Eight principal components were found responsible for the data structure and 67.7% of the total variance of the data set in clusters were formed. DA managed to discriminate 16 and 19 water quality variables thru forward and backward stepwise significant variables that influence Muar River as well as sources of apportionment pollution. HACA observed three spatial sources of pollution. ANNs model was used to predict linear relationship between water quality variables, the most (PCA) and the factor analysis (FA) to study the spatial variations of water quality variables and to determine the origin hierarchical agglomerative cluster analysis (HACA), the discriminant analysis (DA), the principal component analysis using the Environmetric Techniques and the Artificial Neural Networks (ANNs) model. Environmetric method, the ABSTRACT

VOL. 11, NO. 11, JUNE 2016 ISSN 1819-6608

WATER QUALITY ASSESSMENT OF MUAR RIVER USING ENVIRONMETRIC TECHNIQUES AND ARTIFICIAL NEURAL NETWORKS

Putri Shazlia Rosman
Faculty of Engineering Technology, Universiti Malaysia Pahang, Tun Razak Highway, Kuantan Pahang, Malaysia
E-Mail: putrishazlia@ump.edu.my

ABSTRACT

The pollution discharge has influence the chemical composition of Muar River where studied was carried out using the Environmetric Techniques and the Artificial Neural Networks (ANNs) model. Environmetric method, the hierarchical agglomerative cluster analysis (HACA), the discriminant analysis (DA), the principal component analysis (PCA) and the factor analysis (FA) to study the spatial variations of water quality variables and to determine the origin sources of pollution. ANNs model was used to predict linear relationship between water quality variables, the most significant variables that influence Muar River as well as sources of apportionment pollution. HACA observed three spatial clusters were formed. DA managed to discriminate 16 and 19 water quality variables thru forward and backward stepwise. Eight principal components were found responsible for the data structure and 67.7% of the total variance of the data set in PCA/FA analysis. ANNs analysis, strong relationship correlation was observed between salinity, conductivity, DS, TS, Cl, Ca, K, Mg and Na (r = 0.954 to 0.997), moderate relationship observed between COD and E.coli (r = 0.449) and Cd and Pb (r = 0.492) and others variables have no significant correlation, pH was the most significant variables (51.6%) and Fe was less significant variables (-0.52%). The major sources of pollution of the river were due to natural degradation / natural process that affecting the pH value of the river. Other pollution contribution was from anthropogenic sources such as agricultural runoff, industrial discharge, domestic waste, natural erosion, livestock farming and present of nitrogenous species. The ANNs showed better prediction in identified most significant variable compare to Environmetric techniques. ANNs is an effective tool in decision making and problem solving for local/global environmental issues.

Keywords: environmetric techniques, artificial neural networks, hierarchical agglomerative cluster analysis, discriminant analysis, principal component analysis.

1. INTRODUCTION

Rivers has become an important to people in daily use for drinking, bathing, transportation, recreational activities, agriculture purpose etc. and to support life of many endangered species animals and aquatic. Muar River has riches with natural river ecosystem where the local community people can gain their income by providing boat ride for the tourist to see the beauty of the riverside of Muar River.

According to Malaysia Environmental Quality Report (2009), about 521 rivers were categorized as polluted rivers. From polluted rivers, about 152 rivers were polluted by high BOD level, 183 rivers were polluted by high ammonia cal Nitrogen concentration and 186 rivers were polluted by Suspended solid concentration. The present status of Muar River Basin for overall Water Quality Index (WQI) is slightly polluted that is 81 based on the Water Quality Monitoring Program that has been carried out by Department of Environment with 17 stations in year 2002 until 2008 (DOE, 2009).

The problem that occurs in Muar River is from the agriculture activities. The agricultural land uses remain dominant feature of the entire basin where the water quality in the Muar River basin has become considerably poor and polluted due to diffuse pollution originating from agriculture activities. Other than that water pollution in Muar River derived from numerous sources of pollution that may adversely affect water quality. These major sources are from agriculture wastewater and minor sources pollutions derived from domestic and industrial wastewater.

However, Muar River Basin was categories as slight polluted rivers in Malaysia Environmental Quality Report 2009 by Department of Environment. About 11 rivers stated in Muar River basins were under class II and class III. The rivers are Gemas River, Gemencheh River, Labis River, Merbodu River, Merlimau River, Muar River, Palong River, Senarut River, Serom River, Simpang Loi River and Tenang River.

1.1. General description of environmetric technique and ANN model

Environmetric is also known as chemometrics that uses multivariate statistical modeling and data treatment (Simeonov et al. 2000). Environmetric methods is used in exploratory data analysis tools for classification (Brodnjak-Voncina et al.2002; Kowalkowski et al. 2006) of sampling stations or samples identification of pollution sources (Massart et al.1997, Vega et al. 1998; Shresthaa and Kazama 2007) and drawing information from environmental data. Environmetric method has become important tools to evaluate and reveal complex relationships in environmental science area through variety of environmetric technique application. Application of environmetric techniques, viz, hierarchical agglomerative cluster analysis (HACA), the discriminant
analysis (DA), the principal component analysis (PCA) / factor analysis (FA) were employed to identify spatial variations of most significant water quality variables and to determine the origin of pollution sources at the Muar River Basin. Discriminant analysis is operates on raw data by using standard, backward and forward stepwise modes in order to construct a discriminant function (DF). Discriminant function for each group was build up from DA technique. DFs are calculated using equation:

$$f(G_i) = k + \sum_{j=1}^{n} w_j x_j$$

(1)

The principle component (PC) can be expressed as

$$z_i = \alpha_{i1}x_1 + \alpha_{i2}x_2 + ... + \alpha_{in}x_n$$

(2)

The basic concept of FA is expressed as

$$z_i = \alpha_{i1}x_1 + \alpha_{i2}x_2 + ... + \alpha_{in}x_n + \epsilon_i$$

(3)

ANNs has become common use in many researchers study. ANNs are frequently to predict various ecological processes and phenomenon related to water resources (Talib et al., 2009). The basic structure of ANN model is usually comprised of three distinctive layers, the input layer, where the data are introduce to the model and computation of the weighted sum of the input is preformed, the hidden layer or layers, where data are processed, and the output layer, where the results of ANN are produced (Singh et al., 2009).

1.2. Study area and data acquisition

The Muar River is the major water way in State of Johor. The total of catchment area approximately 6,149 km² which consists of four states, State of Johor with 3,641 km², State of Negeri Sembilan with 2,427 km², state of Pahang with 122 km² and state of Melaka with 8 km². The length of main trunk of Muar River is about 288.20 km with the width is estimated to be 1.6 km. The Muar River has two major tributaries with the principal ones being the Gemencheh River and Segamat River shown in Figure-5. In 2003, about 85% of the total land area within the basin is devoted to agricultural use and about 15% of total land areas are devoted to forests conservation.

The water stations (WS06, WS07, WS08 and WS11) located at Muar River in Negeri Sembilan, while water stations (WS09, WS12, WS16, WS21, WS23, WS24, WS25, WS26, WS27, WS29 and WS32) located at Muar River in Johor.

<table>
<thead>
<tr>
<th>Study code in map</th>
<th>Study Code</th>
<th>Grid Reference</th>
<th>Location</th>
</tr>
</thead>
<tbody>
<tr>
<td>S1</td>
<td>WS06</td>
<td>02 28 102 33 480</td>
<td>Muar River, Negeri Sembilan</td>
</tr>
<tr>
<td>S2</td>
<td>WS07</td>
<td>02 28 102 33 360</td>
<td>Muar River, Negeri Sembilan</td>
</tr>
<tr>
<td>S3</td>
<td>WS08</td>
<td>02 48 102 00 134</td>
<td>Muar River, Negeri Sembilan</td>
</tr>
<tr>
<td>S4</td>
<td>WS11</td>
<td>02 28 102 16 004</td>
<td>Muar River, Negeri Sembilan</td>
</tr>
<tr>
<td>S5</td>
<td>WS09</td>
<td>02 28 102 45 162</td>
<td>Muar River, Johor</td>
</tr>
<tr>
<td>S6</td>
<td>WS12</td>
<td>02 28 102 45 956</td>
<td>Muar River, Johor</td>
</tr>
<tr>
<td>S7</td>
<td>WS16</td>
<td>02 28 102 44 085</td>
<td>Muar River, Johor</td>
</tr>
<tr>
<td>S8</td>
<td>WS21</td>
<td>02 28 102 48 714</td>
<td>Muar River, Johor</td>
</tr>
<tr>
<td>S9</td>
<td>WS23</td>
<td>02 28 102 42 591</td>
<td>Muar River, Johor</td>
</tr>
<tr>
<td>S10</td>
<td>WS24</td>
<td>02 28 102 42 591</td>
<td>Muar River, Johor</td>
</tr>
<tr>
<td>S11</td>
<td>WS25</td>
<td>02 28 102 42 591</td>
<td>Muar River, Johor</td>
</tr>
<tr>
<td>S12</td>
<td>WS26</td>
<td>02 28 102 42 591</td>
<td>Muar River, Johor</td>
</tr>
<tr>
<td>S13</td>
<td>WS27</td>
<td>02 28 102 42 591</td>
<td>Muar River, Johor</td>
</tr>
<tr>
<td>S14</td>
<td>WS29</td>
<td>02 28 102 42 591</td>
<td>Muar River, Johor</td>
</tr>
<tr>
<td>S15</td>
<td>WS32</td>
<td>02 28 102 42 591</td>
<td>Muar River, Johor</td>
</tr>
</tbody>
</table>
The 30 water quality variables involves are Dissolve oxygen (DO), Biochemical Oxygen Demand (BOD), Chemical Oxygen Demand (COD), Suspended Solid (SS), Ammoniacal- Nitrogen (NH$_3$-N), Dissolve Solid (DS), nitrate (NO$_3$), chlorine (Cl), total solid (TS), zinc (Zn), potassium (K), magnesium (Mg), phosphate (PO$_4$), temperature (T), pH, calcium (Ca), sodium (Na), iron (Fe), salinity (Sal), conductivity (COND), turbidity (TUR), oil and grease (OG), cadmium (Cd), lead (Pb), copper (Cr), mercury (Hg), arsenic (As), MBA/BAS, Escherichia coli (E.coli), and coliform.

2. RESULT AND DISCUSSION

a) Correlation matrix between variables

By measured simple correlation coefficient (Pearson) matrix ($r$), the degree of a linear association between variables indicates suspended solid and turbidity are strongly and significantly correlated each other ($r = 0.820$). Strongly relationship correlations also can be observed between salinity, conductivity, dissolve solid, total solid, chloride, calcium, potassium, magnesium and sodium ($r = 0.954$ to $0.997$) which responsible for water mineralization. Moderate relationship between COD and $E.coli$ ($r = 0.449$) and between Cd and Pb ($r = 0.492$) were indicates significant anthropogenic contribution as sources of contaminants. While DO, BOD, pH, NH$_3$, temperature, NO$_3$, PO$_4$, As, Cr, Zn, Fe, oil and grease, MBAS and coliform showed no significant correlation with any other variables.

b) HACA: Classification of sampling station on historical water quality data

The result HACA analysis shown in Figure-3 where sampling stations were places in three cluster. The similarities of natural background and characteristics of these stations causes the clustering procedure generated 3 cluster/group in a very convincing ways. Cluster 1 (stations WS06, WS07, WS08 and WS11) correspond to regions of low pollution sources (LPS), cluster 2 (stations WS16, WS21) represents the moderate pollution sources (MPS) and cluster 3 (WS09, WS12, WS23, WS24, WS25, WS26, WS27, WS29 and WS32) represents the high pollution sources (HPS) from sub basin Muar River.

c) Spatial variations of river water quality

Water quality parameters were treated as independent variables while the groups (C1, C2 and C3) were treated as dependent variables. The DA method carried out was via standard, backward stepwise and forward stepwise. The accuracy of spatial classification using standard, backward stepwise and forward stepwise mode DFA were 80.67% (29 discriminant variables), 80.41% (19 discriminant variables) and 80.41% (16 discriminant variables), respectively (Table-2).

d) Data structure determination

Eight PCs were obtained in this study with eigenvalues larger than 1 where almost 69.7% of the total variance in the data set. Among the eight VFs coefficient that having strong correlation were VF1 accounts for 30.3% of the total variance, shows strong positive loadings on conductivity, salinity, dissolve solid, total solid, chloride and mineral salt such as calcium, potassium, magnesium, and sodium.

e) Artificial Neural Network in prediction of water quality samples

30 parameters water quality from 776 stations were trained, tested and validated by using ANN on spatial distribution.
Table 2. Contribution percentage sources identification from ANNs analysis.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>$R^2$</th>
<th>Different (R2AP) $R^2$</th>
<th>Contribution %</th>
</tr>
</thead>
<tbody>
<tr>
<td>All parameter</td>
<td>0.901848</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Conductivity, Salinity, TS, Cl, Ca, K, Mg, Na</td>
<td>D1</td>
<td>0.890989</td>
<td>0.001975</td>
</tr>
<tr>
<td>pH</td>
<td>D2</td>
<td>0.895888</td>
<td>0.003260</td>
</tr>
<tr>
<td>SS, TUR</td>
<td>D3</td>
<td>0.828842</td>
<td>0.042006</td>
</tr>
<tr>
<td>pH</td>
<td>D4</td>
<td>0.757977</td>
<td>0.143871</td>
</tr>
<tr>
<td>COD, Esch</td>
<td>D5</td>
<td>0.890012</td>
<td>0.011836</td>
</tr>
<tr>
<td>COD, NH3</td>
<td>D6</td>
<td>0.849334</td>
<td>0.053514</td>
</tr>
<tr>
<td>NO3</td>
<td>D7</td>
<td>0.827378</td>
<td>0.018110</td>
</tr>
<tr>
<td>Fe</td>
<td>D8</td>
<td>0.903298</td>
<td>0.001450</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td>0.279104</td>
<td></td>
</tr>
</tbody>
</table>

The result showed pH with 51.55% is the major contribution parameter into Muar River followed by 19.17% of COD and Ammoniacal- Nitrogen (NH3-N). About 15.41% of Suspended Solid and Turbidity, 6.49% of nitrate, 4.24% of COD and Escherichia coli parameter such as conductivity, salinity, dissolved solids, total solid, chloride, calcium, potassium, magnesium and sodium are contribute about 2.49% in the river. The cadmium parameter contributes about 1.17% into the river. However, we found that iron (Fe) contributes about -0.52% meaning the present amount of ion in the river was much lower from all parameter above.

3. CONCLUSIONS

The HACA could helps to reduce the number of sampling station, associated time and cost. DA gives encouraging results for spatial variations in discriminating the fifteen monitoring station with nineteen and sixteen discriminant variables assigning 80% cases correctly using backward stepwise and forward stepwise modes. The PCA/FA results in eight variables responsible for major variation in along Muar River. The main sources of variations come from agriculture farming, livestock waste, natural erosion, excessive of fertilizer, domestic/ commercial waste and the present of nitrogenous species in the river. However, ANNs models were developed to predicting the high and low contributions of variables to the river. pH was identified the most contribution variables into Muar River with the highest percentage among other variables. However, pH has no significant relationship to other variables as result of analysis ANNs. High pH can be determine due to high amount of nitrogenous species in the river where these species. Compare to Environmetric techniques, ANNs model gives better result interpretation of the variables.

ACKNOWLEDGEMENT

The author sincerely thank to The Department of Environment Malaysia for providing database of water quality Muar River and assistance provided by Faculty of Environmental Studies, Universiti Putra Malaysia.

REFERENCES


Networks. 18th World IMACS/MODSIM Congress, Cairns, Australia 13-17 July 2009.


