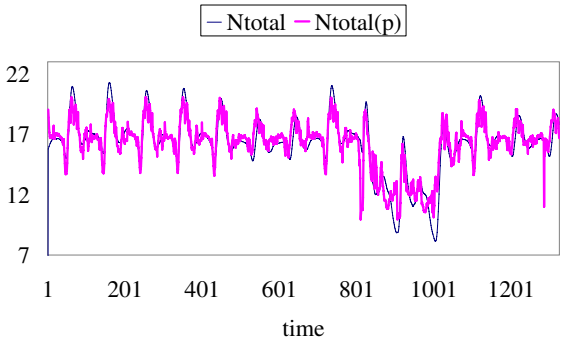


Figure 3: Ntotal real data versus predicted data based on a Kalman filter with genetic algorithm variable selection for NN2

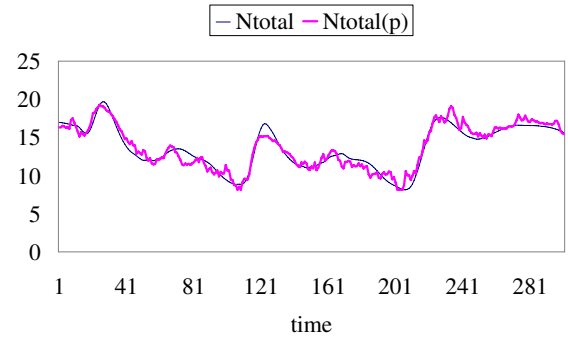


Figure 4: Ntotal real rain data versus predicted rain data based on a Kalman filter with genetic algorithm variable selection for NN1

a Kalman filter with genetic algorithm variable selection which has been applied to 20 new data outside the interval selected for the training, test and validation sets. We can appreciate a good fit between real and predicted data.

The first column of Table 5 shows the fuzzy ranking for each input variable on the prediction of Ntotal.

These values represent a fuzzy subset $\tilde{R}_1$ with membership function

$$\begin{array}{lll} \mu_{\tilde{R}_1}(t) = 0.1 & , & \mu_{\tilde{R}_1}(\mathrm{S_S}) = 0.3 \\ \mu_{\tilde{R}_1}(\mathrm{X_{B,H}}) = 0.8 & , & \mu_{\tilde{R}_1}(\mathrm{X_S}) = 0.7 \\ \mu_{\tilde{R}_1}(\mathrm{X_I}) = 0.8 & , & \mu_{\tilde{R}_1}(\mathrm{S_{NH}}) = 0.8 \\ \mu_{\tilde{R}_1}(\mathrm{S_I}) = 0.8 & , & \mu_{\tilde{R}_1}(\mathrm{S_{ND}}) = 0.3 \\ \mu_{\tilde{R}_1}(\mathrm{X_{ND}}) = 0.8 & , & \mu_{\tilde{R}_1}(\mathrm{Q}) = 1 \\ \mu_{\tilde{R}_1}(\mathrm{AirE}) = 0.2 & , & \mu_{\tilde{R}_1}(\mathrm{PumpE}) = 0.9 \\ \mu_{\tilde{R}_1}(\mathrm{WAS}) = 0.8 & & \end{array}$$

Taking $\theta = 0.7$ as the minimum value necessary to select a variable we get the subset $A_1$ of selected variables

$$A_1 = \{\mathrm{X_{BH}}, \mathrm{X_S}, \mathrm{X_I},$$

$$\mathrm{S_{NH}}, \mathrm{S_I}, \mathrm{X_{ND}}, \mathrm{Q}, \mathrm{PumpE}, \mathrm{WAS}\}$$

Table 6 and Table 7 show the average contribution -for some methods- for each input variable on the prediction of the outcome depending on the architectures NN1 and NN2. These values have been fuzzified from very little contribution (1) to very high contribution (7). Genetic algorithm variable selection and fuzzy variable selection coincides in 70% of the variables but without preserving their contributions.

Table 8 and Table 9 give the sensitivity (measured in absolute value) for the Total Nitrogen with respect to the inputs for NN1 and NN2 respectively.

## 3.2 Total Kjeldahl Nitrogen

Table 10 and Table 11 show the results of the different methods applied for NN1 and NN2. Globally, we can assert that the adaptative gradient with genetic algorithm variable selection and the Kalman filter with genetic algorithm variable selection are the most performing.

The second column of Table 5 shows the fuzzy ranking for each input variable on the prediction of TKN. These values represent a fuzzy subset $\tilde{R}_2$ with membership function

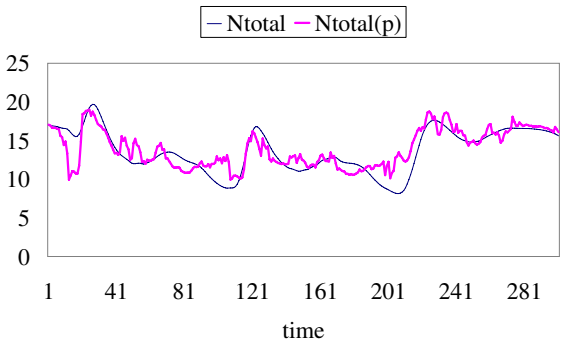


Figure 5: Ntotal real rain data versus predicted rain data based on a Kalman filter with genetic algorithm variable selection for NN2

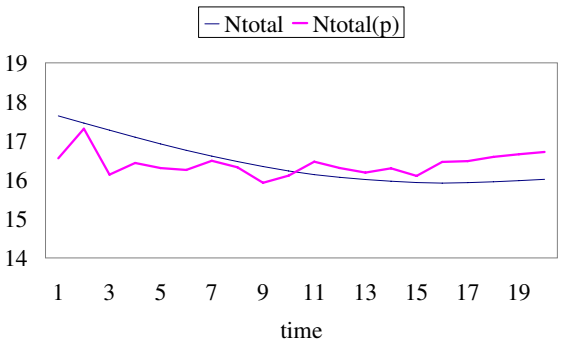


Figure 7: Ntotal: real 20 new data versus predicted 20 new data based on a Kalman filter with genetic algorithm variable selection for NN2

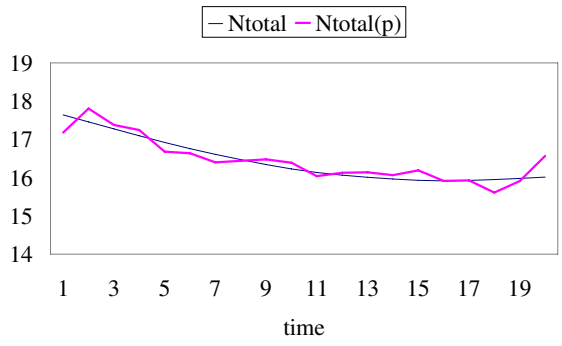


Figure 6: Ntotal: real 20 new data versus predicted 20 new data based on a Kalman filter with genetic algorithm variable selection for NN1

Table 5: Fuzzy ranking for Ntotal and TKN

| | Ntotal | TKN |
|---|---|---|
| $t$ | 0.1 | 0.1 |
| $S_S$ | 0.3 | 0.5 |
| $X_{B,H}$ | 0.8 | 0.8 |
| $X_S$ | 0.7 | 0.3 |
| $X_I$ | 0.8 | 0.7 |
| $S_{NH}$ | 0.8 | 0.7 |
| $S_I$ | 0.8 | 0.8 |
| $S_{ND}$ | 0.3 | 0.7 |
| $X_{ND}$ | 0.8 | 0.3 |
| Q | 1 | 0.7 |
| AirE | 0.2 | 0.8 |
| PumpE | 0.9 | 1 |
| WAS | 0.8 | 0.8 |

Table 6: Contribution for Ntotal with NN1

| | AG | AGG | AGF | KG |
|---|---|---|---|---|
| $t$ | 1 | 2 | | 2 |
| $S_S$ | 1 | | | |
| $X_{B,H}$ | 6 | | 2 | |
| $X_S$ | 1 | 4 | 3 | 6 |
| $X_I$ | 7 | 7 | 1 | 3 |
| $S_{NH}$ | 3 | 2 | 2 | 7 |
| $S_I$ | 6 | 7 | 1 | 2 |
| $S_{ND}$ | 7 | | | |
| $X_{ND}$ | 7 | | 3 | |
| Q | 1 | 1 | 7 | 1 |
| AirE | 1 | 4 | | 7 |
| PumpE | 2 | 4 | 1 | 4 |
| WAS | 4 | 7 | 1 | 5 |

Table 7: Contribution for Ntotal with NN2

| | AG | AGG | KG |
|---|---|---|---|
| $t$ | 1 | | |
| $S_S$ | 1 | | |
| $X_{B,H}$ | 6 | | |
| $X_S$ | 1 | | |
| $X_I$ | 7 | 7 | 3 |
| $S_{NH}$ | 3 | | |
| $S_I$ | 6 | | |
| $S_{ND}$ | 7 | | |
| $X_{ND}$ | 7 | | |
| Q | 1 | 1 | 1 |
| AirE | 1 | | |
| PumpE | 2 | 4 | 4 |
| WAS | 4 | 7 | 5 |

Table 8: Sensitivity for Ntotal with NN1

| | AG | AGG | AGF | KG |
|---|---|---|---|---|
| $t$ | 0.08 | 0.12 | | 0.06 |
| $S_S$ | 1.24 | | | |
| $X_{B,H}$ | 0.93 | | 2.13 | |
| $X_S$ | 0.61 | 0.40 | 3.27 | 0.29 |
| $X_I$ | 1.50 | 0.29 | 1.97 | 0.20 |
| $S_{NH}$ | 0.65 | 0.43 | 0.74 | 0.76 |
| $S_I$ | 3.23 | 0.02 | 1.95 | 0.04 |
| $S_{ND}$ | 1.25 | | | |
| $X_{ND}$ | 1.26 | | 2.53 | |
| Q | 0.86 | 0.30 | 0.48 | 0.09 |
| AirE | 0.11 | 0.28 | | 0.13 |
| PumpE | 0.19 | 0.20 | 0.08 | 0.11 |
| WAS | 0.39 | 0.45 | 0.05 | 0.02 |

Table 9: Sensitivity for Ntotal with NN2

| | AG | AGG | KG |
|---|---|---|---|
| $t$ | 0.05 | | |
| $S_S$ | 1.07 | | |
| $X_{B,H}$ | 1.02 | | |
| $X_S$ | 0.48 | | |
| $X_I$ | 1.17 | 0.22 | 0.25 |
| $S_{NH}$ | 0.36 | | |
| $S_I$ | 2.85 | | |
| $S_{ND}$ | 1.13 | | |
| $X_{ND}$ | 0.86 | | |
| Q | 1.32 | 0.45 | 0.18 |
| AirE | 0.21 | | |
| PumpE | 0.24 | 0.27 | 0.11 |
| WAS | 0.65 | 0.61 | 0.17 |

$$\begin{array}{ll} \mu_{\tilde{R}_2}(t) = 0.1 & , \quad \mu_{\tilde{R}_2}(\mathrm{S_S}) = 0.5 \\ \mu_{\tilde{R}_2}(\mathrm{X_{B,H}}) = 0.8 & , \quad \mu_{\tilde{R}_2}(\mathrm{X_S}) = 0.3 \\ \mu_{\tilde{R}_2}(\mathrm{X_I}) = 0.7 & , \quad \mu_{\tilde{R}_2}(\mathrm{S_{NH}}) = 0.7 \\ \mu_{\tilde{R}_2}(\mathrm{S_I}) = 0.8 & , \quad \mu_{\tilde{R}_2}(\mathrm{S_{ND}}) = 0.7 \\ \mu_{\tilde{R}_2}(\mathrm{X_{ND}}) = 0.3 & , \quad \mu_{\tilde{R}_2}(\mathrm{Q}) = 0.7 \\ \mu_{\tilde{R}_2}(\mathrm{AirE}) = 0.8 & , \quad \mu_{\tilde{R}_2}(\mathrm{PumpE}) = 1 \\ \mu_{\tilde{R}_2}(\mathrm{WAS}) = 0.8 & \end{array}$$

Taking $\theta = 0.7$ as the minimum value necessary to select a variable we get the subset $A_2$ of selected variables

$$A_2 = \{\mathrm{X_{B,H}}, \mathrm{X_I}, \mathrm{S_{NH}}, \mathrm{S_I}, \mathrm{S_{ND}}, \mathrm{Q}, \mathrm{AirE}, \mathrm{PumpE}, \mathrm{WAS}\}$$

Genetic algorithm variable selection and fuzzy variable selection coincides in 70% of the variables but without preserving their contributions.

It is relevant to notice that the errors for NN1 have increased significantly with respect the results for Ntotal. Total nitrogen is the addition of the total Kjeldahl nitrogen with the concentration of $NO_2$ and $NO_3$. During the period of rain the peaks of TKN become more accentuated and, on the other hand, the concentrations of $NO_2$ and $NO_3$ diminish, making the curve of Ntotal more smoothing then more easily to predict.

In Figure 8, Figure 9 and Figure 10 are plotted the real data versus the predicted data for the Kalman filter with genetic algorithm variable selection with NN1 for the whole time interval, in the period of rain and for the 20 new data. We do not present the figures for NN2 because follow a pattern already seen several times.

Table 12 and Table 13 show the average contribution -for some methods and depending on NN1 and NN2- for each input variable on the prediction of the outcome fuzzified from 1 (very little) to 7 (very high). Genetic algorithm variable selection and fuzzy variable selection coincides (as in the case of Total Nitrogen) in 70% of the variables but in this case the differences of contributions are less important.

Table 14 and Table 15 show the sensitivity (measured in absolute value) of the total Kjeldahl nitrogen with respect to the inputs for NN1 and NN2 respectively.

Table 10: Main analytical results for TKN with NN1

| | AG | AGG | AGF | KG | KF |
|---|---|---|---|---|---|
| R | 0.902 | 0.884 | 0.871 | 0.9305 | 0.8920 |
| RMS | 1.062 | 1.141 | 1.225 | 0.884 | 1.107 |
| E | 6.92 | 6.16 | 7.92 | 6.16 | 7.04 |
| E1 | 6.88 | 6.00 | 7.76 | 5.32 | 6.92 |
| E2 | 6.84 | 6.02 | 8.08 | 5.88 | 7.08 |
| E3 | 7.16 | 6.76 | 8.20 | 5.96 | 7.24 |
| EG | 5.28 | 4.44 | 7.27 | 4.56 | 6.12 |
| Erain | 7.50 | 7.12 | 8.65 | 7.23 | 8.13 |

Table 11: Main analytical results for TKN with NN2

| | AG | %Δ | AGG | %Δ | KG | %Δ |
|---|---|---|---|---|---|---|
| R | 0.865 | -4.1 | 0.837 | -5.3 | 0.870 | -6.5 |
| RMS | 1.513 | 42.5 | 1.592 | 39.5 | 1.09 | 1.315 |
| E | 9.99 | 44.5 | 9.03 | 46.6 | 8.61 | 41.2 |
| E1 | 10.9 | 58.3 | 9.38 | 56.4 | 8.05 | 51.3 |
| E2 | 9.61 | 40.5 | 8.41 | 39.7 | 8.70 | 47.9 |
| E3 | 9.92 | 38.5 | 9.93 | 46.9 | 8.32 | 39.6 |
| EG | 8.41 | 59.2 | 7.93 | 78.7 | 8.13 | 78.2 |
| Erain | 11.3 | 50.9 | 12.5 | 75.3 | 12.1 | 66.8 |

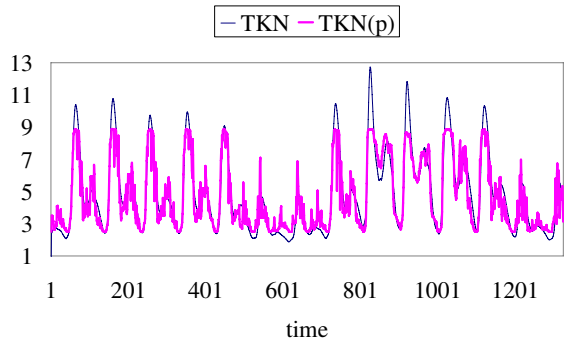


Figure 8: TKN real data versus predicted data by means a Kalman filter with genetic algorithm variable selection for NN1

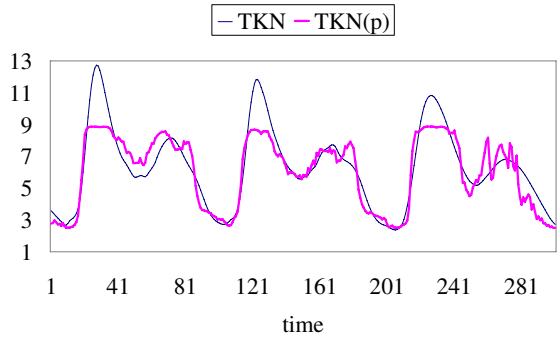


Figure 9: TKN real rain data versus predicted rain data based on a Kalman filter with genetic algorithm variable selection for NN1

Table 12: Contribution for TKN with NN1

| | AG | AGG | AGF | KG |
|---|---|---|---|---|
| $t$ | 1 | | | |
| $S_S$ | 3 | 5 | | 5 |
| $X_{B,H}$ | 3 | | 2 | |
| $X_S$ | 1 | | | |
| $X_I$ | 1 | 1 | 3 | 2 |
| $S_{NH}$ | 2 | 3 | 1 | 3 |
| $S_I$ | 4 | 2 | 2 | 3 |
| $S_{ND}$ | 4 | | 5 | |
| $X_{ND}$ | 2 | | | |
| Q | 2 | | 1 | |
| AirE | 4 | 7 | 3 | 7 |
| PumpE | 7 | 7 | 7 | 7 |
| WAS | 1 | 7 | 1 | 1 |

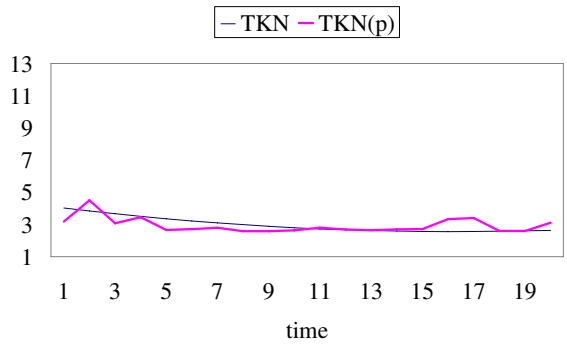


Figure 10: TKN: real 20 new data versus predicted 20 new data based on a Kalman filter with genetic algorithm variable selection for NN1

Table 13: Contribution for TKN with NN2

| | AG | AGG | KG |
|---|---|---|---|
| $t$ | 1 | | |
| $S_S$ | 2 | | |
| $X_{B,H}$ | 4 | | |
| $X_S$ | 1 | | |
| $X_I$ | 1 | 4 | 4 |
| $S_{NH}$ | 3 | | |
| $S_I$ | 4 | | |
| $S_{ND}$ | 4 | | |
| $X_{ND}$ | 4 | | |
| Q | 5 | 5 | 6 |
| AirE | 5 | | |
| PumpE | 3 | 4 | 4 |
| WAS | 7 | 7 | 7 |

Table 14: Sensitivity for TKN with NN1

| | AG | AGG | AGF | KG |
|---|---|---|---|---|
| $t$ | 0.05 | | | |
| $S_S$ | 0.34 | 0.99 | | 1.39 |
| $X_{B,H}$ | 0.29 | | 0.23 | |
| $X_S$ | 0.63 | | | |
| $X_I$ | 0.01 | 0.05 | 0.14 | 0.06 |
| $S_{NH}$ | 0.26 | 1.20 | 0.20 | 0.99 |
| $S_I$ | 1.11 | 0.44 | 1.18 | 0.41 |
| $S_{ND}$ | 0.32 | | 0.79 | |
| $X_{ND}$ | 0.24 | | | |
| Q | 0.05 | | 0.16 | |
| AirE | 0.97 | 0.57 | 0.20 | 0.43 |
| PumpE | 0.86 | 0.75 | 0.82 | 1.18 |
| WAS | 0.32 | 0.34 | 0.11 | 0.10 |

Table 15: Sensitivity for TKN with NN2

| | AG | AGG | KG |
|---|---|---|---|
| $t$ | 0.04 | | |
| $S_S$ | 0.73 | | |
| $X_{B,H}$ | 0.89 | | |
| $X_S$ | 0.52 | | |
| $X_I$ | 0.93 | 0.45 | 0.39 |
| $S_{NH}$ | 0.77 | | |
| $S_I$ | 1.65 | | |
| $S_{ND}$ | 0.65 | | |
| $X_{ND}$ | 0.77 | | |
| Q | 1.56 | 1.12 | 0.76 |
| AirE | 0.34 | | |
| PumpE | 0.32 | 0.45 | 0.31 |
| WAS | 1.12 | 1.09 | 0.87 |

# 4 Conclusions

The results explained in section 3 show that Neural Networks are a confident tool to predict the behaviour of the nitrogenous variables especially for the total nitrogen. Considering to implement an unique Neural Network for predicting all the variables that describe the water quality at the effluent of a Wastewater Treatment Plant induce a significant increment of the errors. Selection variables techniques of soft computing have improved the statically results in a significant value for some methods.

**Acknowledgements:** The Environmental Chemical Engineering Laboratory of the University of Girona has collaborated in those aspects strongly related with Wastewater Treatment Plants.

*References:*

[1] I. Antunes, L. Dionisio, M.C. Silva, M.S. Valente, J.J. Borrego, Coliforms as Indicators of Efficiency of Wastewater Treatment Plants, *Proceedings of the 3rd IASME/WSEAS International Conference on Energy, Environment, Ecosystems and Sustainable Development*, 2007, pp. 26–29.

[2] L.A. Belanche, J.J. Valdés, J. Comas, I.R. Roda, M. Poch, Towards a model of input-output behaviour of wastewater treatment plants using soft computing techniques, *Environmental Modellin and Software.* 14, 1999, pp. 409–419.

[3] C.M. Bishop, *Neural Networks for Pattern Recognition,* Clarendon Press 1995.

[4] M. Campolo, N. Clara, C. Morabito, Environmental Time Series Prediction by Improved Classical Feed-Forward Neural Networks, *Neural Nets, Lecture Notes in Computer Science,* Springer, 2006, pp. 171–178.

[5] I.A.C.P. Espirito Santo, E.M.G.P. Fernandes, M.M. Arajo, E.C. Ferreira, Contribution of a Simple Sludge Treatment in a WWTP Optimization Procedure, *Proceedings of the 5th WSEAS International Conference on System Science and Simulation in Engineering,*, 2006, pp. 54–59.

[6] I.A.C.P. Espirito Santo, E.M.G.P. Fernandes, M.M. Arajo, E.C. Ferreira, Optimization and Simulation of Secondery Settler Models, *Proceedings of the 6th WSEAS International Conference on Simulation, Modelling and Optimization,*, 2006, pp. 240–245.

[7] S.E. Fahlman, C. Lebiere, *The Cascade-Correlation Learning Architecture. Advances in Neural Information Processing Systems 2,* Morgan Kaufmann 1988.

[8] E. M. Fernandes, M. M. Arajo, E. C. Ferreira, Contribution of a Simple Sludge Treatment in a WWTP Optimization Procedure, *Proceedings of the 5th WSEAS International Conference on System Science and Simulation in Engineering,*, 2006, pp. 54–59.

[9] M. Fiter, J. Colprim, M. Poch, I. Rodriguez-Roda, Enhancing biological nitrogen removal in a small wastewater treatment plant by regulating the air supply, *Proceedings of the 5th Specialised Conference on Small Water and Wastewater Treatment Systems.* 2, 2002, pp. 793–800.

[10] M. Fiter, D. Güell, J. Comas, J. Colprim, M. Poch, I. Rodriguez-Roda, Fuzzy Logic Control for saving energy in a biological nitrogen removal process, *FAIA- Recent Advances in Artificial Intelligence research and development.* 113, 2004, pp. 185–192.

[11] D.E. Goldberg, *Genetic Algoritms for Search, Optimitzation* & *Machine Learning,* Addison-Wesley 1989.

[12] M. Henze, Jr. Grady, W. Gujer, G.v.R. Marais, T. Matsuo, *Activated Sludge Model No. 1. IAWQ. Scientific nad Technical Report No. 1*, IAWQ, London 1987.

[13] J. Hert, A. Krogh, R.G Palmer, *Introduction to the Theory of Neural Computation*, Addison-Wesley, Redwood City 1991.

[14] J.R. Koza, *Genetic Programming,* MIT Press 1993.

[15] P. Makvandi, J. Jassbi, S. Khanmohammadi, Application of Genetic Algorithm and Neural Network in Forecasting with Good Data, *WSEAS Transactions on Systems,* 4 (4), 2005, pp. 337–342.

[16] G. Mappa, R. Tagliaferri, D. Tortora, On line Monitoring based on Fuzzy Neural Techniques applied to existing hardware in Wastewater Treatment Plants, *Proceedings of the International Sympsosium on Intelligent Systems, AMSE-ISIS'97,* F. C. Morabitol et al. Ed.s, IOS Press, 1997, pp. 339–342.

[17] M. Mladenov, M. Dejanov, Application of Neural Networks for Seed Germination Assessment, *Proceedings of the 9th WSEAS International Conference on NEURAL NETWORKS,*, 2008, pp. 67–72.

[18] C.F. Morabito, M. Versaci, The Use of Fuzzy Curves for the Reconstruction of the Plasma Shape and the Selection of the Magnetic Sensors, *Fusion Technology,* 1, 1996, pp. 937–940.

[19] NeuralWare, *Neural Computing: A Technology Handbook for Professional II/PLUS, and NeuralWorks Explorer,* NeuralWare, 1993.

[20] H. T. Shandiz, H. G. Tehrani, H. Hadizadeh, Using Multi Layer Perceptron Network to Classify Road Cracks, *WSEAS Transactions on Systems,* 4 (4), 2005, pp. 355–358.

[21] I. Skrjanc, Monitoring of Waste-water Treatment Plant Based on Fuzzy Model, *WSEAS Transactions on Systems,* 6 (3), 2007, pp. 637–643.

[22] I. Takacs, G.G. Patry, D. Nolasco, A dynamic model of the clarification thickening process, *Wat. Res,* 25, 1991, pp. 1263–1271.

[23] S. Yordanova, P. Rusanka, V. Mladenov, Neuro-Fuzzy Control for Anaerobic Wastewater Treatment, *WSEAS Transactions on Systems,* 3 (2), 2004, pp. 724–730.

[24] H.J. Zimmermann, *Fuzzy Set Theory and its applications,* Kluver Academic Publishers 1992.