Nature Environment and Pollution Technology
An International Quarterly Scientific Journal

p-ISSN: 0972-6268
e-ISSN: 2395-3454

Vol. 18 | No. 3 | pp. 719-733 | 2019

Original Research Paper | Open Access

Forecasting of Heavy Metal Contamination in Coastal Sea Surface Waters of the Karachi Harbour Area by Neural Network Approach

Muhammad Ayaz† and Nasir-Uddin Khan
Department of Mathematics, University of Karachi, Karachi 75270, Pakistan
†Corresponding author: Muhammad Ayaz

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

Received: 10-11-2018
Accepted: 04-02-2019

Key Words:
Artificial neural network
Karachi coastal area
Seawater pollution
Heavy metals

ABSTRACT

The major and overriding factors affecting water quality and the aquatic ecosystems in the coastal areas are sewage and nutrient inputs from municipal and industrial wastewater, depletion of seaside contrivances, risks of public health as well as loss of biodiversity. The coastal area of the Karachi harbour is most heavily polluted due to these reasons. In this study, we proposed the artificial neural network (ANN) models to monitor and control the sea surface water quality of the Karachi coastal area along the harbour. Recently, various types of ANN have been successfully applied in hydrological fields. In this study, Nonlinear Auto Regressive eXogenous Neural Network (NARX-NN) shall be applied to predict the concentration of heavy metals in coastal sea surface water of the Karachi harbour area. This method provides significant insight into the comparative study of two different training functions of NARX-NN, namely, Levenberg-Marquardt (LM) and Scale Conjugate Gradient (SCG). The physical parameters like sea surface temperature (SST), salinity, tides and pH are taken as an input and the chemical parameters chromium, copper, lead, nickel and zinc are taken as output individually for all six locations. The performance of the model was evaluated by statistical criteria that include a correlation coefficient (r) and mean square error (MSE). The prediction results indicated that the LM training function is superior to SCG training function. Hope this study is helpful for local authorities and policy makers to develop a new infrastructure and install a water treatment plant to reduce the water pollution of the harbour area.

INTRODUCTION

Marine pollution is an important issue of the coastal areas; it generally affects the health of human beings, pollutes the surrounding environment and is harmful to aquatic life. The water quality of a sea surface in a region generally relies on the nature and extent of the industry, agriculture and other anthropogenic activities in the catchment. The impairment of seashore water eminence has triggered the commencement of serious controlling endeavours. In many countries, tourism, recreation and fishing requires at least an acceptable level of seawater variables (Palani et al. 2008). The Karachi coastal area stretches 100 km, and its harbour is situated along the west coast of Karachi which is about 35 km long (Majeed et al. 2010). This is the natural harbour, which has evolved over hundreds of years by a process of dredging and reclamation. Although there are rivers and drains in the harbour, in fact, there is a little rainwater most of the time; contamination and a substantial quantity (possibly one third) of the siltation in the harbour are affected by the discharge of garbage, raw sewage and industrial runoff from the city. There are also organic waste and oil spills from the Karachi fish harbour. The main sources of pollution in the Karachi harbour are industrial and municipal wastewaters from the Liyari River. It receives from the Sindh Industrial and Trade Estate (SITE) as well as central and northern districts and discharges into the harbour water near Manora channel. This is the area where the fish harbour, main harbour and the Karachi shipyard is located. The aggregate region around the Karachi port, including the backwaters is about 25 km^2 and around 50 billion m^3 of marine water enters and leaves the port region throughout a tidal cycle. An expected 3000 vessels visit the Karachi port consistently and around 5000 mechanical ships make their fish dockings at the fish harbour and are spotted inside the port zone of Karachi (Karachi port trust (KPT) 1995).

Researchers investigated the concentration of zinc (Zn), copper (Cu), magnesium (Mn) and iron (Fe) in water and sediment samples of the Karachi harbour. They found higher concentrations of heavy metals in sediment and water. They further concluded that the untreated effluents of industrial and sewage wastes, being discharged through Lyari River into the harbour increases the level of pollution (Khan & Saleem 1988). The concentration of lead (Pb), Zn and Cu at Baba channel, Chari Kundi channel and Manora Channel in seawater surface samples are high (Beg 1992). The concentration of nickel (Ni), zinc (Zn), copper (Cu), lead (Pb), cadmium (Cd) and chromium (Cr) in seawater and surface sediments of the Karachi harbour was found higher near the outfall of the Lyari River (Saleem & Kazi 1995). The concentration of heavy metals (Cr, Cu, Pb, Ni and Zn) is high in seawater along the Karachi harbour compared to other Karachi coastal areas because of municipal and industrial wastewater, untreated tannery waste and bad handling of cargo and vessels (World Wide Fund 2002). Their studies indicate that the Karachi harbour is highly polluted and they observed that the trace metals are notably greater compared to a number of areas in the world. The high concentration of heavy metals during the low tides indicates high metal concentration due to sewage and industrial wastewater spreading during the low tide formation, while during the high tide, river water mixed with the open seawater creates a low concentration of heavy metals.

Nowadays, the artificial neural network techniques are used to predict the data in bioscience, applied sciences, social science, economics and environmental sciences. ANNs are nonlinear, and have for quite some time been in the realm of linear statistics. The conventional methodologies of time arrangement forecast, for example, the Box-Jenkins or ARIMA technique (Box & Jenkins 1976, Pankartz 1983). In this study, ANN models have been used to predict and quantify the presence of pollutants, particularly heavy metals, in seawater of the Karachi harbour and its surrounding coastal. ANN is used as a modelling tool for the prediction of seawater quality parameter like salinity, temperature, dissolved oxygen and chlorophyll-*a* (Palani et al. 2008). There are two different methods of ANN namely, Multilayer perceptron (MLP) and Radial base Function (RBF), which are applied to predict the surface water quality parameters such as TDS, electrical conductivity and turbidity in Johar river of Malaysia. Its model can reveal the complex non-linear relationship between the input variables and showed that MLP and RBF results are good as compare to the linear regression model (LRM) (Najah et al. 2013). Its model can reveal the complex non-linear relationship between the input variables. ANNs estimate model is good for forecasting the ion concentration of water. There are three different training algorithms, namely Levenberg-Marquardt (LM), scale conjugate gradient (SCG) and conjugate gradient backpropagation (CGP) used to predict calcium, chloride, sulphate and sodium ion concentration of the Kajar River, Tehran. The study revealed that the best training algorithm is Levenberg-Marquardt for calcium ion concentration with input variables, temperature, pH, hardness, TDS, electric conductivity (EC) and turbidity, as well as the calcium ions concentration model result is similar with three inputs, namely hardness, TDS and EC as compared to the model with six inputs by LM training function (Movaghar-nejad et al. 2017).

Research in this area is very costly affair, and as such not much has been done in this specific field. A team of UNO was tasked to undertake this issue, quantify the damage caused and proposed the remedial measures to control the menace along the Karachi coastal area. Moreover, in the past few decades, Pakistani scientists and research with their meagre resources effectively addressed the problem, and their positive contribution helps a lot to control the alarming situation of the coastal area. In this study, the NARX-NN models have been developed to predict and quantify the presence of pollutants, particularly heavy metals such as chromium (Cr), copper (Cu), lead (Pb), nickel (Ni) and zinc (Zn), in sea surface water of the Karachi coastal area along the Karachi harbour at different six locations. The physical parameters include sea surface temperature, salinity, tides and pH that have been used as input variables in this model.

MATERIALS AND METHODS

Study area description and data: The coastal zone of the Karachi harbour is about 100 km long. The area being close to coast is considered to be the polluted region. There are a number of parameters that specify the quality of water. The parameters include sea surface temperature, pH, salinity, tides and heavy metals, namely Cr, Cu, Ni, Pb and Zn. The recorded data were taken ranging from January 2008 to December 2013 at six different locations (Fig. 1) of the Karachi coastal area along the Karachi harbour to study the pollution level. These data were provided to the National Institute of Oceanography (NIO), Karachi and Karachi Port Trust (KPT).

Six different sites of the study area (Fig. 1) were taken to analyse the heavy metals in seawater . Clifton is considered as a location one (L1) with latitude 24°46’24.37” N and longitude 67°01’07.30” E; in this location sewage water and industrial waste from tanneries and car painting industries are directly poured into the beach water. Human activities are also a main cause of organic and inorganic pollution. Location two (L2) is Kemari with latitude 24°26’08.14” N and longitude 66°59’08.00” E; it is an extremely polluted area due to harbour activities such as the spilling of cargos, cargo handling, dredging and dumping of ship waste, and sewage water through tanneries. Manora is taken as a location three (L3) with latitude 24°47’24.52” N and longitude 66°56’34.64” E, which is outfall of the Lyari River and harbour activity. This area receives emissions and disposal of untreated industrial and domestic wastes primarily from

Fig. 1: Map showing locations of the Karachi coastal area along the harbour.

the Lyari River. Further continuation of this beach is known as Sandspit. The location number four (L4), with latitude 24°50'52.88" N and longitude 60°53'08.89" E, is normally polluted because one side is affected by the Lyari River wastewater and other side due to recreation activities. Hawksbay is location five (L5) with latitude 24°49'45.82" N and longitude 66°51'30.40" E is a low polluted area, mainly used for recreational purpose. Paradise point is marked as location six (L6) with latitude 24°48'41.95" N and longitude 66°46'36.14" E, where Karachi Nuclear Power Plant (KANUPP) is in operation and expected to be the source of the change in the chemical composition of beach water. The statistical description of all the data set is represented in Table 1 to Table 7.

Artificial neural network (ANN): The ANN is a scientific structure planned to imitate the system of limits of the neurons in the brain. Through this system information is processed according to modifiable weight bias and numerical exchange capacities. Each system unit gets information from another unit and subsequently imparts a response to the next unit. ANNs are useful when applied to issues whose solutions need to be learnt by using the available information. In this way, the problem can be dealt with using non parametric statistical techniques (White 1989, Ripley 1993, Cheng & Titterington 1994). ANN is a very flexible and versatile tool compared with the conventional methodologies of time arrangement forecast, for example, the Box-Jenkins or ARIMA technique (Box & Jenkins 1976, Pankartz 1983) to solve different real-world problems such as pattern classification, image completion, clustering, forecasting and complexity. There is no need to define the relationship between the input and output for solving the non-linear problem. The ANN model trained with data that had less variation and near normal distribution performed better compared to the model trained with data that had high variation in values and non-normal distributions (Melesse et al. 2011). In the training process, exploratory data are acquainted with a neural network to set up a connection between the input and output of the network under review. The goal of the training algorithm is to lessen the total error between forecasted output, and given target esteems (Altun et al. 2007). Data preprocessing, strategies for deciding sufficient model data sources, and the inside workings of artificial neural networks are from time to time reflected in the model building development. This can bring about the substandard model execution and a failure to precisely think about the performance of various ANN models (Maier & Dany 2001). In this study, we developed ANN forecasting solution based on NARX model.

Nonlinear auto regressive exogenous neural network (NARX-NN): Nonlinear autoregressive with exogenous input (NARX) neural network is a discrete time system and its mathematical expression is:

$$y(n+1) = f\,[y(n), \ldots\ldots.\, y(n-d_y);\; u(n) \ldots\ldots\ldots\ldots.\, u(n-d_u)] \quad \ldots(1)$$

Where, $u(n)\ R$ and $y(n)\ R$ are the input and output of the model at discrete time step n respectively, and $d_u \geq 1$ is the input and $d_y \geq 1\ (d_y \geq d_u)$ is the output delay. The function $f(.)$ is generally unknown and can be approximated. Usually NARX-NN can be predicted and trained through two methods, open loop mode and closed loop mode. In closed loop system, output regressed also include the estimated value of output and in open loop system output regressed only includes the system actual output value. In this study, the network will be created and trained in open loop as shown in Fig. 2. Open loop is more efficient than closed loop training. Open loop allows us to supply the network with correct past output as we train it to produce the connect current output.

The general NARX-NN equation with one input and one

Table 1: Correlation coefficient (r) and mean square error (MSE) of training and testing data set of chromium for all six locations of the Karachi coastal area along the harbor by LM training function.

Model	Training dataset		Testing dataset		Total
	MSE	r	MSE	r	r
CrL1	1.78E-06	0.999	0.0169	0.860	0.947
CrL2	5.40E-04	0.993	0.0205	0.865	0.929
CrL3	1.53E-03	0.984	0.0239	0.942	0.943
CrL4	4.89E-04	0.997	0.0093	0.921	0.971
CrL5	7.20E-03	0.918	0.0109	0.914	0.910
CrL6	2.20E-03	0.977	0.0229	0.857	0.935

Table 2: Correlation coefficient (r) and mean square error (MSE) of training and testing data set of chromium for all six locations of the Karachi coastal area along the harbor by SCG training function.

Model	Training dataset		Testing dataset		Total
	MSE	r	MSE	r	r
CrL1	0.0150	0.837	0.0234	0.805	0.812
CrL2	0.0046	0.948	0.0203	0.988	0.923
CrL3	0.0077	0.915	0.0082	0.965	0.914
CrL4	0.0046	0953	0.0048	0.924	0.944
CrL5	0.0120	0.822	0.0081	0.900	0.840
CrL6	0.0056	0.925	0.0094	0.939	0.927

Table 3: Correlation coefficient (r) and mean square error (MSE) of training and testing data set of copper for all six locations of the Karachi coastal area along the harbor by LM training function.

Model	Training dataset		Testing dataset		Total
	MSE	r	MSE	r	r
CuL1	0.0067	0.958	0.0301	0.892	0.902
CuL2	0.0009	0.993	0.0174	0.913	0.957
CuL3	0.0008	0.990	0.0068	0.865	0.944
CuL4	0.0035	0.953	0.0566	0.842	0.876
CuL5	0.0027	0.963	0.0157	0.864	0.906
CuL6	0.0008	0.993	0.0292	0.843	0.913

Table 4: Correlation coefficient (r) and mean square error (MSE) of training and testing data set of copper for all six locations of the Karachi coastal area along the harbor by SCG training function.

Model	Training dataset		Testing dataset		Total
	MSE	r	MSE	r	r
CuL1	0.0128	0.898	0.0386	0.844	0.873
CuL2	0.0137	0.889	0.0114	0.867	0.899
CuL3	0.0060	0.917	0.0209	0.909	0.901
CuL4	0.0065	0.920	0.0090	0.841	0.907
CuL5	0.0080	0.900	0.0097	0.943	0.893
CuL6	0.0095	0.912	0.0084	0.941	0.915

Table 5: Correlation coefficient (r) and mean square error (MSE) of training and testing data set of nickel for all six locations of the Karachi coastal area along the harbor by LM training function.

Model	Training dataset		Testing dataset		Total
	MSE	r	MSE	r	r
NiL1	0.0047	0.941	0.0315	0.737	0.863
NiL2	0.0021	0.983	0.0241	0.848	0.934
NiL3	0.0047	0.955	0.0107	0.888	0.933
NiL4	0.0031	0.960	0.0066	0.883	0.932
NiL5	0.0066	0.931	0.0410	0.835	0.892
NiL6	0.0562	0.874	0.0190	0.937	0.883

Table 6: Correlation coefficient (r) and mean square error (MSE) of training and testing data set of nickel for all six locations of the Karachi coastal area along the harbor by SCG training function.

Model	Training dataset		Testing dataset		Total
	MSE	r	MSE	r	r
NiL1	0.0155	0.849	0.0206	0.811	0.843
NiL2	0.0079	0.933	0.0144	0.906	0.917
NiL3	0.0080	0.900	0.0148	0.854	0.879
NiL4	0.0097	0.856	0.0148	0.845	0.850
NiL5	0.0110	0.885	0.0135	0.873	0.879
NiL6	0.0058	0.954	0.0413	0.840	0.941

Table 7: Correlation coefficient (r) and mean square error (MSE) of training and testing data set of lead for all six locations of the Karachi coastal area along the harbor by LM training function.

Model	Training dataset		Testing dataset		Total
	MSE	r	MSE	r	r
PbL1	0.0076	0.917	0.010	0.940	0.904
PbL2	0.00002	0.999	0.017	0.897	0.945
PbL3	0.0089	0.914	0.034	0.835	0.867
PbL4	0.0067	0.935	0.017	0.823	0.904
PbL5	0.0053	0.938	0.024	0.841	0.881
PbL6	0.0064	0.935	0.019	0.842	0.896

Table 8: Correlation coefficient (r) and mean square error (MSE) of training and testing data set of lead for all six locations of the Karachi coastal area along the harbor by SCG training function.

Model	Training dataset		Testing dataset		Total
	MSE	r	MSE	r	r
PbL1	0.0087	0.878	0.0131	0.876	0.874
PbL2	0.0108	0.855	0.0091	0.897	0.851
PbL3	0.0141	0.873	0.0141	0.862	0.860
PbL4	0.0085	0.900	0.0161	0.812	0.880
PbL5	0.0138	0.841	0.0150	0.860	0.853
PbL6	0.0106	0.873	0.0842	0.710	0.774

output can be written as:

$$y(n+1) = f_o\left[b_o + \sum_{h=1}^{Nh} w_{ho} f_h\left(b_h + \sum_{i=0}^{d_u} w_{ih} u(n-i) + \sum_{i=0}^{d_y} w_{jh} y(n-j)\right)\right] \quad ...(2)$$

Where, w_{ho}, w_{ih}, $i = 1,2.....d_u$, $j = 1,2.....d_y$, $h = 1,2.....N_h$ are the weights, b_o and b_h are biases and $f_o(.)$ and $f_h(.)$ are output of hidden function.

NARX-NN can be used in multi-time series input and multi-time series output applications. The equation for a general NARX-NN with four inputs and one output can be written as:

$$y(n+1) = f_o\left[\begin{matrix} b_o \\ + \sum_{h=1}^{Nh} w_{ho} \cdot f_h \left(\begin{matrix} b_h + \sum_{i1=0}^{d_{u1}} w_{i1h} u_1(n-i1) + \sum_{i2=0}^{d_{u2}} w_{i2h} u_2(n-i2) + \\ \sum_{i3=0}^{d_{u3}} w_{i3h} u_3(n-i3) + \sum_{i4=0}^{d_{u4}} w_{i4h} u_4(n-i4) + \sum_{j=0}^{d_y} w_{jh} y(n-j) \end{matrix}\right)\end{matrix}\right] \quad ...(3)$$

Where, w_{i1h}, w_{i2h}, w_{i3h} and w_{i4h}, i_1=1, 2......d_{u1}, i_2=1, 2......d_{u2}, i_3=1, 2......d_{u3} *and* i_4=1, 2......d_{u4} are the weights of the connection between the first input units and hidden units, second input units and hidden units, respectively (Zang 2011, Farsa & Zolfaghari 2010). The Nonlinear Auto Regressive eXogenous (NARX) technique of ANN is without a doubt decreased to the time delay neural network (TDNN), keeping in mind the end goal to be connected to time series forecast. Bearing this under usage of the NARX network as a top priority, we propose a straightforward system in view of Take n's implanting theorem that permits the actual structure design of the NARX NN system to be easily, furthermore proficient, and associated with long expression forecast of univariate non-linear time series (Lin 1996).

In the context of this paper, the Nonlinear Auto Regressive eXogenous Neural Network (NARX-NN) architecture is used to develop the models with feed forward neural network with back propagation algorithm (FFNN-BP). There are two second order methods, namely Levenberg Marquardt (LM) and Scale Conjugate Gradient (SCG) training function with FFNN-BP algorithm, used as a training function. A noteworthy concentrate in ANN investigation on the advancement of more proficient training algorithm. Various second-order techniques (e.g., the Levenberg-Marquardt, Shanno and conjugate gradient algorithms) have been considered for optimizing the connection weights of feed forward networks trying to speed up the training process (Battiti 1992, Golden 1996). Generally, the second order methods training functions result in greater convergence speed.

Feed forward neural network with back propagation algorithm (FFNN-BP): In many neural network structures, the feed forward back propagation network (Rumelhart et al. 1986) is most popular and widely used architecture. It has been used for various tasks, such as dynamic modelling, pattern recognition, data mining, time series modelling, function approximation to name just a few. It is consisting of three layers of neurons: input layer, hidden layers and output layers, represented in Fig. 3. These layers consist of processing units called nodes, the input unit is connected to every hidden node and all hidden nodes fully connected to the outputs layers through synoptic weights. Knowledge is usually stored as a set of connection weights; this procedure is called training. After the training process getting the target values (forecast data) and the difference (error) between target-actual is not good, the error signal is then back propagated from the output to input layer and adjust the different weights. This training process continues numerous times and stops when error between the actual and target values is minimal. One cycle of input-hidden-output network is said to be iteration (Epoch). In this study, we used FFNN-BP based training functions, namely Levenberg Marquardt (LM) and scale conjugate gradient (SCG).

Levenberg Marquardt (LM) training function: The training of neural network involves updating the weights of the connections in such a manner that the error between the outputs of the neural network and the actual output is minimized. The LM method was used for training of the given NARX-NN. It is a merger of two optimization methods, namely steepest descent and Gauss Newton methods. The Levenberg-Marquardt training function was aimed to approach second-order training speed but without having to compute the Hessian matrix directly. Hessian matrix (H) can be approximated as:

$$H(x) = J^t(x)J(x) \quad ...(4)$$

Where, J is the Jacobian matrix

$$J(x) = \begin{bmatrix} \frac{\partial e_1(x)}{\partial x_1} & \frac{\partial e_1(x)}{\partial x_2} & \cdots & \frac{\partial e_1(x)}{\partial x_n} \\ \frac{\partial e_2(x)}{\partial x_1} & \frac{\partial e_2(x)}{\partial x_2} & \cdots & \frac{\partial e_2(x)}{\partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial e_N(x)}{\partial x_1} & \frac{\partial e_N(x)}{\partial x_2} & \cdots & \frac{\partial e_N(x)}{\partial x_n} \end{bmatrix} \quad ...(5)$$

And the gradient can be computed as:

$$g(x) = J^t(x)e(x) \quad ...(6)$$

It contains first derivatives of the network errors with respect to the weights and biases and *e(x)* is a vector of

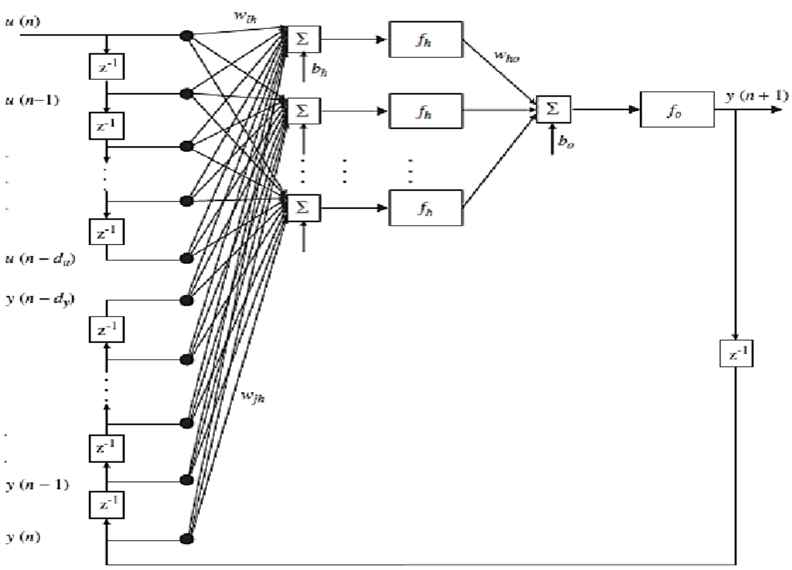


Fig. 2: The architecture of NARX-NN model.

network errors. Suppose that we have a function *f(x)* which we want to obtain minimized with respect to the parameter.

$$\Delta x = -\left(J^t(x)j(x)\right)^{-1} J^t(x)e(x) \quad ...(7)$$

The Gauss Newton method can be developed from Newton's method if the function *f(x)* is a sum of square function

$$f(x) = \sum_{i=1}^{N} e_i{}^2(x) \quad ...(8)$$

If one considers the non-linear case, the gradient and the Hessian of the function of Newton method could be written as:

$$\nabla f(x) = \sum_{i=1}^{N} e_i(x)\, \nabla e_i(x) = J^t(x)e(x) \quad ...(9)$$

$$\nabla^2 f(x) = \sum_{i=1}^{N} e_i(x)\, \nabla^2 e_i(x) + J^t(x)J(x) = S(x) + J^t(x)J(x) \quad ...(10)$$

For the Gauss Newton method, the term $S(x) = 0$, then the equation Gauss Newton method is:

$$x_{k+1} - x_k = [J^t(x)J(x)]^{-1} J^t(x)e(x) \quad ...(11)$$

We defined, the LM method combines two optimization methods, one of them being the gradient decent method, a technique for minimizing a function by using an iterative process. In each step, one adds the gradient of the function multiplied by a negative parameter.

$$x_{k+1} - x_k = -\mu \nabla f(x) \quad ...(12)$$

The Levenberg-Marquardt algorithm uses approximation to the Hessian matrix as a combination of two mentioned algorithms in the following update:

$$x_{k+1} = x_k - [J^t(x)J(x) + \mu I]^{-1} J^t(x)e(x) \quad ...(13)$$

Scale conjugate gradient (SCG) training function: Scaled conjugate algorithm is a supervised learning algorithm. This works for feed forward neural networks. It is a member of class of Conjugate Gradient Methods (CGM). It works with second order information from neural network.

$$s_k = \frac{E^I(w_k + \sigma_k . p_k) - E^I(w_k)}{\sigma_k} + \lambda_k p_k \quad ...(14)$$

SCG has two initial value parameters λ_k and σ_k. $E\,(w_{k+1}) \leq E\,(w_k)$ is a small number that is used to rectify the special case of converging to a function value of exactly zero. SCG is a batch learning method, hence there will be no effect if parameters are being shuffled. Comparing the number of epochs is not relevant in the case of SCG and other algorithms like standard back propagation rather iteration can be checked out for comparing SCG with standard back propagation (BP). One iteration in SCG needs the calculation of two gradients, and in addition to this, it requires one call to the error function, while one iteration in standard back propagation needs the computation of one gradient and one call to the error function. Moller defines a complexity unit (cu) to be equivalent to the complexity of one forward passing of all patterns in the training set. Then computing the

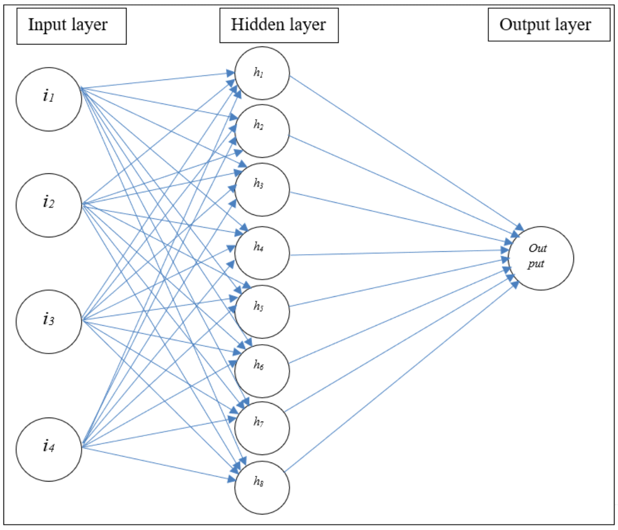


Fig. 3: The feed forward neural network structure of model.

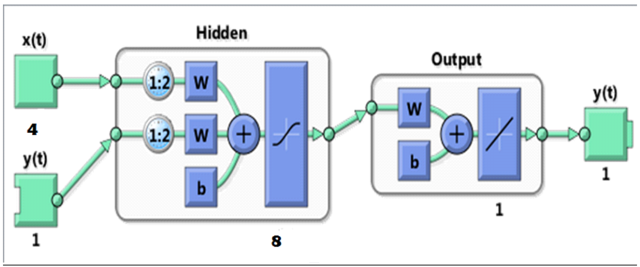


Fig. 4: MATLAB base NARX-NN Model structure.

error costs 1 cu, while computing the gradient can be estimated to cost 3 cu. According to Moller's metric, one iteration of SCG is as complex as around 10^{16} iterations of standard back propagation (BP).

ANN predictive models' structure: The artificial neural network is used in various fields of forecasting the future value. In this research, we constructed the Nonlinear Auto Regressive eXogenous neural network (NARX-NN) model to predict the water quality parameters along the Karachi coastal area near Karachi harbour. Six-year monthly data were used for the modelling purpose. The water quality parameters of interest are sea surface temperature, tides, pH, salinity, zinc (Zn), copper (Cu), chromium (Cr), nickel (Ni) and lead (Pb). In ANN, we first find out the structure of the input parameters associated with the problem. Most authors' conception experiments in imitation of help choosing the widest variety regarding input nodes while others adopt some intuitive or experiential ideas. In this study, the choice of input variables is based on a Pearson's correlation coefficient analysis. The input parameters are tides, sea surface temperature, salinity and pH for all models.

At second stage, normalized the data in range of 0-1 to save from satiation impact that may be due to use of sigmoid function. Furthermore, the data were categorized in three partitions such as training set, validation set and testing set. Seventy two readings of data distributed in the ratio of 70%, 15% and 15% for training, validation and testing set respectively. The training set is used to determine the adjusted weight and biases of networks. The test is used for calibration which prevents over training networks. The validation data will be used to check the performance of the model. Before using the training function select the hidden nodes and number of time delay. The number of hidden layers and nodes play a very important role for many successful applications of neural network. It is the hidden node in the hidden layer that allow NN to detect the feature, to capture the pattern in the data, and to perform complicated mapping between input and output variables (Farsa & Zolfaghari 2010). In the case of the popular one hidden layer networks, several practical guidelines exist. These include using "2*n*+1" (Lippman 1987, Hecht-Nielsen 1990), "2*n*"(Wong 1991), "*n*"(Tang & Fishwick 1993), "*n*/2" (Kang 1991). In this research, we chose 2*n* hidden layers, *n* is a number of inputs. However, none of these heuristic choices work well for all problems. In this study, we selected 2*n* number of hidden layers. The number of output nodes is relatively easy to specify as it is directly related to the problem, the number of output nodes often corresponds to the forecasting horizon. There are two types of forecasting, one step ahead (which use one output node) and multistep ahead forecasting. In this study, we used to forecast two steps ahead and the chemical parameters Cr, Cu, Ni, Pb and Zn used as an output individually for all locations. The Levenberg-Marquardt (LM) back propagation algorithm and Scale Conjugate Gradient (SCG) training function used for training the NARX-NN model. The performance criteria of these models mean square error (MSE) and correlation coefficient (r). In this study, we used MATLAB 2016(a) and Eview 8.1 software. Fig. 4 shows the NARX model structure in MATLAB.

RESULTS AND DISCUSSION

In this study, thirty NARX-NN architectures were developed. Each chemical parameter zinc, lead, copper, chromium and nickel at different six locations was taken as the output and sea surface temperature, tides, salinity and pH as an input for each model. We discuss each parameter location wise.

Chromium model: The NARX-NN models of chromium for all six locations have given good results (Table 1) by LM training function. The best model of chromium is CrL4 by LM training function because the *r* values of training and

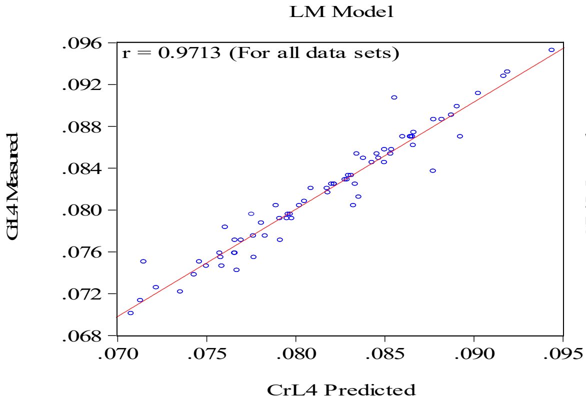


Fig. 5(a): Scatter diagram of LM training of CrL4 function.

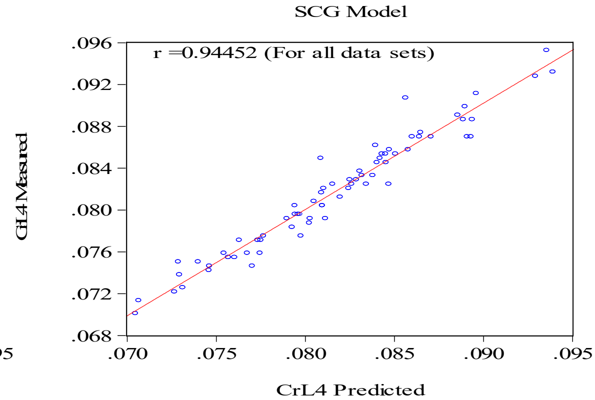


Fig. 5(b): Scatter diagram of SCG training of CrL4 function.

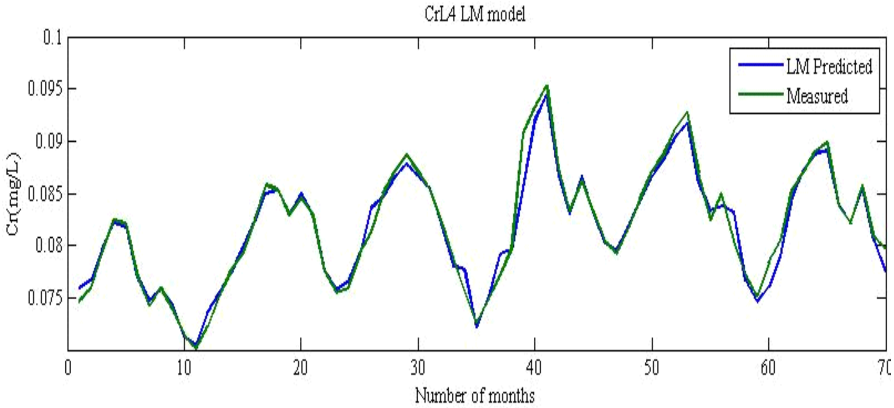


Fig. 5(c): Graphical comparison between measured and LM predicted monthly data of CrL4.

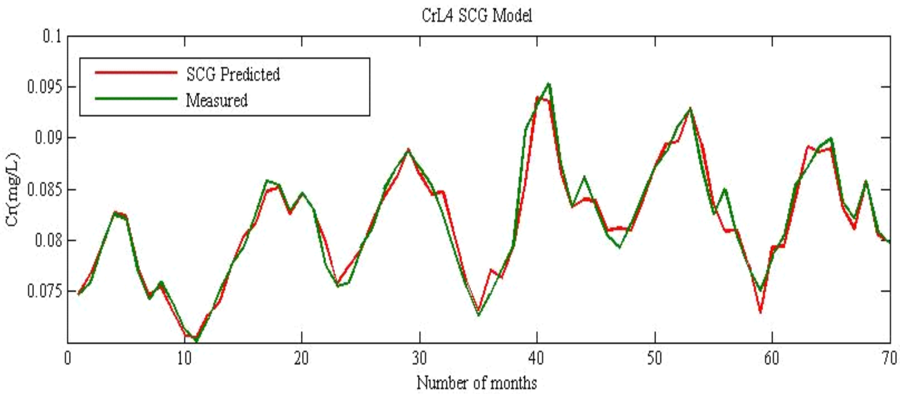


Fig. 5(d): Graphical comparison between measured and SCG predicted monthly data of CrL4.

testing data set are 0.997 and 0.921 respectively, with the *r* value of all data set as 0.971. The mean square error (MSE) for the training and testing dataset is 4.89E-04 and 9.37E-03 respectively. The scattered diagram in Fig. 5(a) shows that the correlation coefficient between all data set of measured and predicted data is 0.971. The predicted result of LM model is illustrated in Fig. 5 (c).

SCG predict the model results of chromium for all six locations as given in Table 2. CrL4 model correlation coefficient is good for all data set as compared to the other location model. The mean square error for testing and training data is 0.0046 and 0.0048 respectively and r value of all data set is 0.944. The scattered plot graph of observed and predicted CrL4 SCG model in Fig. 5(b) represents that the model is well fit under given input and output variables with number of eight hidden layers. The comparison graph in Fig. 5(d) shows the accuracy of SCG model, the simulated values are closed to the measured values of data.

Copper model: The results of all the six locations in LM model for copper are given in Table 3. The CuL2 model is good as compared to others, because correlation coefficient of training, testing and total data set are 0.993, 0.913 and 0.957 respectively. It means that the observed and simulated data accuracy is good by using the LM training function. In Fig. 6(a) the scattered graph of measured and predicted values of all data set for CuL2, obtained from training, testing and validation data, show a very close relationship. The correlation coefficient between both the data is 0.957. The graphical representation in Fig. 6(c) shows that the LM predicted and measured data are good at every point.

Results of scale conjugate gradient training function of all the six locations for copper are given in Table 4. CuL6 model is best on the basis of MSE and the correlation coefficient of training (MSE = 0.0095, r = 0.912) and testing (MSE = 0.0084, r = 0.941) data set as well as the correlation coefficient of the total data set is 0.91586. Scattered plot graph in Fig. 6(b) between the measured and predicted data set illustrate a strong correlation coefficient of 0.915. Fig. 6(d) represents the graphical comparison of SCG predicted and measured data.

Nickel model: All the nickel models by LM training function gave a good result (Table 5). On the basis of performance criterion NiL3 model is best with a correlation coefficient of training, testing and total data set as 0.955284, 0.888431 and 0.93377 respectively. The scattered graph of simulated values and predicted values with linear regression trends are plotted in Fig. 7(a). The visual representation of figure revealed the accuracy of LM models.

The NARX neural network with SCG training function results are revealed in Table 6 for all the locations for nickel. The summary of the table illustrates that the best performance of SCG training function is found to be at NiL6 model because the correlation coefficient between observed and predicted data for the total data set is 0.941. The scattered diagram in Fig. 7(b) shows that the correlation coefficient between all data set of measured and predicted data is close to one. The predicted result of SCG model is illustrated in Fig. 5(c).

Lead model: The LM model performance criterion is given in Table 7, it demonstrated the mean square error and correlation coefficient between measured and predicted data for training, testing dataset and correlation coefficient of all data set. According to the performance criterion, the best model is PbL2 because the *r* value between predicted and original data is 0.957. In Fig. 8(a) *r* value is an indication of the relationship between the simulated and original values of Pb, both data indicate a relatively good fit of the LM training function.

The SCG training function forecasting result of lead models is given in Table 8. In all the models, the PbL4 is a good fitness model because mean square error and correlation coefficient of training and test dataset is good as compared to other locations and *r* value for all data set is 0.880. Scattered plot diagram in Fig. 8(b) defines accuracy of SCG training function. The correlation coefficient between the observed and predicted data is 0.880 for the total data set. The correlation coefficients of measured and simulated values of all the models are best in the training data set as compared to the testing data set. The performance of all models of Pb is good.

Zinc model: The models of Zn by LM training function are represented in Table 9. The existence of a high correlation between predicted and observed data is present in the ZnL4 model compared to other models of Zn. The MSE shows better accuracy of model ZnL4, it is 0.0048 and 0.0107 for training and testing data set respectively. Fig. 9(a) scattered plot show the correlation between LM predicted and measured data.

NARX-NN predicted model of SCG training function is represented in Table 10. ZnL4 model is best on the basis of

Table 9: Correlation coefficient (r) and mean square error (MSE) of training and testing data set of zinc for all six locations of the Karachi coastal area along the harbor by LM training function.

Model	Training dataset		Testing dataset		Total
	MSE	r	MSE	r	r
ZnL1	0.0006	0.994	0.0141	0.943	0.940
ZnL2	0.00009	0.999	0.0194	0.894	0.943
ZnL3	0.0004	0.995	0.0187	0.807	0.941
ZnL4	0.0048	0.958	0.0107	0.874	0.944
ZnL5	0.0047	0.968	0.0200	0.788	0.909
ZnL6	0.0025	0.968	0.0241	0.941	0.942

Table 10: Correlation coefficient (r) and mean square error (MSE) of training and testing data set of zinc for all six locations of the Karachi coastal area along the harbor by SCG training function.

Model	Training dataset		Testing dataset		Total
	MSE	r	MSE	r	r
ZnL1	0.0050	0.943	0.0141	0.914	0.920
ZnL2	0.0066	0.942	0.0111	0.895	0.919
ZnL3	0.0041	0.945	0.0108	0.900	0.920
ZnL4	0.0042	0.966	0.0110	0.931	0.957
ZnL5	0.0065	0.941	0.0105	0.871	0.916
ZnL6	0.0071	0.908	0.0076	0.950	0.899

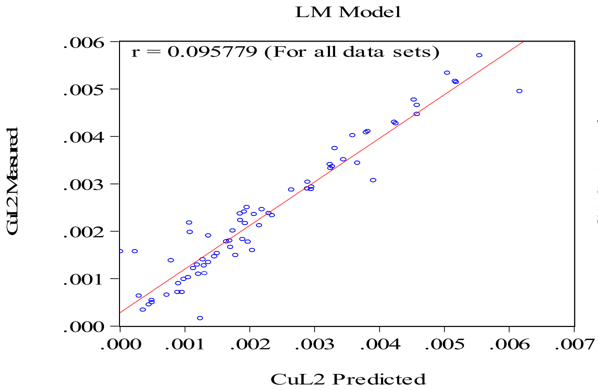


Fig. 6(a): Scatter diagram of LM training of CuL2 function.

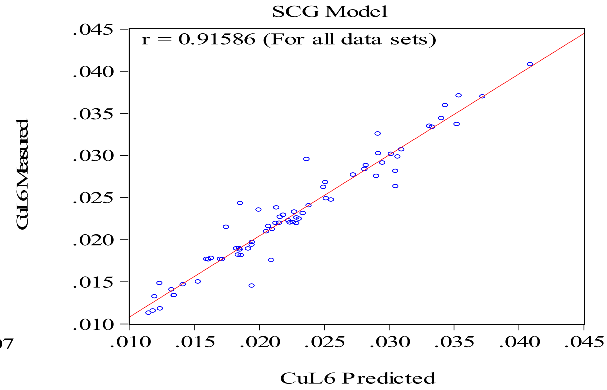


Fig. 6(b): Scatter diagram of SCG training of CuL4 function.

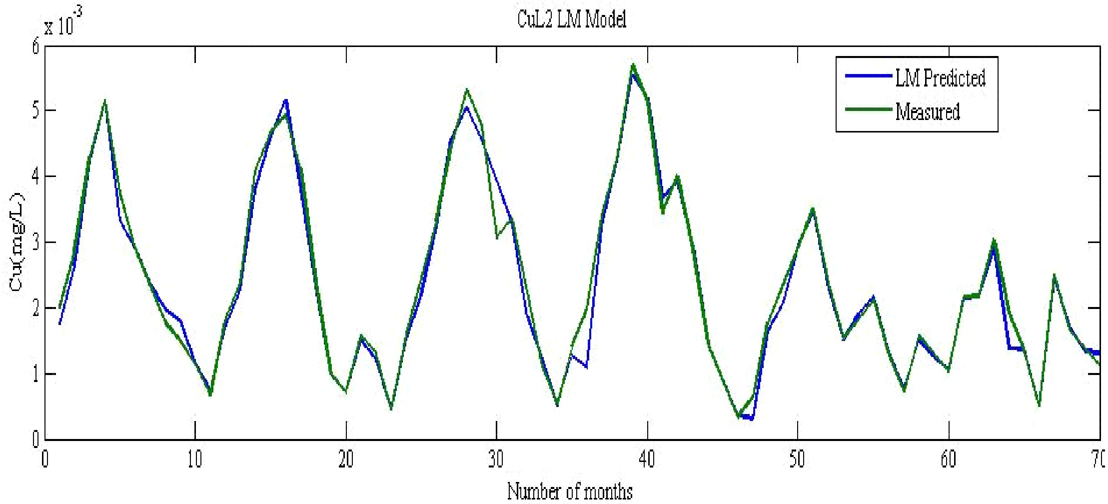
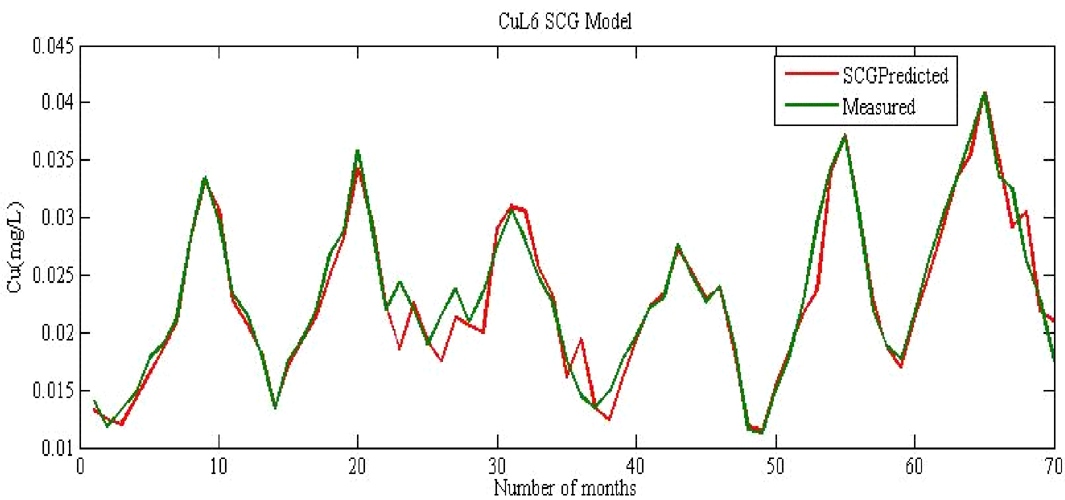


Fig. 6(c): Graphical comparison between measured and LM predicted monthly data of CuL2.



Fi

Fig. 6(d): Graphical comparison between measuredand SCG predicted monthly data of CuL6.

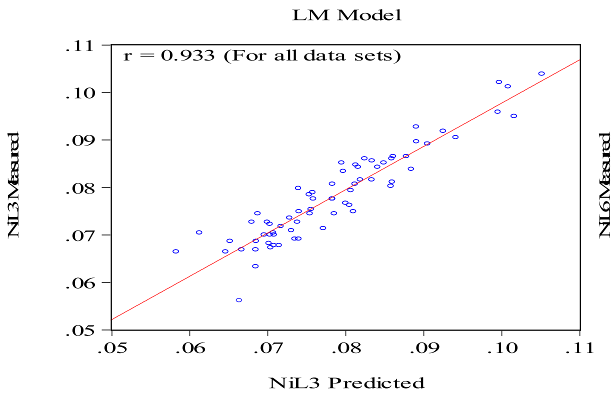


Fig. 7(a): Scatter diagram of LM training of NiL3 function.

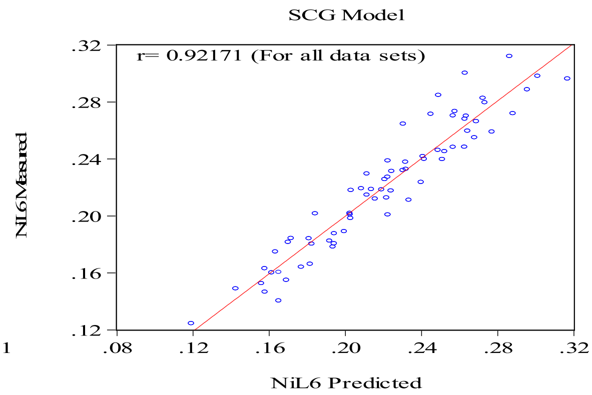


Fig. 7(b): Scatter diagram of SCG training of NiL6 function.

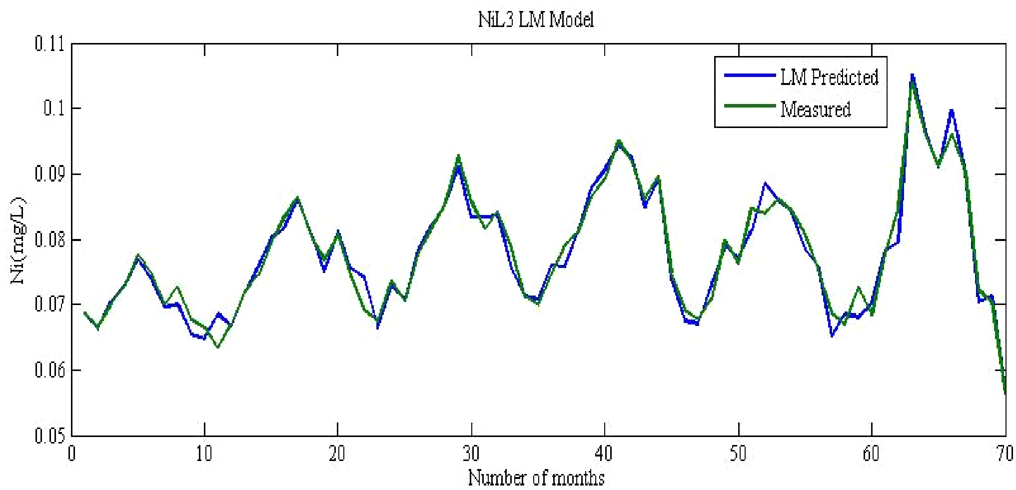


Fig.7(c): Graphical comparison between measured and LM predicted monthly data of NiL3.

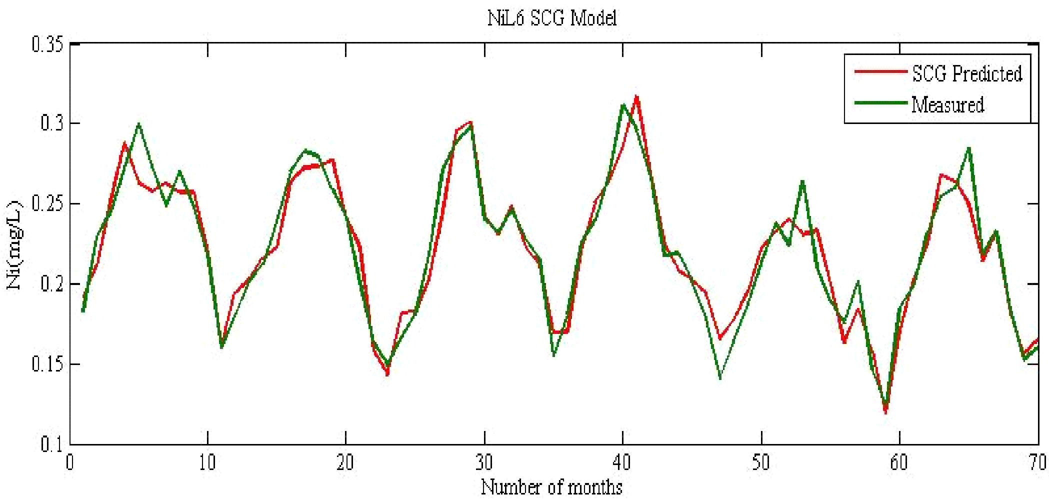


Fig. 7(d): Graphical comparison between measured and SCG predicted monthly data of NiL6.

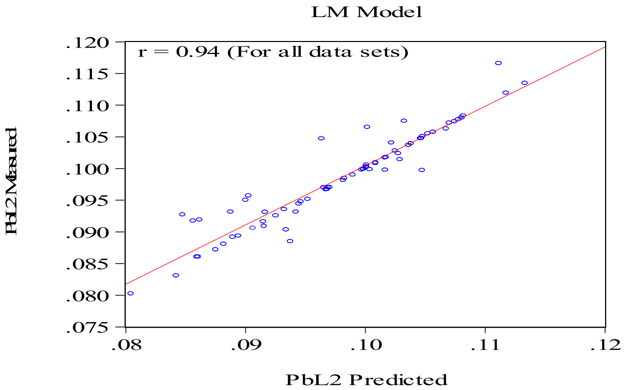


Fig. 8(a): Scatter diagram of LM training of PbL2 function.

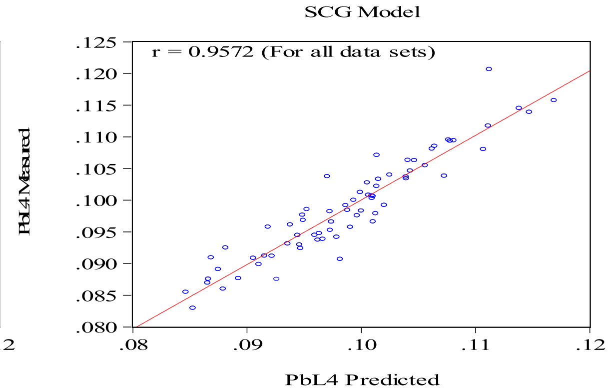


Fig. 8(b): Scatter diagram of SCG training of PbL4 function.

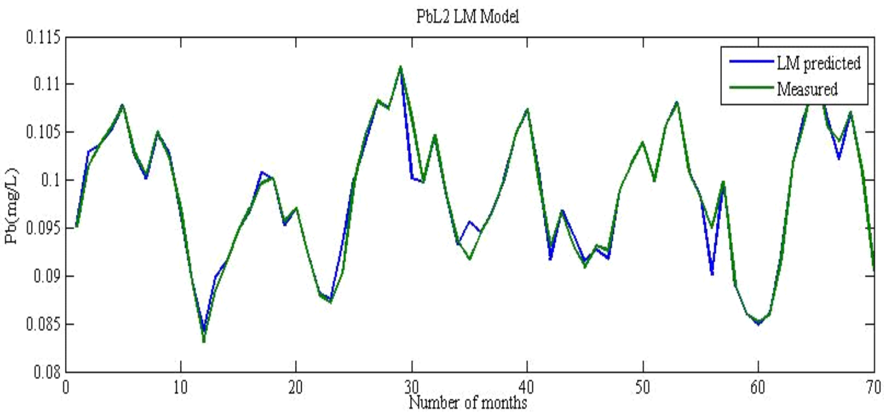


Fig. 8(c): Graphical comparison between measured and LM predicted monthly data of PbL2.

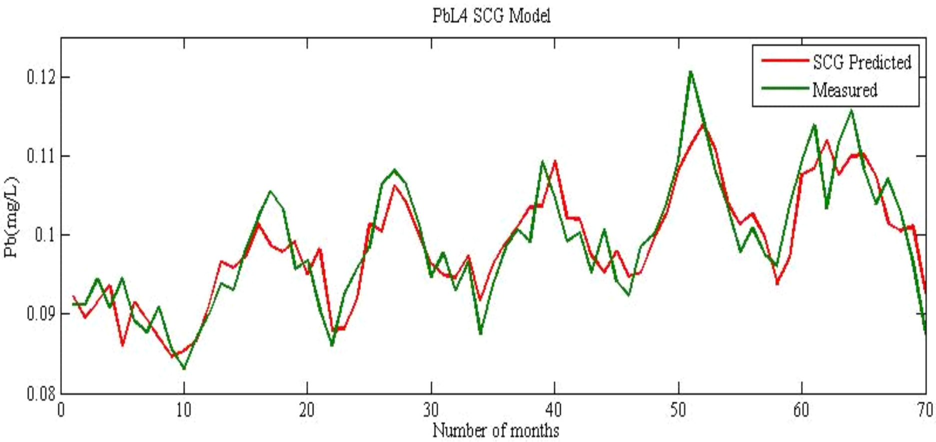


Fig. 8(d): Graphical comparison between measured and SCG predicted monthly data of PbL4.

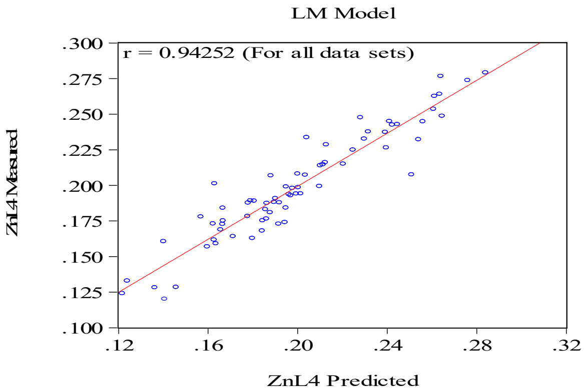


Fig. 9(a): Scatter diagram of LM training of ZnL4 function.

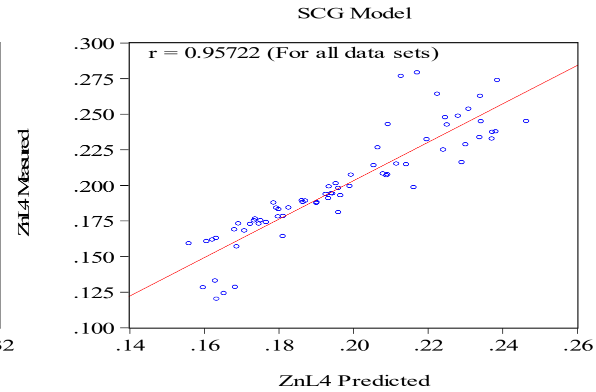


Fig. 9(b): Scatter diagram of SCG ZnL4 training function.

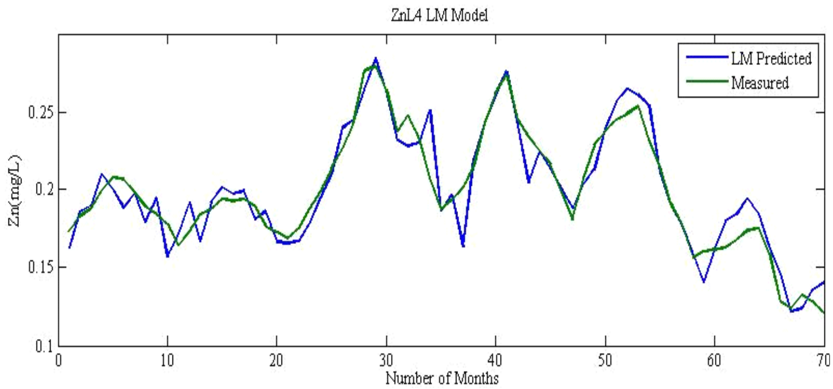


Fig. 9(c): Graphical comparison between measured and LM predicted monthly data of ZnL4.

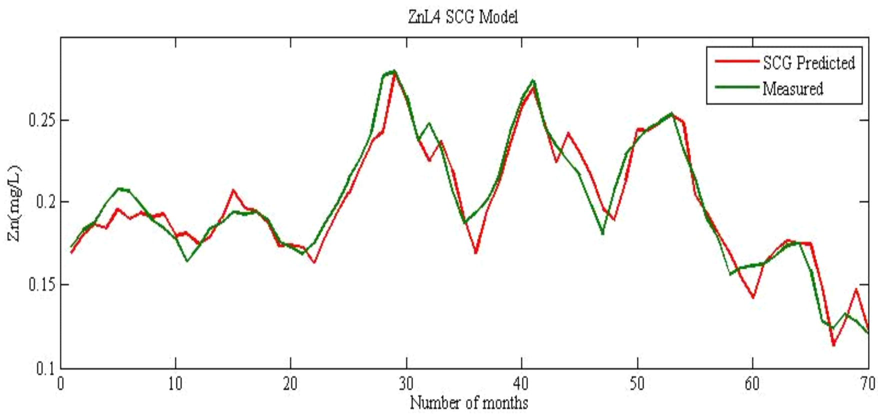


Fig. 9(d): Graphical comparison between measured and SCG predicted monthly data of ZnL4.

MSE and the correlation coefficient of training (MSE = 0.0042, r = 0.966) and testing (MSE = 0.0110, r = 0.931) data set as well as the correlation coefficient of the total data set is 0.957. Scattered plot graph in Fig. 9 (b) between measured and predicted data set illustrate a strong correlation coefficient of 0.915.

CONCLUSIONS AND SUGGESTIONS

There are many classical process-based models used to predict the water quality parameters. These models took a long time to process the data. ANN is a new technique used to predict the future values. It is used in biosciences, physical sciences, environmental sciences, etc. An ANN prediction capability was tested and found to be faster than that from a processed-based model with minimal input required. Many different techniques are used in ANN. In this paper, we have used Nonlinear AutoRegressive eXogenous Neural Network (NARX-NN) technique to predict heavy metals in sea surface water in Karachi harbour. There are two training functions namely, Levenberg Marquardt (LM) and Scale Conjugate Gradient (SCG), which are used in NARX-NN models to predict the heavy metals. Both training functions are supervised feed-forward neural network with the back-propagation learning strategy. The number of hidden layers 2n (n= number of inputs) is best for both the models. The Levenberg Marquardt (LM) and Scale Conjugate Gradient (SCG) training functions were compared to predict the Cr, Cu, Ni, Pb and Zn at six various locations. LM and SCG Model performance criterion based on mean square error (MSE) and correlation coefficient (r). The LM training function model accuracy is better than compared to SCG training function.

In this study, heavy metals in the sea surface of water in Karachi harbour are predicted using several existing factors and mathematical models. An approach based on the anticipated amount of metals in the sea will be invaluable in the conservation of biological diversity which in turn will provide sustainable use of its components. As a corollary, there will be fair and equitable sharing of benefits from use of inherent resources.

- Not degradable, compound half-life of centuries
- Strong perseverance, compound half-life of several years
- Medium perseverance, compound half-life of several months
- Low perseverance, compound half-life less than several months. Using the predicted model will ease in preventive measures of marine pollution. It will help dealing towards sustainable use of natural resources and advancement of the human race and aims at the sustainable development of the ecosystem.
- The dumping of household, hospital, fish harbour, pharmaceutical and industrial waste on sea side should strictly be prohibited because it is a major cause of harbour pollution.
- The future needs planning for municipal sewage water and industrial wastewater inlets, because nowadays all wastewater is discharged directly in coastal water through the Nala's, tanneries and Liyari River without treatment.
- The city needs to install new wastewater treatment plant with new technology and functioning of the installed plants properly.
- This study is useful to EPA and other coastal protection agencies for the protection and conservation of the harbour.
- Efficient and appropriate monitoring efforts are needed in the area as it is most susceptible to water contamination or degradation.
- The current study provides exclusive evidence of large scale deterioration of harbour water quality as measured by selected heavy metal concentrations.
- Artificial oxygenation should be provided in order to maintain marine life. This could be achieved by specially designed air compressor on boats. These measures are expected to produce a healthy effect on the overall environment of the Karachi coastal area and would help in rapid restoration of marine biota.
- Harbour pollution is also impacting the economy of Pakistan because the fishes, prawns and other seafood export affect due to pollution.
- The artificial neural network is good for prediction of heavy metals in coastal water and suggested to other researchers to apply ANN to predict DO, BOD, COD, TDS TTS and other water quality parameters.
- This study revealed that the physical parameters, sea surface temperature, pH, salinity and tides correlated to heavy metals of sea surface water.
- The research demonstrated the Levenberg Marquardt (LM) training function is good for predictive non-linear data as compared to scale conjugate gradient (SCG) training function in NARX-NN.
- Nowadays, the fuzzy neural network and deep learning neural network have been used for hydrological data.

REFERENCES

Altun, H., Bilgil, A. and Fidan, B.C. 2007. Treatment of multi-dimensional data to enhance neural network estimators in regression problems. Expert Systems with Applications, 32: 599-605.

Battiti, R. 1992. First-and second-order methods for learning: Between steepest descent and Newton's method. Neural Computing, 4: 141-166.

Beg, M.A.A., Mahmood, S.N. and Yousufzai, A.H.K. 1992. Heavy metals pollution in the coastal environment of Karachi. Pak. J.

Marine Science, 1(2): 117-126.

Box, G.E.P. and Jenkins, G.M. 1976. Time Series Analysis, Forecasting and Control. Holden-Day, San Francisco, USA.

Cheng, B. and Titterington, D.M. 1994. Neural networks: A review from a statistical perspective. Statistical Science, 9(1): 2-54.

Farsa, M.A. and Zolfaghari, S. 2010. Chaotic time series prediction with residual analysis method using hybrid Elman-NARX neural networks. Neurocomputing, 73: 2540-2553.

Golden, R.M. 1996. Mathematical Methods for Neural Network Analysis and Design. MIT Press, Cambridge.

Hecht-Nielsen, R. 1990. Neurocomputing. Addison-Wesley, Menlo Park, CA.

Kang, S. 1991. An Investigation of the Use of Feed Forward Neural Networks for Forecasting. Ph.D. Thesis, Kent State University.

Khan, S.H. and Saleem, M. 1988. A preliminary study of pollution in Karachi harbor. Proc. Int. Conf. Marine Sciences of the Arabian Sea, 539-547.

KPT 1995. Environmental Impact Studies. Final Report, Vol. 1, Karachi Port Trust, Karachi.

Lin, T., Horne, B.G., Tino, P. and Giles, C.L. 1996. Learning long-term dependencies in NARX recurrent neural networks. JEEE Transactions on Neural Networks, 7(6): 1329-1338.

Lippman, R.P. 1987. An introduction to computing with neural nets, IEEE ASSP Magazine, April, 4-22.

Maier., H.R. and Dandy, G.C. 2001. Neural network based modelling of environmental variables: A systematic approach. Mathematical and Computer Modelling, 33(6-7): 669-682.

Majeed, S., Zaman, S.B., Ali, I. and Ahmed, S. 2010. Situational analysis of Sindh coast-issues and options. Managing natural resources for sustaining future agriculture. Research Briefings, 2(11).

Melesse, A.M., Ahmad, S., McClain, M.E., Wang, X. and Lim, Y.H. 2011. Suspended sediment load prediction of river systems: an artificial neural network approach. Agric. Water Manag., 98: 855-866.

Movagharnejad, K., Tahavvori, A. and Ali, F.M. 2017. Artificial neural network modeling for predicting of some ion concentrations in the Karaj River. Advanced in Environmental Technology Article, 5(2): 109-117.

Najah, A., El-Shafie, A., Karim, O.A. and El-Shafie, H. 2013. Application of artificial neural networks for water quality prediction. Journal of Neural Computation and Application, 22(1): 187-201.

Palani, S., Liaong, S.Y. and Tkalich, P. 2008. An ANN application for water quality forecasting. Marine Pollution Bulletin, 56: 1586-1597.

Pankartz, A. 1983. Forecasting With Univariate Box-Jenkins Models: Concepts and Cases. John Wiley, New York.

Ripley, B.D. 1993. Statistical aspects of neural networks. In: Barndorff-Nielsen, O.E., Jensen, J.L., Kendall, W.S. (Eds.), Networks and Chaos-Statistical and Probabilistic Aspects, London, 40-123.

Rumelhart, D.E., Hinton, G.E. and Williams, R.J. 1986. Learningrepresentations by back-propagating error. Nature, 323: 533-536.

Saleem, M. and Kazi, G.H. 1995. Distribution of heavy metals in the seawater and surficial sediments of the Karachi harbor. The Arabian Sea living marine resources and the environment. Eds. Mary, Frances Thompson, Nasima M. Tirmizi, 659-666.

Tang, Z. and Fishwick, P.A. 1993. Feed forward neural nets as models for time series forecasting. ORSA Journal on Computing, 5(4): 374-385.

World Wide Fund for Nature (WWF) 2002. Study of heavy metal pollution level and impact on the fauna and flora of the Karachi and Gwadar coast. Final Project Report, No. 50022801.

White, H. 1989. Learning in artificial neural networks: A statistical perspective. Neural Computation, 1: 425-464.

Wong, F.S. 1991. Time series forecasting using backpropagation neural networks. Neurocomputing, 2: 147-159.

Zang, C., Huang, S., Wu, M., Du, S., Scholz, M., Gao, F., Lin, C., Guo, Y. and Dong, Y. 2011. Comparison of the relationships between pH, dissolved oxygen and chlorophyll *a* for aquaculture and non-aquaculture waters. Water Air Soil Pollution, 219: 157-174.