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]

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, Fırat 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 *el 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.

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 (Gutterez-Estrada *et al*., 2008).

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

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)
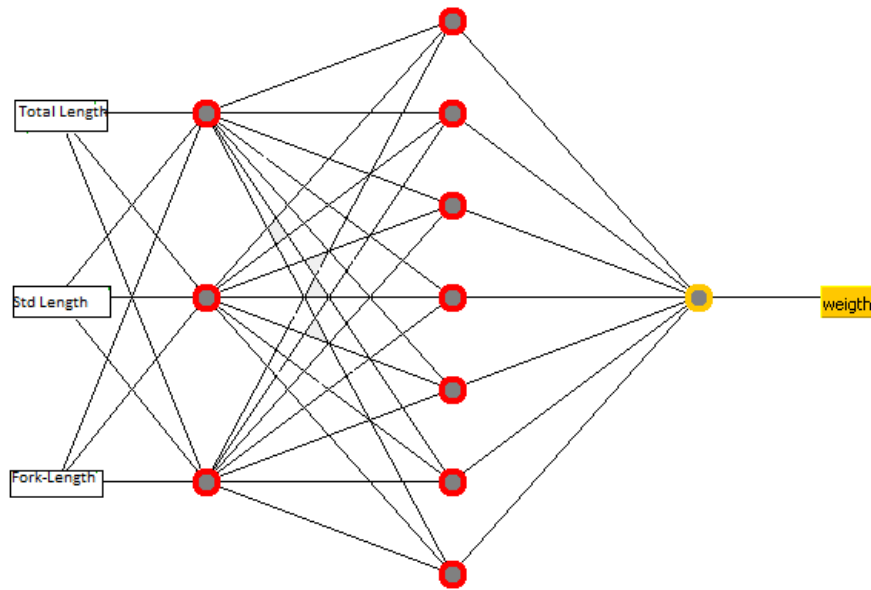
Figure 1: Artificial nerve plexus designed as MLP-(Multi-Layer Perceptron) (H-3.7).

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 1: Min., max., mean and standard deviation values for *Sparus aurata* (sea bream), in total and by sex.

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

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

Scheme: Linear Regression Test mode (split 66.0% train, remainder test)	
Total Instances: 321	w =3.2636 TL +2.0616 FL -53.8062
Female Instances:164	w=3.267 TL+2.0895 FL-54.5483
Male Instances:157	w =3.2562 TL +2.0594FL -53.3502

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)		
	TL > 16.75 TL <= 18.95	W = 1.7283 TL+ 2.4219 SL + 0.2792 FL-- 40.3199 [130/24.433%]
Total Instances: 321	TL <= 17.85 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 = 2.5365TL + 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 = -4.8005 TL + 3.6321SL + 2.3907 FL + 35.0308 [18/20.466%]
	FL > 17.55	W = 3.6883FL- 3.779 [6/59.759%]
		W = + 55.2133 [3/100%]
Scheme: Multi-Layer Perceptron -L 0.3 -M 0.2 -N 500 -V 0 -S 0 -E 20 -H a		
Total Instances: 321	Threshold: -0.6935	Sigmoid Node1 w=02.61167TL-0.24276SL-1.03215-0.1290F
Female Instances: 157	Threshold: -3,856	sigmoid Node1 w=4.51532TL+2.7856SL+1.9684FL-0.01875C
Male Instances: 164	Threshold: 1.1740	Sigmoid Node1 w = -0.9640TL-1.3780SL-0.73361FL-0.0496C

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 4: T-LR (Total–linear regression),T-M5P (Total-M5P), T-MLP (Total multi-layer perceptron), F-LR (Female-linear regression), F-M5P (Female-M5P), F-MLP (Female-multi-layer perceptron), M-LR (Male-linear regression), M-M5P (Male-M5P), M-MLP (Male-multi-layer perceptron), (CC-correlation coefficient, MAE-mean absolute error, RMSE-root mean squared error, RAE-relative absolute error, RRSE-root relative squared error).

	T-LR	T -M5P	T-MLP	F-LR	F-M5P	F-MLP	M-LR	M-M5P	M-MLP
CC	0.9473	0.9735	0.9663	0.9498	0.9596	0.9605	0.9777	0.9722	0.9686
MAE	3.1149	2.1571	2.7074	2.811	2.6412	2.5705	3.2179	2.4003	2.9456
RMSE	3.8979	2.7688	3.2568	3.8304	3.4553	3.4779	4.1188	3.0383	3.7867
RAE	30.36%	21.03%	26.39%	26.08%	24.50%	23.84%	32.41%	24.18%	29.67%
RRSE	31.91%	22.66%	26.66%	31.11%	28.06%	28.24%	32.72%	24.50%	30.08%

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-MLP, M5P, 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 Gamsız (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); Gutterez-Estrada, *et al*., (2008); Türeli-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

References
Akyol, O. and Gamsız, K., 2011. Age and growth of adult gilthead seabream (*Sparus aurata* L.) in the Aegean Sea. *Journal of the Marine Biological Association of the United Kingdom*, 91(**6**), 1255–1259.

Arechavala-Lopez, P., Sanchez-Jerez, P., Bayle-Sempere, J.T., Sfakianakis, D.G. and Somarakis, S., 2012. Discriminating farmed gilthead sea bream *Sparus aurata* and European sea bass *Dicentrarchus labrax* from wild stocks through scales and otoliths. *Journal of Fish Biology*, 80, 2159−2175.

Bahmanzadegan, A.R., Lari K., Fatemi M.R. and Azarsina F., 2013. The pattern determination of sea surface temperature distribution and chlorophyll a in the Southern Caspian Sea using SOM Model. *Iranian Journal of Fisheries Sciences,* 12(**1**), 105-114.

Baran, P., Lek, S., Delacoste, M. and Bleaud, A., 1996. Stochastic models

that predict trout population densities or biomass on macrohabitat scale. *Hydrobioloa*, 337, 1-9.

Brosse, S., Guegan, J., Tourenq, J. and Lek, S., 1999. The use of artificial neural networks to assess fish abundance and spatial occupancy in the littoral zone of a mesotrophic lake. *Ecological Modelling*, 120(**2-3**), 299-311.

Brosse, S., Lek, S. and Townsend, C.R., 2001. Abundance, diversity, and structure of freshwater invertebrates and fish communities: an artificial neural network approach. N. Z. J. *Marine and Freshwater Research*, 35(**1**), 135–145.

Brosse S. and Lek S., 2002. Relationship between environmental characteristics and the density of age-0 Eurasian Perch *Perca fluviatilis* in littoral zone of lake: a nonlinear approach. *Transactions of the American Fisheries Society*, 131, 1033–1043.

Carpene, E., Martin, B. and Libera, L.D., 1998. Biochemical differences in lateral muscle of wild and farmed gilthead sea bream (*Sparus aurata* L.). *Fish Physiology and Biochemistry*, 19, 229-238.

Chaouı, L., Kara, M.H., Faure, E. and Quıgnard, J.P., 2006. Growth and reproduction of the gilthead seabream *Sparus aurata* in Mellah Lagoon (North-Eastern Algeria). *Journal of Marine Science*, 70(**3**), 545-552.

Cherif, M., Zarrad1, R., Gharbı H., Mıssaouı, 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.

Crvello, J.F., Landers, Jr.D.F., and Keser, M., 2005. The contribution of egg-bearing female American lobster (*Homarus americanus*) populations to lobster larvae collected in Long Island sound by compariosn of microsatellite allele frequencies. *Journal of Shellfish Research*, 24(**3**), 831-839.

Engelhard, G.H., Dieckmann, U. and Godo, O.R., 2003. Age at maturation predicted from routinon scale measurement in Norwegian spring-spawning herring (*Clupea harengus*) using discriminant and neural network analyses. *ICES Journal of Marine Science*, 60, 304-313

Engelhard, G.H. and Heino M., 2004. Maturity changes in Norwegian spring-spawning herring before, during, and after a major population collapse. *Fisheries Research*, 66, 299–310.

Gutterez-Estrada, J.C., Vasconcelos, R. and Costa, M.J., 2008. Estimating fish community diversity from enviromental features in the Tagus estury (Portugal): Multiple linear regression and artificial neural network approaches. *Journal of Applied Ichtyology*, 24, 50-162.

Haddon, M., 2001. Modelling and quantitative methods in fisheries. Chapman and Hall/CRC. 406 P.

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.

Hansen, M., Kenchıngton, E. and Nıelsen, E.E., 2001. Assigning indiviual fish to population using microsatellite DNA markers. *Fish and Fisheries*, 2, 93-112

Haralabous, J. and Georgakarakos, S., 1996. Artificial neural networks as a tool for species identification of fish schools. *ICES Journal of Marine Science*, 53, 173–180.

Hardman-Mountford, N.J., Richardson, A.J., Boyer, D.C., Kreiner, A. and Boyer, H.J. 2003. Relating sardine recruitment in the Northern Benguela to satellite-derived sea surface height using a neural network pattern recognition approach, *Progress in Oceanography*, 59(**2–3**), 241–255.

Hwanh, K., Aokı, I., Komatsu, T., Ishızakı, H. and Shıbata, I., 1996. Forecasting fot the catch of jack mackerel in the Komekami set net by a neural network. *Bulletin of the Japanese Society of Scientific Fisheries*, 60, 136-142.

Joy, K.M. and Death, R.G., 2004. Predictive modelling spatil mapping of freshwater fish and decapod assemblages using GIS and neural networks. *Freshwater Biology*, 49(**5**), 1306-1052

Laffaılle, P., Feunteun, E., Baısez, A., Robinet, T., Acou, A., Lagault, A. and Lek, S., 2003. Spatial organisation of European eel (*Anguilla anguilla* L.,) in a small catchment. *Ecology of Freshwater Fish*, 12(**4**), 254-264.

Laffaılle, P., Baısez, A., Rıgaud, C. and Feunteun, E., 2004. Habitat preferences of different European eel size classes in a reclaimed marsh: a contribution to species and ecosteme conservation. *Wetlands,* 24, 642-651.

Lee, K.Y., Chung, N. and Hwang, S., 2016. Application of an artificial neural network (ANN) model for predicting mosquito abundances in urban areas. *Ecological Informatics*, 36, 172–180.

Lek, S., Delacosta, M., Barab, P., Dımopoulus, I., Lauga, J.C. and Aulagnıer, S., 1996. Application of neural networks to modeling nonlinear relationships in ecology. *Ecological Modeling*, 90(**1**), 39-52.

Lek, S. and Guegan, J.F., 1999. Artificial neural networks as a tool in ecological modelling an introduction. *Ecological Modelling*, 120, 65–73.

Manoj, M., Gandhı, R.S., Raja, T.V., Ruhıl, A.P., Sıngh, A. and Gupta, A.K., 2014. Comparison of artificial neural network and multiple linear regression for prediction of first lactation milk yield using early body weights in Sahiwal cattle. *Indian Journal of Animal Sciences,* 84(**4**), 427–430.

Moutopoulos, D.K., Koukou, K., Vavarouta, V., Ramfos, A. and Katselıs, G., 2011. Investigation of length-weight relationships for 10 commercial fish species as a possible trophic state index of coastal

Lagoons. *Acta Adriatica*, 52(**2**), 261 – 268.

Mastrorillo, S., Lek, F., Dauba, A. and Belaud, 1997. The use of artificial neural networks to predict the presence of small-bodied fish in a river. *Freshwater Biology*, 38, 237 246

Oakes, R.M., Gido, K.B., Falke, J.A., Olden, J.D. and Brock, B.L., 2005. Modelling of stream fishes in the Great Plains, USA. *Ecology of Freshwater Fish*, 14, 361–374.

Park, Y.S., Verdonschot, P.F.M., Chon, T.S. and Lek, S., 2003. Patterning and predicting aquatic macro invertabrate diversities using artificial neural network. *Water Research*, 37, 1749-1758.

Scardi, M. and Harding, L., 1999. Developing an emprical model of phytoplankton primary profuction a neural network case study. *Ecological Modelling*, 120(**2-3**), 213-223

Türeli-Bilen, C., Kokcu, P. and Ibrıkcı, T., 2011. Application of artificial neural networks (ANNs) for weight predictions of blue crabs (*Callinectes sapidus* RATHBUN, 1896) using predictor variables. *Mediterranean Marine Science*, 12(**2**), 439-446.

Witten, I., Frank, E., and Hall, M., 2011. Data mining: Practical machine learning tools and techniques. 3rd edition, Morgan Kaufmann Publishers. 664 P.