E-ISSN 2149-0236
Journal of Aquaculture Engineering and Fisheries Research

FULL PAPER

TAM MAKALE

FORECASTING NILE TILAPIA WEIGHT PRODUCTION UNITS IN TEMPERATE ZONES USING AN ANFIS AND LINEAR REGRESSION MODELS

José L. Rodríguez-Álvarez[1], Rogelio López-Herrera[2], Iván E. Villalón-Turrubiates[3], Gerardo Grijalva-Ávila[4], Arturo Soto-Cabral[5]

Cite this article as:

Rodríguez-Álvarez, J.L., López-Herrera, R., Villalón Turrubiates, I.E., Grijalva-Ávila, G., Soto-Cabral, A. (2019). Forecasting Nile Tilapia Weight Production Units in Temperate Zones Using an Anfis and Linear Regression Models. Journal of Aquaculture Engineering and Fisheries Research 5(1), 12-24

[1]Department of Doctoral Program in Engineering Sciences, Instituto Tecnológico y de Estudios Superiores de Occidente, Anillo Perif. Sur Manuel Gómez Morín 8585, Santa María Tequepexpan, 45604 San Pedro Tlaquepaque, Jal., Mexico

[2]Department of Manufacturing Technology Engineering, Universidad Politécnica de Durango, Carr. Durango-México km 9.5, 34300, Durango, Dgo., Mexico

[3]Department of Industrial Engineering, Instituto Tecnológico de Durango, Blvrd. Felipe Pescador 1830, 34080, Durango, Dgo., Mexico

Submitted: 02.05.2019

Accepted: 15.05.2019

Published online: 24.06.2019

Correspondence:

José L RODRÍGUEZ-ÁLVAREZ

E-mail:
ng718130@iteso.mx

Journal abbreviation:

J Aquacult Eng Fish Res

ABSTRACT

This paper proposes two models to estimate individual net weight of the Nile Tilapia (Oreochromis Niloticus), in an experimental intensive production unit located in the temperate zone of the state of Durango, Mexico. As a first approach, it is proposed a model based on the neuro-fuzzy ANFIS system (Adaptative Neuro-Fuzzy Inference System) model, which is used for training phase data size and individual weight obtained from a sampling process during a production period of six months, considering as input variable the precaudal length (centimeters) and as output variable the weight (grams). Different configurations are tested in the antecedent and consequent parameters of the ANFIS network to determine the best fit in the model. The first four months were used to collect data for training, and the remaining two months for validation. As a second approach, a linear regression model using data in the same way (first four months to make the model adjustment and the remaining months to verify the predictive capabilities) is also proposed. The results show that the ANFIS model has greater predictive power, since the error in the forecast inside and outside of the sample is below than the error obtained with the linear regression model.

Keywords: ANFIS, Regression, Neural network, Neuro-fuzzy, Tilapia, Forecast

Introduction

Aquaculture is the cultivation of aquatic animals, such as fish, molluscs, crustaceans, and aquatic plants, in controlled environments. Regional aquaculture production is progressing at rates above world averages, this activity contributes a third of the world's fishing production and is one of the fastest growing sectors since the end of the 80s. Regional consumption of fish is very limited (it is the lowest among continents) and therefore domestic demand has not been a good driver for aquaculture, which depends more on export possibilities and its ability to create employment opportunities, particularly in rural areas (FAO Aquaculture Global and Regional Reviews, 2017). Although this activity is an alternative to satisfy the food needs of many tropical countries, it has numerous effects on the aquatic ecosystems where it develops (World Atlas of Biodiversity (2002) UNEP-WCMC, 2017). These effects range from the affectation of the natural populations of the crop species by the collection of egg, larva, young and adults, to the discharge of the residual water to the effluents of water adjacent. Countries such as Brazil and Mexico are large importers of seafood, even though they still have plenty of opportunities for further aquaculture development. Currently aquaculture is a practice that takes place in various regions of Mexico, and there are several species susceptible to be cultivated with extensive business opportunities or only as food alternatives (National Commission of Aquaculture and Fisheries, Mexico, 2018). This is the case of the state of Durango, where the number of aquaculture units has increased for self-consumption and business, as an alternative to traditional agricultural activities. Once aquaculture is business, it is necessary to have optimal production systems, where the development of estimation models can be a useful tool that, together with an intermittent monitoring of the size of the fish, allows the producer to forecast or validate the performance of the production unit according to the initial planning. In the State of Durango, the species with the greatest demand for the best exploitation in aquaculture are tilapia and trout, however, in the case of tilapia, the low temperatures in the winter and the precocity of this species, are limiting its cultivation. Tilapia dies at a temperature lower than 9°C and its growth decreases considerably when it is lower than 20°C, extending its fattening period up to 8 or 9 months to reach sizes of 250 to 300 g. This situation is a disadvantage for the producers, because it increases their production costs, so the use of models allows to determine the maximum size of harvest to which the producer can obtain the greatest profit and reduce the costs of the crop.

Due to the demand for the cultivation of tilapia, in 2012 an intensive production module of masculine nilotic tilapia was established in Durango, in this module information was collected based on growth of the species for the best adjustment of a mathematical model. It is therefore very important to select the most suitable model, as well as to find the parameters that allow an acceptable adjustment to the data obtained, so that the producer can count on a mechanism to estimate and validate the weight as a function of the length at any instant of time during production.

To model the process of tilapia growth in temperate zones it is necessary to build reliable models that minimize the forecast error. Model developing is usually based on two approaches: models based on neural networks and parametric models. We need to select the model with the best predictive capacity.

A Brief Review on Predictive Models

Artificial neural networks

Artificial neural networks are mathematical structures that try to imitate the functioning of a biological brain. They are composed of a series of nodes or neurons, interconnected with each other, structured in a series of layers. The input layer receives the signals from the outside and is responsible for distributing these signals to the internal layers. The internal layers are the ones that perform the necessary calculations to obtain an output. Finally, it is the output layer that shows the results of the network. These structures are able to extract knowledge from a series of sample data, and then apply it to unknown data (Acuña et al., 2014).

Artificial neural networks have been used to solve numerous problems in many different application fields. In particular, they have been used economic and financial, highlighting to a considerable extent their application in the prediction of time series and their ability to detect and exploit the non-linearity in the data, even in conditions where there is incomplete data or the presence of noise; they also stand out for their performance in solving complex problems, where the recognition of models or behaviors is important (Villada et al., 2014).

A type of network that stands out for its structured method of application and for its effectiveness in both modeling and process control is the ANFIS network (Adaptative Neuro-Fuzzy Inference System), which was developed (Jang, 1993). The adaptive capabilities of the ANFIS networks make them directly applicable to many areas, such as adaptive control, processing and filtering of signals and time series, classification of data, and extraction of characteristics from examples (Gómez et al., 2010).

Models based on neural networks for predictive purposes are being used in different areas of science. In (Villada et al., 2014), a model based on artificial neural networks for the international oil forecast is proposed; (Pérez et al., 2015) use artificial neural networks (ANN) as their main method to make financial prediction for hotels, given their ability to capture past patterns and replicate them to make financial predictions in the hotel sector, in addition to making estimates with new data. or out of sample, as well as infer behaviors and future results. Acuña et al., 2014 proposes two models of artificial neural networks for the prediction of the result of the compression test of a concrete construction after the curing period from easily measurable manufacturing data.

Traditional predictive models

Traditional forecasting models have been widely used for decision making in the areas of production, finance, and energy, such as the ARIMA models, developed in the 70s (Box and Jenkins, 1976).

As part of the traditional models are linear regression models, which are widely used in engineering as they serve to analyze the behavior of input (or regressor) and output (or response) variables, establishing predictions and estimates (Montgomery et al., 2006). In research related to parametric models in animal production, the functions that are normally used to evaluate the growth of distinct species of animal origin are chosen empirically (Gómez et al., 2008) and their choice is made based on the ability of the function to adjust to the data, and sometimes the parameters obtained in these functions do not have a biological interpretation. Growth functions can be characterized through knowledge of some physiological delineation or biological mechanisms (Tidwell and D'Abramo, 2010), allowing the parameters obtained to be analyzed from the productive point of view. A model similar to the one proposed in this work is developed by De Los Santos and Silva, 2008, however, the model focuses on the growth of shrimp.

Materials and Methods

Sampling of the tilapia population to measure the variables of interest was obtained in the aquaculture module located in the facilities of the Valle del Guadiana Experimental Field of the INIFAP-Durango. This module is composed of 4 circular tanks of 30 m^3 capacity, with densities of sowing of 15 kg per cubic meter. Biweekly, samplings of organisms were carried out in each pond from sowing to 550 g in weight. The sample size was taken at random and represented 10% of the total population.

The weight and length data from tilapias were obtained through good management practices, ensuring at all times the avoidance of measurement errors (digital scale and ichthyometer). During this process the water temperature was also recorded, and along with the weight, both variables allow calculating the quantity of food according to tables recommended for the species under study, in this way, the feeding criteria are standardized in the four ponds. By means of a statistical analysis of confidence intervals, the first two variables were validated and then the existing correlation, established from different linear parametric models, was analyzed. The length of the fish in its distinct stages of growth was considered as the only input variable to the model, and the weight as the only output variable, both for the phase of obtaining the model and for its subsequent validation.

Below are the procedures used to obtain the models, first neuro-diffuse and next the linear regression model.

Regression model

For the statistical analysis, the best regression model is adjusted using Minitab. The method used to estimate parameters is the least squares method, which is a method that adds the squares of the residuals or errors around the regression line to minimize them (Walpole et al., 1999). It is estimated that the adjustments of the different models are considered highly reliable since during the sampling process particular care was taken to not violate the assumptions of the model. This assumption is; normality, independence, and constant variance, as suggested by Pulido and De La Vara, 2008. For the choice of the best model, the coefficient of determination R^2 was considered, which can be interpreted as the proportion of the total variation in the y_i that is explained by the variable x in a simple linear regression model (Wackerly et al., 2010). In Açikkar and Sivrikaya, 2018, the calculation of determination coefficient is described by the following equation:

$$R^2=1-\frac{\sum_{i=1}^{m}(Y_i-Y_i')^2}{\sum_{i=1}^{m}(Y_i-\bar{Y})^2} \quad (1)$$

Where Y_i is the measure value, $\bar{Y}$ is the predicted value, $\bar{Y}$ is the mean of the measured values and the number of instances in a test set is given by n.

It is of interest the adjustment of biological models with linear models of higher order (if necessary), a second and a third model are also used to adjust the data.

A frequent practice in linear regression is to find optimal setting condition by transforming the response to a desirability value. The first original transformation of a response to a desirability value was proposed by Harrington (De Jongh et al. 2014). The desirability index function converts all the desirability values, d_j's, into a single composite desirability. The composite desirability is maximized by conventional or unconventional search techniques. The composite desirability value (D) always lies between 0 and 1. An equation developed by Harrington in 1965 and shown in (De Jongh et al. 2014), combines the individual desirability to a composite desirability (D) by the following expression:

$$D = \text{maximize}\left(\Pi_{j=1}^{r} d_j\right)^{\frac{1}{r}} \quad (2)$$

Where, d_j are the desirability values.

Anfis

A fuzzy inference system is basically described in Figure 1. Abbas et al., 2017, mention that fuzzy inference process consists of the following steps: (1) fuzzification, (2) application of fuzzy operation in the antecedent, (3) implication to the consequent, (4) aggregation, and (5) defuzzification. The basic algorithm of the adaptative network to train the input-output data is steepest descent (backpropagation). To make the learning process faster, a hybrid learning algorithm is usually used.

Control System Design and Analysis Tool Box of Matlab is used to build the ANFIS network model, considering two phases. In the first phase, the network is trained from the input and output data set (length and weight) using the data from the first four months and considering the different adjustment functions. In the second phase, the input and output data of the process of the last two months were used to carry out the validation for each membership function.

Results and Discussion

All the terms of the models shown in Table 1 were significant at 0.05 alpha, also in this table we can see that variation (s) and coefficient of determination (R^2) are much better with models of higher order because of, less variation indicates less error and more coefficient of determination (preferably close to 1) indicates that model explains the variability. Because of the difference between second and third order models is minimal and in order to have a simpler model, second order model is chosen. In Figure 2 is shown the assumptions of the model and Figure 3 shows the adjustment to the data (Tables 2-4).

The fitted training data are shown in Figure 4, there we can see that red dots are mostly superimposed on the blue dots (observed data), while the ANFIS structure is shown in Figures 5 and 6.

Once the two neuro-diffuse models and linear regression are available, it is necessary to test their predictive capacity, for which data from the last two months were used. The following evaluation measures to choose the best model are considered in (Acuña et al., 2014): RMSD (root of mean square error) and MAPE (average absolute percentage error), calculated by:

$$RMSE=\sqrt{\frac{1}{n}\sum_{t=1}^{n}(y_t'-y_t)^2} \quad (3)$$

Grade	**Model**
Linear	$Weight = -182.6 + 23.4\ L + \varepsilon$ $R^2 = 84.3\%$ $s = 59.8$
Order 2	$Weight = 53.8 - 15.9\ L + 1.3L^2 + \varepsilon$ $R^2 = 97.9\%$ $s = 21.5$
Order 3	$Weight = 38.8 - 11.1\ L + 0.9L^2 + 0.009L^3 + \varepsilon$ $R^2 = 98.0\%$ $s = 21.4$

Table 1. Regression models, variation and determination coefficient.

Number of inputs	1 (Length)
Input MF	All
Number of outputs	1 (Weight)
Output type	Constant and Linear
Number of input-output pairs (training data)	200
Learning algorithm	Hybrid
Number of epochs	1000
Error tolerance	0
Evaluation method	Training data versus ANFIS output

Table 2. ANFIS information.

Constant	Linear
Trimf (150-20.01)	Trimf (250-20.62)
Trapmf (450-20.08)	Trapmf (250-20.27)
Gbellmf (1000-20.06)	Gbellmf (250-20.31)
Gaussmf (1000-20.43)	Gaussmf (500-20.09)
Gauss2mf (900-20.01)	Gauss2mf (600-20.21)
Pimf (1000-20.02)	Pimf (650-20.18)
Dsigmf (200-20.80)	Dsigmf (300-20.36)
Psigmf (150-20.80)	Psigmf (300-20.36)

Table 3. ANFIS membership functions.

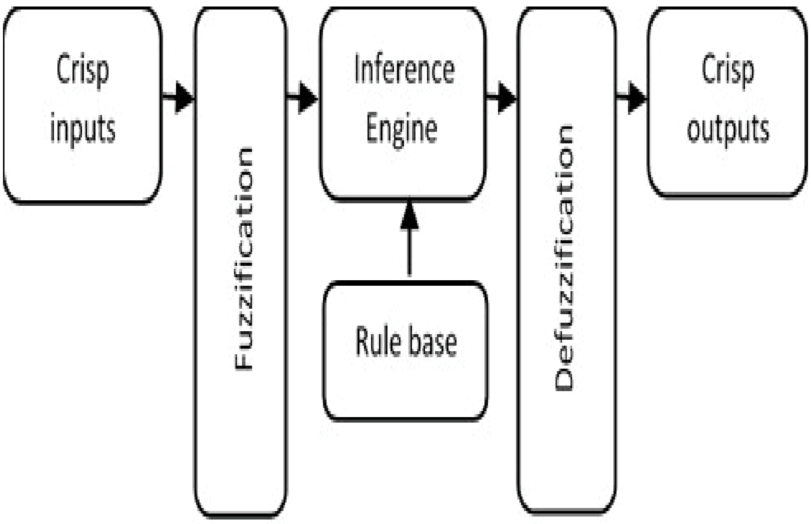

Figure 1. Fuzzy inference system.

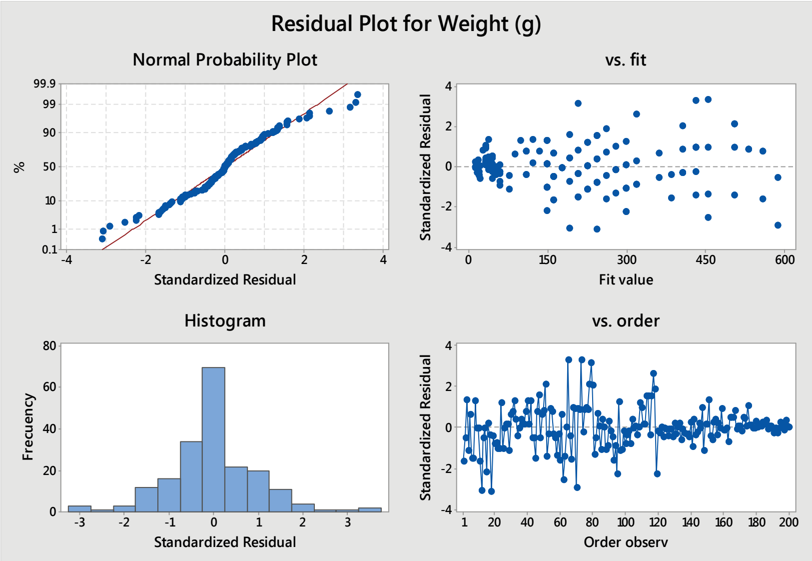

Figure 2. Assumptions of second order regression model.

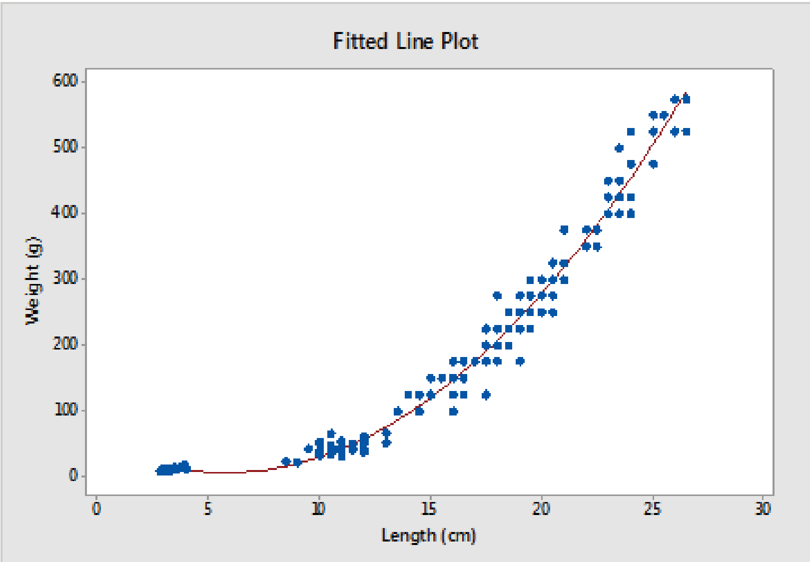

Figure 3. Fitted regression model of second order.

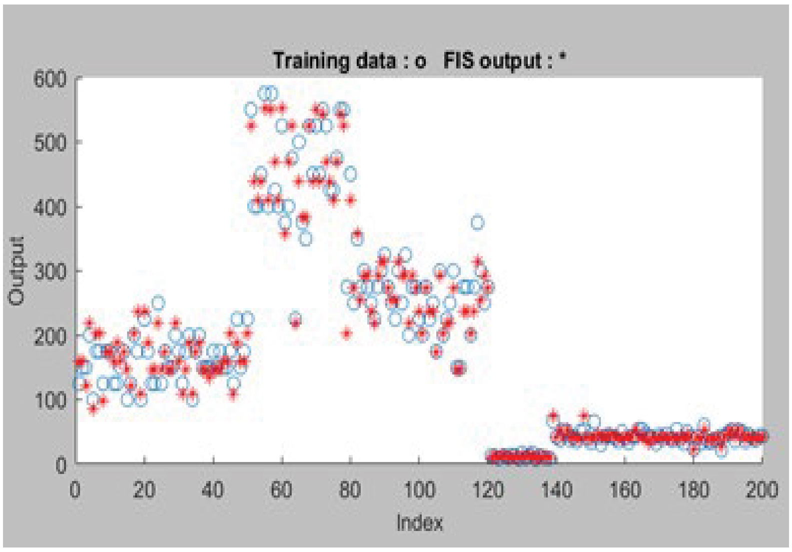

Figure 4. Fitted training data with ANFIS.

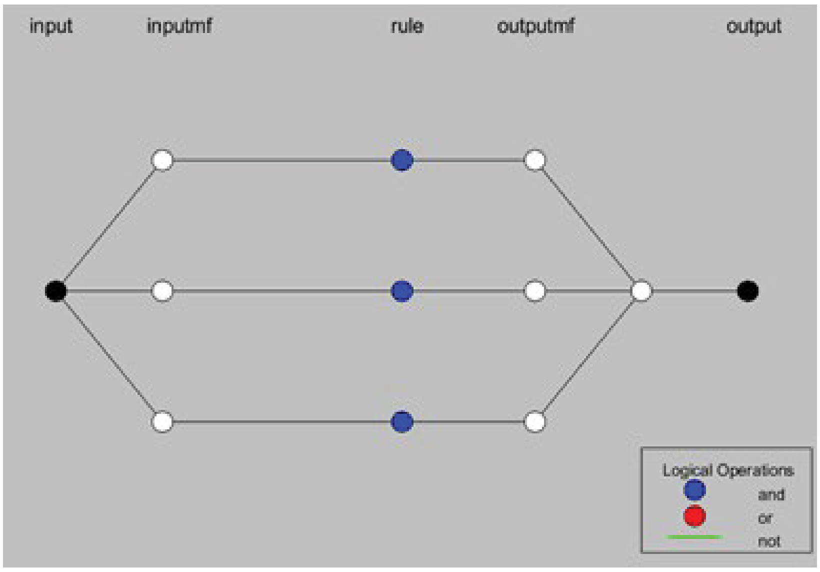

Figure 5. ANFIS structure.

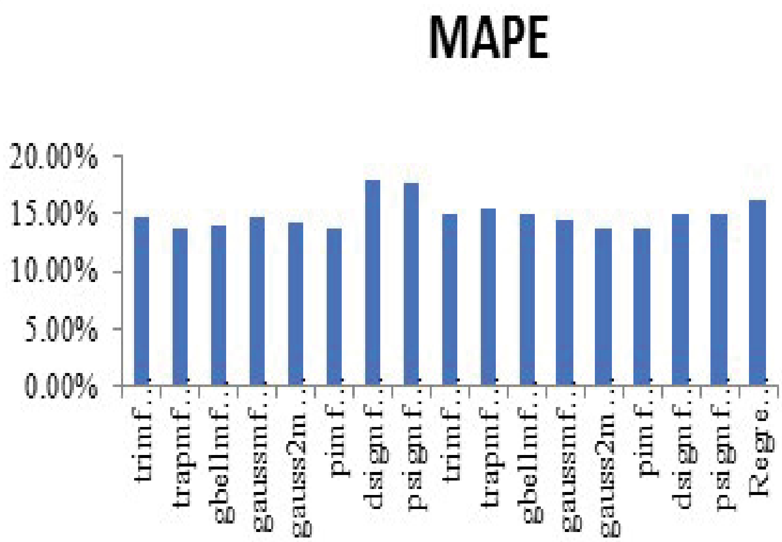

Figure 6. ANFIS and linear regression.

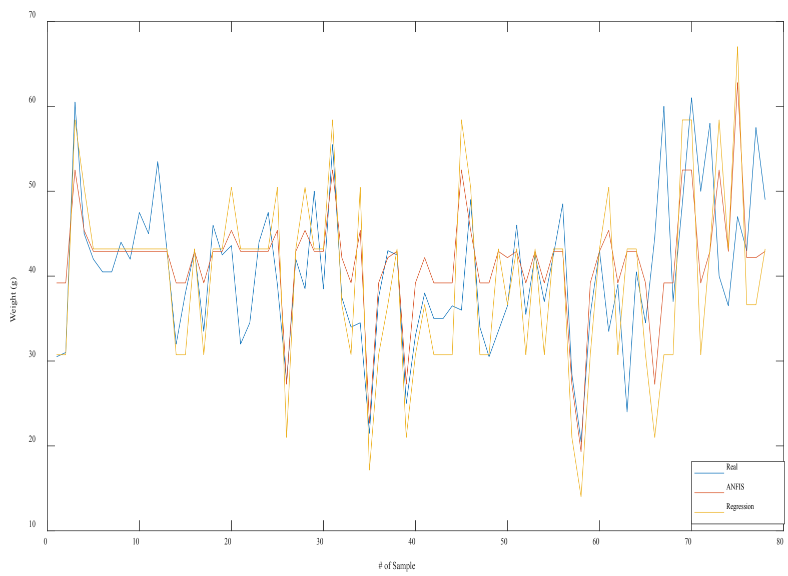

Figure 7. Observed values vs ANFIS and regression.

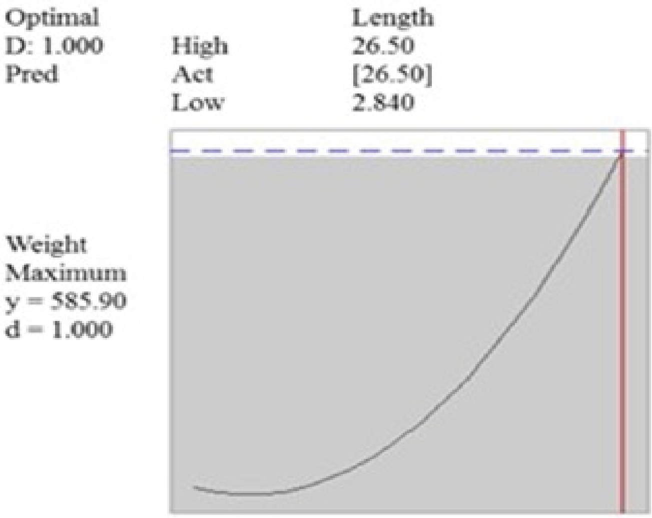

Figure 8. Desirability index for weight.

$$MAPE = \frac{100}{n}\sum_{t=1}^{n}\left|\frac{y_t' - y_t}{y_t}\right| \quad (4)$$

In Figure 7, it is shown a comparison between the results observed and predicted by the regression model and the ANFIS model. It is seen that ANFIS model shows more capacity to make more reliable predictions about tilapia weight.

Next values can be used only as a reference for any producer of Nile tilapia in temperate zones. In Figure 8 is shown the composite desirability of 1.00 for weight and the optimal setting condition are in a length of 26.50 centimeters and expecting a maximum weight of 585.90 grams.

Conclusions

Currently it is very important that the actions for control and improvement of any productive process are based on reliable information supported by predictive models to make the process more profitable.

The regression and ANFIS models can be widely applicable to aquaculture production processes under protected areas. In this work, it is demonstrated their capacity to predict the weight of tilapia, which is very helpful for the producer since this allows acting in a timely manner to improve returns.

Although ANFIS network demonstrated a greater predictive capacity than the linear regression model and confirming it as an excellent alternative when it is required to build models for predictive purposes, a regression model was selected due to its ease of use in practice.

REFERENCES

Abbas, A.T., Alata, M., Ragab, A.E., El Reyes, M.M., El Danaf, E.A. (2017). Prediction model for cutting parameters for turning high strength steel Grade-H: comparative study of regression model vs anfis. *Advances in Materials Science and Engineering*, 2017, 1-12.

Açikkar, M., and Sivrikaya, O. (2018). Prediction of gross value of coal based on proximate analysis using multiple linear regression and artificial neural networks, *Turkish Journal of Electrical Engineering and Computer Sciences*, 26, 2541-2552.

Montgomery, D.C., Peck, E., Vining, G. (2006). Introducción al análisis de regresión lineal México, *Limusa Wiley*, 2006.

Wackerly, D.D., Mendenhall, M.M., Scheafer, R.L. (2010). Estadística Matemática con Aplicaciones México, *CENGAGE Learning*, 2010.

De Jongh, P. J., De Jonghy, E., Pienaarz, M., Gordon-Grantz, H., Oberholzerz, M. and Santanax, L. (2014). The impact of pre-selected variance inflation factor thresholds on the stability and predictive power of logistic regression models in credit scoring. Orion, vol. 31, pp. 17-37.

Divier, A., Gómez, H., Mario, F., Muñoz, L.F., (2008) Modelación de las funciones de crecimiento aplicadas a la producción animal, *Revista Colombiana de Ciencias Pecuarias*, 21, 39-58.

Villada, F., Arroyave, D., Villada, R. (2014). Pronóstico del precio del petroleo mediante redes neuronales artificiales, *Información Tecnológica*, 25, 145-154.

FAO. Aquaculture Global and Regional Reviews. Retrieved October 12, 2017.

Box, G., Jenkins, G. (1976) Time series analysis: forescasting and control. San Francisco: *Holden Day*, 1976.

Pulido, H. G., De La Vara R. S. (2008). Análisis y Diseño de Experimentos. México: McGraw-Hill, 2008.

Tidwell, J.H., D'Abramo, L.R. (2010). Grow out systems culture in temperate zones, in A quantitative theory of organic growth. In: L. B. Von (ed.). *Oxford, UK: Wiley-Blackwell*, 10, 180-193.

Acuña, L., Torre, A., Moromi, I., García, F. (2014). Uso de las redes neuronales artificiales en el modelado del ensayo de resistencia a compresión de concreto de construcción según la norma ASTM C39/C 39M, *Información Tecnológica*, 25, 3-12.

Pérez, L., Garriga, L., Benítez, M. (2015). Sistema de predicción financiera para hoteles mediante redes neuronales artificiales, *Retos Turísticos*, 14, 1-11.

National Commission of Aquaculture and Fisheries, Mexico. (2018). Live aquaculture a new age in Mexico in favor of food security.

Walpole, R. E., Raymond, H., Myers, S. L. (1999). Probabilidad y Estadística Para Ingenieros. *México: Prentice Hall*, 1999, 18.

De Los Santos Romero R. R., Silva, M. E. (2008). Crecimiento de macrobrachium michoacanus con relación al tipo de alimento y densidad de cultivo. *Revista Naturaleza y Desarrollo*, 5, 15-25.

Jang, S. J. (1993). ANFIS: adaptative-network based fuzzy inference systems, *IEEE Transactions on Systems, Man, and Cybernetics*, 23, 665-685.

Gómez, V. E., Socarras, Q. V., Obregón, N. N. (2010). Aplicación del modelo neuro-difuso ANFIS vs redes neuronales, al problema predictivo de caudales medios mensuales del Río Bogotá en Villapinzón, *Revista Tecnura*, 14, 18-29.

World Atlas of Biodiversity. (2002). UNEP-WCMC. Retrieved November 05, 2017.