GIS-based Study on the Susceptibility of Dubai Creek (UAE) to Eutrophication

Tarig A. Ali*, Maruf Mortula, Serter Atabay

Department of Civil Engineering, American University of Sharjah, Sharjah, United Arab Emirates

Received: 29 March 2016
Accepted: 11 June 2016

Abstract

Rapid urbanization in the UAE has led to some environmental implications, one of which concerns Dubai Creek – a major attraction in the city of Dubai. The creek’s water quality monitoring program showed increased concentrations of phosphorus- and nitrogen-based nutrients starting in 2008 and 2009. The creek has since been through redundant eutrophication, which has been attributed to the high levels of nutrients in addition to the creek’s poor flushing and irregular circulation processes. The aim of this study was to (a) assess the susceptibility of the creek to eutrophication considering its principal factors, and (b) to identify the pattern of this process (i.e., seasonal, cyclic, etc.). Principal component analysis was used to identify the principal factors from nitrates, phosphates, total nitrogen, chlorophyll-a, dissolved oxygen, and turbidity, which were collected as quarterly averages in 2012 and 2013. Logistic regression was utilized to assess the susceptibility of the creek to eutrophication considering the principal factors obtained by PCA. The analysis showed that three different factors, which included at least nitrates or phosphates, have contributed to eutrophication in every quarter of the year in the period of study. Further analysis showed weak correlation between principal factors of eutrophication in consecutive quarters. However, strong correlations were observed between these factors when the same quarters over the period of the study were considered, suggesting a possible seasonal pattern.

Keywords: Dubai Creek, nutrients, eutrophication, logistic regression

*e-mail: atarig@aus.edu

DOI: 10.15244/pjoes/63659
developed in the past decade and the vast urbanization has resulted in a significant impact on the environment.

Dubai Creek is a 14-km saltwater body located at the heart of Dubai and is mainly used for trade, transport, and recreation. The major source of the creek’s nutrient pollution is the treated sewage discharge from Al-Aweer Sewage Treatment Plant [17-18]. The Dubai Municipality (DM) monitors the creek’s water quality at 10 stations and the measured water quality parameters include phosphates, nitrates, total nitrogen, chlorophyll-a, turbidity, dissolved oxygen, salinity, and pH. Water quality data recorded in the past few years showed a pattern of increased concentrations of nutrients and chlorophyll-a in the creek to levels that exceeded the seasonal averages set in the Organization for Economic Cooperation and Development (OECD) criteria for assessing eutrophic status and the numeric nutrients standard of the U.S. Environmental Protection Agency’s (EPA) national nutrient strategy [19-20]. This situation has created serious concern in the aftermath of the major algal bloom event of 2008 (aka red-tide), which caused catastrophic fish kills in the creek and devastated the coasts of Dubai and UAE. Therefore, the vulnerability of the creek to eutrophication needs to be assessed. Based on the success of logistic regression (LR) in numerous studies – especially those that were carried out to map the susceptibility of floods [21], landslides [22], and also based on the success in wetland assessment and management using geospatial techniques [23] – this study is unique in the application of LR within the GIS environment to assess the susceptibility of the creek to eutrophication.

The objectives of the study were to identify the principal factors of eutrophication in the creek in every quarter during the period of study, to estimate the probability of occurrence of eutrophication in the creek to produce maps of the susceptibility of the creek to eutrophication, and to identify the processes that are responsible for eutrophication in the creek. LR is a multivariate statistical method that considers all of the independent variables (referred to as conditioning factors in this study) that influence the occurrence of the dependent variable [24]. The principal conditioning factors of eutrophication in the creek were estimated using the Principal Component Analysis (PCA) given the creek’s water quality parameters, which were considered the initial factors before LR was utilized. This was an essential task because LR requires the use of factors that are uncorrelated or in other words have maximum variances.

**Study Area and Data**

Dubai Creek is located in Dubai, United Arab Emirates (UAE) and is spatially identified by a minimum enclosing rectangle that has left top coordinates of 25°16´30.8´´N, 55°17´30´´E and right bottom coordinates of 25°10´58.5´´N, 55°21´6.9´´E. The creek is composed of a narrow channel segment surrounded by urban and vegetation areas on the two banks, while the lagoon part is surrounded by bare lands, a sanctuary, and undeveloped areas. The creek extends for about 14 km and has a surface area of 8 km², of which about 4.6 km² is the lagoon’s area. Major sources of nutrients into the creek include the discharge of the treated sewage from the Al Aweer Sewage Processing Plant and surface runoff from urban and vegetation areas (Fig. 1).

![Fig. 1. The geographic location of Dubai Creek and the water quality monitoring stations in Dubai, UAE.](image-url)
The water quality dataset used in this study was collected at the 10 monitoring stations over a two-year period from 2012 to 2013 and was provided by the Dubai Municipality as monthly averages in the following quarters: January-March (first quarter), April-June (second quarter), July-September (third quarter), and October-December (fourth quarter). The measured water quality parameters include phosphates, nitrates, total nitrogen, chlorophyll-a, turbidity, dissolved oxygen, salinity, and pH. Table 1 shows the water quality dataset of the creek in the first quarter of 2012 for each monitoring station, which is presented herein as a sample of the dataset used in the period of study.

### Methods

Logistic regression (LR) is a multivariate statistical analysis approach that considers all the conditioning factors (independent variables) that may influence the probability of occurrence of the dependent variable, in this case eutrophication [24]. The LR model can help in estimating the probability of occurrence of any event if its independent conditioning factors are known. LR has a function \( f(z) \) of the general simple form of Equation 1 [25]:

\[
f(z) = \frac{1}{1+e^{-z}}
\]  

...where \( f(z) \) is the LR function and \( z \) is the dependent variable such that \( 1 \geq f(z) \geq 0 \). The fact that the logistic function, \( f(z) \), ranges from 0 to 1 is the main reason for LR model uniqueness and popularity as it describes the probability of occurrence of the dependent variable \( z \) as it varies with independent variables.

In the simplest form of the LR model, inclusion of many independent variables makes the model more robust. The main reasons for the selection of the LR method in this study were: (a) the range of the estimates of the LR function \( f(z) \) within 0 and 1, which is useful in defining different levels of susceptibility of the creek to eutrophication; (b) the s-shaped graph of the LR function can help in describing the threshold criteria, which is commonly used in defining the eutrophic status of a water body, and (c) the LR model dependent variable, \( z \), can model the combined effects of several conditioning factors, which describes the problem we are trying to address in this study. That is, the independent variable \( z \) is a linear combination of “k” independent uncorrelated multivariable observations \( x_1, x_2, x_3, \ldots, x_k \). The general equation of the simple LR, which was used in this study, is of the following form:

\[
z = a_0 + a_1x_1 + a_2x_2 + a_3x_3 + \cdots + a_kx_k
\]  

...where \( a_0 \) is the intercept and \( a_1, a_2, a_3, \ldots, a_k \) are LR model coefficients.

The LR model was employed considering the principal conditioning factors of eutrophication in order to estimate the probability of its occurrence in the creek and further define levels of susceptibility to eutrophication. It is important to note that although the LR model is simple, it requires the selection of the “principal” independent variables (i.e., conditioning factors) that are uncorrelated (i.e., have the largest variance).

The general criteria for selecting the independent variables in the LR model require it to have the following characteristics: (a) they have relations with the dependent variable, (b) they have complete representation in the study area, (c) they vary spatially in the study area, (d) they are measurable, and (e) they are not repeated [26-27]. Based on this criterion, the following six water quality parameters collected at the monitoring stations were selected as the initial conditioning factors given their relationships with eutrophication: chlorophyll-a, nitrates, total nitrogen, phosphates, turbidity, and dissolved oxygen. These parameters were originally collected as

<table>
<thead>
<tr>
<th>Stations</th>
<th>Salinity ppt</th>
<th>pH</th>
<th>Chlorophyll-a mg/L</th>
<th>DO Conc mg/L</th>
<th>Turbidity N.T.U</th>
<th>Nitrates mg/L</th>
<th>Total Nitrogen mg/L</th>
<th>Phosphates mg/L</th>
</tr>
</thead>
<tbody>
<tr>
<td>Creek Mouth</td>
<td>38.68</td>
<td>8.05</td>
<td>5.3</td>
<td>6.91</td>
<td>1.5</td>
<td>0.74</td>
<td>2.47</td>
<td>0.11</td>
</tr>
<tr>
<td>Hayat Regency</td>
<td>39.81</td>
<td>8.08</td>
<td>6.8</td>
<td>7.31</td>
<td>1.5</td>
<td>0.47</td>
<td>1.74</td>
<td>0.05</td>
</tr>
<tr>
<td>Abra</td>
<td>38.23</td>
<td>8.04</td>
<td>8.2</td>
<td>6.2</td>
<td>1.5</td>
<td>1.02</td>
<td>2.92</td>
<td>0.14</td>
</tr>
<tr>
<td>Wharfage</td>
<td>37.76</td>
<td>8.14</td>
<td>40.8</td>
<td>7.36</td>
<td>1.5</td>
<td>1.14</td>
<td>3.47</td>
<td>0.16</td>
</tr>
<tr>
<td>Floating Bridge</td>
<td>36.05</td>
<td>8.29</td>
<td>51.2</td>
<td>11.21</td>
<td>1</td>
<td>1.7</td>
<td>5.44</td>
<td>0.16</td>
</tr>
<tr>
<td>Al Garhoud Bridge</td>
<td>36.8</td>
<td>7.96</td>
<td>17.9</td>
<td>7.23</td>
<td>1.5</td>
<td>1.94</td>
<td>6.04</td>
<td>0.19</td>
</tr>
<tr>
<td>Dubai Festival City</td>
<td>34.52</td>
<td>8.05</td>
<td>13.3</td>
<td>7.9</td>
<td>1</td>
<td>3.22</td>
<td>8.2</td>
<td>0.31</td>
</tr>
<tr>
<td>Sewage Treatment Plant (STP) Outfall</td>
<td>36.08</td>
<td>8.08</td>
<td>11.5</td>
<td>7.61</td>
<td>1</td>
<td>8.75</td>
<td>24.7</td>
<td>0.95</td>
</tr>
<tr>
<td>Jaddaf</td>
<td>33.67</td>
<td>8.21</td>
<td>3.5</td>
<td>7.97</td>
<td>1</td>
<td>3.46</td>
<td>11.55</td>
<td>0.28</td>
</tr>
<tr>
<td>Sanctuary</td>
<td>36.54</td>
<td>8.27</td>
<td>6.4</td>
<td>8.88</td>
<td>1</td>
<td>1.72</td>
<td>5.95</td>
<td>0.23</td>
</tr>
</tbody>
</table>
daily averages by Dubai Munciplaity, but were provided as monthly averages and were converted in this study to quarter-based averages in the following quarters: January-March (first quarter), April-June (second quarter), July-September (third quarter), and October-December (fourth quarter).

For high performance of the LR model, it is essential to ensure that the selected conditioning factors are independent and uncorrelated. In order to achieve this requirement and also to satisfy the main characteristics of the conditioning factors presented above, we used principal component analysis (PCA). This can help to avoid multicollinearity, which may lead to non-identifiable model coefficients (Härdle and Simar, 2012). In order to calculate the principal components, consider the “k” independent observations presented in equation (2) above and define covariance between xi and xj as shown below for i, j = 1, 2, 3, ..., k [28-29]:

$$ Cov(x_i, x_j) = \sigma^2 R_{i,j} $$

Let \( \mu_1 > \mu_2 > \mu_3 > \ldots > \mu_k > 0 \) be the eigenvalues of R and let \( q_1, q_2, q_3, \ldots, q_k \) be the corresponding eigenvectors, which are normalized as shown below:

$$ q_j^T q_j = 1 $$

The first principal component (PC_1) is a linear combination of the multivariable observations; \( x \)'s shown in equation (2) that have the largest variance:

$$ PC_1 = \sum_{i=1}^{k} q_{1i} x_i $$

The second principal component (PC_2) has the largest variance among all linear combinations of the multivariable observations, \( x \)'s that are orthogonal to PC_1.

$$ PC_2 = \sum_{i=1}^{k} q_{2i} x_i $$

The remaining principal components (third, fourth, fifth, and six, in this case) are computed using the same procedure shown above for obtaining PC_1 and PC_2.

**Results and Discussion**

The principal component analysis (PCA) was carried out in ArcGIS 10.1 for the initial six conditioning factors in every quarter in the period of study. Sample PCA results, which include the correlation matrix of the initial conditioning factors as well as the principal components, are shown in Tables 2 and 3, respectively, for the first quarter of 2012. In this quarter, the correlations between nitrates, phosphates, and total nitrogen are high (above 0.96). The three principal components with the highest variance values in the first quarter of 2012 were PC_4, PC_5, and PC_6. These are basically the conditioning factors that were later adopted in the LR model as independent variables.

In the data of the first quarter of 2012, the most significant component was PC_4, which illustrated a high positive contribution from nitrates and a negative contribution from total nitrogen. One reason for the negative contribution from total nitrogen could be its recorded high concentration at the STP Outfall station the second station from the bottom of the creek. The second significant principal component was PC_5, which illustrated high positive contribution from turbidity and a negative contribution from phosphates. The record high concentration of phosphates at the STP Outfall station could be the reason for the negative contribution from phosphates. Nitrates and dissolved oxygen have approximately the same level of contribution to this principal component. The third significant principal component was PC_6, which illustrated high positive contribution from total nitrogen and a positive contribution from nitrates. The fourth significant principal component was PC_5, which illustrated high positive contribution from phosphates. The principal components PC_4, PC_5, and PC_6 – which have respective high contributions from nitrates, turbidity, and phosphates – were selected as the principal conditioning factors of eutrophication in this quarter.

The LR model was then run in SPSS V.19 software to establish the relationship between eutrophication occurrence and its conditioning factors in this quarter, and accordingly the LR model equation (2) can be re-written with the coefficients as:

$$ z = -0.545 + 0.212 (PC_4) + 0.084 (PC_5) + 0.187 (PC_6) $$

The probability of occurrence of eutrophication in the creek in this quarter was estimated by running equations (7) and then equation (1) in ArcGIS 10.1 and is shown in Fig. 2. The map of the susceptibility of the creek to eutrophication was then created from the probability of occurrence map using the quantile classification method and is shown in Fig. 3.

The maps of probability of occurrence of eutrophication and susceptibility of the creek to eutrophication in the second, third, and fourth quarters of 2012 are shown in Figs 4 and 5, respectively.

In the second quarter of 2012, high correlations were observed between the following factors: chlorophyll-a and nitrates, chlorophyll-a and phosphates, nitrates and...
Table 2. The correlation matrix for the six initial conditioning factors in the first quarter of 2012.

<table>
<thead>
<tr>
<th>Layers</th>
<th>Chl-a</th>
<th>Nitrates</th>
<th>Phosphates</th>
<th>Total nitrogen</th>
<th>Dissolved Oxygen</th>
<th>Turbidity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chl-a</td>
<td></td>
<td>-0.30726</td>
<td>-0.4028</td>
<td>-0.33871</td>
<td>0.38275</td>
<td>0.42996</td>
</tr>
<tr>
<td>Nitrates</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Phosphates</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total nitrogen</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dissolved Oxygen</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Turbidity</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 3. Principal components (PC) for the six initial conditioning factors in the first quarter of 2012.

<table>
<thead>
<tr>
<th>Layers</th>
<th>PC1</th>
<th>PC2</th>
<th>PC3</th>
<th>PC4</th>
<th>PC5</th>
<th>PC6</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chl-a</td>
<td>0.79666</td>
<td>0.16361</td>
<td>-0.03003</td>
<td>-0.00621</td>
<td>-0.01023</td>
<td>-0.00011</td>
</tr>
<tr>
<td>Nitrates</td>
<td>-0.05097</td>
<td>0.33285</td>
<td>-0.11649</td>
<td>0.98461</td>
<td>0.2045</td>
<td>-0.11362</td>
</tr>
<tr>
<td>Phosphates</td>
<td>-0.00601</td>
<td>0.02902</td>
<td>-0.00196</td>
<td>0.15843</td>
<td>-0.2156</td>
<td>0.99309</td>
</tr>
<tr>
<td>Total nitrogen</td>
<td>-0.15543</td>
<td>0.92813</td>
<td>0.04003</td>
<td>-0.33127</td>
<td>-0.05363</td>
<td>0.01363</td>
</tr>
<tr>
<td>Dissolved Oxygen</td>
<td>0.03133</td>
<td>0.005</td>
<td>0.98186</td>
<td>0.0929</td>
<td>0.16062</td>
<td>0.02272</td>
</tr>
<tr>
<td>Turbidity</td>
<td>0.00622</td>
<td>-0.01188</td>
<td>-0.14097</td>
<td>-0.19539</td>
<td>0.99964</td>
<td>0.2426</td>
</tr>
</tbody>
</table>

phosphates, dissolved oxygen, and total nitrogen, and dissolved oxygen and turbidity. The high correlations between dissolved oxygen on the one hand and total nitrogen and turbidity on the other were notable. The principal components in this quarter were PC2, PC3, and PC5, with major contributions from phosphates, chlorophyll-a, and total nitrogen, respectively. In the third quarter of 2012, no high correlations were observed between the initial factors. The principal components in this quarter were PC3, PC4, and PC5, with major contributions from chlorophyll-a, phosphates, and nitrates, respectively. In the fourth quarter of 2012,
high correlations were observed between phosphates, total nitrogen, and nitrates. Also observed was a rather medium to high correlation between dissolved oxygen and turbidity. The principal components in this quarter were PC$_2$, PC$_4$, and PC$_5$, with major contributions from total nitrogen, turbidity, and nitrates, respectively. Following the LR approach used with the first quarter of 2012 data, the maps of the probability of occurrence of eutrophication in the three quarters of 2012 were created. Accordingly, maps of the susceptibility of the creek to eutrophication were produced. The results suggest that susceptibility to eutrophication in every quarter in this period was directly influenced by three conditioning factors that included at least nitrates or phosphates.

In the first quarter of 2013, high correlations were observed between chlorophyll-a and dissolved oxygen, chlorophyll-a and turbidity, and nitrates and total nitrogen. In the second quarter of 2013, high correlations were observed between phosphates and nitrates and then phosphates and total nitrogen. In the third quarter of 2013, high correlations were observed between dissolved oxygen and turbidity, phosphates and nitrates,
and phosphates and total nitrogen. In the fourth quarter of 2013, high correlations were observed between chlorophyll-a and dissolved oxygen, chlorophyll-a and turbidity, and dissolved oxygen and turbidity. Following the same LR approach used with 2012 data, the maps of the probability of occurrence of eutrophication in the creek in the four quarters of 2013 were created, followed by maps of the susceptibility of the creek to eutrophication. Based on the identified PCs in all four quarters of 2013 and their respective major contributing factors, at least one of nitrates or phosphates was identified as a major contributing factor to the respective principal components in all quarters of 2013.

Comparing the PCA results for 2012 and 2013, it was evident that two of the major contributing factors to the principal components in every quarter of 2012 were the same in the respective quarter in 2013. For example, nitrates and phosphates were two of the major contributing factors to the PCs in quarter 1 in both 2012 and 2013. This can indicate a seasonal pattern that characterizes eutrophication and its major factors in the creek in the period of study.

Further investigation was carried out using linear regression analysis to study the relationships between the probability of occurrence of eutrophication in any quarter (as estimated with LR) in the period of study and the previous quarter’s concentrations of nitrates, phosphates, and chlorophyll-a, considered individually and jointly [30-31]. The selection of these factors in this analysis was based on the following facts: (a) at least one of the nitrates or phosphates was one of the major contributors to the principal components in all eight quarters of 2012 and 2013, and (b) chlorophyll-a is the common parameter used by many threshold-based guidelines for assessing eutrophication in water bodies. Table 4 below lists the resulted R-square values obtained through regression analysis for the relationships between the probability of occurrence of eutrophication in a quarter and the selected factor combinations in the previous quarter. Low values of R-square were obtained for the relationship between the occurrence of eutrophication in a quarter and the levels of each of the three factors in the preceding quarter when considered individually. This is an indication that the use of a single parameter to assess the creek’s eutrophication status can be misleading. However, high correlation coefficients were obtained for this relationship when the combination of all three factors was considered (column 8 in Table 4). This outcome indicated the fact that the assessment of the eutrophication status in the creek would be much more robust when a combination of more than one parameter is used.

### Table 4. R-square values of regression analysis of the relationships between the occurrence of eutrophication in a quarter and the values of the selected factor combinations in the preceding quarter.

<table>
<thead>
<tr>
<th>Quarters in consideration</th>
<th>R-square values (%) of relationships of the occurrence of eutrophication between the consecutive quarters shown in column 1</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>NT¹</td>
</tr>
<tr>
<td>Q1 &amp; Q2-2012</td>
<td>45</td>
</tr>
<tr>
<td>Q2 &amp; Q3-2012</td>
<td>33</td>
</tr>
<tr>
<td>Q3 &amp; Q4-2012</td>
<td>28</td>
</tr>
<tr>
<td>Q4-12¹ &amp; Q1-13³¹</td>
<td>29</td>
</tr>
<tr>
<td>Q1 &amp; Q2-2013</td>
<td>34</td>
</tr>
<tr>
<td>Q2 &amp; Q3-2013</td>
<td>59</td>
</tr>
<tr>
<td>Q3 &amp; Q4-2013</td>
<td>45</td>
</tr>
</tbody>
</table>

¹ Nitrate, ² Phosphate, ³ Chlorophyll-a, 4 2012, 5 2013

**Conclusion**

PCA was used to identify the main conditioning factors of eutrophication in Dubai Creek. LR modeling was used to study the susceptibility of the creek to eutrophication based on the principal conditioning factors identified by PCA. Accordingly, maps of probability of occurrence of eutrophication in the creek as well as the levels of susceptibility of the creek to eutrophication were created. Analysis showed that susceptibility of the creek to eutrophication in any quarter of the year during the period of study was influenced by three conditioning factors, which included nitrates or phosphates. Therefore, the use of threshold criteria based on the value of one or two parameters to assess the “eutrophic” status in the creek can be misleading. The analysis showed weak correlations between principal factors of eutrophication in consecutive quarters of the year during the period of study. Further analysis showed stronger correlations between these factors when same quarters over the two-year period of study were considered, which suggests a possible seasonal pattern. Furthermore, weak correlations were obtained between the occurrence of eutrophication in any quarter and the levels of nitrates, phosphates, and chlorophyll-a in the creek’s water in the following quarter (when considered individually). Stronger correlations were obtained between the occurrence of eutrophication in any quarter and the levels of these factors in the following quarter (when considered jointly).
Acknowledgements

The authors acknowledge Dubai Municipality for supporting this study by providing the water quality monitoring data. The authors also acknowledge support from the American University of Sharjah Office of Research and Graduate Studies through Faculty Research Grants (FRG14-2-16) and (FRG12-2-24).

References

6. HAVENS K., FUKUSHIMA T., XIE P., IWAKUMA T., JAMES R., TAKAMURAN T., HANAZATO T., YAMAMOTO T. Nutrient dynamics and the eutrophication of shallow lakes Kasumigaura (Japan), Donghu (PR China), and Okeechobee (USA). Environmental Pollution, 111 (2), 263, 2001.
21. TEHRANY M., PRADHAN B., JEBUR M. Spatial prediction of flood susceptible areas using rule based decision tree (DT) and a novel ensemble bivariate and multivariate statistical models in GIS. J Hydrol, 504, 69, 2013.