Nature Environment and Pollution Technology
An International Quarterly Scientific Journal

p-ISSN: 0972-6268
(Print copies up to 2016)
e-ISSN: 2395-3454

Vol. 21 | No. 3 | pp. 1431-1438 | 2022

Original Research Paper

https://doi.org/10.46488/NEPT.2022.v21i03.052

Open Access Journal

Salinity Prediction at the Bhairab River in the South-Western Part of Bangladesh Using Artificial Neural Network

Khan Md. Rabbani Rasha†
Department of Civil Engineering, Khulna University of Engineering and Technology, Khulna-9203, Bangladesh
†Corresponding author: Khan Md. Rabbani Rasha; rabbanirashakhan@gmail.com

Nat. Env. & Poll. Tech.
Website: www.neptjournal.com

Received: 20-10-2021
Revised: 16-12-2021
Accepted: 23-12-2021

Key Words:
Salinity
Water quality
Artificial neural network
Bhairab river
Correlation

ABSTRACT

Salinity is a significant ecological element that influences the kind of creatures that reside in a water body. Salinity also determines the types of plants that will grow in a water body or on land that is fed by a water body. Three models were generated using an artificial neural network to estimate the salinity concentrations in the Bhairab River. Different combinations of variables were used to train the model using sample values of temperature, pH, turbidity, electrical conductivity (EC), color, total dissolved solids (TDS), total solids (TS), and suspended solids (SS). The performance of the models was determined using the statistical mechanism root mean square error (RMSE), coefficient of correlation (R), and determination coefficient (DC). ANN-2 model had the best performance which had the input variables electrical conductivity (EC), total dissolved solids (TDS), and total solids (TS). These three input variables were highly correlated with salinity. The correlation between the observed and the predicted values was also very high, the coefficient of correlation is 0.98 in validation. The RMSE value was very low for the model training and the value reduced even more after validation to 0.58.

INTRODUCTION

Khulna is the commercial and port city of Bangladesh, situated in the country's southwest. Nonetheless, the region is surrounded by a lot of industries for geographical, cultural, social, and financial purposes. The water supply system in Khulna is entirely reliant on groundwater. Drinking water is not available in the Khulna region due to high salinity and iron levels. Due to the clayey soil, deep boring is needed to obtain drinking water. Since the upper aquifer is salt water, Khulna's water supply scheme is entirely reliant on groundwater sources. The WHO's (World Health Organization) norm threshold for water quality is used as a benchmark. Clean drinking water is a prerequisite for good health.

The Bhairab, which is thought to be older than its parent river, the Jalangi, splits from it at Bhagwangola (Vidhan Sabha constituency) in Murshidabad district, a few miles north of Karimpur near Akheriganj (in West Bengal). This river has been virtually dead for a long time, with its intake from the Jalangi having silted up. The stagnancy of its water is blamed in large part for the poor environment of Meherpur, which lies on its banks. The Khulna-Ichamati and the Kobadak are the two major branches of the Bhairab River. Bangladesh and India are separated by the Khulna-Ichamati River. The cities of Khulna and Jessore are located on the river's side. The river affected the growth of their settlements and culture. The Bhairab and Atai rivers combine to form the Rupsa River, which flows into the Pasur River.

The cumulative concentration of dissolved inorganic ions in water or soil is referred to as salinity (Williams & Sherwood 1994). The ionic activity of a solution can also be expressed in terms of its ability to transmit electrical current (electrical conductivity (EC), measured in Siemens per meter). As a result, EC is commonly used to determine salinity, and the relationship between the two is dependent on water temperature. The following are the salt content classifications for surface waters (Battaglia 1958): freshwater < 0.5 $g.L^{-1}$; oligohaline 0.5-4.0 $g.L^{-1}$; mesohaline water 5-18 $g.L^{-1}$; polyhaline water 18-30 $g.L^{-1}$; euhaline water 30-40 $g.L^{-1}$; hyperhaline water > 40 gL^{-1} (Cañedo-Argüelles et al. 2013). Salinity and the proportions of the above ions come from three places in the absence of anthropogenic factors. (1) Catchment weathering is influenced by both the geology of the catchment and the amount of precipitation. (2) Sea water, though this is a major source of salts only in coastal areas. (3) As a result of seawater evaporation, small quantities of salts dissolve in rainwater. This third source of salt may be a major source of salt in land areas far from the sea (Herczeg et al. 2001). Because of the underlying geology and high evaporation in semiarid and arid areas, rivers and streams are frequently naturally saline; however, irrigation agriculture is one of the major causes of secondary, or anthropogenic, salinization.

One of the main causes of secondary salinization has been identified as irrigation and rising groundwater tables, especially in arid and semi-arid regions of the world where crop production consumes large amounts of water. Salt concentrates and soil water becomes more acidic as crops consume just a fraction of the salt in irrigation water (Maryoung et al. 2014). These salts can be leached out and end up in the river as a result of run-off. Furthermore, irrigation, which is primarily formed in flat geomorphological bottom areas (arid landscape natural salt sinks), triggers the mobilization of vast fossil salt storages dating from the soil's past marine or otherwise saline geological history (Smedema & Shiati 2002). As a result, irrigation has been blamed for the salinization of many streams, including the Amu Darya and Syr Darya Rivers in Central Asia (Létolle & Chesterikoff 1999, Crosa et al. 2006). South Africa's Breede River (Scherman et al. 2003), Spain's Ebro River (Isidoro et al. 2006), and Turkey's Great Menderes River (Koç 2008). A large portion of groundwater in Australia is saline due to the climatic and geomorphological features of the landscape (Blinn et al. 2004).

As a result, salinization has been identified as one of the most serious stressors facing freshwater habitats worldwide (Douglas 2017). Furthermore, salinization is one of the top 15 causes of stream impairment in the United States, ranking alongside pesticide input (Watershed Assessment, Tracking & Environmental Results 2012). In an Australian survey of river managers, salinity was ranked among the top three most significant environmental pollutants.

Pumping groundwater to bring the level down, then reusing wastewater or rather discharging it into local rivers, wetlands or streams is perhaps the most feasible alternative for controlling salinization (G.o. 1988). Unfortunately, there is a scarcity of scientific evidence to determine the biological consequences of these saline wastewater discharge systems (Brock & Hammer 1987).

An Artificial Neural Network (ANN) is a computational tool inspired by biological organisms' brains and nervous systems. ANNs are mathematical models that are highly idealized representations of our current knowledge of complex systems. The capacity of neural networks to learn is one of their characteristics. A neural network is not configured similar to a traditional computer program; instead, it resembles provided with instances of processes, insights, theories, or any other type of data that it must comprehend. Through the learning (also known as training) phase, the neural network arranges itself to generate an intrinsic collection of characteristics that it utilizes to recognize data. ANNs can handle inaccurate or partial data, provide estimations, and are less susceptible to anomalies than traditional techniques. They are pretty similar, which means that their numerous independent processes may be performed at the same time. The ANN's massively parallel processing architecture allows it to efficiently handle complex computations, making it the leading method for high-speed data processing today (Sarkar & Pandey 2015). These characteristics make ANNs ideal tools for dealing with a variety of water modeling issues. But the number of applications for ANNs is growing, and it has recently been effective in predicting various water-related problems (Daniell 1991).

ARTIFICIAL NEURAL NETWORK (ANN)

Framework of ANN

An artificial neural network (ANN) is a computational device made up of a highly interconnected group of basic information processing elements called units, which are similar to neurons. The neuron receives input from sources and generates output in accordance with a non-linear mechanism that is predetermined. The interconnection of several neurons in a known configuration creates an ANN model.

The training step creates the link between neurons using known inputs and outputs and shows them in a logical sequence to the ANN. An error convergence approach is used to generate the appropriate power given a known input of data to adjust the intensity of these interconnections.

The feed-forward error backpropagation algorithm is used for ANN training in this analysis (Rumelhart et al. 1986). The ANN network utilized in this research is represented in Fig. 1. The input, the covert, and the output layers are the three fundamental layers or tiers of information systems. Each of these levels comprises the processing units of neural network nodes. The neural weight is the relation of nodes in various layers (Daniell 1991).

Training and Validation of ANN

Neural networks learn by analyzing instances with a given "input" and "end," resulting in probability-weighted connections between the two that are stored in the net's data structure. When training a neural network from a given example, the difference between the network's processed output and a target output is generally determined. This is the error. Using this error value and a learning rule, the network then modifies its weighted associations. With each modification, the neural network can provide output that becomes closer and closer to the desired output. After a sufficient number of these adjustments have been made, the training may be ended based on certain circumstances.

This is referred to as supervised learning. The number of input nodes, output nodes, and hidden layer nodes varies

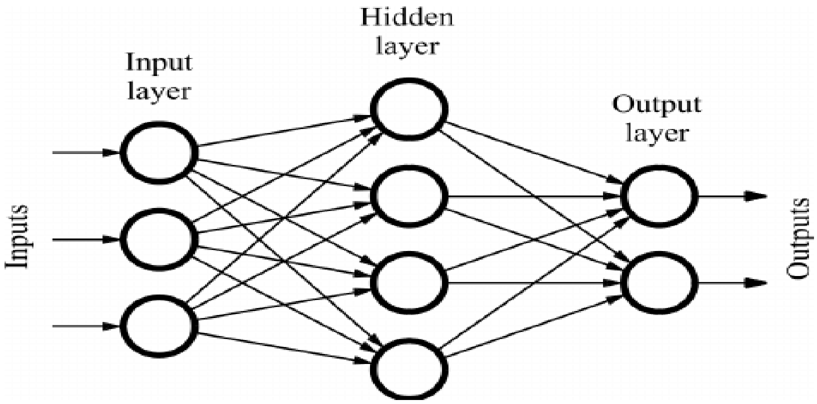


Fig. 1: A multi-layer feed-forward Artificial Neural Network model's layout.

according to the task at hand. If the hidden layer has a small number of nodes, the network will not have enough degrees of freedom to appropriately learn the process. If the number is excessively huge, the training process may take a lengthy time, and the network may over-fit the data in some situations (Sarkar & Pandey 2015).

After training is finished, the ANN's output is verified. Depending on the outcome, the ANN will either need to be retrained or will be able to serve its original function.

Performance Assessment of ANN Model

A wide variety of statistical parameters are available to evaluate the suitability of any particular model. The root mean square error (RMSE), coefficient of correlation (R), and determination coefficient (DC) are the success assessment statistics used in this study for ANN preparation. The following equations were used to calculate these parameters.

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^{n}(Q_i - q_i)^2}{n}} \quad \ldots (1)$$

$$R = \frac{\sum_{i=1}^{n}\left(Q_i - \overline{Q}\right)\left(q_i - \overline{q}\right)}{\sqrt{\sum_{i=1}^{n}\left(Q_i - \overline{Q}\right)^2 \left(q_i - \overline{q}\right)^2}} \quad \ldots (2)$$

$$DC = \frac{\sum_{i=1}^{n}\left(Q_i - \overline{Q}\right)^2 - \sum_{i=1}^{n}\left(Q_i - \overline{q}\right)^2}{\sqrt{\sum_{i=1}^{n}\left(Q_i - \overline{Q}\right)^2}} \quad \ldots (3)$$

Where, $\overline{Q} = \frac{1}{n}\sum_{i=1}^{n} Q_i$, $\overline{q} = \frac{1}{n}\sum_{i=1}^{n} q_i$, Q = observed, q = calculated

MATERIALS AND METHODS

Study Area and Data Collection

The Bhairab River was chosen for determining various water quality parameters and determining whether the river water is appropriate for drinking according to WHO standard permissible limits (World Health Organization). These criteria aid in determining the impurity level of river water, characterizing temporal data variations in water quality, and identifying various trends in data analysis based on tidal conditions, day-to-day differences, and seasonal changes over the course of a year. By measuring the appropriate parameters, it may be possible to understand the impact of seasonal precipitation, surface run-off, and other natural changes in river water parameters. The Bhairab river serves as a natural water source in the Khulna area, carrying away industrial polluted water. This data was collected from an undergraduate thesis study at Khulna University of Engineering and Technology (Chowdhury & Hamidul Bari 2020). Nine water quality parameters were considered for the completion of this study - temperature, pH, turbidity, electrical conductivity (EC), color, salinity, total dissolved solids (TDS), total solids (TS), and suspended solids (SS). Salinity is a major component when you consider the quality of water, especially drinking water. The information demonstrates that the salinity level is far beyond the WHO guideline's normal threshold. A total of 126 samples of data were available.

RESULTS AND DISCUSSION

Standardization of input data is a crucial phase in the data processing process before implementing ANN. The ANN model was used to standardize the input data for a target variable in this analysis. The available data is split into two sections when estimating the parameters of an ANN model. The first is for calibrating the model, and another is for validating it. This is called a "split-sample" test. The number of parameters to be estimated determines the duration of calibration data. The rule of thumb is to use half to two-thirds of the data for training and the rest for testing. Other than salinity itself, the rest of the eight of the nine parameters were used for training and validation to predict salinity.

Various parameter combinations were used in the development of the ANN model: (a) Datasets of temperature, pH, color, turbidity, suspended solids (SS) (b) Dataset of electrical conductivity (EC), total solids (TS), total dissolved solids (TDS) (c) Datasets of temperature, pH, turbidity, electrical conductivity (EC), color, total dissolved solids (TDS), total solids (TS), suspended solids (SS). For the analysis, the ANN model had 5 input variables, 3 input variables, and 8 input variables for the case (a), (b), and (c) respectively. The training of the ANN model epoch had 1000 cycles. A three-layer standard network was used.

Table 1 compares the accuracy of various ANN models for salinity estimation based on RMSE, correlation coefficient, and determination coefficient. It shows that the RMSE, R, and DC values for all built models during calibration range from 0.78 to 2.41, 0.6 to 0.97, and 0.34 to 0.94, respectively. During model validation, the values of RMSE, R, and DC range from 0.58 to 2.55, 0.53 too so 0.98, and 0.28 to 0.96, respectively. Three types of ANN models have been built with different combinations of variables, as discussed in the previous sections. The three ANN models have distinct results. The ANN2 model had the best performance, with RMSE, R, and DC values of 0.78, 0.97, and 0.94 during calibration and 0.58, 0.98, and 0.96 during validation. ANN-2 model has the best performance which has 3 input variables electrical conductivity (EC), total solids (TS), and total dissolved solids (TDS). These three input parameters are highly correlated with salinity so these variables are the optimum variables to correctly measure salinity. ANN-1 model had 5 input parameters pH, color, temperature, turbidity, and suspended solids (SS) which are loosely correlated with the values of salinity so, the model found it tough to find out the values of salinity. Because there are fewer input variables, the performance is significantly reduced when compared to ANN-2. These input variables are unable to identify the underlying physical mechanism. There are many input parameters for the ANN-3 model, which combines all of the parameters, including temperature, pH, turbidity, electrical conductivity (EC), colour, total dissolved solids (TDS), total solids (TS), and suspended solids (SS). These input parameters include both parameters that are highly and weakly correlated with

Table 1: Comparative performance of ANN models.

ANN Model	Training			Validation		
	RMSE	R	DC	RMSE	R	DC
ANN-1	2.39	0.61	0.36	2.51	0.55	0.3
ANN-2	0.78	0.97	0.94	0.58	0.98	0.96
ANN-3	2.41	0.6	0.34	2.55	0.53	0.28

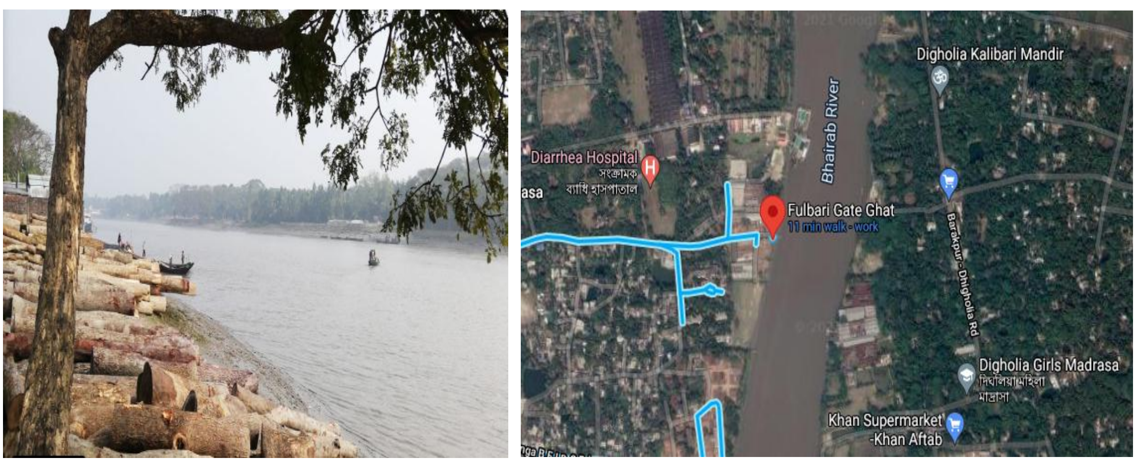


Fig. 2: Sample collection area of the Bhairab River.

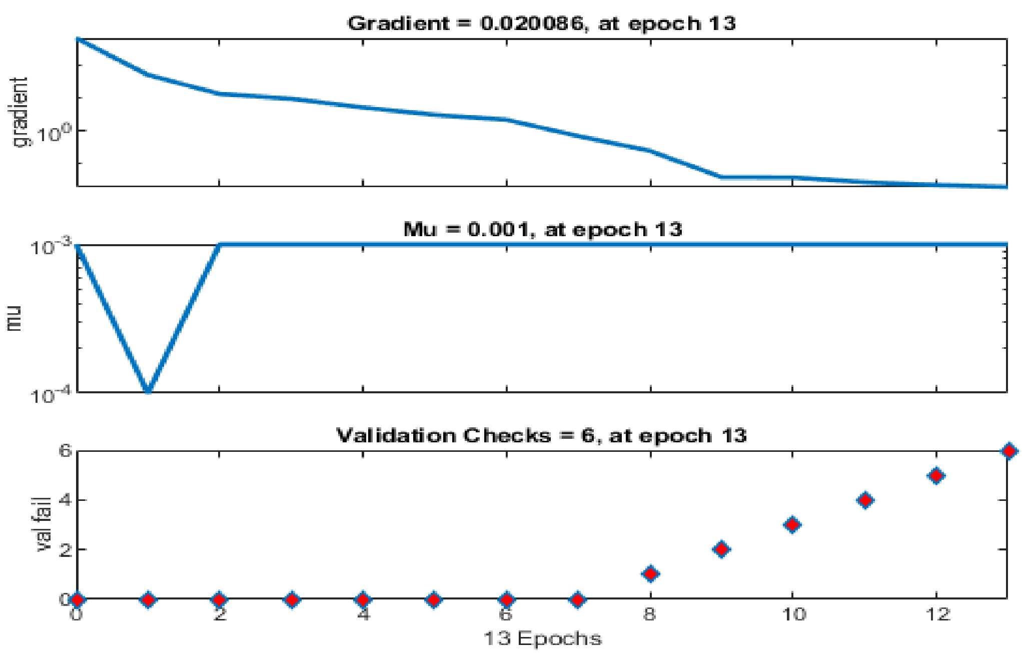
salinity. So, this makes the model more convoluted and may lead to the model overfitting the training data, resulting in bad forecasts (Tokar & Johnson 1999). As a result, it is critical to employ the optimum number of input variables when developing an ANN model, and the findings of this study show that 3 input variables that are highly correlated to salinity, as used in ANN-2, are optimal for salinity simulation at Bhairab river. Fig. 1 shows the graphical results of the top performing ANN, ANN-2, in the form of gradient values, and regression plots, demonstrating a good relationship between the observed and predicted salinity using the ANN approach.

With regard to the unknown weights and biases, the gradient is the gradient of the error function's square (Error = Known target -Variable output). Typically, the training goal is to minimize the sum of squared errors by utilizing the steepest descent technique to optimize the choice of weights and biases. A dynamic component is included to control the descent rate in order to stop the search value from going backward or forward before stopping close enough to it. The gradient value in an artificial neural network tends to be zero and in the 0-1 value range. As shown in Fig. 3, the gradient value during training for the ANN-2 model is close to.02, which is well within the acceptable range. The model is doing very well in the regression part where it achieves more than 90 percent accuracy in all categories. In training, the model predicted the values of salinity very accurately.

When the model is validated using Fig. 4, the gradient value is close to 0.01, which is well within the permitted range and has decreased since the training, indicating that the model has been properly trained and that the error has decreased. The hidden layers were thoroughly checked and from trial and error, the right number of hidden layers were applied to reduce the model from overfitting. It can be observed from the regression and prediction section that the ANN-2 model predicts the salinity values much more accurately than during training, indicating a very accurate model.

CONCLUSION

Water quality modeling using artificial neural networks is hardly used in Bangladesh. Measuring water quality parameters throughout the year or in a very specific time continuously is a very tough and long process. Using an artificial neural network this predictive model could save time and ease the process. Of the model that was created to predict salinity ANN-2 model had the most efficient result. For identifying the performance of the model, the values of RMSE, R, and DC were very helpful. These values presented that the ANN-2 model had highly correlated input parameters that were very good at predicting the salinity values. The quality of the input data and the number of hidden layers used were paramount to better the accuracy of the model.



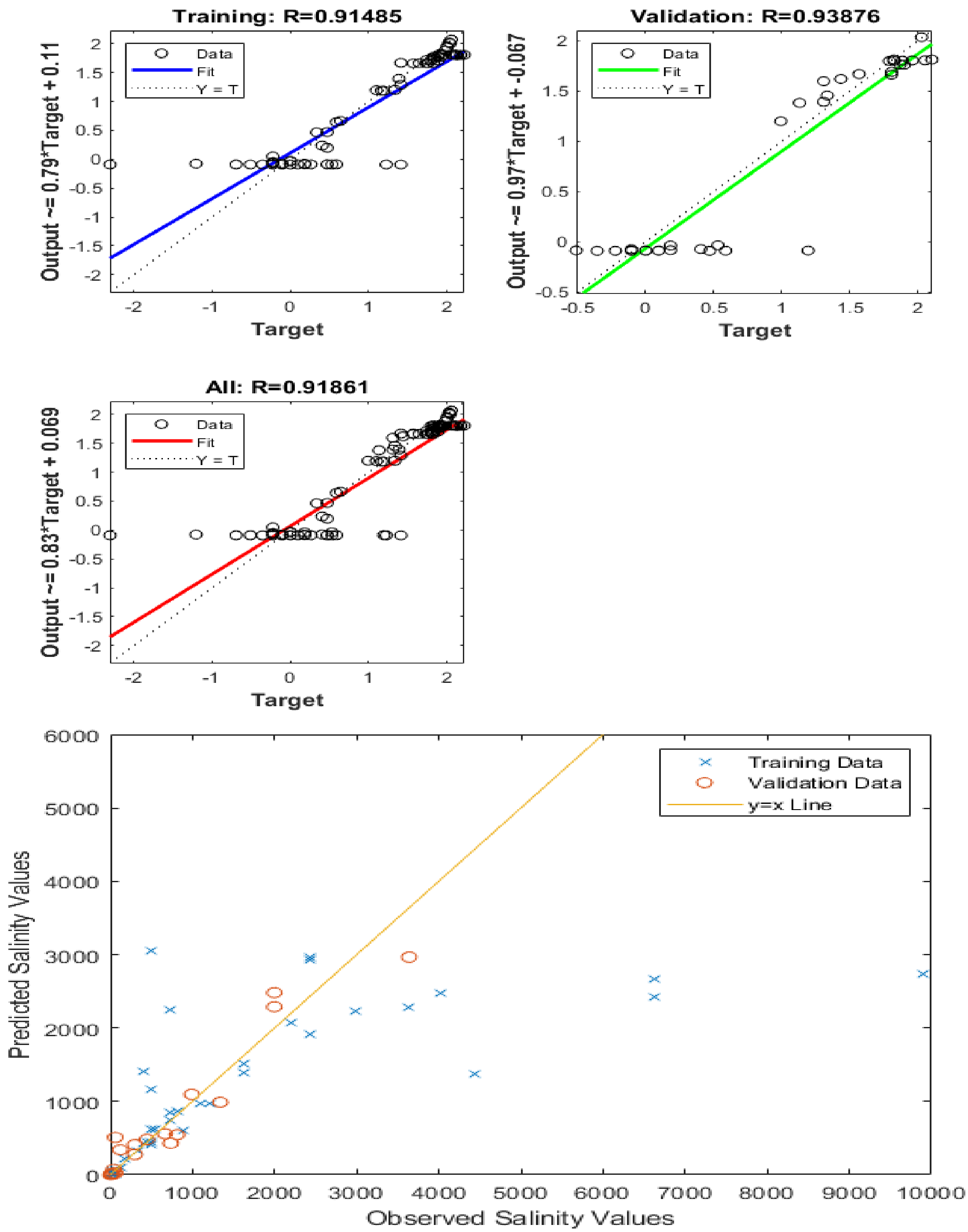
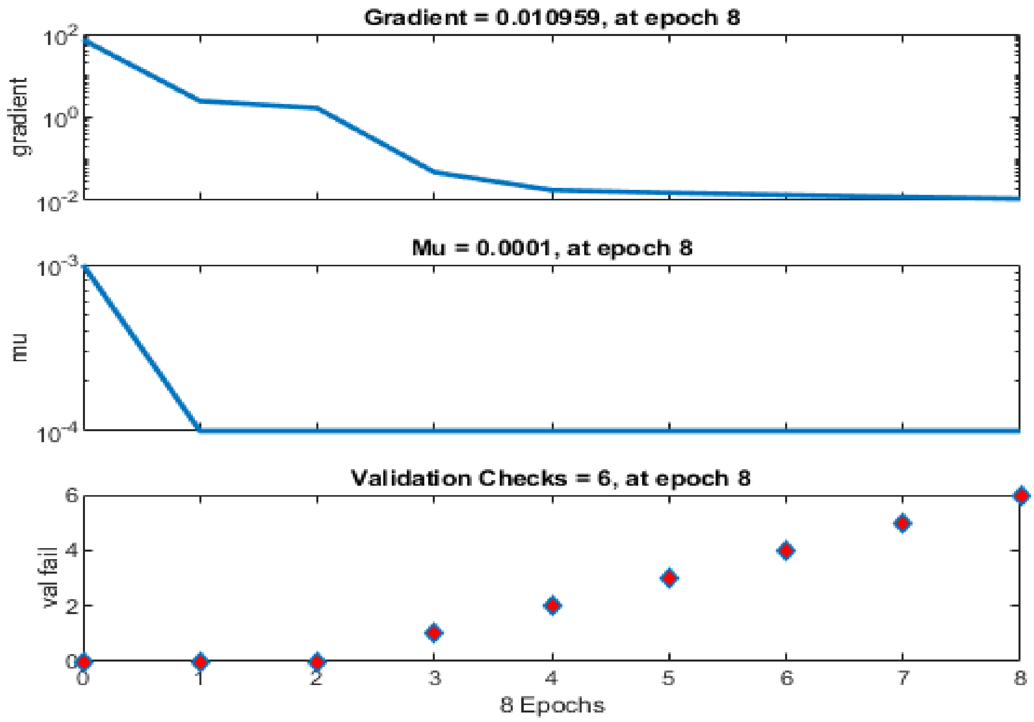
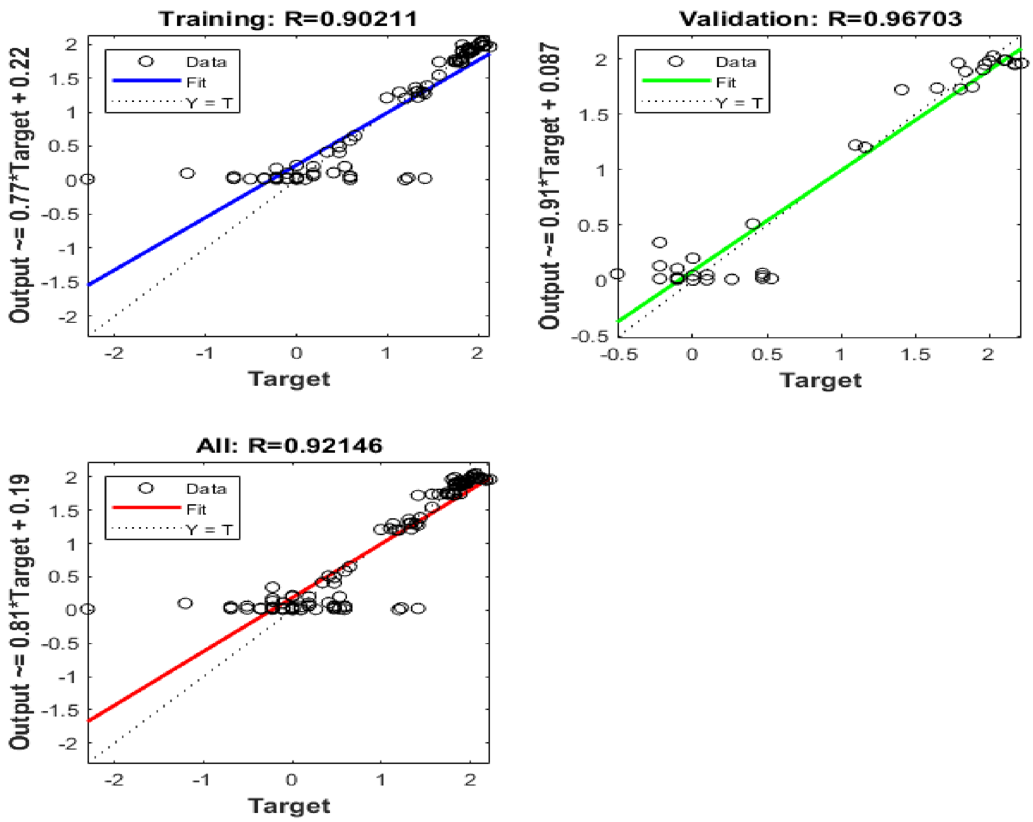


Fig. 3: Gradient values, regression, and predicted values of salinity in the training of the ANN-2 model.


Gradient = 0.010959, at epoch 8
gradient
Mu = 0.0001, at epoch 8
mu
Validation Checks = 6, at epoch 8
val fail
8 Epochs


Training: R=0.90211
Output ~= 0.77*Target + 0.22
Data
Fit
Y = T
Target
Validation: R=0.96703
Output ~= 0.91*Target + 0.087
Data
Fit
Y = T
Target
All: R=0.92146
Output ~= 0.81*Target + 0.19
Data
Fit
Y = T
Target

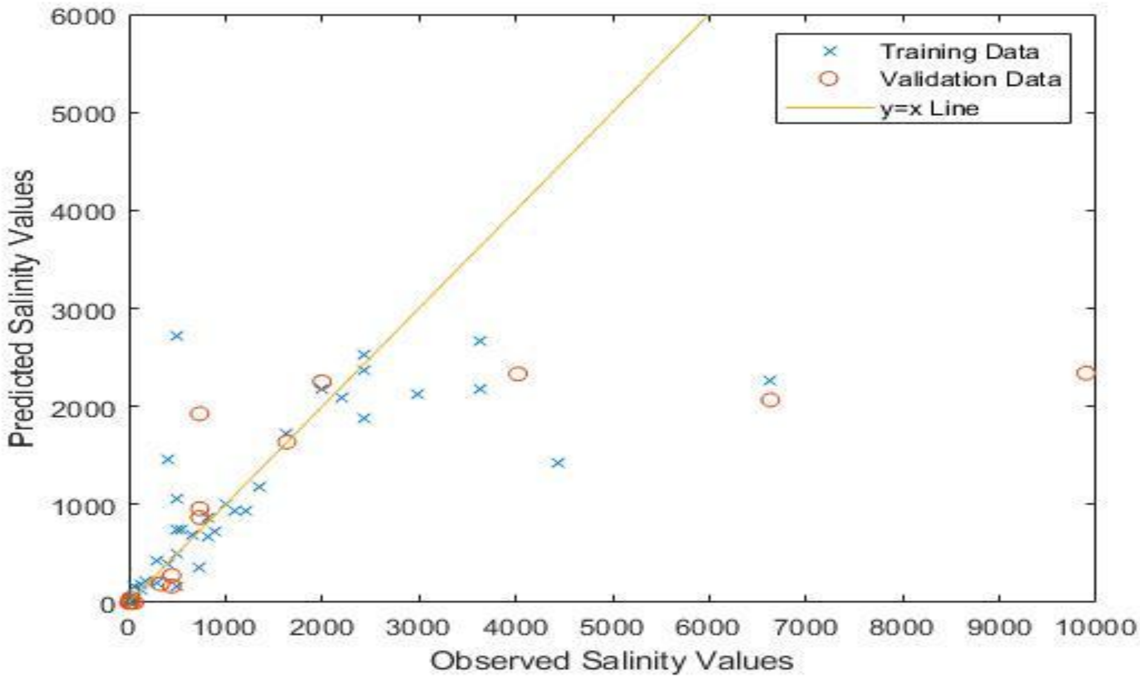


Fig. 4: Gradient values, regression, and predicted values of salinity in the validation of the ANN-2 model.

REFERENCES

Battaglia, B. 1958. Symposium on the classification of brackish waters. Oikos, 9(2): 311.

Blinn, D., Halse, S., Pinder, A. and Shiel, R. 2004. Diatom and micro-invertebrate communities and environmental determinants in the western Australian wheat belt: A response to salinization. Hydrobiologia, 528(1-3): 229-248.

Brock, M.A. and Hammer, U.T. 1987. Saline lake ecosystems of the world. J. Ecol., 75(2): 580.

Cañedo-Argüelles, M., Kefford, B.J., Piscart, C., Prat, N., Schäfer, R.B. and Schulz, C.J. 2013. Salinisation of rivers: An urgent ecological issue. Environ. Pollut., 173: 157-167.

Chowdhury, D. and Hamidul Bari, Q. 2020. Seasonal Variation of Water Quality Parameters in Bhairab River. Khulna University of Engineering & Technology, Khulna.

Crosa, G., Froebrich, J., Nikolayenko, V., Stefani, F., Galli, P. and Calamari, D. 2006. Spatial and seasonal variations in the water quality of the Amu Darya River (Central Asia). Water Res., 40(11): 2237-2245.

Daniell, T.M. 1991. Neural networks. Applications in hydrology and water resources engineering. Natl. Conf. Publ. Inst. Eng. Aust., 3: 797-802.

Douglas, I. 2017. Ecosystems and human well-being. In: Encycl. Anthr., 1-5: 185-197.

G.o. 1988. Salt action, joint action: Victoria's strategy for managing land and water salinity. T2-A2.

Herczeg, A.L., Dogramaci, S.S. and Leaney, F.W.J. 2001. Origin of dissolved salts in a large, semi-arid groundwater system: Murray Basin, Australia. Mar. Freshw. Res., 52(1): 41-52.

Isidoro, D., Quílez, D. and Aragüés, R. 2006. Environmental impact of irrigation in La Violada District (Spain). J. Environ. Qual., 35(3): 766-775.

Koç, C. 2008. The environmental effects of salinity load in Great Menderes Basin irrigation schemes. Environ. Monit. Assess., 146(1-3): 479-489.

Létolle, R. and Chesterikoff, A. 1999. The salinity of surface waters in the Aral sea region. Int. J. Salt Lake Res., 8(4): 293-306.

Maryoung, L.A., Lavado, R. and Schlenk, D. 2014. Impacts of hypersaline acclimation on the acute toxicity of the organophosphate chlorpyrifos to salmonids. Aquat. Toxicol., 152: 284-290.

Rumelhart, D.E., Hinton, G.E. and Williams, R.J. 1986. Learning representations by back-propagating errors. Nature, 323(6088): 533-536.

Sarkar, A. and Pandey, P. 2015. River Water Quality Modelling Using Artificial Neural Network Technique. Aquat. Procedia., 4: 1070-1077.

Scherman, P.A., Muller, W.J. and Palmer, C.G. 2003. Links between ecotoxicology, biomonitoring, and water chemistry in the integration of water quality into environmental flow assessments. River Res. Appl., 19(5-6): 483-493.

Smedema, L.K. and Shiati, K. 2002. Irrigation and salinity: A perspective review of the salinity hazards of irrigation development in the arid zone. Irrig. Drain Syst., 16(2): 161-174.

Tokar, A.S. and Johnson, P.A. 1999. Rainfall-runoff modeling using artificial neural networks. J Hydrol. Eng., 4(3): 232-239.

Watershed Assessment, Tracking & Environmental Results. 2012. https://www.epa.gov/waterdata/waters-watershed-assessment-tracking-environmental-results-system

Williams, W.D. and Sherwood, J.E. 1994. Definition and measurement of salinity in salt lakes. Int. J. Salt Lake Res., 3(1): 53-63.