Principal Component Analysis and Multivariate Index for Assessment of Eutrophication in Southern Part of Caspian Sea

1,2 Ali Ganjian Khenari, 2,3 W.O. Wan Maznah, 2,3 Khairun Yahya, 1 Shaban Najafpour, 4 Ghasem D. Najafpour, 1 Hasan Fazli and 4 Aboulghasem Roohi

1 Ecological Academy of Caspian Sea, Sari, Iran
2 School of Biological Sciences, Universiti Sains Malaysia, 11800 Penang, Malaysia
3 Centre for Marine and Coastal Studies (CEMACS), Universiti Sains Malaysia, 11800 Penang, Malaysia
4 Faculty of Chemical Engineering, Noshirvani University of Technology, Babol, Iran

Abstract: In this study, the principal component analysis (PCA) and multivariate index were used for the quality measurement of variables such as nitrate, nitrite, ammonia and phosphate and phytoplankton biomass from coastal waters of southern part of Caspian Sea (SCS). Totally 4274-4556 water samples were collected in SCS for duration of 1994 to 2007. On these water samples, both nutrient and biological analysis were performed. The first principal component was derived and evaluated as an eutrophication index (EI) on an independent dataset. The EI presented an increasing trend from oligotrophy to eutrophication, its lower value was observed for the typical oligotrophic (O) in 1994 and the highest in 2005-06 (1.22). Long term EI ranged from 0.26 to 1.22 in 1994/1996 and 1996/1997 were grouped in the typical oligotrophic (O) 1995/1996 and 1999/2000 were grouped as the standard mesotrophic (M) 2003/2004, 2004/2005, 2005/2006 and 2006/2007 were grouped with the typical eutrophic (E). The index of a linear combination of the five variables with almost equal weights was found efficient in discriminating levels of eutrophication and critical thresholds characterizing oligotrophy, mesotrophy and eutrophication were set.

Key words: Caspian Sea eutrophication · Mesotrophy · Nutrients · Oligotrophy · Phytoplankton

INTRODUCTION

Eutrophication is the enrichment of water by nutrients, especially nitrogen and/or phosphorus, causing an accelerated growth of algae and higher forms of plant life to produce an undesirable disturbance to the balance of organisms present in the water and to the quality of water concerned [1]. In other words, it is used to describe the inputs and effects of nutrients into aquatic systems. There are two types of eutrophication, a) natural and b) anthropogenic. Over geologic timescales coastal systems throughout the world have at one time or another experienced some eutrophication [2]. Eutrophication is true, especially along the coast where biodiversity is high and small changes in the primary producers can have a ripple effect and cause large changes in the structure of higher trophic levels. It may cause an increase in the rates of primary productivity, changes in algal plant biomass, shifts in species compositions, destruction of corals and sea grasses, death of fish and shellfish populations, decreases in biodiversity, oxygen depleted in bottom waters and an increase in harmful algal blooms [3-6]. Additionally, phytoplankton is affected on turbidity, oxygen depletion and total productivity of the system [7].

Phytoplankton has also gained a new meaning due to the global eutrophication problem and the new European guidelines for surface water quality. The European Union Directive 2000/60/EC [8], also known as the Water Framework Directive (WFD), aims to achieve and maintain a clean and well-managed water environment through the establishment of reference conditions of water quality, based on the evaluation of several biological and chemical quality elements.

Assessment of eutrophication has been admitted to be a complex process [9], since number of variables such as nitrate, nitrite, ammonium ion, phosphorus and chlorophyll that cause response cannot be ignored. Water transparency and dissolved oxygen concentrations may also add useful information, if highly eutrophic environments are under consideration [10].

Corresponding Author: Ali Ganjian Khenari, Ecological Academy of Caspian Sea, P.O. Box 961, Sari, Iran.
Tel: +981513462498, E-mail: aganjian2002@yahoo.com.
A multivariate statistical methodology, often used in many fields of the scientific literatures [11-13] is principal component analysis (PCA). It is used as a dimension reducing technique by extracting a number of principal components accounting for the inter-correlation of the variables involved [14]. The method is most valuable when high correlation is observed among the variable [15] that is often observed in coastal eutrophication.

The aim of the current study is to examine the possibility of developing a multivariate index of eutrophication based on PCA, taking into account nutrient/biomass of phytoplankton and suggest scales for assessing water quality on the proposed index. The eutrophication scaling confronts with the Water Framework Directive 2000/60/EC [8] and is recommended as a multimetric index for assessing levels of coastal marine eutrophication.

MATERIALS AND METHODS

Site Study, Date and Sampling Strategy: Samples were collected during 33 cruises carried out on board the R/V Guilan from summer 1994 to winter 2007 in SCS at 18 transects (1994-2000) and 6 transects (2001-2007). Along each transect four stations were located at depths of 10, 20, 50 and 100 m (38°38’N and 49°54’E), (Figure 1).

Phytoplankton Sampling: Phytoplankton samples were collected along the southern coasts of Caspian Sea using a Van Dorn water bottle sampler [16], from the surface, 2, 5, 10, 20, 50 and 100m depth of column waters. A total of 4556 of phytoplankton samples were collected from 1994 to 2007. The collected samples were held in 0.5L bottles and preserved using buffered formaldehyde to obtain a final concentration of 4% [17]. The samples were kept stagnant for at least 10 days then concentrated to 30 ml by sedimentation and centrifugation (5 minutes with 3000 rpm), in a clinical centrifuge (Hettich- D7200, Tuttingen: Germany). For micro and nanophytoplankton analysis, a 0.1 ml sub-sample taken from the 30 ml sample was counted using a scanned slide (in two steps of quality and one step in quantity) under a phase contrast binocular microscope (Nikon) covering slip 24×24 and with magnification of 100×, 200× and 400×) [17-19]. The volume of each cell was calculated by measuring morphometric characteristics (i.e., diameter, length and width) and geometric shape [20]. Finally, the volume values were converted to biomass. Phytoplanktons were identified to possible taxonomy used in previous studies [21, 22].

Outline of the Method: PCA was applied on a set of data including nutrient (nitrate, nitrite, ammonia and phosphate) and biomass of phytoplankton,
originated from 1994-2007 in southern part of Caspian Sea, characterizing eutrophic, mesotrophic and oligotrophic levels. The first principal component was derived and was further applied and tested as eutrophication index on an independent set of data. Finally a five level water quality scale was developed to be consistent with the Water Framework Directive for assessing eutrophic levels in marine coastal waters. Factor analysis (FA) is designed to transform the original variables into new uncorrelated variables called of the factors, which are linear combinations of the original variables. The FA is a data reduction technique and suggests how many varieties are important to explain the observed variances in the data. Method PCA is used for extraction of different factors. The axis defined by PCA is rotated to reduce the contribution of less significant variables [23, 24]. This treatment provides a small number of factors that usually account for approximately the same amount of information as the original set of observations. The FA can be expressed as:

\[ F_i = a_1x_1 + a_2x_2 + \ldots + a_nx_n \]

Where \( F_i \) is the factor, \( a \) is the loading, \( x \) is the measured value of variable, \( i \) is the factor number, \( j \) is the sample number and \( m \) is the total number of variables. The factor scores can be expressed as:

\[ Z_j = a_1f_1 + a_2f_2 + \ldots + a_nf_n + e \]

Where \( Z \) is the measured variable, \( a \) is the factor loading, \( f \) is the factor score, the residual term accounting for errors or other source of variation.

**Application of the principal component analysis (PCA):**

Nitrate, nitrite, ammonium ion, phosphate and biomass of phytoplankton were used for the analysis, since these variables are considered as the most appropriate to describe the phenomenon of eutrophication in the Mediterranean region [25]. Principal component analysis was applied on the dataset [13] using the correlation matrix and the principal components were derived. The possibility to use the first principal component as an eutrophication index was then tested. The form of this potential eutrophication index (E.I.) is as follows:

\[ E.I. = aC_{POM} + bC_{NOD} + cC_{Am} + dC_{PO4} + eC_{phytoplankton biomass} \]

Where \( C \) denotes the various concentrations and \( a, b, c, d \) and \( e \) the coefficients derived from the PCA analysis for the first principal component after the standardization of variables.

A dataset of nutrient (nitrate, nitrite, ammonia and phosphate) and chlorophyll concentrations/phytoplankton biomass [26-31], described in previous work was used for the application of PCA.

**RESULTS**

Data on the water samples, both nutrient (nitrate, ammonium ion, nitrate and total phosphate) and biological (phytoplankton biomass) showed the maximum and minimum, mean and standard error which were summarized in Table 1. The observed multi-co-linearity of the five variables imposes the need for the application of PCA, a multivariate method that generates new sets of

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Mean ± Std.Error (Min-Max)</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>Phosphate</td>
<td>31.01 ± 0.28 (3.2-232)</td>
<td>4463</td>
</tr>
<tr>
<td>Nitrate</td>
<td>1.41 ± 0.016 (0.04-10)</td>
<td>4437</td>
</tr>
<tr>
<td>Nitrate</td>
<td>19.26 ± 0.42 (0.03-212.1)</td>
<td>4444</td>
</tr>
<tr>
<td>Ammonium ion</td>
<td>19.41 ± 0.39 (0.02-355.4)</td>
<td>4275</td>
</tr>
<tr>
<td>Phytoplankton biomass</td>
<td>100.28 ± 3.43 (0.0004-11560.0)</td>
<td>4355</td>
</tr>
</tbody>
</table>

**Table 2: PCA results for nutrient and phytoplankton biomass in the southern part of Caspian Sea 1994-2007**

<table>
<thead>
<tr>
<th>Component</th>
<th>Total</th>
<th>% of Variance</th>
<th>Cumulative %</th>
<th>Total</th>
<th>% of Variance</th>
<th>Cumulative %</th>
<th>Total</th>
<th>% of Variance</th>
<th>Cumulative %</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>3.558</td>
<td>71.165</td>
<td>71.165</td>
<td>3.558</td>
<td>71.165</td>
<td>71.165</td>
<td>3.391</td>
<td>67.817</td>
<td>67.817</td>
</tr>
<tr>
<td>3</td>
<td>0.106</td>
<td>3.929</td>
<td>96.353</td>
<td>0.106</td>
<td>3.929</td>
<td>96.353</td>
<td>0.858</td>
<td>100.000</td>
<td>100.000</td>
</tr>
<tr>
<td>4</td>
<td>0.141</td>
<td>2.811</td>
<td>99.165</td>
<td>0.141</td>
<td>2.811</td>
<td>99.165</td>
<td>0.141</td>
<td>100.000</td>
<td>100.000</td>
</tr>
<tr>
<td>5</td>
<td>0.042</td>
<td>0.835</td>
<td>100.000</td>
<td>0.042</td>
<td>0.835</td>
<td>100.000</td>
<td>0.042</td>
<td>100.000</td>
<td>100.000</td>
</tr>
</tbody>
</table>

285
Table 3: 1st and 2nd PCA component score coefficient matrix coefficients for five variables in the southern part of Caspian Sea 1994-2007.

<table>
<thead>
<tr>
<th>Variables</th>
<th>Component 1</th>
<th>Component 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Phosphate</td>
<td>0.201</td>
<td>-0.230</td>
</tr>
<tr>
<td>Nitrate</td>
<td>0.527</td>
<td>0.213</td>
</tr>
<tr>
<td>Nitrate</td>
<td>0.261</td>
<td>-0.075</td>
</tr>
<tr>
<td>Ammonium ion</td>
<td>0.301</td>
<td>0.145</td>
</tr>
<tr>
<td>Phytoplankton biomass</td>
<td>0.140</td>
<td>0.873</td>
</tr>
</tbody>
</table>

Fig. 2: Component plot in rotated space of the first and second principal Component of the five variables.

Uncorrelated variables (the principle components) as linear compositions of the original variables. The correlation matrix was further used to apply PCA. It is observed that the first principal component accounts for more than 71% of the total variance, whereas the other four components explain parts of the remaining variation from 21.3% to 0.84% (Table 2). Therefore, a remarkable dimensional reduction is achieved, if the information from the first component is used. The coefficients of the five variables in the first principal component after standardization are given in Table 3. It is observed that all coefficients are positive and almost equal, implying that the five variables participate with equal weights to the formation of the first principle component and therefore to the proposed eutrophication index based on the formula as:

\[ E.I. = 0.201C_{PCL} + 0.261C_{NOX} + 0.227C_{POX} + 0.301C_{NH4} + 0.140CBiomass \]

The ordination biplot of the first and second principal axes and the eigenvectors of the five variables used in the analysis are shown in Figure 2.

In general, two factors were extracted and 92.42% of the total variance was explained as shown in Table 2. The extracted factors were shown in Table 3. It is clear from this table that most of the parameters associated with each factor are well defined and contribute very little to other factors, that helps in the interpretation of factors. Factor 1, has a high positive loading of nitrite, ammonium ion and nitrate. This factor can be labeled as a measurement for conductivity, nitrite, ammonium ion and nitrate which explains 71.12% of the total variation. The contribution of this factor was the average of all parameters related to this factor. Factor 2, had a high positive loading for phytoplankton biomass and a high negative loading for nitrate and total phosphorus. This factor could be labeled as water quality indicator, which explains 21.26% of the total variation.

The validation of the index was carried out on an independent dataset of different years (1994-2007) in SCS. The index value was calculated for each year, along with the values for three eutrophication index typical of eutrophic, mesotrophic and oligotrophic condition in SCS (Table 4). The E.I. presented an increasing trend from oligotrophy to eutrophication, its lower value observed
Table 4: The eutrophication index values for three standards during 1994-2007 in the southern part of Caspian Sea

<table>
<thead>
<tr>
<th>E.I. Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>E (Eutrophic)</td>
</tr>
<tr>
<td>M (Mesotrophic)</td>
</tr>
<tr>
<td>O (Oligotrophic)</td>
</tr>
<tr>
<td>1994-1996</td>
</tr>
<tr>
<td>1996-1997</td>
</tr>
<tr>
<td>1999-2000</td>
</tr>
<tr>
<td>2001-2002</td>
</tr>
<tr>
<td>2003-2004</td>
</tr>
<tr>
<td>2004-2005</td>
</tr>
<tr>
<td>2005-2006</td>
</tr>
<tr>
<td>2006-2007</td>
</tr>
</tbody>
</table>

Table 5: Eutrophication index range for oligotrophy, mesotrophy and eutrophication resulting from the application of frequency distribution analysis

<table>
<thead>
<tr>
<th>Trophic status</th>
<th>Lower limit</th>
<th>Upper limit</th>
</tr>
</thead>
<tbody>
<tr>
<td>Oligotrophy</td>
<td>0.94</td>
<td>0.38</td>
</tr>
<tr>
<td>Mesotrophy</td>
<td>0.37</td>
<td>0.87</td>
</tr>
<tr>
<td>Eutrophication</td>
<td>0.88</td>
<td>1.51</td>
</tr>
</tbody>
</table>

for the typical oligotrophic (0.21) and the highest for year 2005-2006 (1.22). The E.I. of different years ranged from 0.26 to 1.22. It was minimum in 1994-1996 and maximum in 2005-2006. Years of 1994-1996 and 1996-1997 were grouped with the typical oligotrophic (O) and 1999-2000 was grouped with the standard mesotrophic (M) and 2003-2004, 2004-2005, 2005-2006 and 2006-2007 were grouped with the typical eutrophic (E).

Frequency based on data reported in Table 5 a monotonic and close to linear increase of the index was observed from oligotrophy to eutrophication and almost no overlapping between the three groups. Considering the limits of Table 5, the continuum of the eutrophication index can be divided into five classes, according to the requirements of the European water framework directive: (a) less than 0.04, (b) 0.04-0.38, (c) 0.38-0.85 (the average of the lower limit of the eutrophic and the upper limit of the mesotrophic groups), (d) 0.85-1.51 and (e) greater than 1.51. The water quality of the three standard sets (oligotrophic, mesotrophic and eutrophic) used in the previous analysis are characterized as good, moderate and poor water quality on the basis of their chl-a means were 0.11, 0.65 and 1.43 mg/l, respectively. Therefore it can be considered that the five classes of water quality defined above on the E.I. index correspond to the high, good, moderate, poor and bad water qualities of the WFD (Water framework directive).

DISCUSSION

Phytoplankton production can be regulated by many environmental factors such as temperature and light conditions. However, during the growing season phytoplankton communities are mainly limited by nutrients, especially phosphorus and/or nitrogen. A major fraction of nutrient loading to surface waters is derived from anthropogenic sources, a fact that provides for the opportunity to control nutrient load and thus eutrophication of the lakes, rivers and coastal waters [32]. Previous studies have quantified the relationships between nutrients and chlorophyll [33] but these have been largely based on data from NOAA in Caspian Sea.

[34-35] Established lists of indicators and indicator indices to evaluate the trophic range using biomass values. According to [32], changes in biomass and algal class structures are robust measures of either eutrophication or oligotrophication process, whereas changes on the species level are more relevant. This, in turn, elevates the need of experienced investigators and knowledge about the characteristic algal assemblages along the trophic spectrum in relation to the lake type is important [32].

Principal component analysis (PCA) has been shown to be a useful technique in the evaluation of environmental data [36, 37]. Recently, PCA has also been applied as a tool for water quality evaluation and management with special emphasis on the trophic potential [10]. Assessments of coastal eutrophication have also been based on PCA [38]. The main advantage of this method is that the new variables are uncorrelated among themselves [14] and their number is reduced without significant loss of the original information. In the present study the first component accounts for more than 71% of the total variance of the data (Table 2). It was therefore, found useful that the principal component scores could be used to form an index evaluating eutrophic conditions.

One of the problems concerning the assessment of eutrophication is the complexity due to the number of variables linked more or less to both causes (nutrient loading) and effects (phytoplankton parameters). In addition, eutrophication assessment is influenced by the seasonal variations of the variables involved [39]. It is therefore obvious that eutrophication indices based on univariate procedures [40] cannot satisfactorily describe the trophic state of a marine system. Therefore, the aim of the present work was to develop a simple index for assessing eutrophication that is made of a set of data of intercorrelated variables (nutrients and phytoplankton biomass).
There are different views in literature concerning the use of physicochemical or biological parameters for eutrophication assessment [38, 41, 42]. Several authors tend to consider seawater transparency (Secchi disk) and oxygen concentrations as variables characterizing eutrophication. Although these variables may be significant in eutrophic fields such as Baltic and Adriatic Sea, both hypertrophic seas, they do not contribute substantial information in water bodies such as Eastern Mediterranean where mesotrophic and oligotrophic conditions seem to prevail [43]. In the present work, it was found that inorganic nutrient concentrations which are the cause of eutrophication and phytoplankton biomass parameters (the effect of eutrophication), is adequate information to characterize eutrophic levels in water bodies. The invasion of the new evader a ctenophore Mnemiopsis leidyi, which was transported via the ballast water from Black Sea to Caspian have cause various disasters such as phytoplankton bloom after invasion in 2001 [33]. The PCA analysis of these five variables reduced the dimensionality of the variable space into one principal component that was used for the calibration of the index. The index was also found efficient when tested on an independent set of data from coastal stations.

It was realized that scale development for characterizing oligotrophy and eutrophication based on EI index should fulfill three criteria: (a) a large body of data should be used so that statistical estimators would be stabilized, (b) overlapping of ranges between different trophic levels (eutrophic, mesotrophic, oligotrophic) should be minimal and (c) the frequency distribution of the index values should form the basis of a probabilistic system for assessing nutrient/chl-a values. The outcome of this procedure was a rather clear-cut scale defining eutrophication, mesotrophy and oligotrophy which can be easily expanded in a five-level water quality scale to fulfill the WFD requirements [8]. The proposed scale was developed using data from coastal areas of the oligotrophic to eutrophication in the SCS, therefore its applicability is restricted to coastal waters of similar characteristics. However, the proposed methodology has been explicitly presented and therefore, it can be easily followed step-by-step for the development and implementation of the appropriate water quality scales for other European regional seas in the framework of the WFD.

The water quality around in SCS is classified as moderately and high organic polluted water [44]. Considering the applicability of the proposed scaling in an independent set of 1994-2007 in CSC, 2003-2004, 2004-2005 and 2005-2006 are classified as Poor quality and the rest (1994-1996, 1996-1997 and 1999-2000) as Good water quality. The latter is in agreement with the characterization on phytoplankton biomass; however there is a discrepancy of one quality class for 2006-2007, which is characterized as Moderate water quality on phytoplankton biomass (see Table 4). It seems therefore that the phytoplankton biomass scale is more optimistic compared to EI scale which takes into account both nutrients and phytoplankton biomass.

CONCLUSION

A multivariate methodology based on PCA is proposed in the present study for the assessment of the eutrophication level in SCS. The resulting eutrophication index was applied successfully in SCS. A clear-cut discrimination between oligotrophy, mesotrophy and eutrophication was achieved with this index as well as the development of an efficient five-level water quality classification scheme according to the requirements of the European Water Framework Directive. It seems the recent change of the eutrophication to mesotrophy situation was due to decreasing of ctenophore abundance in southern Caspian Sea from a more than 1 kgm⁻² in 2001 to one-third in 2006. However, further application of the proposed scheme is essential towards its establishment as a criterion for coastal water quality and its use in coastal management.

ACKNOWLEDGEMENT

Authors acknowledge Iranian Fisheries Research Organization (IFRO) and Caspian Sea Research Institute of Ecology for their technical and financial support and ecology department colleagues. Special thanks go to the crews of research Vessel (Guilan) trough Phytoplankton physico-chemical sampling. Additional thanks are devoted to Dr Mohammad Vahedi, Dr Pourgholam, Dr. Yousefiyan, Dr. Farabi and Mr. Mahdavi for their kind contributions.

REFERENCES


