Int. J. Bio-Inorg. Hybr. Nanomater., 11(1): 49-56, Spring 2022

Linear and Nonlinear QSAR models on Effective Biological and chemical Parameters of water in Abundance Oligochaetes (P. hammoniensis, O. serpentin and B. sowerbyi) in the Anzali International Wetland

R. Rajabei Nezhad**, R. Sayyadi Kord Abadi, F. Nazarhaghighi

[1] *Department of Fisheries, Bandar Anzali Branch, Islamic Azad University, Bandar Anzali, Iran*
[2] *Department of Chemistry and Chemical Engineering, Rasht Branch, Islamic Azad University, Rasht, Iran*
[3] *Administrative Offices of Fisheries, Ardabil, Iran*

Received: 28 September 2021; Accepted: 30 November 2021

ABSTRACT: Quantitative structure-activity relationship (QSAR) of the Biological and chemical properties of water in the Anzali International Wetland was estimated by means of multiple linear regression (MLR), artificial neural network (ANN), simulated annealing (SA) and genetic algorithm (GA) techniques. The obtained results from MLR-MLR, MLR-SA, SA-ANN and GA-ANN approaches were compared and GA-ANN combination showed the best performance according to its correlation coefficient (R^2) and mean sum square errors (RMSE). A high predictive ability was observed for the GA-ANN model; with the root mean sum square errors (RMSE) of 0.0054 in P. hammoniensis and 0.0020 in B. sowerbyi and 0.001 in O. serpentina, respectively. From GA-ANN simulations, it was found that the total soluble solids (TSS), dissolved oxygen (DO), total nitrogen (TN), NH3 concentration, Sodum chloride (Sali), Nitrat (NO_3), Total dissolved solids (TDS), total organic material (TOM), are the most important Biological and chemical parameters of water that affect abundance of the P. hammoniensis, O. serpentin, B. sowerbyi.

Keywords: *Genetic algoritm, MLR, Oligochaetes, QSAR.*

INTRODUCTION

QSAR study is modelling and optimization approaches that relates the Biological and chemical properties of water to the abundance of Oligochaetes. The data sets for QSAR modeling, which contain the Biological and chemical properties of water, are generated by experimental scientists and available in various data sources [1-3]. There are several variable selection QSAR methods including multiple linear regression (MLR), genetic algorithm (GA), simulated Annealing algorithm (SA) and so on [4-8]. Oligochaete is an incredibly diverse group, with at least 5,000 species that have been described across the world. Oligochaetes provide important functions within ecosystems. One reason there is such an impressive diversity of oligochaetes is that they have adapted to living in a wide range of conditions, including terrestrial, marine, and freshwater ecosystems. In freshwater and marine habitats, oligochaetes live within the substrate, or the sediment,

(*) Corresponding Author - e-mail: reza_igl@yahoo.com

covering the bottom. Here, they burrow within the sediment and feed on dead organic matter [9]. These small relatives of earthworms are widely distributed in different soils and land use forms, where they play an important role due to their burrowing activity, their fecal pellet production as well as their transport, ingestion and mixing of mineral and organic soil particles [10]. In present work, MLR and ANN modelling tools coupled with SA and GA optimization techniques were used to find the best set of physico-chemical parameters of the water (depth, temperature, dissolved oxygen, and pH) that correlate the abundance of Oligochaetes in the Anzali International Wetland.

METHODS

Aquatic Oligochaetes were collected twice a season from 13 sites in the Anzali Wetland from August 2012 to June 2013. The samples were taken with a bottom grab (0.04 m2) from the surface layer of bottom sediments among the submerged macrovegetation. During each sampling period, water temperature was measured with a thermometer with a sensitivity of 0.1°C, dissolved oxygen was measured with an oxygen meter WTW-OXI 330/SET, and pH was determined with a pH meter WTW pH 330/SET-1. Details of Sampling and sample processing of Limnodrilus claparedeianus and Limnodrilus hoffmeisteri were given in ref. [11, 3]. The Sodum chloride in water (Salin), dissolved oxygen (DO), pH, TOM(total organic material), turbidity (turb), electrical conductivity (EC), total soluble solids (TSS), total dissolved solids (TDS), total phosphorus (TP), NO_2, NO_3, NH_3, Viruses and Worms counting (Vw), total nitrogen (TN), Biochemical oxygen demand (BOD), chemical oxygen demand (COD), chlorophyll-a (chla), Photovoltaic (Pv), Non-invasive ventilation (Niv), Part By Weight (Pbw), CrS, PbS, Pressure Sensor (Ps), NiS, Vs were measured at the sampling from 13 sites in the Anzali Wetland in 7 species of aquatic oligochaetes including L. claparedeianus and Limnodrilus hoffmeisteri [11, 3]. The physico-chemical parameters values in Linear and Nonlinear QSAR models were normalized using the equation (1) in Excel program.

$$X_{normalized} = \frac{(X_i - X_{min})}{(X_{max} - X_{min})} \quad (1)$$

SPSS [12] program was employed to select the best physico-chemical parameters. An ideal method is one that has low standard deviation, high correlation coefficient (R^2) and root mean of square error (RMSE) [14], where the RMSE is defined as follows:

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(y_i - y_o)^2}{n}} \quad (2)$$

The correlation coefficient (R^2), root mean of square error (RMSE) were calculated by using the linear method (MLR) approach in the unscramble [13] program. The bio-chemical parameters chosen in the previous primary linear selection were implemented under further screening using QSAR methods including GA-ANN, SA-ANN, MLR-GA as shown in Fig.1. In each run via ANN, the 3-5 bio-chemical parameters selected by the optimization method (GA or SA) were used as the inputs and corresponding values of (abundance of L. claparedeianus) were utilized as the target values (Fig.1). This study also used three neurons in the hidden layer of the ANN approaches 80%, 10% and 10% of data sets in these models were randomly chosen as training, validation and test sets, respectively. The networks were trained by using the TSET members with Levenberg-Marquart Algorithm [14], while Logarithmic sigmoid and linear transfer function were used as the hidden and output transfer function, respectively. Logarithmic sigmoid transfer function is defined as follows:

$$\log \ sigm = \frac{1}{1 + e_{-x}} \quad (3)$$

A genetic algorithm (GA) is a method for solving both constrained and unconstrained optimization problems based on a natural selection process that mimics biological evolution. The algorithm repeatedly modifies a population of individual solutions [15]. SA is a descent algorithm modified by random ascent moves in order to escape local minima which are not global minima. It has been used in statistical physics to choose sample states of a particle system model to efficiently estimate some physical quantities [16]. Matlab 2014a was used for modeling and optimization calculations. A genetic algorithm (GA) is a method for

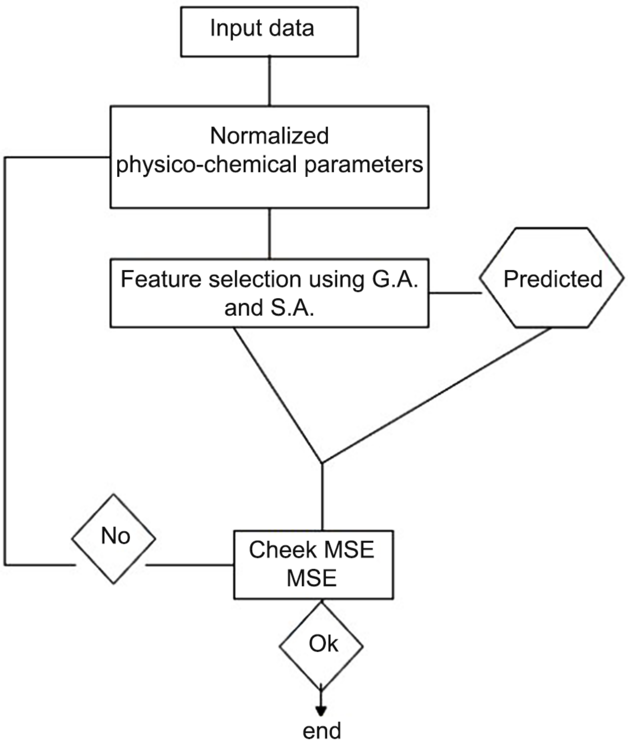

Fig. 1. The employed procedure for finding optimum descriptores of the ANN models.

solving both constrained and unconstrained optimization problems based on a natural selection process that mimics biological evolution. The algorithm repeatedly modifies a population of individual solutions [17]. SA is a descent algorithm modified by random ascent moves in order to escape local minima which are not global minima. It has been used in statistical physics to choose sample states of a particle system model to efficiently estimate some physical quantities [18]. Matlab 2014a was used for modelling and optimization calculations.

RESULTS ANDDISCUSSION

QSAR investigations in Biological and chemical properties of water in the Oligochaetes (P. hammoniensis, O. serpentin, B. sowerbyi) were performed using multiple linear regression (MLR) and artificial neural network (ANN) as modelling tools and simulated annealing (SA) and genetic algorithm (GA) as optimization methods. According to the types of variable section method and feature mapping techniques, these models were shown as MLR-MLR, SA-ANN, MLR-GA and GA-ANN. In MLR-MLR, MLR-PCR and MLR-PLS1 methods, the best physico-chemical parameters were selected using MLR procedure of SPSS [19] software in three steps described in theory and computational methods section. Then the selected physico-chemical parameters were used as input in unscrambles software and statistical parameters were calculated using PCR and PLS1 methods (Tables 2, 4 and 6). The RMSE and the correlation coefficient

Table 1. The best selected bio-chemical parameters of water using QSAR Methods in P. Hammoniensis.

MLR-MLR	SA-ANN	GA-MLR	GA-ANN
COD	COD	TN	TSS
TN	TOM	COD	DO
Sali	EC	Sali	TN

Table 2. Statistical parameters of different linear QSAR models in P. Hammoniensis

QSAR Model	R^2	RMSE
MLR-PLS1	0.2467	0.1378
MLR-MLR	0.2553	0.1370
MLR-PCR	0.3348	0.1561
MLR-ANN	0.5773	0.0107
SA-ANN	0.5323	0.0118
GA-ANN	0.8986	0.0054
GA-MLR	0.5185	0.0122

Table 3. The best selected bio-chemical parameters of water using QSAR Methods in O. Serpentin

MLR-PLS1	SA-ANN	GA-MLR	GA-ANN
sali	TSS	Sali	NH_3
NH_3	BOD	turb	TSS
turb	Sali	NO_3	DO
TOM	TEMP	COD	Sali

Table 4. Statistical parameters of different linear QSAR models in O. Serpentin

QSAR Model	R^2	RMSE
MLR-PLS1	0.5789	0.3182
MLR-MLR	0.632	0.2189
MLR-PCR	0.5103	0.4100
MLR-ANN	0.8375	0.0048
SA-ANN	0.9545	0.0013
GA-ANN	0.9632	0.001
GA-MLR	0.8548	0.0043

Table 5. The best selected bio-chemical parameters of water using QSAR Methods in B. Sowerbyi

MLR-MLR	SA-ANN	GA-MLR	GA-ANN
TDS	NH_3	BOD	TOM
TP	TDS	TDS	TP
NH3	TP	NH_3	NO_3
BOD	TOM	TP	TDS

Table 6. Statistical parameters of different linear QSAR models in B. Sowerbyi

QSAR Model	R^2	RMSE
MLR-PLS1	0.3944	0.1437
MLR-MLR	0.4297	0.1394
MLR-PCR	0.3795	0.1454
MLR-ANN	0.8475	0.0052
SA-ANN	0.9325	0.0023
GA-ANN	0.9754	0.0020
GA-MLR	0.8487	0.0052

(R^2) for predicted the abundance of Oligochaetes in MLR–MLR were found to be [0.1535 0.47015], [0.1140 0.3553], [0.1150 0.3266], [0.2189 0.632] and [0.1394 0.4297], in P. hammoniensis, O. serpentin, B. sowerbyi, respectively. Furthermore, the calculated parameters indicated that MLR–MLR method was better than the two other employed linear methods. The definition of the bio-chemical parameters in the MLR-MLR method is shown in 1, 3 and 5 tables. To establish the SA-ANN and MLR-GA and GA-ANN models, the 28 bio-chemical parameters were fed to the neural network to select the best bio -chemical parameters. Also we used 3 neurons in the hidden layer of the ANN models (Fig.1). The statistical parameters of all QSAR approaches are shown in Tables 2, 4 and 6. In GA-ANN, SA-ANN, MLR-GA methods, 80%, 10% and 10% of data sets were randomly chosen as training, validation and test sets, respectively. The re-

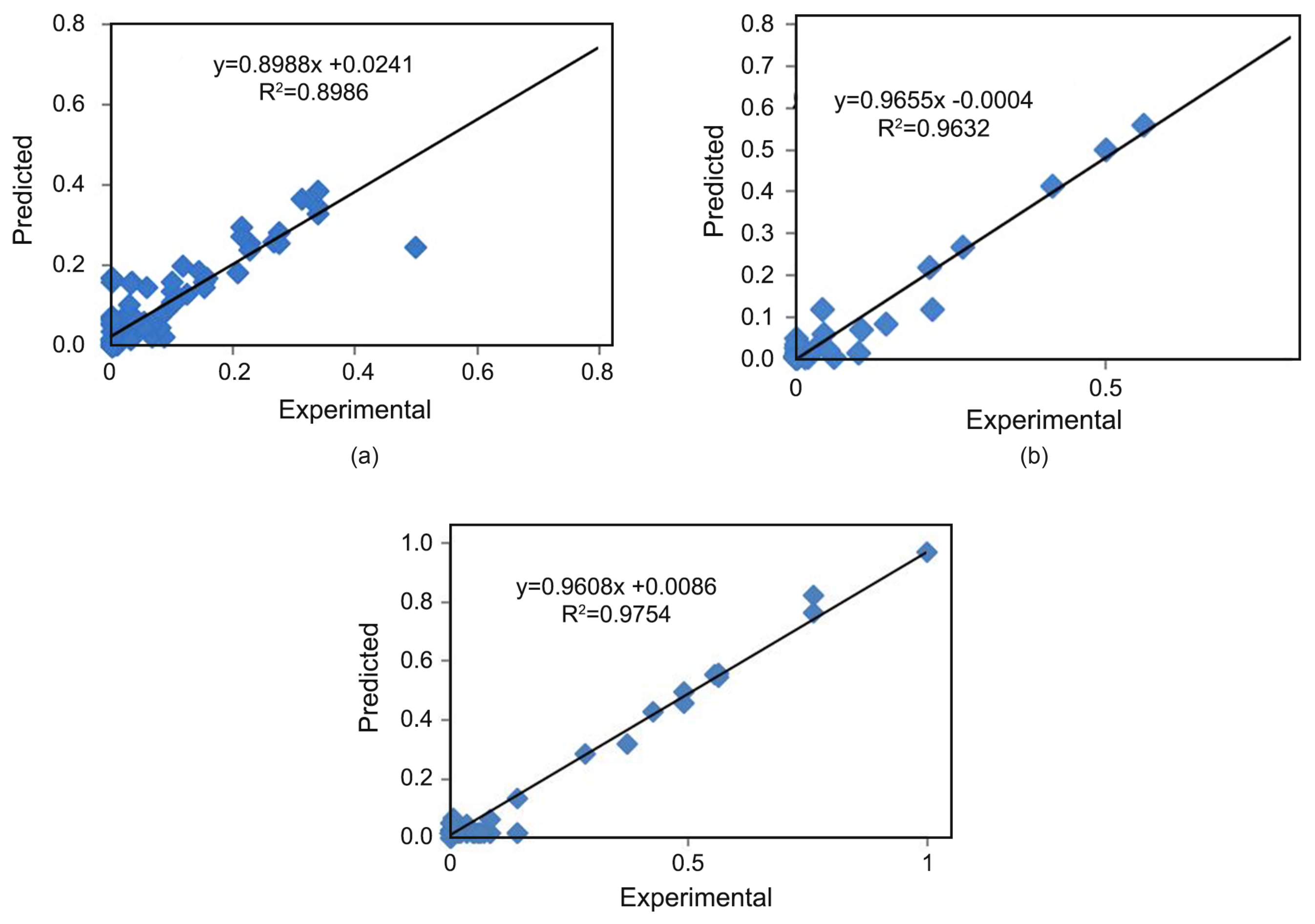

Fig. 2. Correlation between experimental and predicted abundance values calculated by using GA-ANN method in P. Hammoniensis, O. Serpentin and B. Sowerbyi, respectively (A, B, C).

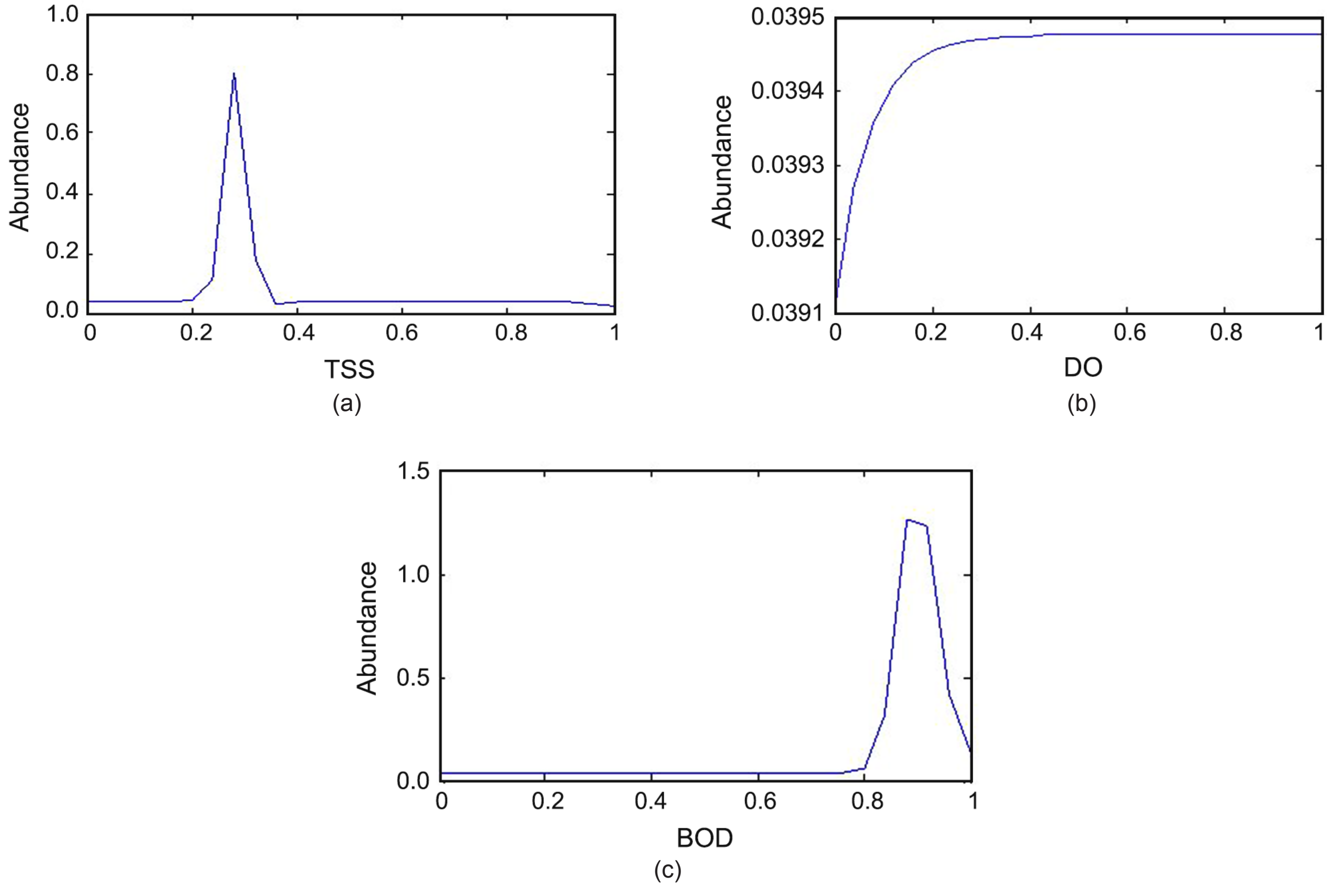

Fig. 3. Experimental values of abundance versus TSS, DO and BOD values in GA-ANN method in P. Hammoniensis.

sults of the QSAR models proved that non-linear feature selection approaches were better than their linear models. The obtained results demonstrated that the GA-ANN method led to better results with good predictive ability than the other QSAR models in the gas phase. Therefore, the selected the best bio -chemical parameters using GA-ANN are discussed here (Tables 1, 3 and 5).

Table 1 shows that in GA-ANN, [TSS, DO.BOD], [NH_3, TSS, DO, Sali] and [TOM, TP, NO_3, TDS] are the best selected Biological and chemical properties of water in P. Hammoniensis, O. Serpentin and B. Sowerbyi, respectively. The plot showing the variation of observed versus

predicted abundance values in P. Hammoniensis, O. Serpentin and B. Sowerbyi are shown in Fig. 2. A good correlation between the calculated and empirical values of abundance can be observed in this figure that approves the appropriateness of the developed model.

The graphs of the found most effective Biological and chemical properties of water in P. Hammoniensis (TSS, DO and BOD), in O. Serpentin(NH_3, TSS, DO and Sali) and in B. Sowerbyi (TOM, TP, NO_3, TDS and Sali) versus the abundance are plotted in Figure 3, 4, 5.

Charts in Fig. 3 showed that the total soluble solids (TSS) and Biochemical Oxygen demand (BOD) increased up to 0.2 and 0.8, the abundance value increased and then, the increased total soluble solids (TSS) and Biochemical Oxygen demand (BOD) decreased the abundance value. Increase in Dissolved oxygen (DO) up to 0.2 resulted in an increase in amount abundance, and then abundance do not change. Fig. 4 showed that an increase in the amount of the NH_3 concentration and total soluble Solids (TSS) brought about a increase in the amount of abundance and then, the increased NH_3 concentration and total soluble solids (TSS) decreased the abundance value. As the Sodum chloride in water (Sali) increased to 0.8, no change in abundance was observed, and then with increase in Sali the abundance value was increased. In increased Dissolved oxygen (DO), constant process

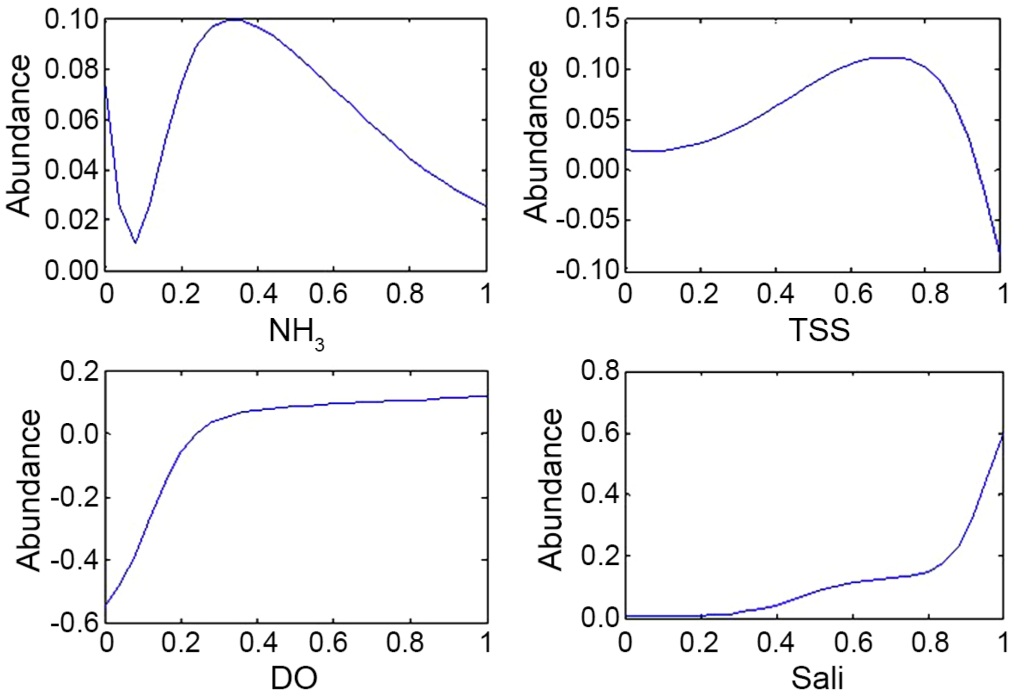

Fig. 4. Experimental values of abundance versus NH_3, TSS, DO and Sali values in GA-ANN method in O. Serpentin.

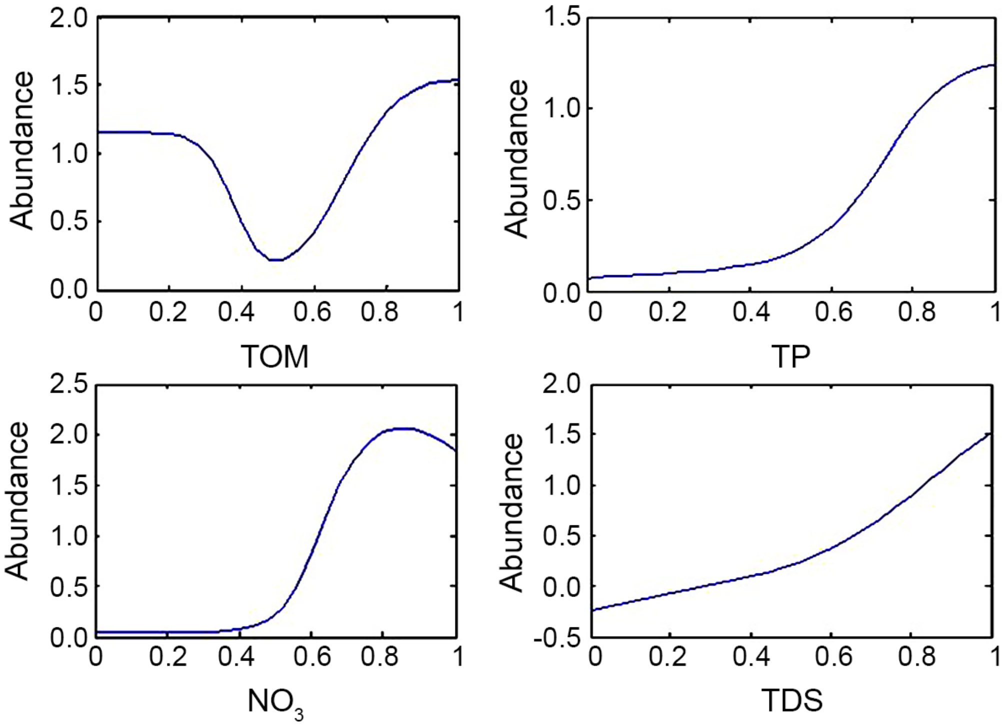

Fig. 5. Experimental values of abundance versus TOM, TP, NO_3 and TDS values in GA-ANN method in B. Sowerbyi.

abundance is seen, and then subsequently increased. Fig. 4 showed that the abundance value in the B. Sowerbyi increases with increasing total phosphorus (TP), total dissolved solids (TDS). By increasing the NO_3 concentration up to 0.8 the response was increased and then reduced. As the total organic material (TOM) increased to 0.5, the abundance value increased and then decreased.

CONCLUSION

The obtained results from QAR models showed that GA-ANN combination were better than the other models used and also proved that total soluble solids (TSS), Biochemical Oxygen demand (BOD) and Dissolved oxygen (DO), NH_3 concentration, Sodum chloride in water (Sali), total phosphorus (TP), NO_3 concentration, total organic material (TOM), total dissolved solids (TDS) were more significant than other Biological and chemical properties of water and predicting abundance values in P. hammoniensis, O. serpentin and B. sowerbyi. It can be concluded that simultaneous use of GA-ANN methods give deeper and more comprehensive knowledge of the effect of Biological and chemical properties of water on the abundance values in P. hammoniensis, O. serpentin and B. sowerbyi.

ACKNOWLEDGEMENTS

The authors gratefully acknowledge the support provided by the Islamic Azad University of Rasht and Administrative Office of Ardabil Fisheries.

REFERENCES

[1] Liao, S.Y.; Chen, J.C.; Qian, L.; Shen, Y.; Zheng, K.C. (2008). QSAR studies and molecular design of phenanthrene-based tylophorine derivatives with anticancer a ctivity. QSAR Comb. Sci., 27, 280-288.

[2] Wang, X.; Yang, W.; Xu, X.; Zhang, H.; Wang, Y. (2010). Studies of benzothiadiazine derivatives as Hepatitis C Virus NS5B polymerase inhibitors using 3D-QSAR, molecular docking and molecular dynamics. Curr. Med. Chem., 17, 2788–2803.

[3] Nazarhaghighi1, F.; Mousavi Nadoushan, R.; Shabanipour, N.; Fatemi, M. R.; A. Mashinchian Moradi, A. (2015). First record of Limnodrilus claparedeianus Ratzel, 1868 (Annelida: Oligochaeta: Tubificidae) from Anzali Wetland, Guilan province, Iran. Caspian J. Environ. Sci., 13, 407-416.

[4] Putta, S., Eksterowicz, J., Lemmen, C., Stanton, R. (2003). A Novel Subshape Molecular Descriptor. J. Chem. Inf. Comput. Sci., 43, 1623.

[5] Gupta, S., Singh, M., Madan, A.K. (1999). Superpendentic index: a novel topological descriptor for predicting biological activity. J. Chem. Inf. Comput. Sci., 39, 272.

[6] Consonni, V., Todeschini, R., Pavan, M. (2002). Structure/Response Correlations and Similarity/Diversity Analysis by GETAWAY Descriptors. 2. Application of the Novel 3D Molecular Descriptors to QSAR/QSPR Studies. J. Chem. Inf. Comput. Sci., 42, 693-705.

[7] Kirkpatrick, S., Gelatt, Jr. C. D., Vecchi, M. P. (1983). Optimization by Simulated Annealing Science, 220, 671–680.

[8] Cerný, V. (1985). Thermodynamical approach to the traveling salesman problem: An efficient simulation algorithm. J. Optim. Theory Appl., 45, 41–51.

[9] Blay, J, Jr.; Dongdem, F. (1996). Preliminary observations on the benthic macrofauna of a polluted coastal lagoon in Ghana (West Africa). Tropical Ecology, 37, 127-133.

[10] Suriani, AL.; Franca, RS.; Pamplin, PAZ.; Marchese, M.; Lucca, JV.; Rocha, O. (2007). Species richness and distribution of oligochaetes in six reservoirs on Middle and Low Tiete river (SP, Brazil). Acta Limnologica. Brasilia, 19, 415-426.

[11] Nazarhaghighi, F.; Timm, T.; Shabanipourd, N.; Nadoushan, R.M. (2014). Oligochaetes (Annelida, Clitellata) in the Anzali International Wetland, north-western Iran. Estonian J. Ecol., 63, 130–144.

[12] Jalali-Heravi, M., Parastar, M.F. (2000). Use of artificial neural networks in a QSAR study of anti-HIV activity for a large group of HEPT derivatives. J. Chem. Inf. Comput. Sci. 40,: 147.

[13] Unscrambler v.9.7, www.camo.com.

[14] Levenberg, K. (1994). A Method for the Solution of Certain Non-Linear Problems in Least Squares. Q. Appl Math., 2, 164–168.

[15] Mauri, A.; Consonni, V.; Pavan, M.; Todeschini,

R. (2006). Dragon Software: an Easy Approach to Molecular Descriptor Calculations. MATCH Commun. Math. Comput. Chem., 56, 237-248.

[16] Kirkpatrick, S., Gelatt, Jr. C. D., Vecchi, M. P., (1983). Optimization by Simulated Annealing Science, 220, 671–680.

[17] Kennedy, C.R. (1965). The distribution and habitat of limnodrilus claparède (oligochaeta: tubificidae). Oikos., 16, 26–38.

[18] Lafont, M.; Vivier, A. (2006). Oligochaete assemblages in the hyporheic zone and coarse surface sediments: their importance for understanding of ecological functioning of watercourses. Hydrobiologia, 564, 171–181.

[19] SPSS, Version 19, (2010). Available at http://www.spssscience.com.

AUTHOR (S) BIOSKETCHES

Reza Rajabei Nezhad, Assistant Professor, Department of Fisheries, Bandar Anzali Branch, Islamic Azad University, Bandar Anzali, Iran, *Email: reza_igl@yahoo.com*

Robabeh Sayyadi Kord Abadi, Assistant Professor, Department of Chemistry and Chemical Engineering, Rasht Branch, Islamic Azad University, Rasht, Iran, *Email: sayyyadi_04@yahoo.com*

Fatemeh Nazarhaghighi, Assistant Professor, Administrative Offices of Fisheries, Ardabil, Iran, *Email: hiva582005@yahoo.com*