Nonlinear Modelling of Water Quality in Shrimp Ponds based on Artificial Neural Networks

José Juan Carbajal[1], Luís Pastor Sánchez [1], Jorge Hernández[2]

[1]Centro de Investigación en Computación, Instituto Politécnico Nacional,
Av. Juan de Dios Batiz s/n, Colonia Nueva Industrial Vallejo, México D.F.
juancarvajal@sagitario.cic.ipn.mx, lsanchez@cic.ipn.mx
[2]Laboratorio de Análisis Integral Acuícola, Centro de Investigaciones Biológicas del Noreste,
Centenario Nte. 53, P. del Centenario, Hermosillo, Sonora, México
Jhlopez04@cibnor.mx

Abstract. Water quality is an important factor for a good maturing and reproduction of the organisms in a water environment. A model that describes its behaviour in shrimp ponds is used to determine good or bad situations that could generate a low growing or lethal problems. This work proposes a non-linear model of water quality using various types of architectures of artificial neural networks (ANN) for recognizing the status into the pond. ANN's are used to establish a relationship among environmental variables that affect the shrimp habitat. The results show a good performance of our approach, giving an alternative tool for the aquaculture field.

Keywords: Water quality, neural networks, aquaculture.

1 Introduction

The water quality in oceanic research is a problem that affects daily the activities of many people that practice fishing activities. The quantity of biological data is increasing each day and it is needed to create models and use some functional features.

The shrimp farming is an important activity in many countries because their economy, that is why there exists many companies that practice it. The farming shrimp is made in different ways; in intensive, semi intensive or extensive ponds. The shrimp production is determined by two main factors: 1) the capacity of maturing in the organisms and 2) the capacity of the environment. The capacity of the environment is referred to the conditions that allow a growing and reproduction, whose would be the best in good environment conditions. The water quality into the farming ponds determines the capacity of the environment that makes influence in the life of the organism [1], [2].

Water quality indicators have been grouped in three categories: physical, chemical and biological, each of them containing a significant number of water quality variables. There are variables that have more significance because they can affect the growing and the surviving of the organism, for example: temperature, salinity, dissolved oxygen, pH, Alkalinity, ammonia and nitrites. The values and the combinations of these variables, allow creating normal patterns. The Table 1 shows the range of the environmental

© L. Sánchez, O. Pogrebnyak and E. Rubio (Eds.)
Industrial Informatics
Research in Computing Science 31, 2007, pp. 233-242

variables and its classifications, these are the most common ranges used by the biologists in the Gulf of California (Páez, O. 2001).

Table 1. Classification of the water quality status using the combination of environmental variables.

GOOD	
Temperature	23 – 30 °C
Salinity	15 – 25 (mg/l)
PH	7.6 – 8.6
Dissolved oxygen	6 – 10 (mg/l)
Turbidity	35 – 45 (mg/l)
RISK	
Salinity	Greater than 25 (mg/l)
PH	Less than 7.6 or greater than 8.6
Dissolved oxygen	Less than 6 or greater than 10 (mg/l)
Turbidity	Less than 35 or greater than 45 (mg/l)

In general, the environmental variables have nonlinear relations, which have been observed and proven experimentally. The equations that represent them, have been formulated, which is very hard to do, for example in the Gulf of California the pH in seawater is a function of total CO_2 (Ct), total alkalinity (At) and pressure (P):

$$\partial \mathrm{pH} = \left(\frac{\partial \mathrm{pH}}{\partial \mathrm{Ct}}\right)\partial \mathrm{Ct} + \left(\frac{\partial \mathrm{pH}}{\partial \mathrm{At}}\right)\partial \mathrm{At} + \left(\frac{\partial \mathrm{pH}}{\partial \mathrm{P}}\right)\partial \mathrm{P} \tag{1}$$

Where the quantities in brackets are the partial derivatives of pH with respect Ct, At and P, respectively [2], [3], [4], [5].

We can represent these nonlinear relations following the equation 2:

$$\text{Habitat status (water quality)} = f(\text{Temp, Sal, Turb, At, Do} \ldots)\,. \tag{2}$$

Where:

f: nonlinear function; *Temp*: temperature; *Sal*: salinity; *Turb*: turbidity and *Do*: dissolved oxygen.

The need for appropriated techniques to manage the importance of water quality variables, the interpretation and the method used to integrate those variables involved in the evaluation is recognized (Fig. 1). In this sense some methodologies have been developed from the artificial intelligence. These methodologies have been based mainly on artificial neural networks (ANN), fuzzy logic and mathematical reasoning. The main problem with the mathematical reasoning is that an equation that describes the behavior of the water quality is very hard to develop, and some techniques like ANN are a great solution [6], [7].

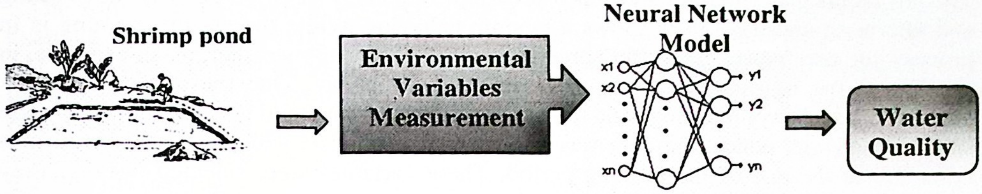


Fig. 1. The measurements of the environment will be analyzed by a model that describes the relationship among them, obtaining a status of water quality.

The objective of this study was to develop a model that describes a relationship between the environmental variables through the use of non-linear system identification techniques. Some variables are measured more frequently than other because they have more impact in the status of the habitat; these variables were taken as an input to the system to asses the water quality (temperature, salinity, pH, dissolved oxygen and turbidity). Since there are not methodologies for water quality with ocean water in shrimp ponds, this work proposes the use of a neural network for recognizing problems in the ponds related with the environment.

2 Methods

2.1 Data Collection

The shrimp ponds are systems where the environment must to be in equilibrium, a controlled habitat generates a lower stress level, high growing and high resistance for sickness; for example, for shrimp maturation and spawning [8], [9] almost all hatcheries require availability of oceanic-quality water on a 24-h basis. In the habitat, there exist environmental variables that make a bigger degradation of the water quality than others salinity and temperature are important water parameters impacting production of shrimp in the hatchery, and must be maintained in a narrow range, between 27 and 36 ppt salinity and 28 C (82 F) plus or minus two degrees for most Shrimps penaeids. According to information obtained from the shrimp ponds located in Sonora, Mexico, the environmental variables with a bigger impact are Temperature, PH, Salinity, Dissolved Oxygen and Turbidity [10], [11]. A bad control of this set of variables will generate chemical reactions that affect other variables. Although there are a big set of variables that affects the water quality, they no generate a bigger impact as the first, in this work the data collecting is based in those environmental variables.

Data collection was made on the farming shrimp ponds from Sonora, Mexico. Almost 400 measurements were obtained in a farming period using manual techniques for example, the DO, pH, salinity and temperature were measured with electronic sensors by an operator and turbidity was measured with laboratories instruments. For each pond, the environmental variables measured were pH, temperature, salinity, turbidity and dissolved oxygen.

The environmental variables were measured in a frequency of twice per day: at morning and afternoon and the values shows a random behavior. While the shrimp farming is in process, the parameters can suffer some variations because the weather, as an example, in the Fig. 3 the oxygen has a big variation during the day. Due these variations this parameter are always in change and the combination of them could generate a risk or good situation into the pond. The data was collected by biologists and were transferred to a computer at the end of the farming period. These variables were selected because they produce the biggest impact and they could generate conditions not appropriated for the good maturing and reproduction of the organisms. The nature of this parameter and their relationship with the water quality make them an excellent set for their processing with an ANN.

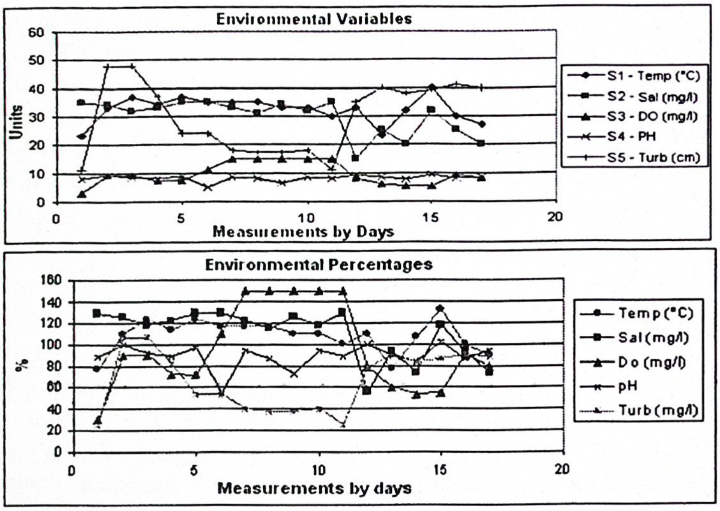


Fig. 2. Measurements and percentages in a month of the environmental variables into a shrimp pond in a farming , S1- Temperature, S2-Salinity, S3-DO, S4-pH and S5-Turbidity .

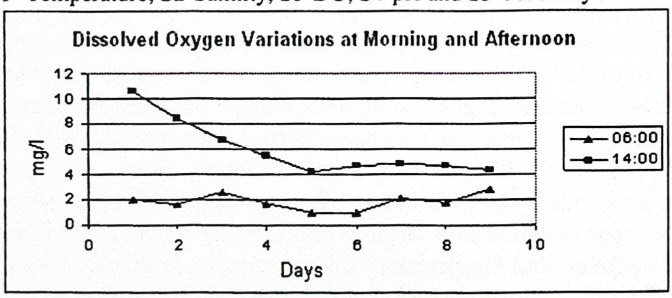


Fig. 3. The measured values of DO during a day show some variations at morning and afternoon, for example in the day one, there is a variation of 8 units, and DO with 2 mg/l generate hypoxia to the organisms.

The water quality status is obtained using a vector of the environmental variables, for example in the day one (Fig. 2) we have the vector v[Temp, Sal, Do, Turb]=[23, 35, 3, 8, 11], this vector is used as an input of the ANN for obtaining the status in the moment of the measurement. In new researches we are using a matrix of values per week as an input of the neural network.

2.2 Pre-processing

Neural networks training can be made more efficient if certain preprocessing steps are made to the data collected. The preprocessing was applied to each of the input variables (pH, temperature, dissolved oxygen, salinity) before were used for the ANN, this way allows transforming the data [12], [13]. All variables has a different range of values, for example, the temperature is given in Celsius (25°, 30°, 45°), the pH is given in a lower range for 6-9 units, and the oxygen is given from 2 to 10 ml/g. It is not convenient to use the data without a pre-processing, because the data with bigger values will have more impact in the result [7], [14]. For standardization, the data were transformed to achieve zero man and unity standard deviation following the equation (3):

$$X_1 = \frac{X - \overline{X}}{\sigma} \tag{3}$$

All variables were scaled to [0, 1] range according to:

$$\text{var} = \frac{x - \min(x)}{\max(x) - \min(x)} \tag{4}$$

Where x is the environment variable and var is the new normalized variable.

The universe of data information that involves all cases of water quality (good and risk) is extremely huge as it is shown in Table 2, but only a smaller set is measured in real time, in order to have a complete set, an artificial test patterns was made, additionally the measured patterns was added for creating a new complete set, more than 1000 patterns were used for training an ANN.

Table 2. Number of patterns used for defining only the set of a good water quality.

Variable	Temp °C	Sal (mg/l)	DO (mg/l)	PH	Turb (mg/l)
Range	23 – 30	15 – 25	6 – 10	7.6 – 8.6	35 – 45
Minimal value measured	.5	1	.1	.1	1
Total measurements per Variable	14	10	40	10	10
Total measurement per all variables	**560000**				

The patterns classification is made with an ANN to recognize and to quantify two statuses within the pool:

a) **Good**: Everything is well, the parameters are on range or some of them are out or range but it do not generate any problem, for example when temperature is one or two grades below the range.

b) **Risk**: One, some or all environmental variables are out of range, which generates problems into the pond.

2.3 Neural Network Model

Neural networks have been used as a powerful tool in the analysis of complex and uncertain data, like non-linear models. A neural network is a system of interconnected processing elements (neurons), based on the structure of the brain. The networks are constructed by nodes connected together by parameters called weights. The units in an ANN are usually accommodated in layers [7], [15].

The artificial neural network theory was applied in this study providing a non-linear relationship between input sets (Environmental variables) and output targets (water quality status). The environmental variables have dependences among them, and trying to represent them with a mathematical equation is a very hard task [6]. There are other methods for classifying patterns, in this study the classification of the status is given by a ANN, another way is the regression linear method, but it can not be applied in this work because its goal is to find parameters of the best linear approximation to the input and the desired response, and the goal of a neural network is to separate the data as well as possible into classes. Other feature of an ANN is its power to generalize, which means when a value that never was learned by the ANN, the network can give an approximation about a desired response. In this work the ANN recognize the status and quantify it, in others words, how well or risky it is, which allows finding trend of the water quality [14], [7].

In this study we use a Multilayer Perceptrons (MLPs). The number of input is given by the number of environmental variables: pH, temperature, salinity, dissolved oxygen and turbidity. The number of outputs is given by the number of status of the water quality: good and risk, and therefore the size of this layer depend on the data representation. The topology used in this work was a feed forward network; it is shown in Fig. 4.

There are many learning algorithms for the ANN; the gradient descent algorithm is very slow because it requires small learning rates for stable learning, the momentum algorithm is faster but it is still too slow for many practical applications. This study uses the Levenberg-Marquardt training algorithm because it gives a better performance for small and medium size networks with enough memory available. If memory is a problem, then there are a variety of other fast algorithms available like Resilient Backpropagation algorithm or the Scaled Conjugate Gradient [7], [16].

3 Results

The ANN was tested with different sets of measurements. Two topologies were used for this work, one of them was a 5-4-1 layered ANN, the output of this ANN responds with a 1 if water quality is good and a -1 if water quality status is bad (risk). The second network is a 5-4-2 layered ANN's, and the output is given by two neurons, whose the first represents the good status when its output is 1 and the output of the second is -1; and the second neuron represents the risk status when its output is 1 and the output of the first is -1, when one neuron has a value of 1 the other has -1 and viceversa.

The training of the MPL's was carried out with the software package MATLAB. The patterns classification given by the environmental variables confirms its nonlinear interrelation. The trained feed forward neural network represents suitably the nonlinear relations between the environmental variables. The good pattern classification of the ANN is shown in the Fig 6.

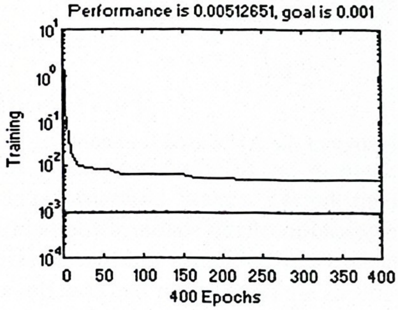


Fig. 4. Performance of the ANN (mse) and epochs obtained on the training.

There were some cautions in the Training of the ANN, when a learning rate is too large, the learning could be unstable or memorize too much and do not generalize. In other hand, a learning rate too small could not memorize. The performance of the training is showed n the Fig. 4, where the curve shows that the maximum performance obtained at 400 epochs, without an overfitting [7].

ANN's are sensitive to the number of neurons in their hidden layers. Too few neurons can lead to underfitting. Too many neurons can contribute to overfitting, in which all training points are well fit, but the fitting curve takes wild oscillations between these points. We used many topologies, training ANN with one, two and three hidden layers and with different number of neurons; however these topologies did not show a better performance, also the ANN was trained with different activations functions as tangential, logarithmic, exponential and lineal functions, finding a better performance with a tangential activation function in the first and the hidden layer, and a linear activation function in the output layer [7], [16].

The 5-4-1 ANN showed a good performance with more epochs for training but the response was not effectively, thus the 5-4-2 layered ANN showed the best performance with low epochs obtaining the best classification pattern. One computational system was designed using this ANN. Fig. 5 presents the topology with better results for this application, and the operation of output neurons with linear activation function in the output neurons.

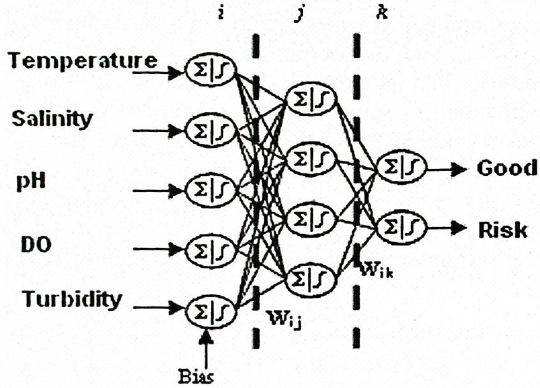
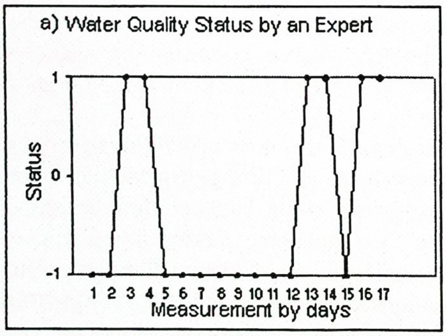


Fig. 5. Topology of the ANN used for analyzing water quality.

Some patterns are given by an expert biologist, such patterns build a set of measurements with its classification of the status of the habitat, in order to compare the classification of the ANN, the figure 6 shows the two classifications showing a similar behavior, however some patterns are not good or bad exactly, and this is because the ANN gives a response of patterns that were not included into the training set (generalize) and this result is closer to a similar patter that exist in the training set.



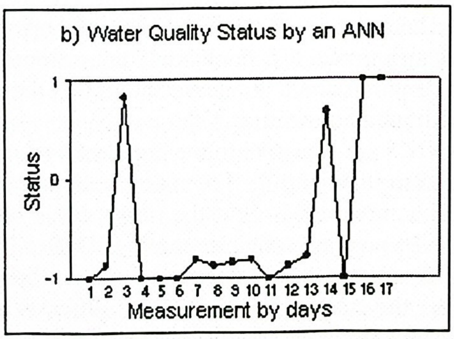


Fig. 6 a) water quality status defined by an expert, when status is 1, means good, and when is -1 means bad water quality (risk); and b) water quality status defined by the ANN., the classification gives a deteriorated status, for example when the variables are not exactly in the normal range but they do not represent a bad status the classification is not entirely 1 or -1.

3.1 Post-processing

Post-processing steps are needed for measuring the performance of the trained ANN. We used a regression analysis between the response and the targets, using the linear equation $y = mx + b$ [7], [16].

For the first neuron (good) the values obtained were: $m = 0.9548$, $b = 0.0077$, $r = 0.9771$ and, for the second neuron (risk) the values obtained were: $m = 0.9675$, $b = 0.0181$ and $r=0.9836$. Where m corresponds to the slope, b the intercept with the y axis and r is the correlation coefficient between outputs and targets. A perfect fit is obtained when the slope is 1, the y-intercept is 0 and the correlation coefficient is 1. We can see that the numbers are very close.

The validation of the ANN was made using experimental data information taken over a three-month period of farming shrimp and using some experimental ponds in laboratories, demonstrating the reliability of the network. The results demonstrated that the ANN model trends very well when all the environmental variables are measured in real time or when a set of information is analyzed.

4 Conclusions and Future Work

In this paper, we present a decision model in the form of neural network for water management in shrimp ponds located in Sonora, Mexico. A mathematical relationship between the environmental variables is too hard to develop. A big research about the shrimp habitat was made for determining what kind of environmental variables affect the water quality into the farming ponds.

An advantage for using the ANN's as classifiers is their sensibility to recognize patterns that were not used in the training set, that is called generalize. A disadvantage is when the classification depends of amount of patterns in the training set, and if this set is incomplete the ANN will not recognize a status in the right way.

According with the results shown in Fig. 6, the ANN shows a good classification of the status of water quality; also it gives a degradation of the status depending of the conditions into the pond.

From the results obtained, we can conclude that our model is effective for resolution of problems related with water quality into shrimp farming ponds; also it is a powerful tool for the analysis that are practiced all days by biologists that are searching a better way for farming shrimp.

Based on tour experimental results, the main advantage of our model over other evaluating methods are: speed on the assessment of water quality and, its portability and adaptability to others aquaculture systems.

As future work, we are researching in other way to obtaining more effectively a status of the water quality. Also we are working in a way for predicting the water quality status into the pond using ANN's and fuzzy systems during a farming period, working with the best model developed to assess the water quality in the present. The model for evaluating and for predicting the water quality will be integrated to a computational system that will give a complete analysis for the ponds.

References

1. Casillas, R., Magallón, F., Portillo, G., Osuna, P.: Nutrient Mass Balances in Semi-intensive Shrimp Ponds from Sonora, Mexico Using Two Feeding Strategies: Trays and Mechanical Dispersal. Aquaculture, Vol. 258. Elservier (2006) 289-298.
2. Páez O, F.: Camaronicultura y Medio Ambiente. Instituto de Ciencias del mar y Limnología. UNAM. pp. 271-298, México, 2001.
3. Angulo Corrales, Adán y Angulo Corrales, Uriel, "Estudio de Calidad del Agua y su Relación con el Crecimiento del Camarón Blanco (Litopenaeus Vannamei), en la Granja Camaronera Agua Verde, S.A. de C.V. en Rosario, Sin.", Tesis, Universidad Autónoma de Sinaloa, 2003.
4. Martínez Córdova Rafael M., "Cultivo de Camarones Pendidos, *Principios y Practicas*", Ed. AGT Editor S.A., 1994.
5. Hernández, J., Zirino, A., Marione, S., Canino, R., Galindo, M.: PH-density relationships in seawater. Ciencias Marinas, Vol. 29 (2003) 597-508.
6. Ocampo, W., Ferré, N., Domingo, J., Schuhmacher, M.: Assessing water quality in rivers with fuzzy inference systems: A case study. Environment International, Vol 32. Elservier (2006) 733-742.
7. Romero, C., Shan, J.: Development of an artificial neural network-based software for prediction of power plant canal water discharge temperature. Expert Systems with Application, Vol 29. Elservier (2005) 831-838.
8. Granvil D. Treece, Shrimp Maturation And Spawning, Texas A&M University, Sea Grant College Program. UJNR Technical Report No. 28, 2001
9. Granvil D. Treece, Shrimp Maturation and Spawning. In Proceedings of the 28th US-Japan Natural Resources Aquaculture Panel. UJNR Technical Report No. 28. pp 121-134. TAMU-SG-01-814. 2001
10. Li, Y., Li, J., Wang, Q.: The effects of dissolved Oxygen Concentration and Stocking Density on Grown and Non-Specific Immunity in Chinese Shrimp, Fenneropenaeus Chinensis. Aquaculture, Vol. 256. Elservier (2006) 608-616.
11. Wahab, A., Bergheim, A., Braaten, B.: Water Quality and Partial Mass Budget in Extensive Shrimp Ponds in Bangladesh. Aquaculture, Vol. 218. Elservier (2006) 413-423.
12. Campos, M., López, D., "Neural Network Approach to Locate Motifs in Biosequences", Progress in Pattern Recognition, Image Analysis and Applications, Springer (2005), LNCS 3773, 214-221.
13. Lek, S., Guiresse, M., Giraudel, J.: Predicting stream nitrogen concentration from watershed features using neural networks. Water Resources, Vol. 33. Pregamon (1999) 3469-3478.
14. Principe, José, Luliano, Neil y Lefebvre W Curt, 2000, Neural and Adaptative systems: *Fundamentals through Simulations,* 1ª Edition, John Wiley & Sons.
15. Sharma, V., Negi, S., Rudra, R., Yang, S.: Neural network for predicting nitrate-nitrogen in drainage water. Agricultural Water Management, Vol 63. Elservier (2003) 169-183.
16. Chacón, M., Blanco, C., Panerai, R. and Evans, D., "Non linear Modelling of Dynamic Cerebral Autoregulation Using Recurrent Neural Networks", Progress in Pattern Recognition, Image Analysis and Applications, Springer (2005), LNCS 3773 205-213.