Mediterranean Marine Science

Vol 12, No 2 (2011)

Application of Artificial Neural Networks (ANNs) for Weight Predictions of Blue Crabs (Callinectes sapidus RATHBUN, 1896) Using Predictor Variables

C. TURELI BILEN, P. KOKCU, T. IBRIKCI

doi: 10.12681/mms.43

Mediterranean Marine Science
Indexed in WoS (Web of Science, ISI Thomson) and SCOPUS
The journal is available on line at http://www.medit-mar-sc.net

Research Article

Application of Artificial Neural Networks (ANNs) for Weight Predictions of Blue Crabs (*Callinectes sapidus* RATHBUN, 1896) Using Predictor Variables

C. TURELI BILEN, P. KOKCU and T. IBRIKCI

[1] Department of Basic Science, Fisheries Faculty, Çukurova University, Adana, Turkey
[2] Department of Electrical and Electronics Engineering, Engineering and Architecture Faculty, Çukurova University, Adana, Turkey

Corresponding author: cantur@cu.edu.tr

Received: 19 March 2010; Accepted: 23 September 2011; Published on line: 14 October 2011

Abstract

An evaluation of the performance of artificial neural networks (ANNs) to estimate the weights of blue crab (*Callinectes sapidus*) catches in Yumurtalık Cove (Iskenderun Bay) that uses measured predictor variables is presented, including carapace width (CW), sex (male, female and female with eggs), and sampling month. Blue crabs (n=410) were collected each month between February 1997 and January 1998. Sex, CW, and sampling month were used and specified in the input layer of the network. The weights of the blue crabs were utilized in the output layer of the network. A multi-layer perception architecture model was used and was calibrated with the Levenberg Marguardt (LM) algorithm. Finally, the values were determined by the ANN model using the actual data. The mean square error (MSE) was measured as 3.3, and the best results had a correlation coefficient (R) of 0.93. We compared the predictive capacity of the general linear model (GLM) versus the ANN for the estimation of the weights of blue crabs from independent field data. The results indicated the higher performance capacity of the ANN to predict weights compared to the GLM (R=0.97 vs. R=0.95, raw variable) when evaluated against independent field data.

Keywords: Artificial Neural Network; Blue Crab; *Callinectes sapidus*; Prediction of Weight; Yumurtalık Cove (Iskenderun Bay, Northeastern Mediterranean – Turkey).

Introduction

The blue crab, *Callinectes sapidus*, is endemic in the western Atlantic basin. It is a widely distributed, estuarine-dependent species that ranges from South America to the Caribbean and the Gulf of Mexico and to the eastern seaboard of North America as far north as New England (WILLIAMS, 1974). Throughout this range, the blue crab is an important component of estuarine food webs (HINES, 2007); in addition to its ecological importance, blue crabs support the most important commercial and recreational fisheries. Within the United States, commercial fisheries exist in coastal states from

Texas to New York (MILLER *et al.,* 2005).

In addition to this endemic range, the species have become established as non-native species in the Mediterranean basin (HOLTHUIS, 1961), as reported in 15 lagoon systems along the Mediterranean coast of Turkey (ENZENROSS *et al.,* 1997). Due to its high economic value, the importance of commercial and recreational fishing of the blue crab is increasing; moreover, the ecological significance of the blue crab is arousing more and more interest along the Mediterranean coast of Turkey (GOKCE *et al.,* 2006).

Commonly used terms for dimensions of crustaceans include carapace width, body length, body width and wet weight. It can be convenient to be able to convert to the desired length measurement when only one of the other length measurements is known, and the length-weight regression may be used to estimate length from weight (AYDIN & AYDIN, 2011). Information about the individual body weight-length/width relationships in population characteristics is generally of great importance for estimating the population size of the stock for the purpose of its exploitation. The weight is used in a given geographic region to observe species growth. Weight data and weight-carapace length/width relationships constitute useful and standard results of sampling studies (ATAR & SECER, 2003). Relationships between variables in these relationships (i.e., weight-carapace length/width relationships) are often non-linear or unknown. The independent variables are transformed by linear regression. Despite these manipulations, the results often remain disappointing and offer poor predictive value. However, ANNs are nonlinear-type models. They do not necessitate the transformation of variables and can yield better results (LEK *et al.,* 1996; YANEZ *et al.,* 2010). Traditional methods of statistical analysis (i.e., linear regression models, both single and multiple) may be inadequate for quantification (MARAVELIAS *et al.,* 2003).

ANNs offer a promising alternative to traditional statistical approaches for predictive modeling when non-linear patterns exist (JOY & DEATH, 2004). ANNs could be used to substitute for regression analyses, particularly those involving non-linear relationships (MASTRORILLO *et al.,* 1997).

Recently, ANNs have been used in biology and in various disciplines of aquatic ecology rather than in physical or chemical sciences. Applications of ANNs have included predicting the distributions of demersal fish species (OLDEN & JACKSON, 2002; MARAVELIAS *et al.,* 2003), predicting the presences of small-bodied fish in a river (MASTRORILLO *et al.,* 1997), predicting aquatic macro-invertebrate diversities (PARK *et al*., 2003), modeling population dynamics of aquatic insects (OBACH *et al*., 2001) and modeling and spatially mapping freshwater fish and assemblies of decapods (JOY & DEATH, 2004). Most of these studies have demonstrated that ANNs performed better than classical linear and non-linear modeling methods, such as linear regression and generalized additive models (BROSSE *et al.*, 2001).

In this paper, we assess the capacity of ANNs to predict the weights of individual blue crabs as related to three predictor variables (i.e., carapace width (CW), sex, and sampling month) in Yumurtalık Cove-Iskenderun Bay. Finally, the application of the ANN was compared to a conventional linear approach (i.e., the general linear model (GLM)). Model-predicted and observed values are compared by different statistical parameters.

Materials and Methods

This study was conducted in Yumurtalık Cove (Iskenderun Bay, Northeastern Mediterranean – Turkey). Blue crabs were sampled monthly in the Yumurtalık Cove between February 1997 and January 1998 using small shrimp trawls (with a 15 m head rope length on the bottom trawl and a 14 mm mesh cod end). Each monthly survey sampled crabs at 6 randomly assigned stations. Trawl shots of approximately 30 minutes were performed at depths between 5 and 30 m (Fig. 1). All crabs collected during the sampling were transported to the laboratory. In the laboratory, sex was determined by the differences in external anatomy between males and females (i.e., the shape and color of the abdomen) (JIVOFF *et al.,* 2007). The specimens were enumerated, weighed (W) and measured for carapace width (CW, to the nearest mm) using Vernier calipers.

In this work, ANN and conventional (GLM) techniques were implemented. Each independent variable was tested, and the function that returned the best correlation with the dependent variable was retained. The GLM was used as a check. The global statistical significances of the relationships between the dependent variable y and the independent variables were analyzed using the analysis of variance (ANOVA, $\mu = 0.05$). Calculations were performed using the SPSS 10 software.

ANNs are simulations of biological nervous systems using mathematical models. They are networks with simple processor units, interconnections, adaptive weights and scalar measurement functions (e.g., summation and activation functions) (RUMELHART *et al.*, 1986).

For ANN, a multilayer feed-forward neural network was used. A schematic representation of a typical ANN is shown in Figure 2 and consists of 4 interconnected layers of ‘nodes’ or ‘neurons’, including an input layer containing 1 node per independent variable (i.e., sampling month, CW, and sex of the blue crab), the first and second hidden layers, and finally, an ‘output layer’ with 1 node (i.e., the weight of the blue crab). Each layer connected to another layer with interconnections and adaptive weight values. The neurons were connected to next-layer neurons with adjustable weights. Training the network consisted of using a training data set to adjust the connection weights to minimize the error between observed and

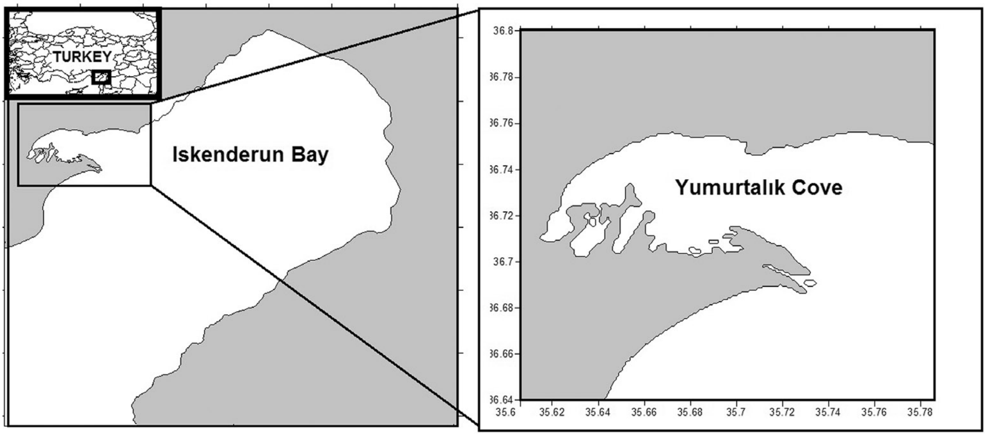

Fig. 1: Yumurtalık Cove (Iskenderun Bay, Northeastern Mediterranean-TURKEY).

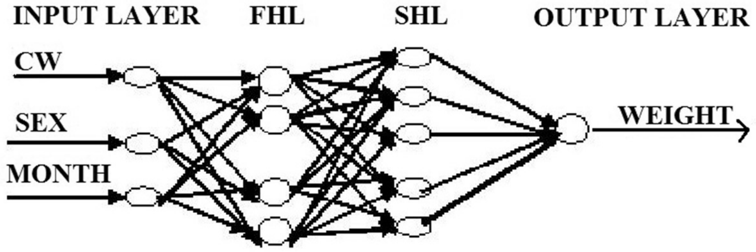

Fig. 2: An ANN consisted of an input layer with 3 nodes (predictor variables), two hidden layer (first hidden layer (FHL), second hidden layer (SHL)) and an output layer with 1 node (Weight) to be predicted.

predicted values. This training was performed according to a back-propagation algorithm (RUMELHART *et al.*, 1986). This is a second-order nonlinear optimization algorithm with very fast convergence and is recommended by several authors (MARAVELIAS *et al.*, 2003; GUTIERRAZ-ESTRADA *et. al.,* 2008). The computational program was written in Matlab (MATLAB, 2006).

Neural Network Toolbox of Matlab was used for the ANN calculations. The Matlab functions were used for "training", "testing", and "validation". They were used randomly: 70% in training, 15% in testing, and 15% in the validation of the 410 blue crabs. The training data set had 286 pattern values for training the network regarding different numbers of epochs. For training, the "*Trainlm*" training function was used, which is a network training function that updates weights and bias values according to the *Levenberg-Marquardt (LM)* optimization, which is the most widely used optimization algorithm (MATLAB, 2006).

LM outperforms simple gradient descent and other conjugate gradient methods in a wide variety of problems. Another function that can be used is the *"Learngdm"* function, which is the gradient descent with momentum weight and bias learning functions. The MSE function is used as the "Network Performance Function", which measures the performance of the network according to the mean of squared errors with *Equation* (MATLAB, 2006). The equation could be written as:

$$E = \frac{1}{n}\sum_{i=1}^{n}\| n_i - t_i \|^2$$

Where "o" and "*t*" are output and desired output, respectively. The W and b values will be updated until the epoch number is reached.

In addition, the coefficient correlation between the actual and estimated weight values for training and testing functions were determined as described previously (BILGILI *et al.*, 2007).

Results

The study classified blue crabs according to sex as females, males, and females with eggs. The collected sample set had 180 females, 150 males, and 80 females with eggs. The ANN used crabs, sex, carapace width, and sampling month as input data to predict the weights of blue crabs with three different predictor values. The ANN was trained with the Levenberg-Marquardt (LM) algorithm for 3000 epochs; then, its results were tested (Fig. 3). As seen in Figure 3, the prediction of the ANN is consistent with the actual values.

The performance value, such as the mean square error (MSE) and the R of the relationships between observed and estimated values for training and testing data are given in Table 1. A significant result shown in this table is that the MSE and R values of the testing data were better than those of the training data. The MSE was measured as 3.3 for validation. Moreover, because of the validation samples, it could be clearly understood how the weights reached the closest calculated results for the blue crabs. The validation results show that the identified ANN weights were correct for testing data.

Table 2 shows the standardized correlation coefficients (R values) and the mean square errors (MSEs) of the relationships between observed and estimated values obtained for both ANN and linear regression (LR). The ANN model demonstrated a higher predictability than did the regression. On the basis of high R and low MSE calculations, the best performances of the ANN were the predicted weights of the blue crabs.

Discussion

In this study, the ANN was trained to estimate the weights of blue crabs based on month caught and prediction data such as sex and carapace width. Growth in the benthic species tends toward increases in size and weight or relatively greater increases in weight rather than in size. The relative weight of the benthic species may be an important adaptive factor, and strictly benthic species such as the brachyuran crab (COMPANY& SARDA, 2000). These researchers claimed that the size-weight relationships differed significantly as functions of the life habits of the species. The most frequently used dimensions for crustaceans include carapace length, width, body length, body width, and wet weight (AYDIN & AYDIN, 2011). It can be convenient to be able to calculate a particular length measurement when only one of the other length measurements is known. The length-weight regression may be used to estimate lengths from weights and vice-versa (ATAR & SECER, 2003; AYDIN & AYDIN, 2011).

Regressions or ANNs can be used to

Table 1
Correlation coefficients (R) and Mean Square Error (MSE) of the relations between observed and estimated values for the training and testing set for blue crabs.

Results	**N**	**MSE**	**R**
Training	286	8.9	0.93
Testing	62	7.3	0.94
Validation	**62**	**3.3**	**0.93**

Table 2
Correlation coefficients (R) and Mean Square Error (MSE) of the relations between observed and estimated values for linear regression and ANN's for blue crabs.

Models	**MSE**	**R**
Linear Regression	53.6	0.94
Artificial Neural network	18.8	0.97

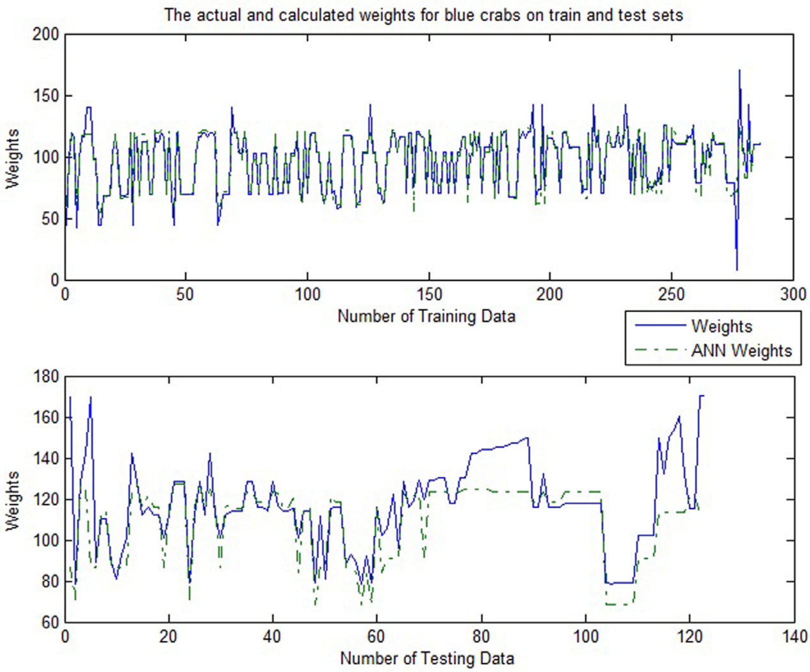

Fig. 3: The actual and calculated weights for blue crabs on train and test sets.

predict the weights of blue crabs from the sampled prediction variables. This study represents a successful attempt to validate the prediction of individual blue crab weights by the application of a neural network to the prediction variables. The return of low error levels in the testing and validation sets of the model suggests that the objective was accomplished. The ANN technique successfully returned an R of 0.93, and the MSE was calculated as 3.3 (Table 1); the prediction of the ANN was consistent with the actual values (Fig. 3). The validation results show that the weights obtained from the ANN were correct for testing data. It can be concluded that the ANN has potential for predicting the weights of blue crabs, depending on the selection of references for measuring prediction variables.

The ANN returned a much higher correlation coefficient between the observed and the predicted values than did the LR. We preferred to use the MSE to assess the model prediction performance. The MSE value of the ANN was clearly lower than it was for the LR (Table 2).

The fact that the ANN provides a better model was highlighted by better predictions for lower values, the normality of the residuals and their independence from the predicted variable. Several authors have reported greater performances of ANNs compared to linear regressions (Sun, 2009). The advantage of ANNs over multiple linear regression (MLR) models is that ANNs can directly take into account any non-linear relationships between the dependent variables and each independent variable (LEK *et al.,* 1996; SUN *et al.,* 2009). ANNs have another advantage in that the ANN modeling approach is fast and flexible (BROSSE, *et al.,* 1999). In this study, the ANN has demonstrated a new and alternative approach for its application in predicting the relative growth and weight of benthic species, especially strictly benthic species such as brachyuran crabs.

It can be concluded that the ANN method is a powerful tool for predicting missing weights of blue crabs using other field data Because weight is a limiting morphological factor for decapod crustaceans, the advantage of this model is that if the required predictor data, such as carapace width, sex and sampling month, are available, the weights of the blue crabs can also be predicted quickly and accurately.

References

ATAR, H. H. & SECER, S., 2003. Width/Length-Weight relationships of the Blue crab (*Callinectes sapidus* Rathbun 1896) population living in Beymelek lagoon lake. *Turkish Journal of Veterinary & Animal Sciences*, 27: 443-447.

AYDIN, I. & AYDIN, C., 2011. Length-Length and length-weight relationships in *Nephrops norvegicus* (Linnaeus, 1758) from the Aegean Sea. *Mediterranean Marine Science*, 12 (1): 121-128.

BILGILI, M., SAHIN, B. & YASAR, A., 2007. Application of artificial neural networks for the wind speed prediction of target station using reference stations data. *Renewable Energy*, 32 (14): 2350-2360.

BOSSE, S., LEK, S. & TOWNSED, C., 2001. Abundance, diversity, and structure of freshwater invertebrates and fish communities: an artificial neural network approach. *New Zealand Journal of Marine & Freshwater Research*, 35 (1): 135-145.

BROSSE, S., GUEGAN, J., TOURENQ, J. & 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.

CAMPELL, N., MACKENZİE, K., ZUUR, A.F., IENO, E.N. & SMITH, G.M., 2007. Fish stock identification through neural network analysis of parasite fauna.p.450-462. In: *Analysis Ecological Data*. Zuur, A.F., Ieno, E.N & Smith, G.M. (Eds). Springer-Business media. NY, USA.

COMPANY, J.B. & SARDA, F., 2000. Growth parameter of deep-water decapod crustaceans in the Northwestern Mediterranean Sea: a comparative approach. *Marine Biology*, 136 (1): 79-90.

ENZENROSS, V.L., ENZENROSS, R. & BINGEL, F., 1997. Occurrence of Blue crab, *Callinectes sapidus* (Rathbun 1896) on the Turkish Mediterranean and adjacent coast and its size distribution in the Bay of Iskenderun. *Turkish Journal of Zoology*, 21 (2): 113-122.

GOKCE, G., ERGUDEN, D., SANGUN, L., CEKIC, M. & ALAGOZ, S., 2006. Width/length-weight and relationships of the blue crab (*Callinectes sapidus*, Rathbun, 1896) population living in Camlik Lagoon Lake (yumurtalik). *Pakistan Journal of Biological Sciences*, 9 (8): 1460-1464.

GUTIERREZ-ESTRADA, J.C., VASCONCELOS, R. & COSTA, M.J., 2008. Estimating fish community diversity from environmental features in the Tagus estury (Portugal): Multiple Linear Regression and Artificial Neural Network approaches. *Journal of Applied Ichthyology,* 24: 150-162.

HINES, A.H., 2007. Ecology of juvenile and adult blue crabs. p.565-654. In: *The Blue Crab Callinectes sapidus*. V.S. Kennedy & L.E. Cronin (Eds). Maryland Sea Grant College, College Park, MD.

HOLTHUIS, L.B., 1961. Report on a Collection Crustacea decapoda and Stomatopoda from Turkey and Balkans. *Zoologische Verhandelingen,* 47 (1): 1-67.

JIVOFF, P., HINES, H.A. & QUACKENBUSH, S.L., 2007. Reproduction biology and embryonic development.p.255-286. In: *The Blue Crab Callinectes sapidus.* V.S.Kennedy & L.E. Cronin (Eds). Maryland Sea Grant College, College Park, MD.

JOY, K.M. & DEATH, R.G., 2004. Predictive modelling and spatial mapping of freshwater fish and decapod assemblages using GIS and neural Networks. *Freswater Biolgogy*, 49 (8): 1306-1052.

LEK, S., DELACOSTA, M., BARAN, P., DIMOPOULOS, I., LAUGA, J. & AULAGNIER, S., 1996. Application of neural networks to modeling nonlinear relationships in ecology. *Ecological Modelling,* 90 (1): 39-52.

MARAVELIAS, C.D., HARALABOUS, J. & PAPACONSTANTINOU, C., 2003. Predicting demersal fish species distributions in the Mediterranean Sea using artificial neural networks . *Marine Ecology. Progress Series,* 255: 249-258.

MASTRORILLO, S., LEK, S., DAUBA, F. & BELAUD, A., 1997. The use of artificial neural networks to predict the presence of small-bodied fish in river. *Freshwater Biology* 38: 237-246.

MATLAB, 2006. *The MathWorks, Inc*. Matlab Help. MATLAB.

MILLER, T.J, MARTELL, S.J.D., BUNNELL, D.B., DAVIS, G., FEGLEY, L.A., SHAROV, A.F., BONZEK, C.F., HEWITT, D.A., HOENIG, J.M. & LIPCIUS, R.N., 2005. *Stock asessment for blue crab in Chesapeake Bay 2005.* Final Report. Ref (UMCES) CBL 05-077. Chesapeake Biological Laboratory, Solomon, MD, 162 pp.

OBACH, M., WAGNER, R., WERNER, H. & SCHMIDT, H-H., 2001. Modelling population dynamics of aquatic insects ith artificial neural networks. *Ecological Modeling*, 146: 207-217.

OLDEN, J.D. & JACKSON, D.A., 2002. A comparison of statical approches for modelling fish species distiributions. *Freshwater Biology*, 47: 1976-1995.

PARK, Y.-S., VERDONSCHOT, P.F.M., CHON, T.-S. & LEK, S., 2003. Patterning and predicting aquatic macro invertabrate diversities using artificial neural network. *Water Research*, 37: 1749-1758.

RUMELHART, D.E., HINTON, G.E. & WILLIAMS, R.J., 1986. Learning internal representations by error propagation in Parallel Distributed Processing. *Explorations in the Microstructure of Cognition MIT Press,* 1: 318-362.

SACMA, S., 2009. *Estimation of the buckling Loads of helical coil springs with the help of the neural networks.* MSc Thesis. Cukurova University, Turkey.

SUN, L., XIAO, H., LI, S. & Yang, D., 2009. Forecating fish stock recruitment and planning optimal harvesting strategies by using neural network. *Journal of Computers,* 4 (11): 1075-1082.

WILLIAMS, A.B., 1974. The swimming crabs of the genus *Callinectes* (Decapoda: Portunidae). Fishery Bulletin, 72: 685-798.

YANEZ, E., PLAZA, F., GUTIERREZ-ESTRADA, J.C., RODRIQUEZ, N., BARBIERI, M.A., PULIDO-CALVO, I. & BORQUEZ, C., 2010. Anchovy (*Engraulis ringens*) and sardine (*Sardinops sagax*) abundance forecast off northern Chile: A multivariate ecosystem neural network approach. *Progress in Oceanography,* 87: 242-250.