Aquacultural Engineering 48 (2012) 19–30

Contents lists available at SciVerse ScienceDirect

Aquacultural Engineering

journal homepage: www.elsevier.com/locate/aqua-online

Modeling inflow rates for the water exchange management in semi-intensive aquaculture ponds

Juan Carlos Gutiérrez-Estrada[a,*], Inmaculada Pulido-Calvo[a], Ignacio de la Rosa[a,b], Bruno Marchini[b]

[a] Dpto. Ciencias Agroforestales, Universidad de Huelva, Escuela Técnica Superior de Ingeniería, Campus Universitario de La Rábida, 21819 Palos de la Frontera, Huelva, Spain
[b] Langostinos de Huelva S.A., Carretera Pozo del Camino s/n, Redondela-Isla Cristina, Huelva, Spain

ARTICLE INFO

Article history:
Received 13 May 2011
Accepted 13 December 2011

Keywords:
Multiple linear regression
Generalized additive model
Multiple fuzzy regression
Evolutionary algorithm
Artificial neural network
Semi-intensive fishfarm
Sparus aurata
Southern Spain

ABSTRACT

Several linear and non-linear models for centralized remote-control systems that can support decision making of semi-intensive aquaculturists concerning the inflow rates to the ponds were evaluated. These models were: multiple linear regressions (MLRs), generalized additive models (GAMs), artificial neural networks (ANNs) and fuzzy logic controllers (FLCs). These modeling techniques were applied in a semi-intensive gilthead seabream (*Sparus aurata*) fishfarm located in southern Spain. The water temperature, ammonia concentration, turbidity and dissolved oxygen concentration in the ponds were measured and used as independent variables. Of all the approaches employed to simulate the actual water exchange operation in the ponds, the best fits were obtained using ANN and FLC models with only three input variables (turbidity measured at the input of the ponds and dissolved oxygen measured at the input and output of the ponds). These models provided levels of correlation between 0.73 and 0.75. In contrast, the best GAM and MLR models provided correlation coefficients of only 0.38 and 0.33, respectively. In spite of the results being statistically significant, the explained variance levels obtained indicate how difficult it is to capture the experience and knowledge of the aquaculturist concerning the operation of the water exchange in the ponds for maintaining the water quality in these production systems.

1. Introduction

Earth ponds are the unit most widely used for fish production worldwide, more than 40% of aquaculture production being carried out in ponds (Lekang, 2007). In Southern Spain (Andalucía region), 60% of aquaculture production is derived from semi-intensive facilities with diversion ponds which are usually built by excavating pits and building dikes. Water for these ponds is diverted from another water source, such as a stream or lake. The source water can either be supplied by a pump or by gravity. Usually, channels are used to distribute the water from the source to the ponds (Kelly and Kohler, 1997).

One of the main difficulties with production in ponds is maintaining control over the water quality and, consequently, over the fish yield. The various chemicals dissolved in water, its temperature and other physical attributes all combine to form what is called water quality. Accordingly, for aquaculture systems, changes in water characteristics that improve the production of an aquatic crop would be considered improvements in water quality, while those reducing production would be considered degradation of water quality. Also, what are considered good water quality characteristics will vary considerably between species. Thus, water quality must be viewed in the context of the species cultured (Diana et al., 1997).

On the other hand, many water quality parameters may affect the well-being of the species cultured, but, normally just a few play a decisive role. As a general rule, aquaculturists only measure the variables that are important to their stock, that can be interpreted and that can be altered through management (Boyd and Tucker, 1998). So, like in others fish rearing systems (Soto-Zarazúa et al., 2011), in the case of the production of gilthead seabream (*Sparus aurata*) in semi-intensive ponds in Southern Spain, the variables usually measured are water temperature, ammonia, turbidity and dissolved oxygen concentration.

In Southern Spain (in particular, in the region of Andalucía), the practice of water exchange is used widely in semi-intensive aquaculture ponds for improving water quality. Normally, these ponds are filled from brackish streams or lakes with relatively high water exchange rates which in some cases may exceed 20% of pond volume per day. Currently, water exchange management in many semi-intensive ponds is carried out depending only on the experience and knowledge of the aquaculturist. Despite this, there is a need to have a centralized remote-control system that can support decision making of semi-intensive farming aquaculturists and administrators to enable efficient water management policies to develop.

* Corresponding author. Tel.: +34 959 217528.
E-mail address: juanc@uhu.es (J.C. Gutiérrez-Estrada).

0144-8609/$ – see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.aquaeng.2011.12.009

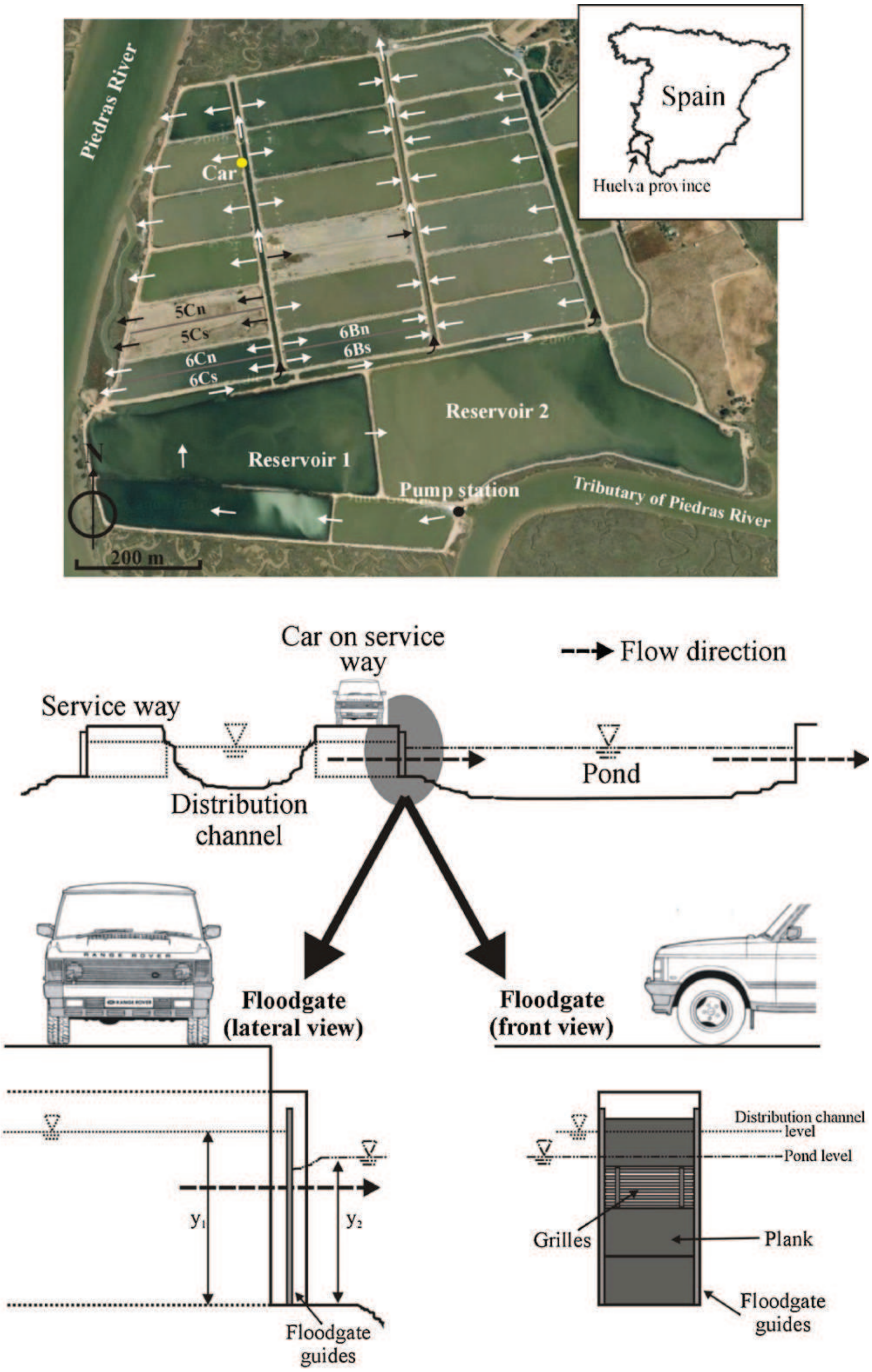

Fig. 1. Study area and schematic representation of the water supply to the ponds.

In this paper, two classical (multiple linear regressions – MLRs – and generalized additive models – GAMs) and two heuristic (artificial neural networks – ANNs – and fuzzy logic controllers – FLCs) methodologies are presented for the modeling of the water exchange in aquaculture ponds which is the basis for automatic control of these systems. As far as linear models are concerned, MLRs have frequently been used in a wide range of different science and engineering applications (Gutiérrez-Estrada et al., 2004; Steeby et al., 2004; Ruiz-Velazco et al., 2010). GAMs, on the other hand, have been applied to solve complex problems in which it is relevant to explore non-linear relationships between variables (Gutiérrez-Estrada et al., 2009; Murase et al., 2009; Li et al., 2011). In recent years, powerful soft-computing methods have been used to simulate, control and forecast the behavior of many different types of systems. Within these methods, ANNs and FLCs can be extremely effective in comparison with traditional techniques in

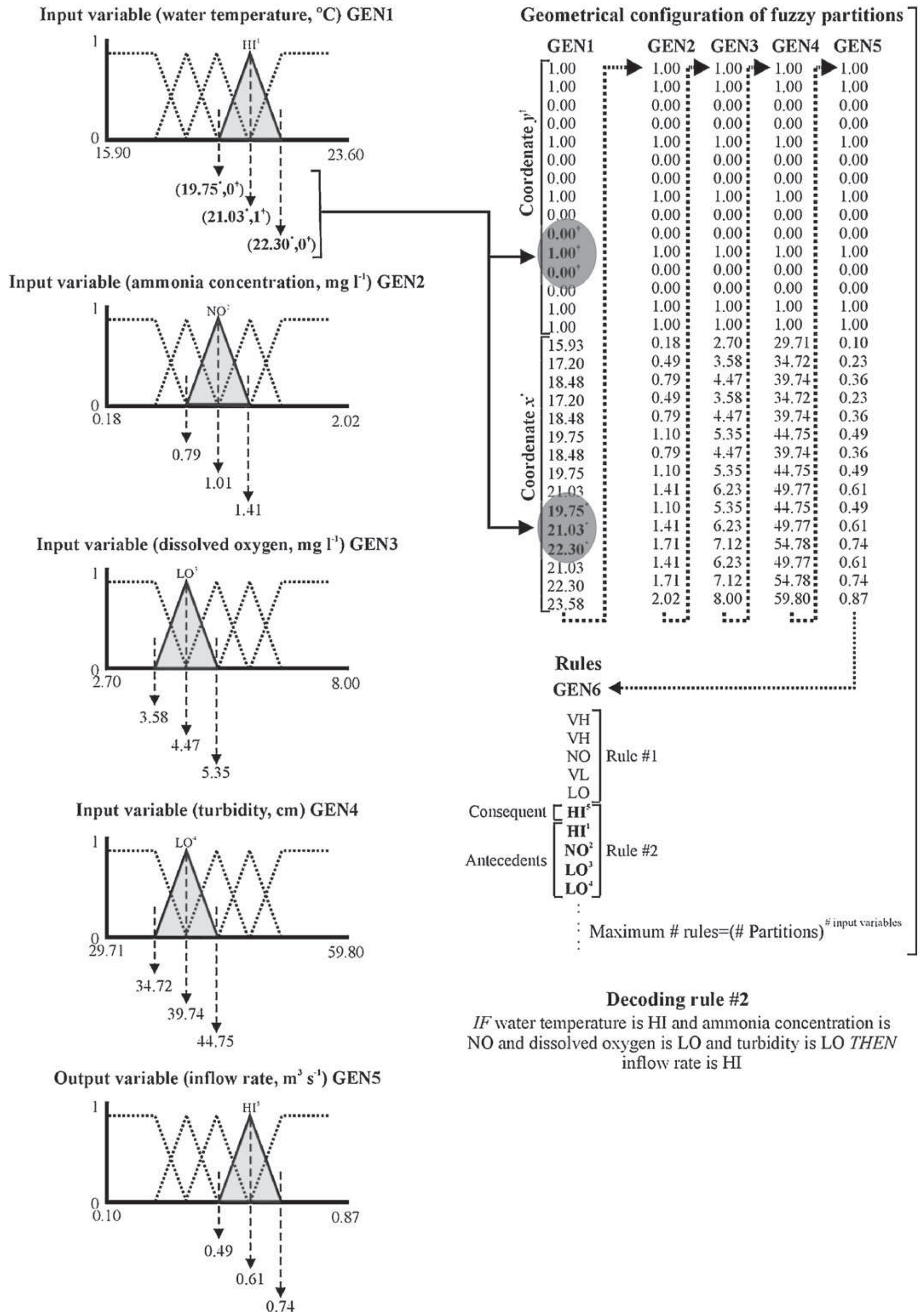

Fig. 2. Example of the codification of four input variables or antecedents (water temperature, ammonia concentration, dissolved oxygen and turbidity) and one output variable or consequent (inflow rate) in a chromosome.

those situations in which the phenomena involved are very complex and not well understood (Lee et al., 2000; Papandroulakis et al., 2000; Yu et al., 2006; Pulido-Calvo and Gutiérrez-Estrada, 2009; Soto-Zarazúa et al., 2011).

Therefore, these four methodologies seem promising approaches for modeling inflow rates to aquaculture ponds and, consequently, for controlling the opening of the ponds gates. The main purpose of applying these models is to capture the knowledge of the human experts who are usually responsible for manual control of the water exchange in aquaculture ponds. The models were tested in a semi-intensive gilthead seabream (*S. aurata*) fishfarm located in southern Spain.

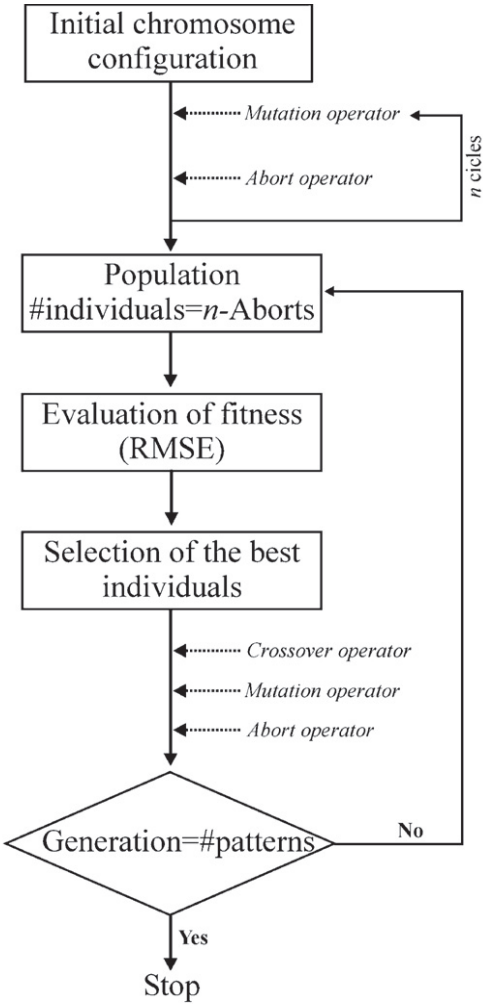

Fig. 3. Flow chart of the FLC model optimization with evolutionary algorithms.

2. Materials and methods

2.1. Study area and data collection

In order to test the methodology developed in this study, ponds of a semi-intensive fishfarm located in southern Spain were selected. The fishfarm, 'Langostinos de Huelva S.A.', is located in the province of Huelva (Southern Spain) and is devoted to gilthead seabream (*S. aurata*) production. The water is pumped from a tributary of the Piedras River to two regulation reservoirs. From these reservoirs, the water is conducted to the ponds by distribution channels (Fig. 1). In the ponds the seabream grow from a weight of 30–100 g to commercial weights (400–500 g and 1000 g).

The sampling period was made to coincide with the moment of the maximum frequency of ponds gates management (mid-April to end-May). This is a consequence of that in previous periods to mid-April the water exchange rate is normally very low (because the maximum temperatures are not very high), and therefore, the control level of ponds gates is very low (the workers have the ponds gates with a similar opening during a large time period). A similar situation can be found after May. At the end of the spring and during the summer, the control level of ponds gates is very low although the water exchange rate is normally very high (the workers have the ponds gates with the maximum opening during a large time period). From mid-April to end-May is a period of transition in which is normal to find environmental conditions that force to change frequently the gates opening.

As noted above, water exchange management is currently carried out manually and guided only on the experience and knowledge of the aquaculturist. The variables that support their decisions about water exchange in the ponds are: water temperature, ammonia, turbidity and dissolved oxygen concentration. From April 21 to May 29, 2008, these variables were measured daily in six selected ponds (Fig. 1: 5Cn, 5Cs, 6Cn, 6Cs, 6Bn, 6Bs), as were the water levels on each side of the ponds gates (y_1 and y_2, Fig. 1). Water temperature (T, °C) and dissolved oxygen concentration (Ox, mg l^{-1}) were recorded using a Hach HQ30d multi-parameter digital meter. The dissolved oxygen was measured at the input (Ox1) and output (Ox2) of each pond. Ammonia concentration (AM, mg l^{-1}) was recorded using the Merckoquant® ammonia test and turbidity (Secchi, cm visibility) with a Secchi disk, at three points: the input (Secchi1), middle (Secchi2) and output (Secchi3) of each pond.

The gates control water discharge from the distribution channel to the ponds. These gates are composed of independent planks and grilles (Fig. 1). The discharge equation for submerged orifices is (Munson et al., 2006):

$$Q = C_d w \sqrt{2gh} \qquad (1)$$

where Q is the flow discharge ($m^3 s^{-1}$), C_d is the discharge coefficient, w is the cross-section of the orifice (m^2), g is the acceleration due to gravity ($m^2 s^{-1}$) and h is the difference between the water levels on the front and back sides of the gate ($h = y_1 - y_2$, m). The discharge coefficient C_d considers geometric and friction effects and it can take a mean value of 0.62 (Munson et al., 2006).

2.2. Multiple linear regression (MLR) and generalized additive model (GAM)

A GAM model is a generalization of an MLR. Specifically, in a linear regression, a linear least squares fit is computed for a set of predictors or q variables, to predict a dependent variable. The generalization of the multiple regression model implies maintaining the additive nature of the model, but replacing the simple terms of the linear equation by a nonparametric function. Thus, instead of a single coefficient for each variable (additive term) in the model, in

Table 1
Cross-correlation matrix between all factors.

	Water temperature (T)	Ammonia concentration (AM)	Dissolved oxygen (Ox1)	Dissolved oxygen (Ox2)	Turbidity (Secchi1)	Turbidity (Secchi2)	Turbidity (Secchi3)
Ammonia concentration (AM)	−0.03						
Dissolved oxygen (Ox1)	−0.35*	−0.06					
Dissolved oxygen (Ox2)	−0.16*	−0.02	0.58*				
Turbidity (Secchi1)	0.08	−0.32*	0.02	0.02			
Turbidity (Secchi2)	0.03	−0.10	0.01	−0.10	0.09		
Turbidity (Secchi3)	0.01	−0.11	0.00	−0.14*	0.15*	0.47*	
Inflow (Q)	0.18*	0.03	0.06	0.18*	0.09	−0.15*	−0.03

* $p < 0.05$.

Table 2
Multiple linear regression considering all physical and chemical variables as independent variables and inflow as the dependent variable. $R = 0.3313$, $r^2 = 0.1098$, adjusted $r^2 = 0.0822$, $F(7,226) = 3.9826$, $p < 0.01$, $N = 234$.

	Beta	Coefficient B	p
Intercept		−0.0002	0.9973
Independent variables			
Temperature	0.2167	0.2167	0.0014*
Ammonia	0.0655	0.0655	0.3260
Ox1	0.0315	0.0315	0.7011
Ox2	0.1889	0.1889	0.0173*
Secchi1	0.1002	0.1002	0.1357
Secchi2	−0.1687	−0.1687	0.0189*
Secchi3	0.0613	0.0613	0.3970

* $p < 0.05$.

GAM models an unspecified (non-parametric) function is estimated for each predictor, to achieve the best prediction of the dependent variable values (Hastie and Tibshirani, 1990). In this study, Gamma distribution models were fitted to the data with logit link functions and cubic spline smoothing function with three degrees of freedom, in the GAM models.

2.3. Artificial neural network (ANN)

ANNs are mathematical models inspired by the neural architecture of biological nervous systems. The most widely studied and used ANN models involve multilayer feed-forward networks, or multilayer perceptrons (Rumelhart et al., 1986). These models 'learn' in an iterative way, whereby the data are introduced a number of times to the neural network until a pre-determined level of error is reached. These supervised ANNs allow the analysis of complex datasets and the assessment of non-linear relationships between dependent and independent variables. Detailed descriptions of the performance of multilayer perceptron ANNs have been published by Hsu et al. (1995), ASCE (2000a, b), Shrestha et al. (2005), Gutiérrez-Estrada et al. (2007), and Pulido-Calvo and Portela (2007). Many methods can be used for calibration or learning with ANNs; in this study, the Levenberg-Marquard algorithm was used (Tan and van Cauwenberghe, 1999; Anctil and Rat, 2005).

Prior to the calibration of any ANN, the dataset was divided into two subsets: (i) the CSS, the calibration subset (CS) + the select subset (SS), which comprised 75% of the data (randomly selected); and (ii) the TS, the test subset with the remaining data. The TS was not used for model calibration or training, but was involved in verification or validation of the models. In the CSS, 25% of the data (randomly selected) composed the select subset (SS), used to avoid the ANN overtraining or over-fitting.

The best method of ensuring that overtraining does not occur is to periodically monitor the sum square error for both the CS and the SS (internal validation) subsets. The sum square error for the CS normally decreases continuously with training. However, this may force the neural network to fit the noise in the CS. To avoid this problem, the sum square error of the SS is calculated at the end of each training iteration (epoch). Training ceases when the sum square error of the SS begins to increase, and at this point the weight of the epoch, which provides a minimum error for the SS, is tested with the TS. This last phase is also called the generalization phase, or external validation. Iyer and Rhinehart (1999) recommend repeating this process at least 30 times for each model, and this recommendation was followed in our work.

The procedure described above was carried out for each neural configuration tested. ANNs with one and two hidden layers were assessed in this study; in each case 5–20 neurons were tested (Gutiérrez-Estrada et al., 2008). ANN models were implemented using STATISTICA 6.0 (Statsoft, Inc., 1984–2002).

All of the CSS data that were used for model training were standardized by subtracting the mean and dividing by the standard deviation. The TS data that were used for model generalization or verification were also transformed using the mean and standard deviation of those variables in the CSS. This procedure was implemented to avoid the masking of features of interest (Ochoa-Rivera et al., 2007; Makkeasorn et al., 2008).

Table 3
Generalized additive model considering all physical and chemical variables as independent variables and inflow as the dependent variable. Distribution = normal, Link function = identity. $r^2 = 0.1491$, $N = 234$.

	Degree of freedom	GAM coefficient	Non-linear p value
Intercept	1.0000	0.0004	
Independent variables			
Temperature	2.9999	0.2167	0.7998
Ammonia	2.9972	0.0570	0.9058
Ox1	2.9959	−0.0615	0.0181*
Ox2	3.0038	0.2681	<0.001*
Secchi1	3.0040	0.1083	0.2273
Secchi2	2.9982	−0.1933	0.4407
Secchi3	3.0013	0.0995	0.5525

* $p < 0.05$.

Table 4
Sensitivity analysis of the best ANN considering all physical and chemical variables.

	Ratio	Ranking
Secchi1	1.4999[a]	1
Ox2	1.4980[a]	2
Ox1	1.4915[a]	3
Secchi3	1.2863	4
Ammonia	1.2805	5
Temperature	1.1368	6
Secchi2	1.1255	7
Average	1.3312	

[a] Selected variables.

2.4. Fuzzy logic controller

Fuzzy logic is an intelligent computational method which does not require a detailed mathematical description of the process to be controlled. It utilizes a form of many-valued logic. Unlike the 'crisp' logic, a fuzzy logic controller has in-between values. That is to say, a fuzzy set is divided into regions by geometric partitions. This allows us to describe a point as a function of its membership to different sets (Zeldis and Prescott, 2000). Fuzzy sets, in contrast to their crisp counterparts, have ambiguous boundaries and therefore gradual transitions between defined sets, allowing for the uncertainty associated with these concepts to be modeled directly. To construct a fuzzy controller, the first step is to define a series of

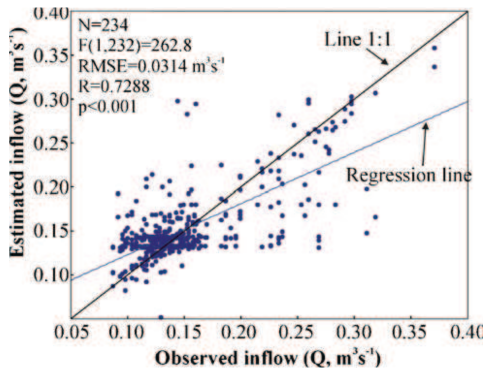

Fig. 4. Scatterplot between observed and estimated inflows for the best ANN. The linear regression parameters and RMSE error are shown.

Table 5
Weights matrix of the best ANN. This model has four layers: the input layer with three neurons, two hidden layers with 10 and 5 neurons, respectively, and the output layer with one neuron. Thresh = bias. The main numbers are layers and subscripts are neurons.

	2_1	2_2	2_3	2_4	2_5	2_6	2_7	2_8	2_9	2_{10}	3_1	3_2	3_3	3_4	3_5	4_1
Thresh	0.37	−1.47	0.44	−1.15	3.02	4.25	−6.61	−3.33	2.23	4.67	−5.40	2.16	1.70	1.07	−2.59	−0.44
1_1	1.23	−3.84	−11.76	2.37	17.54	8.78	−14.71	0.30	−2.73	−6.71						
1_2	−15.02	7.63	6.93	5.66	−3.60	−6.50	−2.25	3.55	−3.31	−1.20						
1_3	12.97	−3.57	1.82	3.75	−6.68	5.16	3.66	1.17	−3.77	13.30						
2_1											−4.49	6.21	−13.13	11.91	−7.90	
2_2											10.72	15.69	−0.15	25.73	12.08	
2_3											−2.02	19.03	5.41	−2.87	7.40	
2_4											4.48	−3.39	2.22	−1.06	8.41	
2_5											16.20	−6.01	10.69	−9.17	11.71	
2_6											2.39	0.39	−12.17	10.54	24.03	
2_7											−22.15	−26.16	14.79	−11.29	−11.65	
2_8											4.20	−5.09	−5.11	3.37	−0.50	
2_9											−5.67	4.88	2.23	3.58	0.60	
2_{10}											14.25	−4.74	1.57	5.81	0.24	
3_1																0.36
3_2																0.27
3_3																−0.07
3_4																−0.08
3_5																−0.36

overlapping fuzzy sets (or geometrical partitions) for each model variable and the mapping of inputs to outputs expressed as a set of IF–THEN rules (Kosko, 1997). Subsequently, it is necessary to define the method that transforms fuzzified inputs to defuzzified or quantitative outputs.

The fuzzy sets and rules are referred to as the knowledge base of the fuzzy system. Crisp inputs to the model are first fuzzified via this knowledge base (called Fuzzy Associative Memory or FAM), and a fuzzy inference engine is then used to process the rules in parallel via a fuzzy inference procedure such as the max–min or max–product operations (Jang et al., 1997). The explicit relationship between the partitions of the input and output fuzzy sets is stored in a FAM, which can be constructed on the basis of expert knowledge or observed data (Dubrovin et al., 2002).

In our study, the fuzzy logic rule-based model can take a maximum of seven input fuzzy sets (water temperature – *T*; ammonia concentration – AM; dissolved oxygen concentration in two points – Ox1 and Ox2; and turbidity in three points – Secchi1, Secchi2 and Secchi3) and produces one output fuzzy set (inflow to the pond – *Q*). All these sets are associated with three or five triangular partitions. In the case of five triangular partitions these were labeled as Very Low [VL], Low [LO], Normal [NO], High [HI] and Very High [VH]. Therefore, the general rule formulation could be as follows:

IF *T* is very low/low/normal/high/very high, AND AM is very low/low/
normal/high/very high, AND Ox1 is very low/low/normal/high/very high, AND Ox2 is very low/low/normal/high/very high, AND Secchi1 is very low/low/normal/high/very high, AND Secchi2 is very low/low/normal/high/very high, AND Secchi3 is very low/low/normal/high/very high, THEN *Q* is very low/low/ normal/high/very high.

The fuzzy solution resulting from the execution of the rule-base is defuzzified to produce the system output. In this work, the method used to obtain the system output was the minimum rule. Also, a function that transforms the fuzzy output into a crisp value is necessary. In this case the defuzzification technique was the center of the area:

$$Y = \frac{\sum_{f=1}^{S} y^{\text{center}\,(f)} \times \mu_{\text{out}}(y^f)}{\sum_{f=1}^{S} \mu_{\text{out}}(y^f)} \tag{2}$$

where the crisp value (Y) is the geometrical center ($y^{\text{center}\,(f)}$) of the output fuzzy values $\mu_{\text{out}}(y^f)$ with $f = 1,\ldots,S$, where S is formed by all contributions of rules whose degree of fulfillment is greater than zero.

2.5. Optimizing the fuzzy logic model with evolutionary algorithms

In the previous section, we have described how to construct a fuzzy logic model that can control the inflow rates to the ponds. The accuracy of this estimation depends of the parameters of the fuzzy logic model: (1) the shape of the fuzzy sets or geometrical partitions; (2) the degree of overlap between fuzzy sets; and (3) the definition of the IF–THEN rules. An evolutionary algorithm was used to find the optimal values of these parameters. Evolutionary algorithms are non-linear search and optimization methods inspired by the biological processes of natural selection and survival of the fittest (Holland, 1975; Goldberg, 1989). This algorithm differs from traditional search methods because it considers many points in the search space simultaneously and therefore has a low probability of converging to local optimum.

In an evolutionary algorithm the basic unit is the gene. Various genes contain the information required to define a chromosome whose decoding is interpreted as an individual. In this case, the parameters of the model were coded as genes in the chromosome (Fig. 2) (Jacob, 2001).

Once the initial information has been coded, three types of operators (reproduction, crossover and mutation) were used in order to evolve toward an optimal fuzzy configuration. Reproduction is a process in which chromosomes with high fitness values in generation t yield a high number of sons (copies) in the next generation ($t+1$). Crossover is an operator that mixes two chromosomes through a random process to take advantage of the best qualities of each chromosome. Lastly, the mutation operator changes the values of bits associated to a gene with a very low probability, which can produce unsuitable configurations (which are aborted) or pre-adaptive fuzzy configurations that generate better solutions. In this study, the probability of reproduction for a good individual was proportionate to the selection. The probability given to the mutation process was 0.01, and the mutation radius was selected as the 10% of the maximum value considered for each fuzzy set, while in the case of FAM the mutation radius was 2.

A fitness function (a term used in evolutionary algorithms to refer to an objective function) is required to apply the evolutionary algorithm (Chen et al., 2000). In this work, the RMSE error between observed and estimated inflow rates was used. The steps in the optimization of the model parameters are summarized in Fig. 3. In order to facilitate the application of this methodology, the

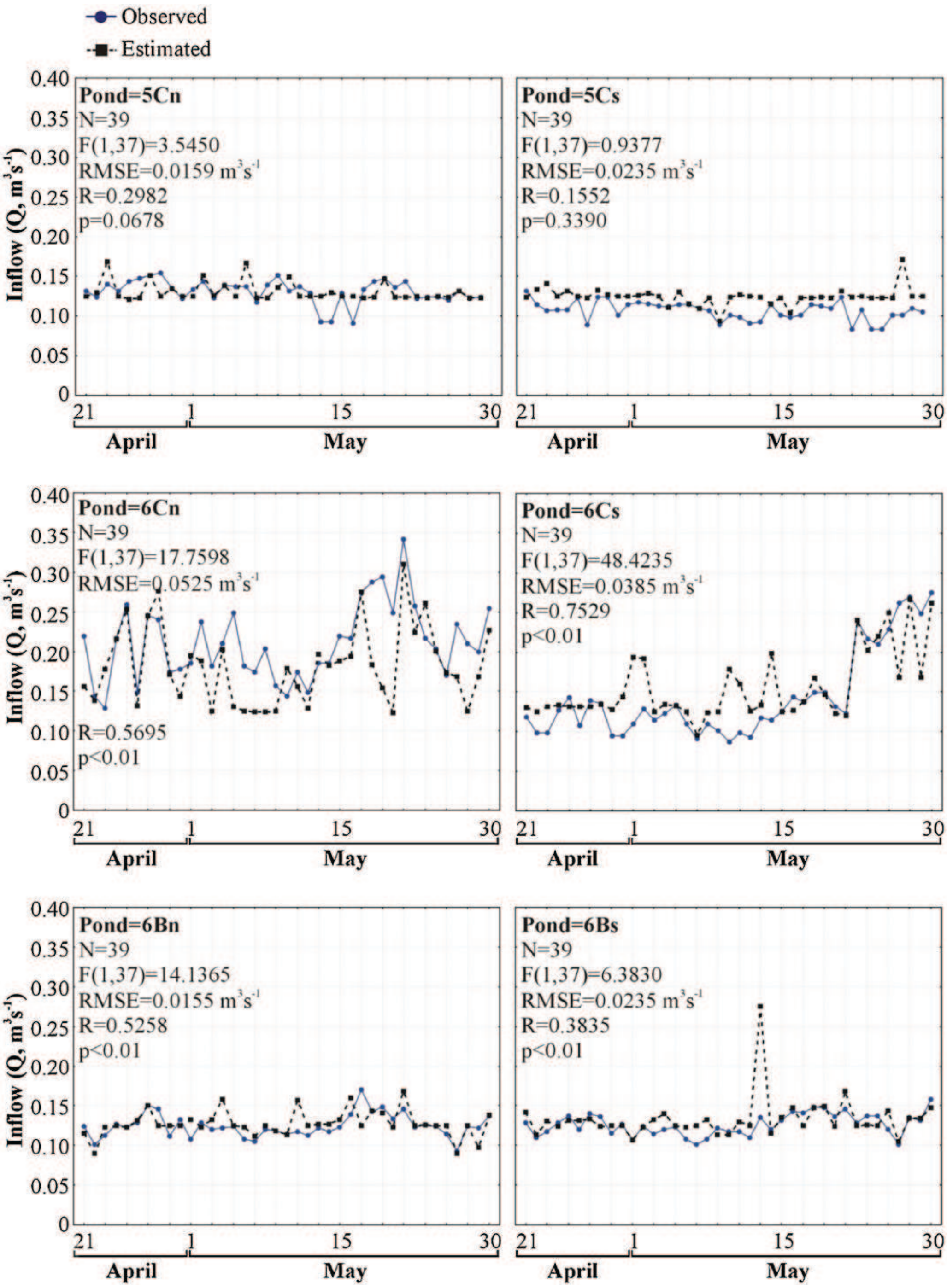

Fig. 5. Daily observed and estimated inflow to each pond for the best ANN.

NUBECILLA 1.0 computer program, written in Microsoft Visual Basic®, was specifically designed for this study.

2.6. Sensitivity analysis and measures of accuracy

An alternative form of sensitivity analysis, based on the approach of the missing value problem, was used for each ANN and FLC model evaluated. The analysis was carried out by replacing each selected input variable with missing values and assessing the effect of this on the output error. The newly calculated error was compared with the original error to obtain a ratio (error of the model with an input variable with missing values/error of the model with all selected input variables). Thus, for any input variable x, a ratio equal to or very close to 1 indicated that this variable had a very low weight in the general structure of the model (Hunter et al., 2000).

Two accuracy measures were calculated in the calibration and validation phases for each model evaluated: the coefficient of determination (r^2) and the square root of the mean square error (RMSE) (Legates and McCabe, 1999; Pulido-Calvo and Portela, 2007).

3. Results

A total of 234 samples were collected during the study period. Each sample was composed of a measurement of the following parameters: ammonia concentration (AM), water temperature (T), oxygen concentration at the input (Ox1) and the output (Ox2) of the pond, turbidity at the input gate (Secchi1), in the middle of the pond (Secchi2) and by the output (Secchi3), as well as the water inflow to the pond (Q). A preliminary analysis of the linear correlations between the variables indicated that there were hardly any significant linear correlations between the physical and chemical variables and the inflow. Only in the cases of the water temperature, the concentration of dissolved oxygen at the output of the pond and the turbidity in middle of the pond were correlation found to be significant, although even these were extremely weak (T: $R = 0.18$, $p < 0.05$; Ox2: $R = 0.18$, $p < 0.05$; Secchi2: $R = -015$, $p < 0.05$) and to a certain extent contradictory to the direction of the pond regulation.

With regards to the cross-correlations among the physical and chemical parameters, statistically significant correlations were found between both measurements of the oxygen concentration taken in the pond as well as between the three measurements of water turbidity (Table 1). Additionally, a significant negative correlation was found between water turbidity at the input grille and ammonia concentration ($R = -0.32$, $p < 0.05$) and between the measurements of dissolved oxygen and water temperature (T-Ox1: $R = -0.35$, $p < 0.05$; T-Ox2: $R = -0.16$, $p < 0.05$).

3.1. Multiple linear regression (MLR) and generalized additive model (GAM)

The results obtained with an MLR in which all available physical and chemical variables were used as independent variables are shown in Table 2.The level of variance explained by this model was extremely low (adjusted $r^2 = 0.08$), despite the fit being statistically significant [$F(7,226) = 3.9826$, $p < 0.05$)]. Of all the independent variables considered only water temperature (T), the oxygen concentration at the output (Ox2) and turbidity in the middle of the pond (Secchi2) had significant beta coefficients. A new model based only on water temperature (T), output oxygen concentration (Ox2) and water turbidity at the center of the pond (Secchi2) showed similar results [adjusted $r^2 = 0.08$, $F(3,230) = 8.0849$, $p < 0.05$].

A GAM model gave significantly better results, using a normal distribution, identity link functions and splines with 3 degrees of freedom. This model reached a level of explained variance close to 24% and a RMSE of $0.04\,m^3\,s^{-1}$. However, only two variables (Ox1 and Ox2) had a significant weight (Table 3). Given this, a new GAM model in which the variables Ox1 and Ox2 were considered as independent variables was calibrated and with this the level of explained variance was close to 15% and the RMSE $0.04\,m^3\,s^{-1}$.

3.2. Fuzzy logic controller (FLC) and artificial neural network (ANN)

Overall, considering all the physical and chemical parameters as input variables, a total of 240 ANNs were calibrated and validated (8 configurations of hidden layers × 30 repetitions), of which a model with one hidden layer with 15 neurons gave the best result. With this model in the validation phase, a level of explained variance above 37% was obtained, which meant a correlation coefficient slightly higher than 0.6 ($R = 0.6084$). The sensitivity analysis of this model showed that only the variables Secchi1, Ox2 and Ox1 had a significant weight in the estimation of the water inflow to the pond (Table 4). Therefore, a new group of 240 neural networks was calibrated and validated taking into account only these three input parameters.

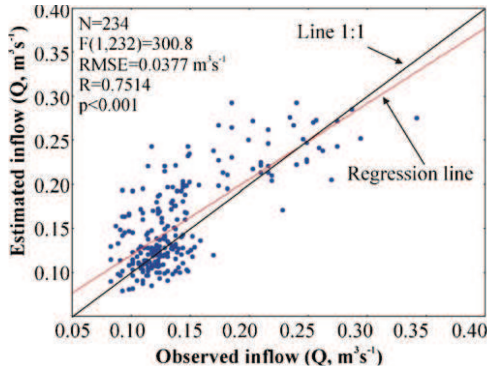

Fig. 6. Scatterplot between observed and estimated inflows for the best FLC. The linear regression parameters and RMSE error are shown.

Starting from this new configuration the best result was obtained using a neural network with two hidden layers with 10 and 5 neurons per layer (the weights matrix is shown in Table 5). Overall, this model explained 53% of the variance, meaning a level of error of $0.0314\,m^3\,s^{-1}$ (Fig. 4). A detailed analysis per pond showed that the model provided an adequate response for ponds 6Cn, 6Cs and 6Bn (6CN: $R = 0.5779$, $RMSE = 0.0525\,m^3\,s^{-1}$; 6Cs: $R = 0.7342$, $RMSE = 0.0385\,m^3\,s^{-1}$; 6Bn: $R = 0.5258$, $RMSE = 0.0155\,m^3\,s^{-1}$), while there were lower explained variances for ponds 5Cn, 5Cs and 6Bs (5Cn: $R = 0.2982$, $RMSE = 0.0159\,m^3\,s^{-1}$; 5Cs: $R = 0.1553$, $RMSE = 0.0235\,m^3\,s^{-1}$; 6Bs: $R = 0.3836$, $RMSE = 0.0255\,m^3\,s^{-1}$) (Fig. 5).

As in the neural network models, the sensitivity analysis of the fuzzy logic controller (FLC) considering all the physical and chemical parameters indicated that the variables with the most weight were Secchi1, Ox2 and Ox1 (Secchi1: ratio = 1.0212, ranking = 1; Ox2: ratio = 1.0172; Ox1 = 1.0000, ranking = 3). In this FLC, the input variables were composed of three partitions. The best approach for this configuration gave an explained variance of close to 56%, corresponding to a correlation coefficient slightly higher than 0.75 ($R = 0.7514$) and an RMSE of $0.0377\,m^3\,s^{-1}$ (Fig. 6). The analysis per pond revealed that the best behavior of the model was found in ponds 6Cn, 6Cs, 6Bn and 6Bs (6CN: $R = 0.5519$, $RMSE = 0.0492\,m^3\,s^{-1}$; 6Cn: $R = 0.6255$, $RMSE = 0.0614\,m^3\,s^{-1}$; 6Bn: $R = 0.3487$, $RMSE = 0.0236\,m^3\,s^{-1}$; 6Bs: $R = 0.4868$, $RMSE = 0.0220\,m^3\,s^{-1}$), while the goodness-of-fit for ponds 5Cn and 5Cs was very low (5Cn: $R = 0.0487$, $RMSE = 0.0280\,m^3\,s^{-1}$; 5Cs: $R = 0.0223$, $RMSE = 0.0223\,m^3\,s^{-1}$) (Fig. 7).

In this last FLC model, the flow was regulated using 24 management rules of a maximum of 27 (3 partitions for Secchi1 × 3 for Ox2 × 3 for Ox1) (Table 6). A detailed evaluation of the associative memory indicated a reasonable behavior, higher inflows being observed as the conditions become more extreme. This becomes most clear in the case of the input variable with the greatest weight (Secchi1). For this variable, the group of rules for a "Very Low" (VL) inflow in the pond (Output rule set = 1, Table 6) has an average level of 1.80. This average increases to 1.92 using the "Normal" (N) rules set for inflow to the pond (Rule set = 2, Table 6) and reaches a maximum for the "Very high" (VH) rules set for the inflow (Rule set = 3, Table 6). The fuzzy partitions obtained from the optimization process for each antecedent or input variable (Secchi1, Ox2 and Ox1), and for the consequent or output variable (Pond water inflow, Q) are shown in Fig. 8. In all cases, a clear asymmetry can be observed in the partitions and a high level of uncertainly associated with them, as can be expected for a high degree of overlapping between partitions.

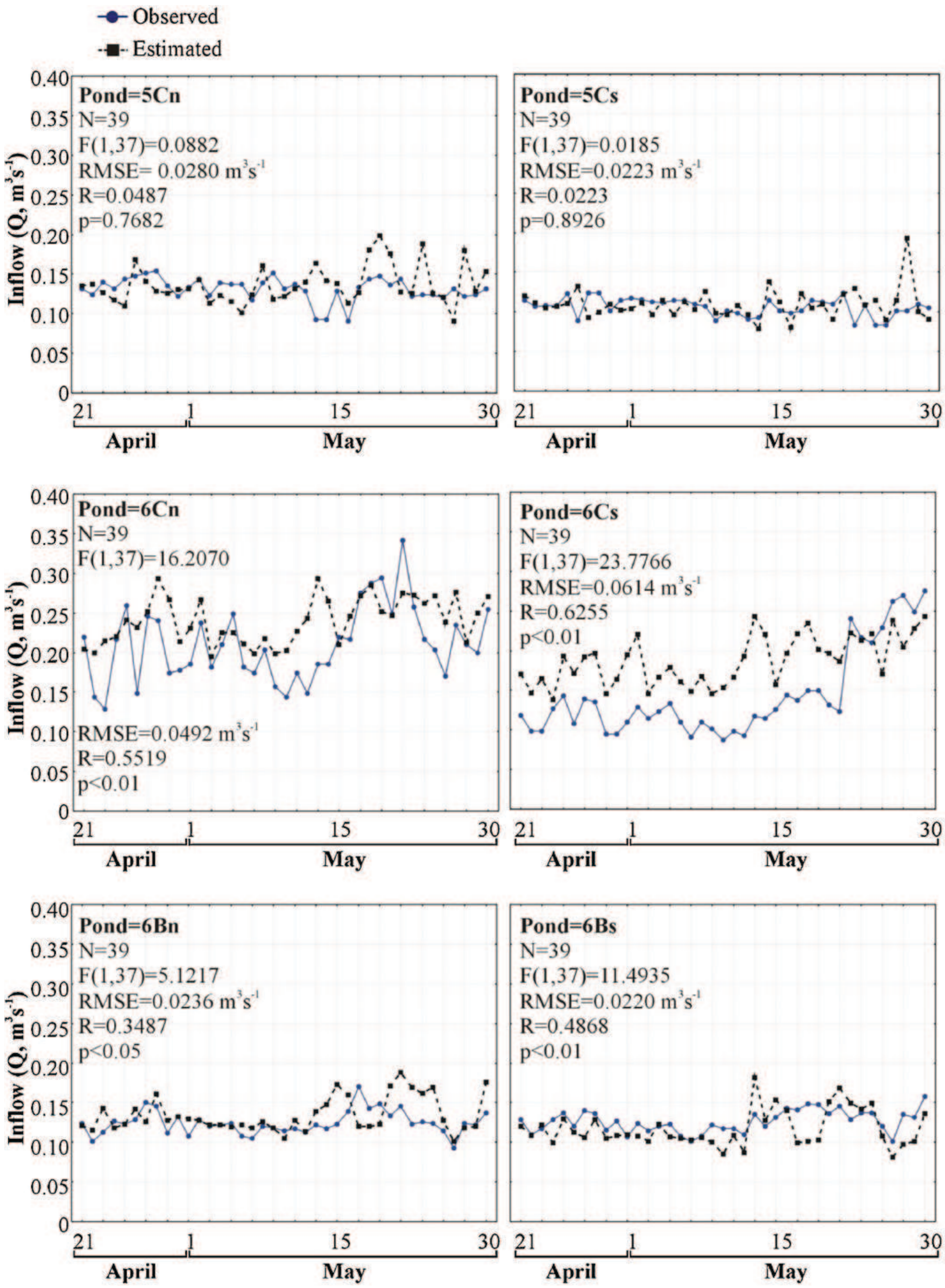

Fig. 7. Daily observed and estimated inflow to each pond for the best FLC.

4. Discussion and conclusions

Several different techniques were used for modeling the water exchange process in gilthead seabream (*S. aurata*) semi-intensive aquaculture systems in Southern Spain. In this context, heuristic approaches, namely Artificial Neural Networks (ANNs) and Fuzzy Logic Controllers (FLCs), were compared with more traditional methods, namely Multiple Linear Regressions (MLRs) and Generalized Additive Models (GAMs). Overall, it is possible to assert that the performance of the heuristic models was good and clearly better than that provided by classical approaches. These results were expected since similar conclusions have been obtained in other studies in which variables of a different nature were correlated or forecasted using classical and heuristic models (Gutiérrez-Estrada et al., 2004; Yu et al., 2006).

One of the critical steps in the development of control models is the selection of an appropriate set of input variables from the available candidates. This is because the performance of data-driven techniques is highly sensitive to the selected input variables. Accordingly, if relevant inputs are omitted the model is unable to capture the desired input-output relationships, and if the model includes superfluous inputs then the extraction of physical meaning from calibrated models can be more difficult and more data is necessary for calibration (Fernando et al., 2009). In our work,

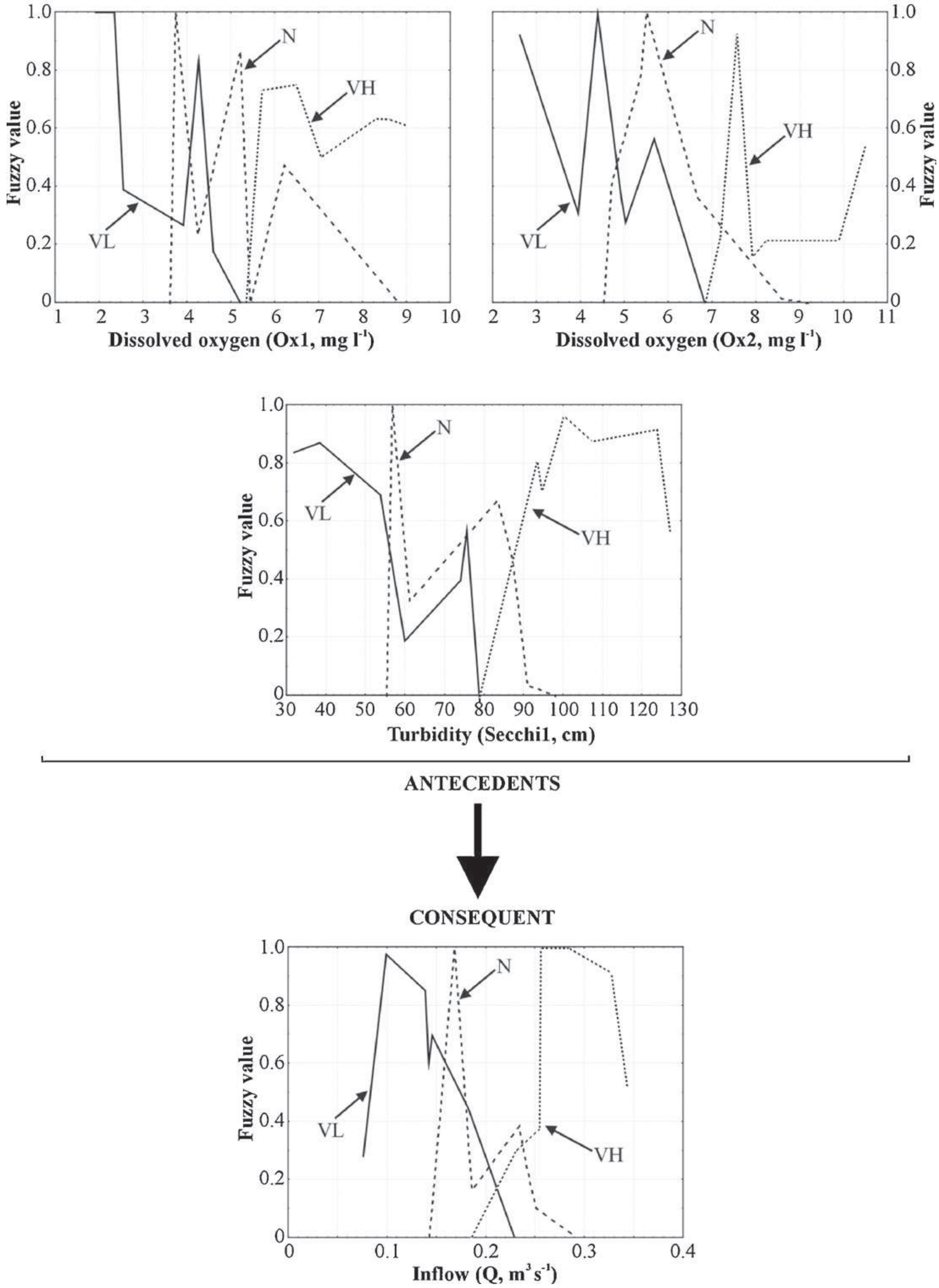

Fig. 8. Fuzzy partitions for each antecedent (Secchi1, Ox1 and Ox2) and the consequent (Q) obtained from the optimization process for the best FLC.

the results suggest that two variables, namely turbidity and oxygen concentration, have a significant weight on the control of the system which might indicate that the inflow management carried out by the aquaculturist is based mainly on the gradients of these two variables. In fact, personal comments of the responsible technicians of 'Langostinos de Huelva S.A.', confirmed us that these variables are the most important in inflow control for the workers.

In the case of dissolved oxygen, its selection seems to be evident because the effects of this factor upon fish response are well known (Poxton and Allouse, 1982) and very easy to identify (for example, superficial respiration and swimming). On the other hand, although there is a clear relationship between dissolved oxygen and water temperature, the selection of the dissolved oxygen by the models rather than the water temperature variable may be due to the fact that oxygen levels change more dramatically over time than temperature (Diana et al., 1997). Another consideration to keep in mind is that dissolved oxygen is a proxy of other factors not considered in this work, such as primary yield or nutrient concentration present in water.

Table 6
Fuzzy associative memory (FAM) of the best FLC. VL, very low; N, normal; VH, very high. The averaging process was carried out considering VL = 1, N = 2 and VH = 3.

Order		Antecedent			Consequent	
		Secchi1	Ox2	Ox1	Q	Output rule sets
1		VH	VL	VL	VL	1
2		VL	VL	N	VL	1
3		VL	VH	N	VL	1
4		VL	N	VH	VL	1
5		VH	N	VH	VL	1
	Average	1.80	1.80	2.20		
6		N	VL	VL	N	2
7		N	N	VL	N	2
8		VH	N	VL	N	2
9		N	VL	N	N	2
10		VH	VL	N	N	2
11		VL	N	N	N	2
12		N	N	N	N	2
13		VL	VL	VH	N	2
14		N	VL	VH	N	2
15		N	N	VH	N	2
16		VL	VH	VH	N	2
17		N	VH	VH	N	2
	Average	1.92	1.75	2.17		
18		VL	VL	VL	VH	3
19		VL	N	VL	VH	3
20		VH	N	N	VH	3
21		N	VH	N	VH	3
22		VH	VH	N	VH	3
23		VH	VL	VH	VH	3
24		VH	VH	VH	VH	3
	Average	2.29	2.14	2.00		

The weight of the turbidity variable is even easier to understand than the dissolved oxygen concentration since the turbidity is a physical property of water which is easily identifiable and qualitatively assessed by the aquaculturist. When the turbidity is high the aquaculturist trends to increase the inflow rates to the ponds because they know that the turbidity often limits production and growth of the fish (Yi et al., 2003). This trend was captured by the neural network and fuzzy models which denotes the capacity of these models to explain the underlying processes in the water exchange management. Therefore, the results reported in this paper suggest that the monitoring of the turbidity and dissolved oxygen variables could be sufficient to enable automatic control of gates to be efficient and, consequently, improve the cost-effectiveness of the management of ponds.

In spite of the overall performance of neural networks and fuzzy models being generally good, it can be inferred from the analysis of the accuracy measures in each pond that neither type of models captured the expert thinking in certain exceptional circumstances. In particular, the sources of the highest RMSE error in both models were ponds 6Cn and 6Cs, and the lowest explained variance was found in ponds 5Cn and 5Cs. This weakness could be solved by constructing a particular rule base in the case of the fuzzy logic controller and the calibration of a neural network for each pond. This solution has been proposed by Dubrovin et al. (2002) who constructed fuzzy models for real-time multipurpose reservoir operation. Nevertheless, this solution could make the implementation of the models for the gates control more complicated. On the other hand, it is necessary to bear in mind that a relatively small database from a single limited time period has been used which could explain the unexpected behavior of the models in some ponds.

In summary, the ANN and FLC models proved to be useful tools that, with relatively small data requirements, may be very suitable for the development of policies on water exchange management in aquaculture ponds which are key to the well-being of the species cultured and also to minimize operation costs in these aquaculture systems.

References

Anctil, F., Rat, A., 2005. Evaluation of neural network streamflow forecasting on 47 watersheds. J. Hydrol. Eng. 10 (1), 85–88.

ASCE Task Committee on Application of Artificial Neural Networks in Hydrology, 2000a. Artificial neural networks in hydrology. I. Preliminary concepts. J. Hydrol. Eng. 5 (2), 115–123.

ASCE Task Committee on Application of Artificial Neural Networks in Hydrology, 2000b. Artificial neural networks in hydrology. II. Hydrologic applications. J. Hydrol. Eng. 5 (2), 124–137.

Boyd, C.E., Tucker, C.S., 1998. Pond Aquaculture Water Quality Management. Kluwer Academic Publishers, MA, USA.

Chen, D.G., Hargreaves, N.B., Ware, D.M., Liu, Y., 2000. A fuzzy logic model with genetic algorithm for analyzing fish stock–recruitment relationship. Can. J. Fish. Aquat. Sci. 57, 1878–1887.

Diana, J.S., Szyper, J.P., Batterson, T.R., Boyd, C.E., Piedrahita, R.H., 1997. In: Egna, H.S., Boyd, C.E. (Eds.), Water Quality in Ponds. Dynamics of Pond Aquaculture, 3. CRC Press, Boca Raton, FL, USA, pp. 53–71.

Dubrovin, T., Jolma, A., Turunen, E., 2002. Fuzzy model for real-time reservoir operation. J. Water Resour. Plan. Manage. 128 (1), 66–73.

Fernando, T.M.K.G., Maier, H.R., Dandy, G.C., 2009. Selection of input variables for data driven models: an average shifted histogram partial mutual information estimator approach. J. Hydrol. 367, 165–176.

Goldberg, D., 1989. Genetic Algorithms in Search, Optimization and Machine Learning. Addison-Wesley, Reading, MA, USA.

Gutiérrez-Estrada, J.C., de Pedro-Sanz, E., López-Luque, R., Pulido-Calvo, I., 2004. Comparison between traditional methods and artificial neural networks for ammonia concentration forescasting in an eel (*Anguilla anguilla* L.) intensive rearing system. Aquacult. Eng. 31, 183–203.

Gutiérrez-Estrada, J.C., Silva, C., Yáñez, E., Rodríguez, N., Pulido-Calvo, I., 2007. Monthly catch forecasting of anchovy *Engraulis ringens* in the north area of Chile: non-linear univariate approach. Fish. Res. 86, 188–200.

Gutiérrez-Estrada, J.C., Vasconcelos, R., Costa, M.J., 2008. Estimating fish community diversity from environmental features in the Tagus estuary (Portugal): multiple linear regression and artificial neural network approaches. J. Appl. Ichthyol. 24, 150–162.

Gutiérrez-Estrada, J.C., Yáñez, E., Pulido-Calvo, I., Silva, C., Plaza, F., Bórquez, C., 2009. Pacific sardine (*Sardinops sagax*, Jenyns 1842) landings prediction: a neural network ecosystemic approach. Fish. Res. 100, 116–125.

Hastie, T.J., Tibshirani, R.J., 1990. Generalized Additive Models. Chapman & Hall/CRC, Boca Raton, FL, USA.

Holland, J., 1975. Adaptation in Natural and Artificial Systems. MIT Press, Cambridge, MA, USA.

Hunter, A., Kennedy, L., Henry, J., Ferguson, I., 2000. Application of neural networks and sensitivity analysis to improved prediction of trauma survival. Comput. Methods Programs Biomed. 62, 11–19.

Hsu, K., Gupta, H.V., Sorooshian, S., 1995. Artificial neural network modeling of the rainfall-runoff process. Water Resour. Res. 31 (10), 2517–2530.

Iyer, M.S., Rhinehart, R.R., 1999. A method to determine the required number of neural-network training repetitions. IEEE Trans. Neural Netw. 10 (2), 427–432.

Jacob, C., 2001. Illustrating Evolutionary Computation with Mathematica. Morgan Kaufmann Publishers, San Francisco, CA, USA.

Jang, J.S.R., Sun, C.T., Mizutani, E., 1997. Neuro-Fuzzy and Soft Computing. Prentice Hall, NJ, USA.

Kelly, A.M., Kohler, C.C., 1997. Climate, site, and pond design. In: Egna, H.S., Boyd, C.E. (Eds.), Dynamics of Pond Aquaculture, 5. CRC Press, Boca Raton, FL, USA, pp. 109–133.

Kosko, B., 1997. Fuzzy Engineering. Prentice Hall, London, UK.

Lee, P.G., Lea, R.N., Dohmann, E., Prebilsky, W., Turk, P.E., Ying, H., Whitson, J.L., 2000. Denitrification in aquaculture systems: an example of a fuzzy logic control problem. Aquacult. Eng. 23, 37–59.

Legates, D.R., McCabe Jr., G.J., 1999. Evaluating the use of 'goodness-of-fit' measures in hydrologic and hydroclimatic model validation. Water Resour. Res. 35 (1), 233–241.

Lekang, O.I., 2007. Aquaculture Engineering. Blackwell Publishing, Oxford, UK.

Li, Y., Jiao, Y., He, Q., 2011. Decreasing uncertainty in catch rate analyses using Delta-AdaBoost: An alternative approach in catch and bycatch analyses with high percentage of zeros. Fish. Res. 107, 261–271.

Makkeasorn, A., Chang, N.B., Zhou, X., 2008. Short-term streamflow forecasting with global climate change implications—a comparative study between genetic programming and neural network models. J. Hydrol. 352, 336–354.

Munson, B.R., Young, D.F., Okiishi, T.H., 2006. Fundamentals of Fluid Mechanics. John Wiley & Sons, USA.

Murase, H., Nagashima, H., Yonezaki, S., Matsukura, R., Kitakado, T., 2009. Application of a generalized additive model (GAM) to reveal relationships between environmental factors and distributions of pelagic fish and krill: a case study in Sendai Bay, Japan. ICES J. Mar. Sci. 66, 1417–1424.

Ochoa-Rivera, J.C., Andreu, J., García-Bartual, R., 2007. Influence of inflows modeling on management simulation of water resources system. J. Water Resour. Plan. Manage. 133 (2), 106–115.

Papandroulakis, N., Markakis, G., Divanach, P., Kentouri, M., 2000. Feeding requirements of sea bream (*Sparus aurata*) larvae under intensive rearing

conditions—development of a fuzzy logic controller for feeding. Aquacult. Eng. 21 (4), 285–299.
Poxton, M.G., Allouse, S.B., 1982. Water quality criteria for marine fisheries. Aquacult. Eng. 1, 153–191.
Pulido-Calvo, I., Gutiérrez-Estrada, J.C., 2009. Improved irrigation water demand forecasting using a soft-computing hybrid model. Biosyst. Eng. 102, 202–218.
Pulido-Calvo, I., Portela, M.M., 2007. Application of neural approaches to one-step daily flow forecasting in Portuguese watersheds. J. Hydrol. 332, 1–15.
Ruiz-Velazco, J.M.J., Hernández-Llamas, A., Gómez-Muñoz, V.M., 2010. Management of stocking density, pond size, starting time of aeration, and duration of cultivation for intensive commercial production of shrimp *Litopenaeus vannamei*. Aquacult. Eng. 43, 114–119.
Rumelhart, D.E., Hinton, G.E., Williams, R.J., 1986. 'Learning' representations by back-propagation errors. Nature 323, 533–536.
Shrestha, R.R., Theobald, S., Nestmann, F., 2005. Simulation of flood flow in a river system using artificial neural networks. Hydrol. Earth Sits. Sci. 9 (4), 313–321.
Soto-Zarazúa, G.M., Penuche-Vera, R., Rico-García, E., Toledano-Ayala, M., Ocampo-Velázquez, R., Herrera-Ruiz, G., 2011. An automated recirculation aquaculture system based on fuzzy logic control for aquaculture production of tilapia (*Oreochromis niloticus*). Aquacult. Int. 19, 797–808.
Steeby, J.A., Hargreaves, J.A., Tucker, C.S., Cathcart, T.P., 2004. Modeling industry-wide sediment oxygen demand and estimation of the contribution of sediment to total respiration in commercial channel catfish ponds. Aquacult. Eng. 31, 247–262.
Tan, Y., van Cauwenberghe, A., 1999. Neural-network-based d-step-ahead predictors for nonlinear systems with time delay. Eng. Appl. Artif. Intel. 12 (1), 21–25.
Yi, Y., Lin, C.D., Diana, J.S., 2003. Techniques to mitigate clay turbidity problems in fertilized earthen fish ponds. Aquacult. Eng. 27, 39–51.
Yu, R., Leung, P., Bienfang, P., 2006. Predicting shrimp growth: artificial neural network versus nonlinear regression models. Aquacult. Eng. 34, 26–32.
Zeldis, D., Prescott, S., 2000. Fish disease diagnosis program—problems and some solutions. Aquacult. Eng. 23, 3–11.