Estimation of body weight of *Sparus aurata* with artificial neural network (MLP) and M5P (nonlinear regression)–LR algorithms

Sangün L.1*; Güney O.İ.1; Özalp P.1; Başusta N.2

Received: February 2017 | Accepted: July 2017

Abstract
In this study, morphometric features such as total length, standard length, and fork length obtained from a total of 321 *Sparus aurata* samples, including 164 females and 157 males, captured between 2012 and 2013 from İskenderun Bay were used as input value, while weight was used as an output value. The Artificial Neural Network (MLP-Multi-Layer Perceptron) as well as the M5P algorithm and Linear Regression (LR) algorithm from version 3.7.11 of the WEKA Program were applied. When coefficients of correlation were assessed, the MLP algorithm for males, females and the total were calculated as 0.9686, 0.9605 and 0.9663, respectively; the M5P algorithm for males, females and the total were calculated as 0.9722, 0.9596 and 0.9735, respectively; and the LR Model for males, females and the total were calculated as 0.9777, 0.9498 and 0.9473, respectively. With respect to the Mean Absolute Error (MAE) calculations, the MLP algorithm MAE values for males, females and the total were calculated as 2.94, 2.57 and 2.7074, respectively; the M5P algorithm MAE values for males, females and the total were calculated as 2.400, 2.641 and 2.157, respectively; and the LR Model MAE values for males, females and the total were calculated as 3.217, 2.811 and 3.11, respectively. It can also be concluded from the study that, in order to predict ANN interactions Nonlinear Regression model is more effective and has better performance than the conventional models.

Keywords: Weka 3.7.11, Artificial Neural Network-MLP, M5P, *Sparus aurata*, Morphometric feature, İskenderun Bay

---

1-Vocational School of Adana, University of Çukurova, , Çukurova-Adana, Türkiye.
2-Fisheries Faculty, Firat University, TR-23119, Elazığ, Türkiye.
*Corresponding author's Email: leventsangun@gmail.com
**Introduction**
There is constant change in the biomass and distribution of fish populations of economic importance. Mathematical methods are widely used to analyze these changes, since it is necessary to manage the fishing industry effectively by analyzing reasons behind the different problems that may occur in populations (Haddon, 2001).

The objective of scientific studies is to reach valid and accurate results, and to convert data into high-quality information, by applying different statistical analysis techniques onto input gathered from populations.

The collection, processing and evaluation of data from populations in sampling studies are phases that involve certain expenditures. Gathering information on a population through sample data, and using the said data to make generalizations constitutes the basis of most studies on fishery biology. In recent studies, artificial neural network is commonly used as an alternative to traditional statistical methods.

Among the empirical approaches, the artificial neural networks (ANN), in particular, multi-layer perceptron (MLP), were widely applied in the last decades in the fields of bioinformatics, ecology, and environmental engineering (Lee *et al*., 2016). Also, in fisheries, various studies involving the use of artificial neural networks and multi-layer perceptron have been applied in recent years. These include Baran *et al*., 1996; Haralabous and Georgakarakos, 1996; Hwanh *et al*., 1996; Lek *et al*., 1996; Mastrorillo *et al*., 1997; Scardi and Harding, 1999; Brosse *et al*., 1999; Brosse *et al*., 2001; Hansen *et al*., 2001; Brosse and Lek, 2002; Hardman-Mountford *et al*., 2003; Engelhard, *et al*., 2003; Engelhard and Heino, 2004; Joy and Death, 2004; Laffaille, *et al*., 2003,2004; Crvello *et al*., 2005; Oakes *et al*., 2005; Bahmanzadegan *et al*., 2013.

Artificial neural network, which is a sub-branch of artificial intelligence, has in recent years become a noteworthy method owing to its ability to solve complicated problems and learn relations between data. The most significant reason for the widespread use of artificial neural network is that it offers an effective alternative to solving complicated problems with classical statistical techniques and performs well in function approximation and pattern recognition (Lek and Guegan, 1999; Manoj *et al*., 2014).

In estimation studies, regression and artificial neural network methods use independent variables to estimate the dependent variables from data sets with different attributes. The Regression Tree is a commonly used method in estimation studies involving the Artificial Neural Network. The neural network consists of a set of processing elements, also known as neurons or nodes whose functionality is loosely based on biological neurons. These units are organized in layers that process the input information and pass it to the following layer (Manoj *et al*., 2014). It allows models to be formed in accordance with the input parameters, and ensures that high estimation values can be achieved. ANN works in
different ways so that the data will be presented successfully (Bahmanzadegan, et al., 2013).

The WEKA program, developed by Waikato University in New Zealand, is an open code data mining program with a functional graphic interface, hosting machine learning algorithms together (Witten et al., 2011).

Gilthead sea bream, the fish species we used in our study, is a euryhaline and eurythermal species living in the Mediterranean Sea and along the Eastern Atlantic coasts (Carpene et al., 1998; Haffray et al., 2005). Sea bream rearing in the Mediterranean has increased in recent years, and now constitutes a large portion of today’s gross production (Arechavala-Lopez et al., 2012).

The aim of this study was to analyze and comparatively interpret the growth model of fish with respect to total length, standard length, fork length and weight values obtained from measurements from S. aurata using nonlinear regression models and mlp-Multi Linear Perceptron algorithms in the WEKA program.

Materials and methods
In this study, a total of 321 Gilthead sea bream, including 157 males and 164 females, were captured in commercial trawl surveys between September 2012 and August 2013 from the north-eastern Mediterranean coast of Turkey. The gilthead sea bream was weighed with a digital balance to an accuracy of 0.01g and measured with a precision of 0.01 cm for their total length, standard length and fork length.

The computational program was written in Weka 3.7.11. They were used randomly: 321 (Test mode:split (66.0% train, remainder test)) Gilthead sea bream. Total instances were 321; 157 males and 164 females were used.

\textit{M5P Algorithm}\
Scheme: weka. classifiers.trees. M5P-M 4.0. The learning process shows that an internal validation phase is carried out at the end of epoch. If the internal validation error in epoch is lower than the error of internal validation in epoch, then save the weights (Gutterrez-Estrada et al., 2008).

\textit{LR algorithm}\
Scheme: weka. classifiers. functions. Linear Regression -S 0 -R 1.0E-8 -num-decimal-places 4 was used.

\textit{Artificial neural network (MLP-multi layer perceptron) algorithm}\
Scheme: weka. classifiers functions. Multi-Layer Perceptron –L0.3-M 0.2-N 500-V0-S0-E20-H- ‘3.7’-G-R 4 attributes (Total length, standard length, fork length and weight) was used in male and females.

In this study; total, artificial neural network in male and female individuals, total length (TL), standard length (SL), fork-length (FL) were used as input, while weight (W) was estimated as an output value. MLP (Multi-Layer perceptron) is designed as (H-3.7) (Fig 1)
Sangün et al., Estimation of body weight of *Sparus aurata* with artificial neural network (MLP) and…

**Results**

The minimum, maximum, mean and standard deviation values for total length, standard length, fork length and weight value of sea bream used in our study are given in Table 1.

<table>
<thead>
<tr>
<th></th>
<th>Total length (cm)</th>
<th>Standard length (cm)</th>
<th>Fork length (cm)</th>
<th>Weight (g)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male (N=157)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Min-Max</td>
<td>11.10-23</td>
<td>9.90-20.60</td>
<td>9.3-19.1</td>
<td>8.31-69.56</td>
</tr>
<tr>
<td>Mean-SD</td>
<td>16.80-2.5</td>
<td>15.37-2.32</td>
<td>13.94-2.11</td>
<td>30.054-13.21</td>
</tr>
<tr>
<td>Female (N=164)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Min</td>
<td>12.40</td>
<td>11.20</td>
<td>10.4</td>
<td>11.5</td>
</tr>
<tr>
<td>Max</td>
<td>21.10</td>
<td>19.60</td>
<td>17.4</td>
<td>55.62</td>
</tr>
<tr>
<td>Mean</td>
<td>17.00</td>
<td>15.57</td>
<td>14.12</td>
<td>30.50</td>
</tr>
<tr>
<td>Standard Dev.</td>
<td>2.01</td>
<td>1.92</td>
<td>1.75</td>
<td>10.97</td>
</tr>
<tr>
<td>Total (N=321)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Min</td>
<td>11.1</td>
<td>9.90</td>
<td>9.3</td>
<td>8.31</td>
</tr>
<tr>
<td>Max</td>
<td>23</td>
<td>20.60</td>
<td>19.1</td>
<td>69.56</td>
</tr>
<tr>
<td>Mean</td>
<td>16.9</td>
<td>15.47</td>
<td>14.04</td>
<td>30.28</td>
</tr>
<tr>
<td>Standard Dev.</td>
<td>2.307</td>
<td>2.13</td>
<td>1.93</td>
<td>12.09</td>
</tr>
</tbody>
</table>

**Table 2:** Growth model of individuals in accordance with Linear Regression model.

<table>
<thead>
<tr>
<th>Scheme: Linear Regression Test mode (split 66.0% train, remainder test)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total Instances: 321</td>
</tr>
<tr>
<td>w = 3.2636 TL + 2.0616 FL - 53.8062</td>
</tr>
<tr>
<td>Female Instances: 164</td>
</tr>
<tr>
<td>w = 3.267 TL + 2.0895 FL - 54.5483</td>
</tr>
<tr>
<td>Male Instances: 157</td>
</tr>
<tr>
<td>w = 3.2562 TL + 2.0594 FL - 53.3502</td>
</tr>
</tbody>
</table>
Table 3: Models obtained from M5P Decision Trees (nonlinear) algorithm classification diagram of individuals and Multi-Layer Perceptron (MLP) algorithm, Total length (TL), Standard length (SL), Fork length (FL), Class - Female (F), Male (M).

Scheme: M5Rules - M 4.0 (Test mode: split 66.0% train, remainder test) (Attributes:5)

| Total Instances: 321 |  
|---------------------|------------------------|
| TL > 16.75 TL <= 18.95 | W = 1.7283 TL + 2.4219 SL + 0.2792 FL - 40.3199 [130/24.43%] |
| TL <= 18.85 FL <= 12.95 | W = 0.65 TL + 2.6632 FL - 0.1329F - 24.0259 [91/13.179%] |
| TL > 17.85 TL <= 20.45 | W = -3.2874 TL + 1.8938 SL + 3.07 FL - 0.6294class=female=59.081 [41/11.849%] |
| TL <= 18.6 | W = 1.9888 TL + 3.5433 SL - 0.6294 F - 59.081 [41/11.849%] |
|                     | W = 5.2467 FL - 11.0079F - 32.4407 [15/42.972%] |

| Female Instances: 164 |  
|----------------------|------------------------|
| TL > 16.7 TL <= 18.95 | W = 2.5365 TL + 1.6622 SL + 0.2824FL - 42.2064 [64/26.252%] |
| TL <= 17.8 FL <= 12.95 | W = 0.4709TL - 0.6139SL + 4.3181 FL - 33.0617 [44/12.837%] |
| TL > 17.8 | W = 1.8975 SL + 1.3642FL - 10.6733 [32/31.616%] |
|                     | W = 4.0895SL - 35.3064 [24/81.16%] |

| Male Instances: 157 |  
|-------------------|------------------------|
| TL > 16.05TL <= 18.95 | W = 0.4037TL + 4.0343 SL + 0.6832FL - 49.0835 [79/21.95%] |
| TL <= 17.5 | W = 3.0073TL - 26.075 [51/12.173%] |
| TL <= 20.35 | W = 4.8005TL + 3.6321SL + 2.3907FL + 35.0308 [18/20.466%] |
| FL > 17.55 | W = 3.6883FL - 3.779 [659.759%] |
|                     | W = e + 55.2133 [3/100%] |

The Correlation Coefficient, Mean Absolute Error, Root Mean Squared Error, Relative Absolute Error and Root Relative Squared Error values in the total and by sex are given in Table 4.

<table>
<thead>
<tr>
<th>T-LR</th>
<th>T-M5P</th>
<th>T-MLP</th>
<th>F-LR</th>
<th>F-M5P</th>
<th>F-MLP</th>
<th>M-LR</th>
<th>M-M5P</th>
<th>M-MLP</th>
</tr>
</thead>
<tbody>
<tr>
<td>CC</td>
<td>0.9473</td>
<td>0.9735</td>
<td>0.9963</td>
<td>0.9498</td>
<td>0.9596</td>
<td>0.9605</td>
<td>0.9777</td>
<td>0.9722</td>
</tr>
<tr>
<td>MAE</td>
<td>3.1149</td>
<td>2.1571</td>
<td>2.7074</td>
<td>2.811</td>
<td>2.6412</td>
<td>2.5705</td>
<td>3.2179</td>
<td>2.4003</td>
</tr>
<tr>
<td>RAE</td>
<td>30.36%</td>
<td>21.03%</td>
<td>26.39%</td>
<td>26.08%</td>
<td>24.50%</td>
<td>23.84%</td>
<td>32.41%</td>
<td>24.18%</td>
</tr>
<tr>
<td>RRSE</td>
<td>31.91%</td>
<td>22.66%</td>
<td>26.66%</td>
<td>31.11%</td>
<td>28.06%</td>
<td>28.24%</td>
<td>32.72%</td>
<td>24.50%</td>
</tr>
</tbody>
</table>

Discussions

In our study, artificial neural network Multi-Layer Perceptron algorithm to use artificial neural network method in Weka program, and as alternatives, M5P for nonlinear regression analysis and LR algorithm for linear regression were applied. The aim of the study was to demonstrate the estimation power of different methods (Artificial Neural Networks-M5P, MLP, LR) in estimating the individual weights of S. aurata based on morphometric features.
In Table 1, when total length and standard length are reviewed, it can be seen that female individuals are larger than male individuals. When growth models are assessed in accordance with the Linear Regression model by sex and in the total, it can be observed that total length and fork length contribution are positive (+) and Standard length is not included in the model. (Table 2)

When the models formed according to the M5P algorithm are reviewed by sex and in the total, it is possible to see that the Decision Tree makes classifications with regard to different length groups; that females contributed more to the model than males in the estimation of weights. An examination of the models also shows that TL, FL and SL also contributed to the model. (Table 3) Model outputs can also be seen composed for MLP (Multi-Layer Perceptron) algorithm in the same table.

When coefficients of correlation given in Table 4 were assessed, the MLP algorithm for males, females and in total were calculated as 0.9686, 0.9605 and 0.9663, respectively; the M5P algorithm for males, females and in total were calculated as 0.9722, 0.9596 and 0.9735, respectively; and the LR Model for males, females and in total were calculated as 0.9777, 0.9498 and 0.9473, respectively. With regards to the Mean Absolute Error (MAE), the MLP algorithm MAE values for males, females and in total were calculated as 2.94, 2.57 and 2.7074, respectively; the M5P algorithm MAE values for males, females and in total were calculated as 2.400, 2.641 and 2.157, respectively; and the LR Model MAE values for males, females and in total were calculated as 3.217, 2.811 and 3.11, respectively. Moutopoulos et al. (2011) previously found a correlation coefficient value between fork length and weight of 0.94 in the Klisova lagoon, and of 0.83 in the Papas lagoon. These values are less than values we found with both the MLP and other methods. Akyol and Gamsiz (2011) and Chaoui et al. (2006) have identified correlation coefficient values closer to the value in our study (0.95 and 0.96). Cherif et al. (2008) have determined a correlation coefficient of 0.92 in males, 0.93 in females, and 0.91 in total. These values are less than the values we found using our artificial neural network.

In conclusion, Nonlinear Regression proved to be more effective at capturing ANN interactions and attained a better predictive performance than conventional models. It may thus prove to be promising in other situations where the association between the dependent and independent variables is complex and non-linear. It can also possibly contribute to the fishing management and the protection of biodiversity. ANN approach is completely different from conventional statistical methods, which need a specified algorithm to be transformed by a computer programme (Manoj et al., 2014).

Brosse et al. (2001); Laffaille et al. (2003, 2004); Park et al. (2003); Engelhard and Heino (2004); Gutierrez-Estrada, et al., (2008); Türel-Bilen et al. (2011); previously investigated and compared the artificial neural network with other traditional methods.
Calculation of dependent variables from independent variables is a regression problem. However, the estimation of dependent variables from independent variables through machine learning, and the estimation of dependent variables solely by independent variables is a problem inherent to artificial neural networks. When calculating the growth in fishes with different methods (i.e. Ford-Walford, Gulland and Holt or von Bertalanffy), the relation between length and weight in fish can be calculated in a faster and more accurate way by using an alternative method such as the artificial neural network. Foundation of MLP and Regression model is based on statistics. Mathematical models in algorithms are similar.

However, short calculation period and absence of need for any mathematical conversion during analysis are preferred by the user. MLP method is a data-based method.

Good quality data should be selected, since Weka program is a data-based program. In fishery biology studies, many problems such as the Fulton Condition Factor (K), the fecundity-length weight and the CPUE are used in artificial neural networks for estimation-prediction.

When performing stock studies, data sets need to be arranged by keeping in mind that abiotic factors can have an effect on stock during stock studies, even if indirectly. In fecundity studies, it is important to bear in mind during the evaluation phase that the egg properties of a population (e.g. the diameter) may change according to different areas. Making estimations using the MLP method is important for the effective management of fisheries, and for the determination of the policies and measures which need to be followed.

Acknowledgments

This study was supported the Resource Fund of University of Çukurova, (Turkey) and we thank them for their financial support (With AMYO2011BAP4).

References


**Cherif, M., Zarrad1, R., Gharbı H., Missaouı, H. and Jarbouı O., 2008.** Length-weight relationships for 11 fish species from the gulf of Tunis (Sw Mediterranean Sea, Tunisia), *Pan-American Journal of Aquatic Sciences*, 3(1), 1-5.


**Haffray, P., Bruant, J.S., Facqueur, J.M. and Fostier, A., 2005.** Gonad development, growth, survival and
quality traits in triploids of the protandrous hermaphrodite gilthead seabream *Sparus aurata* (L.) Elsevier, 247(1–4), 107–117.


