

## RESEARCH ARTICLE

## DEVELOPMENT OF TRACE METALS CONCENTRATION MODEL FOR RIVER: APPLICATION OF PRINCIPAL COMPONENT ANALYSIS AND ARTIFICIAL NEURAL NETWORK

Fikriah Faudzi<sup>a</sup>, Mohd Fuad Miskon<sup>b</sup>, Azman Azid<sup>c</sup>, Nur Shuhada Tajudin<sup>a</sup>, Shazlyn Milleana Shaharudin<sup>d</sup>

<sup>a</sup> *Kulliyah of Science, International Islamic University Malaysia, Jalan Sultan Ahmad Shah, Bandar Indera Mahkota, 25200 Kuantan, Pahang, Malaysia,*

<sup>b</sup> *Institute Oceanography and Maritime Studies, Kulliyah of Science, International Islamic University Malaysia, Jalan Sultan Ahmad Shah, Bandar Indera Mahkota, 25200 Kuantan, Pahang, Malaysia,*

<sup>c</sup> *Faculty of Bioresources and Food Industry, Universiti Sultan Zainal Abidin, Besut Campus, 22200 Besut, Terengganu, Malaysia,*

<sup>d</sup> *Department of Mathematics, Faculty of Science and Mathematics, Universiti Pendidikan Sultan Idris, 35900 Tg Malim, Perak, Malaysia*

\*Corresponding Author Email: [fikriahf@iiu.edu.my](mailto:fikriahf@iiu.edu.my)

*This is an open access journal distributed under the Creative Commons Attribution License CC BY 4.0, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited*

## ARTICLE DETAILS

## Article History:

Received 06 June 2023

Revised 11 July 2023

Accepted 22 August 2023

Available online 25 August 2023

## ABSTRACT

Rapid development along the Kuantan River was long perceived as the rivers serve many communities in terms of drinking water source, domestic, fisheries, recreation, and agricultural purposes. Due to the rapid changes in technology and upsurge in chemical usage, pollutant alterations turn out to be more drastic with respect to space and time. Research on the trace metals in river water is quite limited in Malaysia, probably due to their ppb-level existence and the need for special handling techniques. Hence, the aim of this study is to forecast heavy metals concentration in Kuantan River waters using a collective of 10 years (2007 – 2016) dataset of heavy metals that provided by the Department of Environment, Malaysia. Principal Component Analysis (PCA) was used to compute the data, which showed that As, Cr, Fe, Zn and Cd explain 67.3% of the total variance through three principal components. For ANN computation, those significant metals extracted from rotating PCA was selected and used in ANN model. The developed approaches were trained and tested using 80% and 20% of the data, respectively. Then, the coefficient of determination ( $R^2$ ) was executed to calculate the model performance. Out of five metals, only As shown acceptable  $R^2$  for ANN models with 0.8690 and 0.8088 for training and testing, respectively, probably due to the model's limitation. Generally, this study illustrates the usefulness of PCA and ANN for analysis and interpretation of complex data sets and understanding the temporal and spatial variations in the Kuantan River for effective river water management.

## KEYWORDS

Neural networks, Estuaries, Chemometrics, Heavy metals, Forecasting

## 1. INTRODUCTION

Urbanization, industrialization, logging, and mining were undertaken in order to meet the requirements of a growing population. Heavy metals levels in river waters and sediments have risen as a result of these measures (Suthar et al., 2009). Metals are mobile in the water column and travel through channels from upstream to downstream and vice versa, eventually sinking in riverbeds. Metals will accumulate in plants and fauna across the food chain, and natural deterioration will occur (Rossi et al., 2013). The metals were bound with silicates and minerals in an undisturbed biosphere, making them nearly immovable (Medici et al., 2011).

Principal Component Analysis (PCA) and Artificial Neural Network (ANN) are used to characterize and evaluate the surface water condition and manage to verify temporal and spatial variations caused by natural and anthropogenic factors (Singh et al., 2005; Yidana et al., 2008; Juahir et al., 2013). The metals were bound with silicates and minerals in an undisturbed biosphere, making them nearly immovable (Medici et al., 2011).

anthropogenic pollutants are responsible for the high variation in the water quality of the river water (Han et al., 2009). Similarly, some researchers applied PCA in the surface water quality data of Ceyhan River (Tanriverdi et al., 2010). Three PCs were significantly identified corresponding to areas close to cities, which presented low dissolved oxygen contents and high concentrations of physicochemical parameters, suggesting anthropogenic inputs. The stations in the vicinity of industries have higher pollution due to the discharge of wastewater from industries and domestic activities.

For the past decades, ANNs have been widely applied in various research areas including upgrading the water quality management (Nasri 2010; Chitra et al., 2012; Mutalib et al., 2013; Azid et al., 2014; Kuo et al., 2006; Keskin et al., 2014). This ultimately allows ANNs to model environmental systems without prior specification of the algebraic relationships between variables. Now, ANN has also become an indispensable tool for environmental management in Malaysia, as many researchers have engaged ANN to many water resources applications such as forecasting and modeling (Toriman et al., 2010; Toriman et al., 2011; Khan et al., 2012; Uca et al., 2018).

## Quick Response Code



## Access this article online

## Website:

[www.watconman.org](http://www.watconman.org)

## DOI:

10.26480/wcm.02.2023.122.127

Several comprehensive studies using these techniques were reported which includes research on air pollution, such as air quality pattern recognition in Selangor, Malacca and Sarawak and air pollution and modelling around Malaysia (Mutalib et al., 2013; Azid et al., 2014; Ahmad Isiyaka and Azid, 2015; Azid et al., 2017). Besides, studies also used PCA and ANN to forecasting dissolved oxygen, modelling river discharges, prediction of water quality index and spatial water quality assessment of Langat River Basin, water quality variation on rivers and lakes in Malaysia, water quality in Terengganu estuaries, source apportionment in Perlis River Basin and many more (Juahir et al., 2003a; 2004a; 2004b; 2011; Low et al., 2016; Khalit et al., 2017; Samsudin et al. 2017b). Chau has reviewed the development and current progress of the integration of artificial intelligence into water quality modeling, while has identified the appropriate measures to improve the river water quality in Juru River (Chau, 2006; Toriman et al., 2011). These characteristics render ANNs to be very suitable tools for handling various hydrological modelling problems.

Water quality involves several aspects, including significant non-linear relationships with the variables, making traditional data analysis less reliable in understanding the situation (Goethals et al., 2007). Artificial neural networks (ANNs), on the other hand, are capable of simulating basic human brain features such as self-organization, error tolerance, and self-compliance, and have been widely used for model recognition, analysis, and prediction, as well as system identification in order to improve design (Maier and Dandy, 1998). Furthermore, ANN modelling is likely to shorten the computation time and lower the likelihood of errors in the produced model.

According to the DOE's 2015 Environmental Quality Report, river water quality was monitored across 158 river basins constantly (DOE, 2015). The continuous assessment carried out by DOE resulted in the mass of environmental data sets. However, the data sets were not fully utilized since lack of advanced statistical techniques to extract all sorts of information. Ten years historical data matrix acquired from DOE continuous assessment, from year 2007 to 2016 were presented in this paper. The data sets were performed to the receptor models' techniques involves principal component analysis (PCA) and varimax rotation and numerical models from ANN. This study attempts to predict heavy metals concentration in Kuantan River water.

## 2. MATERIALS AND METHODS

### 2.1 Study Site

Kuantan River is situated contiguous to the Tanjung Lumpur mangrove in Kuantan district, Pahang. The watershed covers an expanse of 1,586 km<sup>2</sup> and is about 80 km long and varied depth between 2 m to 10 m (DOE 2010). Kuantan River flows through two mukims i.e., Ulu Kuantan and Kuala Kuantan and flowing out to the South China Sea.

### 2.2 Pre-Treatment Data

Six physicochemical parameters and seven heavy metals were detected at monitoring stations along the Kuantan River. The monitoring locations were selected based on the ten-year data sets published from 2007 to 2016, but some stations are missing in the raw data, and some data is insufficient, possibly due to technical failure of measuring equipment and lab work. For source apportionment and model construction, a total of 3900 annotations were introduced. The parameters consist of salinity, pH, turbidity, temperature, dissolved oxygen (DO), conductivity, as well as heavy metals, specifically As, Cd, Cr, Pb, Hg, Zn and Fe.

Firstly, data were organized following the station and year of survey. Variables with values below the detection limit were replaced with one half of the detection limit value to ensure that no missing data was present. The normalcy test was computed using XLSTAT software based on the Anderson - Darling test. Data which is not regularly distributed was subjected to the log-scaling procedure, which included a mix of centering and normalisation (Felipe-Sotelo et al., 2007). Then, using XLSTAT software and JMP10, statistical calculations for PCA and ANN were performed, respectively.

### 2.3 Principal Component Analysis

A thorough pretreatment data set was carried out to deliver a clearer information about the intricate data, due to the fact that the PCA is profound to outliers, loss of information and a poor linear correlation between the variables of the inadequately presented variables (Sarbu and Pop, 2015). PCA is developed to reduce a set of variables of interest into a lower number of components while maintaining the majority of the relevant data (Shaharudin et al., 2013; Panigrahi et al., 2007; Khalit et al.,

2018; Sulaiman et al., 2018). This is accomplished by transforming a set of possibly correlated observations into a set of linearly uncorrelated variables known as principle components. (Shaharudin et al., 2018). This analysis will clarify these variations in the conditions of their mutual underlying dimension by defining empirical approximations of the structure of the variables (Hair et al., 1995; Mutalib et al., 2013; Low et al., 2016). The first principal component explains as much variation in the primary data as possible. Then, subject to being uncorrelated with the prior component, each subsequent component explains for as much of the remaining variation as possible. PCA can be computed using Eq. (1).

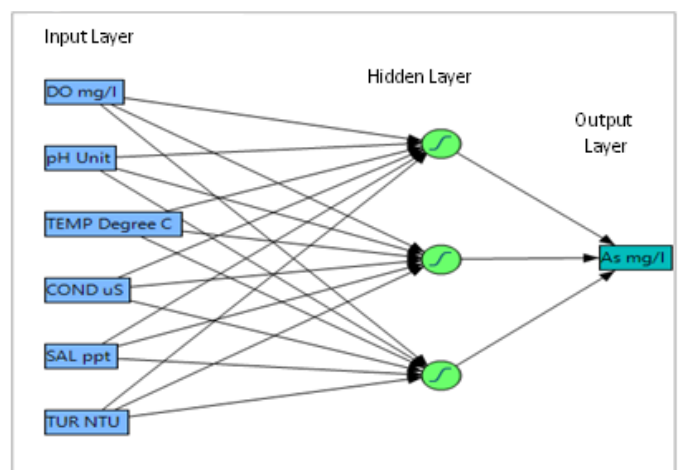
$$f_{ij} + f_{j1}z_{i1} + f_{j2}z_{i2} + \dots + f_{jm}z_{im} + e_{ij} \quad (1)$$

where  $j$  is the measured variable,  $f$  is the factor loading,  $z$  is the factor score,  $e$  is the residual term accounting for errors,  $i$  is the number of samples, and  $m$  is the total number of factors.

In this study, Varimax rotation was used to increase the value of PCA by rotating the eigenvalues (Helena et al., 2000). The main goal of applying the varimax rotational is to achieve a much more concise and eloquent exemplification of the main genes, resulting in a new group of variables known as varimax factors (VFs) (Chou et al., 2009; Samsudin et al., 2017a). Varimax rotation was used to reduce the dimensionality of the data and identify highly significant new variables when the PCs created by PCA were not ready for interpretation (Ismail et al., 2016; Samsudin et al., 2017b; Samsudin et al., 2017c). Factor loading is considered highly significant when the Varimax factor (VF) coefficient has a correlation of greater than 0.75. (Liu et al., 2003). While moderate and weak factor loading are defined as correlations between 0.75-0.50 and 0.5, respectively (Nair et al., 2010).

### 2.4 Artificial Neural Networks

ANNs can use a portion of the data set's input and output training patterns to characterise nonlinear and complex relationships. These methods create a nonlinear relationship between inputs and outputs (Hornik et al., 1990; Samsudin et al., 2017c; Azid et al., 2017; Rani et al., 2018). An ANN can be built using an architecture that expresses the node-to-node association form, connecting weights method identification, and activation function (Kisi, 2008). Since ANN can infer a system's dynamics from data, it can address large-scale, complicated issues (ASCE, 2000; Azaman et al., 2015). The most often used ANN model is the multilayer perceptron network (MLPN) model, which is based on one of the neural network topologies. A three-layered MLPN is made up of neurons in each layer as well as components that connect them (Haykin, 1999; Rani et al., 2018). The optimization process for weights is used to find the right weights to minimise inaccuracies throughout network training; this operation continues until the values of the output layer are close enough to the real outputs (Hornik et al., 1990). The weights were tuned using a training algorithm in this investigation (Kisi, 2008). Figure 1 depicts a feed-forward network with one hidden layer and multiple nodes between the input and output layers in this investigation.



**Figure 1:** The neural network model for estimating heavy metals concentrations of Kuantan River

ANN have been effectively practiced in different research areas of environmental problems. The DOE data was split into two sections: training and testing (80 percent and 20 percent, respectively). Since there are no criteria in ANN modeling regarding the number of hidden nodes, choosing the optimal number of hidden nodes is an unmanageable job. Here, a three-layer MLPN with one hidden layer, were compute using the

trial-and-error procedure to select the number of hidden nodes (Kisi, 2009; Azid et al., 2017). For the hidden and output node activation functions, sigmoid and linear functions were selected, respectively. For all the heavy metal concentrations, the ANN models were first developed using the data in the training sets to get the optimized set of learning coefficients and then examined. RMSE and determination coefficients ( $R^2$ ) were used as evaluation criteria. For the ANN simulations, program codes were written in JMP10 software.

The subsequent statistical indicators were designated in the performance assessment ANN models:

Root-mean-square error (RMSE) Eq. (2)

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

Coefficient of determination ( $R^2$ ) Eq. (3)

$$R^2 = \frac{[\sum_{i=1}^n (O_i - \bar{O}_i) \cdot (P_i - \bar{P}_i)]^2}{\sum_{i=1}^n (O_i - \bar{O}_i) \cdot \sum_{i=1}^n (P_i - \bar{P}_i)} \quad (3)$$

Where  $n$  is the total number of data, and  $P_i$  and  $O_i$  are the heavy metal concentrations predicted by the ANN methods and measured values, respectively.

### 3. RESULTS AND DISCUSSION

Descriptive statistics of trace metals in water samples collected from the Kuantan River indicated in Table 1. The average levels of As, Hg, Cd, Cr, Pb, Zn and Fe were  $0.0016 \pm 0.0034$  mg L<sup>-1</sup>,  $0.0001 \pm 0.0005$  mg L<sup>-1</sup>,  $0.0005 \pm 0.0003$  mg L<sup>-1</sup>,  $0.0008 \pm 0.0010$  mg L<sup>-1</sup>,  $0.0050 \pm 0.0000$  mg L<sup>-1</sup>,  $0.0321 \pm 0.0421$  mg L<sup>-1</sup> and  $0.7328 \pm 1.0941$  mg L<sup>-1</sup>, respectively. The results of statistical analysis showed that the mean concentrations of As, Hg, Cd, Cr, Pb, Zn and Fe were lower than the permissible limit of based on INWQS that is, 0.05 mg L<sup>-1</sup>, 0.001 mg L<sup>-1</sup>, 0.01 mg L<sup>-1</sup>, 0.05 mg L<sup>-1</sup>, 0.05 mg L<sup>-1</sup>, 5.00 mg L<sup>-1</sup> and 1.00 mg L<sup>-1</sup>, respectively (DOE, 2010).

Generally, the water quality data set reveals three varifactors, which defined the 67.3% loadings of the total variance gained over the rotated PCA. Table 2 shows the eigenvalues and factor loadings. Varifactor 1 (VF1) had significant positive loadings on As and Cr, accounting for 26.6 percent of the total variance. VF2 accounts for 23.6 percent of total variation and has significant positive loadings on Zn and Fe, whilst VF3 accounts for 17.1 percent of total variance and has high positive loadings on Cd. These five metals can be clustered together for their common source and controlling factors. Kuantan River and the adjoining regions attribute to small and medium-scale industrial areas that employed these components for countless commercial product and discarded their untreated wastes into the river. Through their activities, the densely populated urban area along the Kuantan River may also contribute significant amounts of metals into the waters, such as agricultural, boating and recreational activities, effluents from the nearby hospital, domestic sewages and drainage run-off.

**Table 1:** Descriptive statistic of heavy metals in Kuantan River (mg L<sup>-1</sup>)

Variable	Minimum	Maximum	Mean	Std. deviation
As	0.0005	0.0373	0.0016	0.0034
Hg	0.0001	0.0081	0.0001	0.0005
Cd	0.0005	0.0040	0.0005	0.0003
Cr	0.0005	0.0082	0.0008	0.0010
Pb	0.0050	0.0050	0.0050	0.0000
Zn	0.0050	0.3263	0.0321	0.0421
Fe	0.0050	9.4500	0.7328	1.0941

The relevant factor loadings obtained from the PCA were investigated, and various input combinations were examined and chosen in order to create the ANN models. The data was then split into training and testing stages in the second step. The best network architecture was chosen through trial and error, and the heavy metals concentration was calculated using the best ANN model. Fundamentally, to define the best model, the highest value of  $R^2$  and RMSE must be achieved. Training data proportion is a very important element for the efficiency of an ANN model; Insufficient neural network training could be caused by a low ratio of training data; hence, providing an appropriate amount of data can optimize the model's accuracy during the training and testing periods.

**Table 2:** Eigenvalues from principal component analysis shows variability, cumulative and factor loadings.

	F1	F2	F3
Eigenvalue	1.598	1.416	1.024
Variability (%)	26.626	23.608	17.059
Cumulative %	26.626	50.234	67.293
As mg L <sup>-1</sup>	0.869	0.075	0.131
Hg mg L <sup>-1</sup>	0.180	0.308	-0.361
Cd mg L <sup>-1</sup>	-0.058	0.018	0.908
Cr mg L <sup>-1</sup>	0.889	-0.001	0.029
Pb mg L <sup>-1</sup>	0.000	0.000	0.000
Zn mg L <sup>-1</sup>	-0.122	0.797	0.205
Fe mg L <sup>-1</sup>	-0.027	0.825	-0.096

The ANN model that was built up to predict as concentration are formulated in Eq. 4. This formula yields an exact representation, indicated by the statistical values such as determination coefficient and root mean square error statics, over the entire range of operation conditions.

$$\begin{aligned} \Theta 1 = & -0.0003 * [\tanh(.5 * (-45.6830 * DO\_mg\_l + 0.5705 * pH\_Unit + \\ & -0.2189 * TEMP\_Degree\_C + -0.0001 * COND\_uS + -0.1447 * SAL\_ppt + \\ & 0.0304 * TUR\_NTU + 282.3177))] + 0.0489 * [\tanh(.5 * (1.1993 * DO\_mg\_l + \\ & -18.3027 * pH\_Unit + -2.4354 * TEMP\_Degree\_C + 0.0004 * COND\_uS + \\ & -0.7817 * SAL\_ppt + 0.0594 * TUR\_NTU + 181.0156))] + -0.0490 * \\ & [\tanh(.5 * (1.3158 * DO\_mg\_l + -18.4603 * pH\_Unit + -2.4432 * TEMP\_Degree\_C \\ & + -0.0002 * COND\_uS + 0.0600 * SAL\_ppt + 0.0620 * TUR\_NTU + 181.4276))] \\ & + 0.0017 \end{aligned} \quad (4)$$

where,  $\Theta 1$  = Predicted As<sub>mg\_1</sub>

To simplify an expression, all the variables in the model (4) are denotes as follow:

$$x = DO_{mg}; y = pH_{Unit}; k = TEMP_{Degree_C}; w = COND_{uS}; t = SAL_{ppt}; g = TUR_{NTU}$$

$$\begin{aligned} \Theta 1 = & -0.0003 * [\tanh(.5 * (-45.6830 * x + 0.5705 * y + -0.2189 * k + -0.0001 * w \\ & + -0.1447 * t + 0.0304 * g + 282.3177))] + 0.0489 * [\tanh(.5 * (1.1993 * x + \\ & -18.3027 * y + -2.4354 * k + 0.0004 * w + -0.7817 * t + 0.0594 * g + 181.0156))] \\ & + -0.0490 * [\tanh(.5 * (1.3158 * x + -18.4603 * y + -2.4432 * k + 0.0002 * w \\ & + 0.0600 * t + 0.0620 * g + 181.4276))] + 0.0017 \end{aligned} \quad (5)$$

And to make it clear, equation (5) is illustrated as follows:

$$\begin{aligned} \Theta 1 = & -0.0003 * [\tanh(-22.8415x + 0.2853y - 0.1095z - 0.0005w - \\ & 0.0724t + 0.0152f + 141.1589)] + 0.0489 * [\tanh(0.5997x - 9.1514y - \\ & 1.2177z + 0.0004w - 0.3909t + 0.0297f + 90.5078)] - 0.0490 * \\ & [\tanh(0.6579x - 9.2302y - 1.2216z - 0.0001w + 0.0300t + 0.0310f + \\ & 90.7138)] + 0.0017 \end{aligned} \quad (6)$$

The performance of this model in terms of  $R^2$  and RMSE during the training and testing period is presented in Table 3. The  $R^2$  values attained while training and testing the as model were 0.8469 and 0.8084 respectively, which indicate the acceptable forecasting performance. Nonetheless, the  $R^2$  computed when training and testing the Cd, Cr, Zn and Fe model were weak and demonstrated greater ranges between the training and testing periods, which was unacceptable.

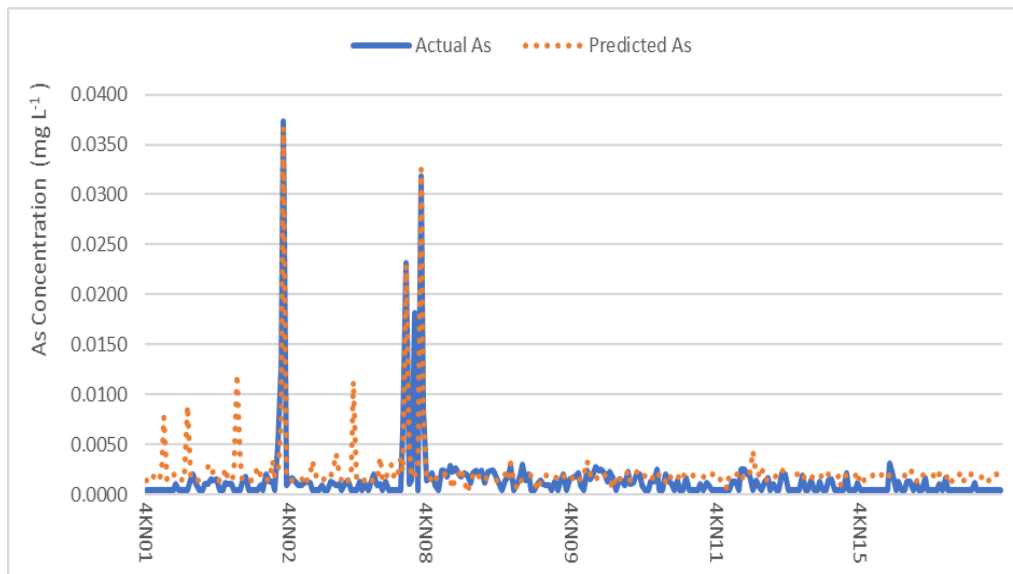
**Table 3:** Comparative performance of ANNs for As, Cd, Cr, Zn and Fe

Metal	Training		Testing	
	RMSE	R2	RMSE	R2
As	0.0014	0.8469	0.0029	0.8084
Cd	0.0002	0.9393	0.0011	-0.7618
Cr	0.0008	0.7629	0.0013	0.1885
Zn	0.0390	0.2646	0.0450	0.1012
Fe	1.0490	0.1150	0.9304	0.1588

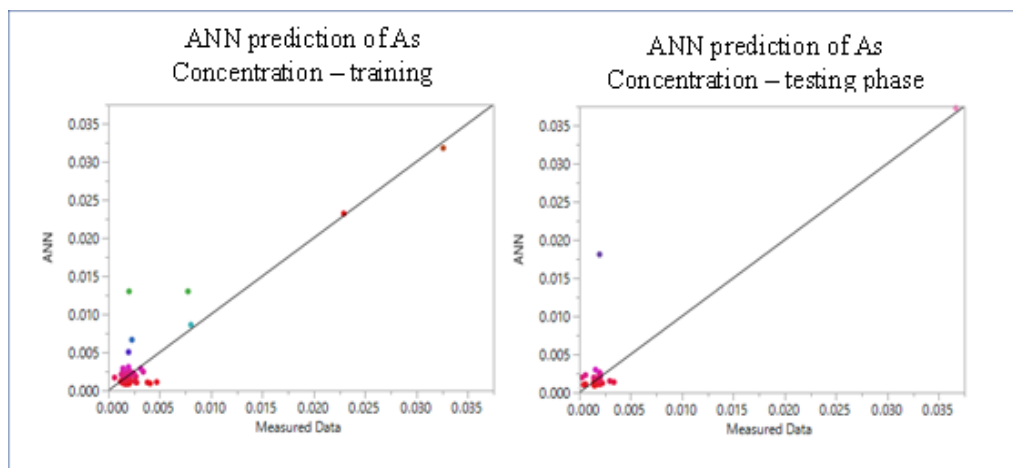
The extent of the match between the actual and predicted concentration of ANN models is shown in Figure 2. The graphs presented comparable trends between actual and predicted As except for some points, predicted as distributed far from the actual As. Figure 3 further emphasizes the better performance of the as model through the scatter plots of observed As versus simulated As concentrations. The left side represents the model training period, and the right side represents the reliable results for the

testing period. The scatterplots illustrate that the simulated data show an agreement with the observed data of as concentration. The results of this study have shown that ANN was effective for predicting the as

concentration of Kuantan River. It is clear that the model's performance was consistent, and the Kuantan River model accurately predicted metal concentrations.



**Figure 2:** Comparison of the actual as concentration values and As predicted values in testing phase



**Figure 3:** Observed and simulated as concentrations by ANN model during training (left) and testing (right) phases

#### 4. CONCLUSION

ANNs are capable of predicting heavy metals concentrations using historical information. Rather than straightforwardly using the mass data set, the PCA employment in this model is observed as wise resolution, since PCA lowered the number of inputs and narrowed down the model complexity. Significant factors generated by PCA/FA were As, Cd, Cr, Zn and Fe were used for model training and testing. The performance of these five metals was evaluated and it was found that only as shown good RMSE and  $R^2$  were very consistent. The performance of PCA and ANN has successfully developed a prediction model for arsenic concentration. Thus, verified that the PCA-ANN approach can be implemented by DOE or another public agency to monitor and manage the environment efficiently. Sampling, labors, and cost of chemical used in the analyses definitely can be minimized by using these developed models.

#### ACKNOWLEDGEMENT

Many thanks to Department of Environment for kindly providing the data used in this study. This research was funded by the Research Management Centre Grant (RMC20-001-0001) of International Islamic University Malaysia (IIUM).

#### REFERENCES

Abernathy, C.O., Liu, Y.P., Longfellow, D., Aposhian, H.V., Beck, B., Fowler, B., Goyer, R., Menzer, R., Rossman, T., Thompson, C., Waalkes, R., 1999. Arsenic: Health effects, mechanisms of actions and research issues. *Environ Health Perspect*; 107, Pp. 593–597.

Ahmad, I.H., and Azid, A., 2015. Air Quality Pattern Assessment in Malaysia Using Multivariate Techniques. *Malaysian Journal of Analytical Sciences*, 19 (5), Pp. 966 – 978.

ASCE Task Committee on Application of Artificial Neural Networks in Hydrology. 2000. Artificial neural networks in hydrology: preliminary concepts. *J Hydrol Eng*, 5 (2), Pp. 115-23.

Azaman, F., Azid, A., Juahir, H., Mohamed, M., Yunus, K., Toriman, M.E., Mustafa, A.D., Amran, M.A., Hasnam, C.N.C., Umar, R., Hairoma, N., 2015. Application of artificial neural network and response surface methodology for modelling of hydrogen production using nickel loaded zeolite. *Jurnal Teknologi*, 77 (1), Pp. 109-118.

Azid, A., Juahir, H., Toriman, M.E., Kamarudin, M.K.A., Saudi, A.S.M., Hasnam, C.N.C., Aziz, N.A.A., Azaman, F., Latif, M.T., Zainuddin, S.F.M., and Osman, M.R., 2014. Prediction of the level of air pollution using principal component analysis and artificial neural network techniques: A case study in Malaysia. *Water, Air, & Soil Pollution*, 225 (8), Pp. 2063-2077.

Azid, A., Rani, N.A.A., Samsudin, M.S., Khalit, S.I., Gasim, M.B., Kamarudin, M.K.A., Yunus, K., Saudi, A.S.M., and Yusof, K.M.K.K., 2017. Air Quality Modelling Using Chemometric Techniques. *Journal of Fundamental and Applied Sciences*, 9 (2S), Pp. 443-466.

Chau, K.W., 2006. A review on integration of artificial intelligence into water quality modelling, *Marine Pollution Bulletin*, 52, Pp. 726-733.

Chitra, V., Ravichandran, K.S., and Varadarajan, R., 2012. Artificial Neural



- Network in Field Oriented Control for Matrix Converter Drive. *World Appl. Sci. J.*, 16 (4), Pp. 560-567.
- Cho, K.H., Park, Y., Kang, J., Ki, S.J., Cha, S., Lee, S.W., 2009. Interpretation of seasonal water quality variation in the Yeongsan Reservoir, Korea using multivariate statistical analyses. *J. Hydroinform*, 59 (11), Pp. 2219-2226.
- DOE, 2015. Environmental Annual Report. 2015. Kuantan. Department of Environment Pahang, Ministry of Natural Resources and Environment.
- Felipe-Sotelo, M., J.M. Andrade, A. Carlosena, Tauler, R., 2007. Temporal characterisation of river waters in urban and semi-urban areas using physico-chemical parameters and chemometric methods. *Analytica Chimica Acta*, 583, Pp. 128-137.
- Goethals, P.L.M., Dedecker, A.P., Gabriels, W., Lek, S., and De Pauw, N., 2007. Applications of artificial neural networks predicting macroinvertebrates in freshwaters, *Aquatic Ecology*, 41 (3), Pp. 491-508.
- Hair, J.F., Anderson, R.E., Tatham, R.L., William, C., 1995. *Multivariate data analysis with readings*. Prentice Hall, Englewood Cliffs.
- Han, S., Kim, E., and Kim, S., 2009. The water quality management in the Nakdong River watershed using multivariate statistical techniques. *Korean J Civ Eng*, 13 (2), Pp. 97-105.
- Haykin, S., 1999. *Neural Network: A Comprehensive Foundation*. Englewood Cliffs, NJ: Prentice-Hall.
- Helena, B., Pardo, R., Vega, M., Barrado, E., Fernandez, J., and Fernandez, L., 2000. Temporal evolution of groundwater composition in an alluvial aquifer (Pisuerga River, Spain) by principal component analysis. *Water Res.*, 34 (3), Pp. 807-816.
- Hornik, K., Stinchcombe, M., and White, H., 1990. Universal approximation of an unknown mapping and its derivatives using multilayer feed-forward networks. *Neural Netw*; 3 (5), Pp. 551-60.
- Ismail, A., Toriman, M.E., Juahir, H., Zain, S.M., Habir, N.L.A., Retnam, A., Kamaruddin, M.K.A., Umar, R., and Azid, A., 2016. Spatial assessment and source identification of heavy metals pollution in surface water using several chemometric techniques. *Marine Pollution Bulletin*, 106 (1), Pp. 292-300.
- Juahir, H., Md. Zain, S., Jaafar, M.N., and Ahmad, Z., 2004a. An Application of Second order backpropagation method in Modeling River Discharge at Sungai Langat, Malaysia. *Water Environmental Planning: Towards integrated planning and management of water resources for environmental risks. IJUM Journal*. Pp. 300-307.
- Juahir, H., Md. Zain, S., Toriman, M.E., and Mokhtar, M., 2004b. Application of Artificial Neural Network Model In the Predicting Water Quality Index. *Jurnal Kejuruteraan Awam*, 16 (2), Pp. 42-55.
- Juahir, H., Md. Zain, S., Toriman, M.E., Jaafar, M.N., and Klaewtanong, W., 2003a. Performance of autoregressive integrated moving average and neural network approaches for forecasting dissolved oxygen at Langat River Malaysia. *Urban Ecosystem Studies In Malaysia: A study of change*. Universal Publishers, Pp. 145-165.
- Juahir, H., Zain, S.M., Yusoff, M.K., Hanidza, T.I.T., Armi, A.S.M., Toriman, M.E., Mokhtar, M., 2011. Spatial water quality assessment of Langat River Basin (Malaysia) using environmetric techniques, *Environ. Monit. Assess.*, 173, Pp. 625-641, doi: 10.1007/s10661-010-1411-x.
- Keskin, T.E., Dugenci, M., and Kacaroglu, F., 2014. Prediction of water pollution sources using artificial neural networks in the study areas of Sivas, Karabük and Bartın (Turkey). *Environ Earth Sci.*, 73 (9), Pp. 5333-5347. doi: 10.1007/s12665-014-3784-6.
- Khalit, S.I., Samsudin, M.S., Azid, A., Yunus, K., Zaudi, M.A., Sharifuddin, S.S., Husin, T.M., 2017. A preliminary study of marine water quality status using principal component analysis at three selected mangrove estuaries in East Coast Peninsular Malaysia. *Malaysian Journal of Fundamental and Applied Sciences*, 13 (4), Pp. 764-768.
- Khan, R.A., Zain, S.M., Juahir, H., Yusoff, M.K., and Tg Hanidza, T.I., 2012. Using Principal Component Scores and Artificial Neural Networks in Predicting Water Quality Index; *Chemometrics in Practical Applications*.
- Kisi, Ö., 2008. Constructing neural network sediment estimation models using a data-driven algorithm. *Math Comput Simul.*, 79 (1), Pp. 94-103.
- Kiş, Ö., 2009. Daily pan evaporation modelling using multi-layer perceptrons and radial basis neural networks. *Hydrol Process.*, 23 (2), Pp. 213-23.
- Kuo, J.T., Wang, Y.Y., and Lung, W.S., 2006. A hybrid neural-genetic algorithm for reservoir water quality management. *World Appl. Sci. J.*, 10 (12), Pp. 1493-1500. 30.
- Liu, C.W., Lin, K.H., and Kuo, Y.M., 2003. Application of factor analysis in the assessment of groundwater area in Taiwan. *The Science of the Total Environ.*, 313, Pp. 77-89.
- Low, K.H., Koki, I.B., Juahir, H., Azid, A., Behkami, S., Ikram, R., Mohammed, H.A., and Zain, S.M., 2016. Evaluation of water quality variation in lakes, rivers, and ex-mining ponds in Malaysia. *Desalination and Water Treatment*, 57 (58), Pp. 28215-28239.
- Maier, H.R., and Dandy, G.C., 1998. The effect of internal network parameters and geometry on the performance of back-propagation neural networks: an empirical study. *Environmental Modelling and Software*, 13, Pp. 193-209.
- Medici, L., Bellanova, J., Belviso, C., Cavalcante, F., Lettino, A., Ragone, P.P., Fiore, S., 2011. Trace metals speciation in sediments of the Basento River (Italy). *Appl. Clay. Sci.*, 53, Pp. 414-442.
- Mutalib, S.N.S.A., Juahir, H., Azid, A., Sharif, S.M., Latif, M.T., Aris, A.Z., Zain, S.M., and Dominick, D., 2013. Spatial and temporal air quality pattern recognition using environmetric techniques: a case study in Malaysia. *Environmental Science: Processes & Impacts*, 15 (9), Pp. 1717-1728.
- Najah, A.A., Elshafie, O.A., Karim and Jaffar, O., 2009. Prediction of Johor River Water Quality Parameters Using Artificial Neural Networks. *European Journal of Scientific Research*, 28 (3), Pp. 422-435.
- Nasri, M., 2010. Application of artificial neural networks (ANNs) in prediction models in risk management, *Water Research*, 40, Pp. 1367-1376.
- Panigrahi, S., Acharya, B.C., Panigrahy, R.C., Nayak, B.K., Banarjee, K., Sarkar S.K., 2007. Anthropogenic impact on water quality of Chilika lagoon RAMSAR site: a statistical approach. *Wetlands Ecol Manage.*, 15 (2), Pp. 113-126.
- Rani, N.A., Azid, A., Khalit, S.I., and Juahir, H., 2018. Prediction Model of Missing Data: A Case Study of Pm10 Across Malaysia Region. *Journal of Fundamental and Applied Sciences*, 10 (1S), Pp. 182-203.
- Rashid, M.H., and Mridha, A.K., 1998. Arsenic contamination in groundwater in Bangladesh. In: *Sanitation and Water for All*, 24th WEDC Conference, Islamabad, Pakistan, Pp. 162-165.
- Rossi, L., Chevre, N., Fankhauser, R., Margot, J., Curdy, R., Babut, M., Barry, D.A., 2013. Sediments contamination assessment in urban areas based on total suspended solids. *Water. Res.*, 47, Pp. 339-350.
- Samsudin, M.S., Azid, A., Khalit, S.I., Saudi, A.S.M., and Zaudi, M.A., 2017a. River water quality assessment using APCS-MLR and statistical process control in Johor River Basin, Malaysia. *International Journal of Advanced and Applied Sciences*, 4 (8), Pp. 84-97.
- Samsudin, M.S., Khalit, S.I., Azid, A., Juahir, H., Saudi, A.S.M., Sharip, Z., and Zaudi, M.A., 2017b. Control limit detection for source apportionment in Perlis River Basin, Malaysia. *Malaysian Journal of Fundamental and Applied Sciences*, 13 (3), Pp. 294-303.
- Samsudin, M.S., Khalit, S.I., Azid, A., Yunus, K., Zaudi, M.A., Badaluddin, N.A., and Saudi, A.S.M., 2017c. Spatial Analysis of Heavy Metals in Mangrove Estuary at East Coast Peninsular Malaysia: A Preliminary Study. *Journal of Fundamental and Applied Sciences*, 9 (2S), Pp. 680-697.
- Sarbu, C., and Pop, H.F., 2005. Principal component analysis versus fuzzy principal component analysis a case study: the quality of Danube water (1985-1996) *Talanta*, 65, Pp. 1215-1220.

- Shaharudin, S.M., Ahmad, N., and Yusof, F., 2013. Improved Cluster Partition in Principal Component Analysis Guided Clustering. *International Journal of Computer Applications*, 75 (11), Pp. 22-25.
- Shaharudin, S.M., Ahmad, N., Zainuddin, N.H., and Mohamed, N.S., 2018. Identification of Rainfall Patterns on Hydrological Simulation Using Robust Principal Component Analysis. *Indonesian Journal of Electrical Engineering and Computer Science*, 11 (3), Pp. 1162-1167.
- Singh, K.P., Malik, A., and Sinha, S., 2005. Water quality assessment and apportionment of pollution sources of Gomti river (India) using multivariate statistical techniques—A case study. *Analytica Chimica Acta*, 538 (1-2), Pp. 355-374.
- Sulaiman, N.H., Khalit, S.I., Sharip, Z., Samsudin, M.S., and Azid, A., 2018. Seasonal variations of water quality and heavy metals in two examining lakes using chemometric assessment approach. *Malaysian Journal of Fundamental and Applied Sciences*, 14 (1), Pp. 67-72.
- Suthar S., Nema, A.K., Chabukdhara, M., Gupta, S.K., 2009. Assessment of metals in water and sediments of Hindon River, India: impact of industrial and urban discharges. *J. Hazard. Mater.*, 171 (1-3), Pp. 1088-1095.
- Tanriverdi, Ç., Alp, A., Demirkıran, A.R., and Üçkardes, F., 2010. Assessment of surface water quality of the Ceyhan River Basin, Turkey. *Environ Monit Assess*, 167 (1), Pp. 175-184.
- Toriman, M.E., Abdullah, M.P., Mokhtar, M.B., Gasim, M.B., Karim, O., 2010. Penilaian hakisan permukaan dan muatan sedimen dari kawasan tadahan sungai anak bangi selangor. *Malaysian Journal of Analytical Sciences*, 14 (1), Pp. 12-23.
- Toriman, O.E., Hashim, N., Hassan, A.J., Mokhtar, M., Juahir, H., Gasim, M.B., and Abdullah, M.P., 2011. Study on the impact of tidal effects on water quality modelling of Juru River, Malaysia. *Asian Journal of Scientific Research*, 4 (2), Pp. 129-138. doi: 10.3923/ajsr.2011.129.138.
- Uca, Toriman, E., Jaafar, O., Maru, R., Arfan, A., and Ahmar, A.S., 2018. Daily Suspended Sediment Discharge Prediction Using Multiple Linear Regression and Artificial Neural Network. *Journal of Physics: Conference Series*, 954 (1), [012030]. DOI: 10.1088/1742-6596/954/1/012030.
- Yidana, S.M., Ophori, D., and Banoeng-Yakubo, B., 2008. A multivariate statistical analysis of surface water chemistry data – the Ankobra Basin, Ghana, *J. Environ. Manager*. 86, Pp. 80-87.

