

Signal Analysis for Assessment and Prediction of the Artificial Habitat in Shrimp Aquaculture

José Juan Carbajal Hernández, Luis Pastor Sanchez Fernandez,
José Luis Oropeza Rodríguez, and Edgardo Manuel Felipe Riverón

Centre of Computer Research – National Polytechnic Institute, Av. Juan de
Dios Bátiz, Col. Nueva. Industrial Vallejo, México D.F., México
juancarvajal@sagitario.cic.ipn.mx,
{lsanchez, joropeza, edgardo}@cic.ipn.mx

Abstract. This paper presents a novel work for prediction of artificial habitat in shrimp aquaculture based on environmental signal analysis. The physical-chemical variables that are involved into the system are studied for modeling and predicting environmental patterns. The prediction model is built using AR models that reconstruct a partial section of a particular measured signal. The physical-chemical variables are classified based on the negative ecological impact using a new statistical model that calculates the frequency and the deviation of the measurements. A fuzzy inference system processes the level classifications using aquaculture rules that define all the cases calculating the condition of the shrimp habitat.

Keywords: fuzzy inference systems, prediction, signal analysis, Assessment.

1 Introduction

The main purpose on water management and aquaculture systems is to control and maintain the optimal conditions for the surviving and growing of the organisms in normal farming conditions [1]. The early detection of potential problems can be decisive in the organism health and the economical activity of the farm. The negative impact of a set of physical-chemical variables can be assessed when they occur, but estimate the future condition of the ecosystem can be a difficult task since they are not tools for solving this problem. In other hand, if the concentrations levels are predicted, potential danger situations could be avoided before they appear [2]. There is a lack of methodologies for prediction and assessment of water quality; the methods actually developed have several weaknesses were the lack of a reasoning process in the assessment of the information is the main problem [3], [4], [5], in addition, the prediction process usually is confused as a present condition to be predominant for the rest of the day [6].

2 Data Collection

A set of physical-chemical variables compounds the ecosystem of the shrimp; this set must be under control and in optimal ranges. As a result of this condition, the features

of the variables are studied with the objective to determine the frequency and importance of their behaviors [2].

The measurements of the variables depend of the exactitude of how a supervisor monitors the aquaculture system. A complete farm was monitored in Rancho Chapo located in Sonora, Mexico. The higher impact variables measured were temperature, dissolved oxygen, salinity and pH, using a sensor device for each variable. The period of monitoring was of 15 minutes. The data set contains four months of measurements; it means a register of 9312 values per variable. The classification levels of the physical-chemical variables (status) are defined in Table 1, for dissolved oxygen we chosen “hypoxia”, “low” and “normal”, for the temperature and salinity variables we chosen “low”, “normal” and “high”, and for the pH variable we chosen “acid”, “low”, “normal”, “high”, and “alkaline”.

Table 1. Classification levels, tolerances (Tol) and limits (Lim) of physical-chemical variables

Variables	Hypoxia Acid	Low	Normal	High	Alkaline	Tol.	Lim.
Temp (°C)	-----	0 - 23	23 - 30	30 - ∞	-----	±1	±1
Sal (mg/L)	-----	0 - 15	15 - 25	25 - ∞	-----	±1	±1
DO (mg/L)	0 - 3	3 - 6	6 - 10	10 - ∞	-----	±0.5	±0.5
PH	0 - 4	4 - 7	7 - 9	9 - ∞	10 - 11	±0.5	±0.5

3 Series Prediction

3.1 Preprocessing

Smoothing

The variables signals have several peaks values, this behavior can be generated due a failed device, human error or environmental situations. The four signals of the physical-chemical variables are smoothed in order to be more easily for processing. A moving average weighted filter works using an average of signal points (measured concentrations) for producing new output points of the new filtered signal and smoothing it [7]. The smoothing process of the physical-chemical variables can be calculated as follows:

$$y(n) = \sum_{i=0}^N b_i x(n - i) \tag{1}$$

where $x(n)$ is the original signal, $y(n)$ is the new output signal, N is known as the filter order, b_i are the Spencer 15 terms coefficients defined as $\frac{1}{320} [-3, -6, -5, 3, 21, 46, 67, 74, 67, 46, 21, 3, -5, -6, -3]$. The smoothing process using a moving average weighted filter is:

$$y(n) = -\frac{3}{320}x(n) - \frac{6}{320}x(n - 1) - \frac{5}{320}x(n - 2) + \dots - \frac{3}{320}x(n - 14) \tag{2}$$

The Fig. 1 shows examples of the original and smoothed measured variables, where the random perturbations are suppressed.

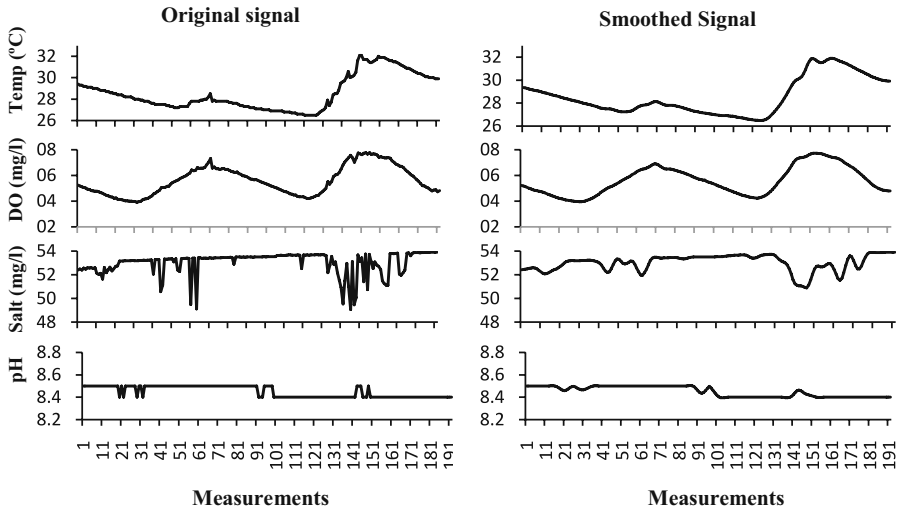


Fig. 1. Original and smoothed signal of the physical-chemical variables using a moving average filter

Detrending

The environmental series usually contain some constant amplitude offset components or trends. The amplitudes of these trends sometimes corrupt the results of series modeling. Therefore, it is needed to remove them before performing further analysis [8]. The trend is calculated using the linear regression method, where the equation of the estimated trend for a particular variable can be expressed as follows:

$$y = a_0 + a_1x + E \quad (3)$$

where a_0 a_1 are coefficients that represent the intersection with the abscise axis and the pendent respectively, y is the physical-chemical variable (temp, salt, DO and pH) and E is the error between the modeled and the observed values. The coefficient a_1 can be calculated using:

$$a_1 = \frac{n \sum x_i y_i - \sum x_i \sum y_i}{n \sum x_i^2 - (\sum x_i)^2} \quad (4)$$

where n is the number of points of the series, and x_i is the i^{th} measurement. For the a_0 coefficient:

$$a_0 = \bar{y} - a_1 \bar{x} \quad (5)$$

3.2 Autoregressive Model (AR)

The AR model of a series allows predicting the current value x_t , based on past values $x_{t-1}, x_{t-2}, \dots, x_{t-n}$ and a prediction error. The n parameter determines the number of past values that are used for predicting the current value (model order). The model order can be estimated using an error estimator, it is known as Akaike criterion [9]:

$$AIC = E_n \left(1 + \frac{2n}{L} \right) \tag{6}$$

where L is the number of points in the time series, n is the model order and E_n is the prediction error. The AR models that describe the physical-chemical variables using the model order estimated with the AIC are:

$$temp_t = \sum_{i=1}^{79} a_i temp_{t-i} + e_t \tag{7} \quad DO_t = \sum_{i=1}^{80} a_i DO_{t-i} + e_t \tag{8}$$

$$Salt_t = \sum_{i=1}^{20} a_i Salt_{t-i} + e_t \tag{9} \quad pH_t = \sum_{i=1}^{56} a_i pH_{t-i} + e_t \tag{10}$$

where a_i is the i^{th} AR coefficient, e_t is the predicted error and p is the model order.

The Fig. 2 shows the reconstruction of the physical-chemical signals, where a total of 96 points (one day/24 hours) were predicted using the AR(p) model suggested.

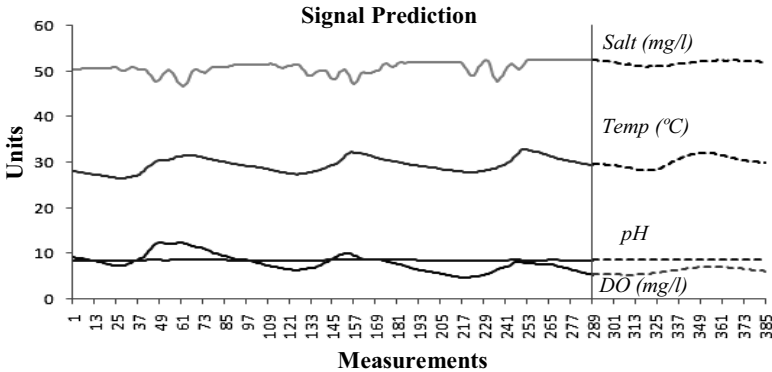


Fig. 2. Prediction of the physical-chemical variables. The AR model predicts 96 measurements (24 hours).

4 Assessment

Historical Water Quality Index

The **Historical Water Quality Index (HWQI)** assesses aquaculture shrimp systems using historical information (variables set measurements). The result of the assessment is a status given by the behavior of the variables measured. The levels of classification of the HWQI are defined as *Excellent*, *Good*, *Regular* and *Poor*.

The HWQI works in two phases; first it calculates the physical-chemical index, which classifies the effect level of a variable value in the ecosystem. The second phase consists on evaluate the result information of the Γ index by a reasoning process using a fuzzy inference system.

Physical – chemical assessment index (Γ)

The fuzzyfication process is done using the physical-chemical index (Γ) and it comprises three factors; frequency, amplitude and deviation of the failed tests [5]. The frequency (α) of failed tests in a set of 96 measurements is calculated using the next expression:

$$\alpha = \frac{m_f}{m_T} = \frac{m_f}{96} \quad (11)$$

where m_T is the total number of measurements and m_f is the number of measurements out of the desired range. The average of the number of deviations (β) of failed test is calculated using:

$$\beta = \frac{\sum_i^n e_i}{m_T} = \frac{\sum_{i=1}^{96} e_i}{96} \quad (12)$$

where e is the deviation of the i^{th} failed measurement (out of the desired level), $i: 1, 2, \dots, n$, n is the number of the deviations and m_T is the number of total tests. The variable e can be expressed as follows:

when the value must not fall below the level:

$$e = \frac{l_b - t_b}{l_b - m} \quad (13)$$

when the value must not exceed the level:

$$e = \frac{m - l_a}{t_a - l_a} \quad (14)$$

where m is the value of the test, l_a is the upper limit of the range to evaluate, t_a is the upper tolerance, l_b is the lower limit of the evaluated range and t_b is the lower tolerance of the range (Table 1).

The Physical – chemical index can be expressed as follows:

$$\Gamma = \sqrt{\frac{\left(\frac{m_f}{96}\right)^2 + \left(\frac{\sum_i^n e_i}{96}\right)^2}{2}} \quad (15)$$

Finally the membership input equation for the fuzzy inference system is defined as follows:

$$\mu_{variable} = \begin{cases} 0 & 1 < \Gamma \\ 1 - \Gamma & 0 \leq \Gamma \leq 1 \end{cases} \quad (16)$$

Reasoning process

The level classifications of the particular variables are processed using a set of rules that involves all the cases of the habitat condition. There are some expressions that are frequently used by experts in water management, these kinds of expressions construct the fuzzy language of the FIS, and these rules can be expressed as follows:

Rule 1: If *Temp* is normal and *Salt* is normal and *pH* is normal y *DO* is normal then *WQI* is Excellent

Rule 2: If *Temp* is normal and *Salt* is normal and *pH* is normal y *DO* is low then *WQI* is Good

The size of the rule set depends of the number of rules that are involved in the environment; a total of 135 rules have been used in this case.

Predicted Water Quality Index (PWQI)

The reasoning process generates one output membership by one rule, and this result is used to establish the membership of the output function (Fig. 3).

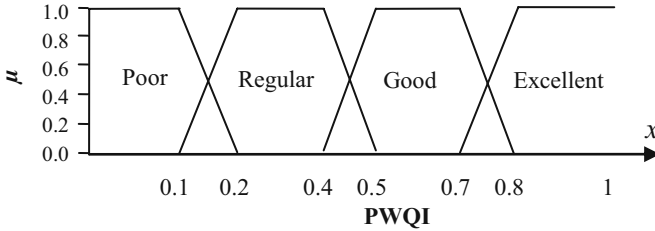


Fig. 3. Membership functions for PWQI

For transforming the final results (the indices results processed by the reasoning process) in a real output value (condition of the shrimp habitat), it is needed to realize the aggregation process, where all the resulting membership functions generate one final function ($\mu_{out}(x)$). The defuzzification process is done when the center of gravity is calculated using $\mu_{out}(x)$, to do this the centroid method is used as a defuzzification solution:

$$PWQI = \frac{\int x\mu_{out}(x)dx}{\int \mu_{out}(x)dx} \tag{17}$$

The final score for the PWQI index have a range from the center of the *poor* function to the center of the *excellent* function [0.078, 0.87], therefore the different status values are located inside this range; *bad* is 0.078, *regular* is 0.3, *good* is 0.6 and *excellent* is 0.87.

5 Results

A predictability analysis shows the perform of the system. Prediction tests were done using one and two days of information. The PWQI (predicted day) results were compared with the HWQI (evaluated day) showing some interesting remarks (Fig. 4). The left column shows the predicted values versus the real values, where the curves have similar results. A second prediction analysis is showed in right column, where a relationship between estimated and real values was done, the nearest points to the diagonal line are a closer prediction to the real assessment of the water quality. The results show a closer relationship between HWQI and PWQI indices.

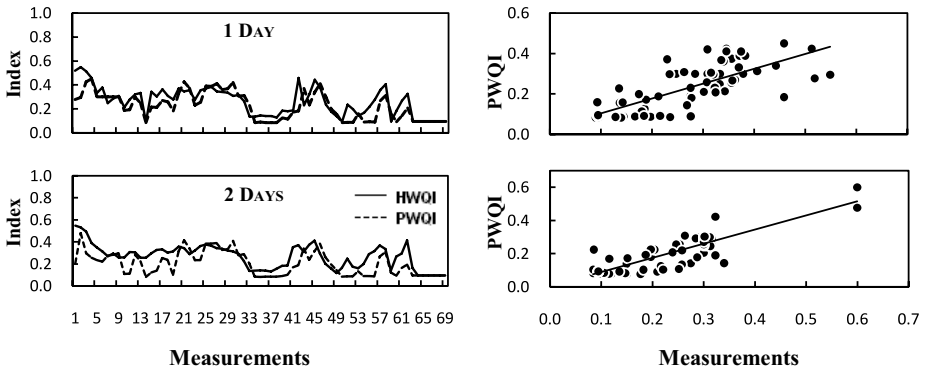


Fig. 4. Results of the prediction and assessment of the artificial habitat. The left column shows a comparison with different prediction horizons (24 and 48 hours) between the real and predicted data. The right column shows the relationship of predicted (PWQI) and real values (HWQI), where the diagonal line represents the exactitude of the prediction.

When the number of measurements to predict is high, the prediction error increase, this behavior can be observed in Fig. 5, where error estimation curve is modeled.

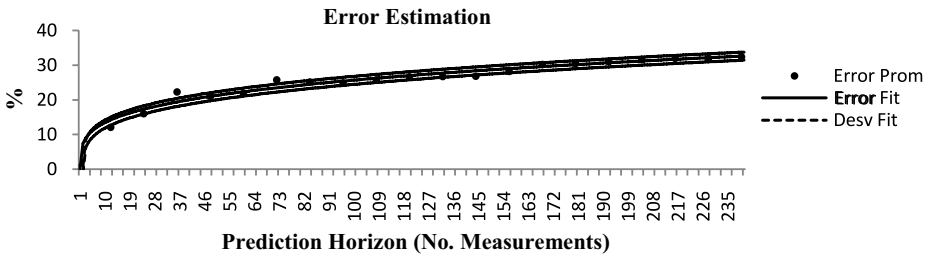


Fig. 5. Error estimation for different prediction horizon

The error prediction can be estimated with the next equation:

$$e = 7.40362x^{0.270686} \pm 1.16313 \quad (18)$$

where e is the predicted error, x is the prediction horizon expressed as the number of predicted points, and the term 1.16313 is the average deviation of the prediction error.

6 Conclusions

In this work a model for predicting the status of the ecological environment for aquaculture systems was developed. The model to predict the water quality status was built in two phases: the first predicts a section of the four signals; the second analyzes the four sections with the HWQI index in order to analyze the predicted signals and to create the PWQI index. A comparison between model (HWQI and PWQI) shows a good performance of the prediction process and the error analysis shows how a bigger

horizon prediction increases the error with a light deviation. The model proposed in this research is a powerful tool in the decision support for monitoring future environmental problems in aquaculture shrimp systems.

References

1. Martínez Córdova Rafael, M.: Cultivo de Camarones Pendidos. In: Principios y Practicas, AGT Editor S.A. (1994)
2. Hirono, Y.: Current practices of water quality management in shrimp farming and their limitations. In: Proceedings of the Special Session on Shrimp Farming. World Aquaculture Society, USA (1992)
3. [ACA] Agencia Catalana del Agua, Catalonia, Spain (2005), <http://www.mediambient.gencat.net/aca/ca/inici.jsp> (accessed August 2007)
4. [NSF] National Sanitation Foundation International (2005), <http://www.nsf.org> (accessed August 2007)
5. [CCME] Canadian Council of Ministers of the Environment (Canada). An assessment of the application and testing of the water quality index of the Canadian Council of Ministers of the Environment for selected water bodies in Atlantic Canada. National indicators and reporting office (2004), <http://www.ec.gc.ca/soer-ree/N> (accessed August 2007)
6. Kenneth, H.: Water Quality Prediction and Probability Network Models. North Carolina State University (1998)
7. Emmanuel, C.: Digital signal processing: a practical approach. Addison-Wesley, Reading (1993)
8. Chapra, S., Canale, R.: Métodos Numéricos para Ingenieros. McGraw-Hill, México (1999)
9. Cohen, L.: Time-frequency signal analysis. Prentice Hall PTR, Englewood Cliffs (1995)