Environmetric approaches to estimate pollution impacts on a coastal area by sediment and river water studies

G. Mihailov*, V. Simeonov**, N. Nikolov* and G. Mirinchev***

* Department of Water Supply, Sewerage, Water & Wastewater Treatment, Faculty of Hydrotechnics, University of Architecture, Civil Engineering & Geodesy, 1421 Sofia “Chr. Smirnensky” blvd. 1, Bulgaria
** Department of Analytical Chemistry, Faculty of Chemistry, University of Sofia “St. Okhridski”, 1164 Sofia, J. Bourchier Blvd. 1, Bulgaria

Abstract This paper represents an effort to demonstrate the opportunities of some environmetric methods like regression analysis, cluster analysis and principal components analysis. Their role for data modeling is stressed and the basic theoretical principles are given. The application of the multivariate statistical methods is illustrated by two major examples: Assessment of metal pollution based on multivariate statistical modeling of “hot spot” sediments from the Black Sea; and a trend study of Kamchia River water quality. In the first part of the study the environmetric approach makes it possible to separate three zones of the marine environment with different levels of pollution (Bourgas gulf, Varna gulf and lake buffer zone). Further, the extraction of four latent factors offers a specific interpretation of the possible pollution sources and separates the natural factors from the anthropogenic ones, the latter originating from contamination by chemical and steel-works and an oil refinery. In the second part of the study nine sampling sites along Kamchia River were considered as sources for water quality monitoring data. Trends for all parameters are calculated by the use of linear regression analysis and special attention is paid to a specific coastal site. Then five latent factors were extracted from the monitoring data set in order to gain information about some structural characteristics of the set.

Keywords Black Sea; coastal area; environmetrics; river basin; sediments; water quality

Introduction

Modern concepts of estimation of pollution impacts, especially in a coastal region, require not only careful organization of monitoring of the chemical composition of surface waters, coastal sediments and benthic organisms, seawaters and plants but also a reliable monitoring data treatment, classification and modeling. Since the analytical data from the coastal ecological systems possess a multivariate nature, the only successful approach for their interpretation is the environmetric one. Environmetrics is a relatively new branch of chemometrics or application of multivariate statistical methods in analytical chemistry.

There is growing concern all over the world at the extent of sediment and coastal river estuaries contamination (Env.Prot.Agency, Proceedings, 1992; National Research Council, 1989). Most contaminants discharged into coastal waters rapidly become associated with marine particulate matter and incorporated in sediments. Although diagenetic processes in the sediments modify and redistribute contaminants between solid and water phases, immobilization by the sedimentation process dominates for most of the typical pollutants. As already indicated the contaminant metal accumulation in coastal sediments provides a record of the spatial and temporal history of pollution (Martin and Whitfield, 1983). That is why sediment monitoring can deliver important information on various pollution events.

The monitoring of the coastal river waters is another possibility to check not only the surface water quality in a certain region but to estimate the impact of the pollution of the
coastal area caused by the water inlet. The combination of sediment and estuary water monitoring gives an opportunity to investigate the assessment of pollution for a coastal region in a better way.

There is another very important point in the monitoring data estimation. There are a variety of approaches for reducing natural variability and improving the statistical power in data intercomparison. For instance, metal concentrations, analysed in sediments are often normalized to a conservative component (often aluminum or iron) whose levels are unaffected by contaminant inputs (DeGroot et al., 1976; Forstner and Wittmann, 1981; Loring, 1990). Baseline regression models are introduced based on observed covariation of elements (Hanson et al., 1993; Daskalakis and O’Conor, 1995). Trend studies for surface water quality also contribute to some extent to the understanding of the river water pollution impacts (Simeonov et al., 2000).

The multivariate statistical analysis, however, seems to deliver more substantial information on relationships between sampling sites, pollutant concentrations, latent factors responsible for the data set structure and pollutant sources apportioning (Einax et al., 1997). The complexity and the large variance of environmental sets limit the use of common statistical methods for the assessment of the state of pollution, so the application of multivariate statistical approaches is recommendable.

The Bulgarian Black Sea coast is known as a recreational area, and this requires a constant environmetric control of all marine phases since a large number of industrial “hot spots” are also located near to the shore (oil refinery near to Bourgas Gulf, cement and chemical plants around Varna Gulf, steel-works, etc.). The Kamchia River estuary is also a preferred touristic area and checking of the water quality could contribute to the national tourist policy. Usually, the reaction of the authorities to monitoring data is slow due to the lack of clear information about the real state of pollution or to the temporary and local character of the pollution events. The present study therefore aims to perform a chemometric analysis of metal concentrations in coastal sediments collected at different “hot spot” sites of the Bulgarian Black sea coast and of the water quality of Kamchia river and its tributaries from the coastal area in order to offer a more informative and rapid assessment of the state of pollution.

**Experimental**

Sediment samples were taken from four different sampling sites: Lake Beloslavsko (10 sites, sample number 1–10, close location of glass production factory), Lake Varnensko (11 sites, sample number 11–21, close to steel-work), Varna Gulf, close to Lake Varnensko (7 sites, sample number 22–28), Varna Gulf near to coast (7 sites, sample number 29–35; close to cement and chemical plant Solvey Soda) and Bourgas Gulf, near to the waste inlets from the local oil-refinery (4 sites, sample number 36–39). It is worth noting that in the configuration of the coastal line, the two lakes (Beloslavsko and Varnensko) serve as a natural buffer zone between the industrial zone and the gulf of Varna. For the gulf of Bourgas no such zone exists, and there is a direct inlet of contaminated waters into the sea.

The sampling was performed with a standard bottom grab of Smith – McIntyre (Hanson et al., 1993) and the elements measured throughout this study were Cu, Pb, Mn, Zn, Co, Cd, Cr, Fe, Ni and As. Digestion in concentrated hydrofluoric acid and subsequent analysis by atomic absorption were used for quantification. ETAAS (graphite furnace AAS, Perkin Elmer Z/3030) was the analytical method to determine Cu, Pb, Co, Cd, Cr, Ni, As and flame AAS (Perkin Elmer 603) was used for Mn, Fe and Zn. Certified reference materials (MESS-1, BCSS-1 and NBS 1646) were run with each series of samples. Precision for Mn, Fe and Zn was ≤ 5% (as relative standard deviation); for the other analyses the RSD (relative standard deviation) was ≤ 10%. The experimental data set is available on request.
A second data set from different sampling sites along Kamchia River was also available. In the present study nine sampling sites were considered. In Figure 1 the location of the sites along the Kamchia River are indicated.

Site 66 is a typical coastal site. Sites 61 and 63 are urban sites located close to the settlements like Preslav and Shumen; site 67 is located on the tributary Luda Kamchia just before the inlet to Tsonevo dam. Site 62 is a rural site; sites 242 and 371 are on the main stream, close to anthropogenic sources. Finally, sites 373 and 374 indicate positions along the stream of the tributary Luda Kamchia before and after the Kamchia dam.

The sampling was performed for a long period of time, different for the different sites (between 7 and 20 years of monitoring). The data available are average annual values of the major indicators responsible for the water quality: pH, dissolved oxygen, nitrite, BOD5, acidity, dissolved matter, suspended solids, chloride, sulfate, ammonium, nitrate, phosphate, iron, manganese, calcium, magnesium, water hardness. The analytical determinations were routine and according to Bulgarian and European water quality standards (Standard Methods for the Examination of Water and Wastewater, APHA, AWWA, WPCF, 1997).

The multivariate statistical approaches applied to both data sets were as follows:

- Linear regression analysis;
- Cluster analysis;
- Principal components analysis (PCA).

All of these methods are well-known and described in the literature (Einax et al., 1997; Massart and Kaufman, 1983; Malinowski, 1991; Draper and Smith, 1981). Since there are different opinions on the use of multivariate statistical methods in environmental analysis, varying from skeptical acceptance as “complex and to some extent formal” (Hanson et al., 1993) to solid conviction of being “a useful tool for the best evaluation and interpretation of environmental data” (Einax and Soldt, 1999), it was the aim of the present study to compare in a practical way these two extreme judgments. Throughout the study the software package STATISTICA 5.0 was used.

The possibility to analyse multidimensional data sets without information about the spatial locations is the advantage of multivariate statistics. We consider that the basic principles and the theoretical fundamentals are known to a substantial number of the scientists dealing

---

**Figure 1** Kamchia river basin
with environmental analysis and data treatment and there is no special need for a detailed explanation. Many studies offer algorithms of different multivariate statistical approaches. For instance, linear regression analysis is performed to detect trends in the data set. Cluster analysis is carried out to reveal specific linkage between sampling sites being an indication of similarities or dissimilarities between their trace metal contaminations. Further, PCA is applied to detect the hidden structure of the data set, trying to explain the influence of latent factors on the data distribution. The regression analysis in its classical linear performance makes it possible to create respective trends and prediction models for each indicator. Therefore, by means of multivariate statistical analysis a semiquantitative assessment of the polluted area is feasible if coastal sediment monitoring data are treated or a survey of the long-term trends in the surface water quality in the coastal region is achieved.

**Results and discussion**

The cluster analysis results (hierarchical clustering, Ward’s method) of the sampling sites as objects are shown in Figure 2.

Altogether, four clusters could be interpreted divided into two bigger subgroups: the first contains heavily polluted sites from Varna and Bourgas Gulf near to the coastal line and waste inlets (sites 29–39 located near a big chemical and cement plant – Varna and an oil refinery – Bourgas) and several sites from both coastal lakes located near to industrial sources (sites 2, 4, 6, 7, 8, 9 from the Lake Beloslavsko located near to a glass production factory; sites 14, 17, 19 from the Lake Varnensko located near to a steel-work); the second one indicates a moderately polluted buffer zone consisting of lake and near to the lake Varna gulf sites. In both big clusters two subgroups could be found. In the first one they represent the most severely polluted gulf areas (sites from Varna gulf 29–35 and from Bourgas gulf 36–39) and the less contaminated lake industrial inlets (sites 2, 4, 6–9 from Lake Beloslavsko and sites 14, 17, 19 from Lake Varnensko). In the second one they reflect the separation between one (Varnensko lake and non-affected Varna gulf parts, sites 11–13, 15, 16, 18, 20, 21 and sites 22–28, respectively) or another part (Lake Beloslavsko, sites 1, 3, 5, 10) of the buffer zone moderately affected by pollutants.

The next step in the multivariate statistical analysis was application of PCA in order to group the chemical components by the loadings plots and the sites by the score plots. It is interesting to note that the site score plot (Figure 3) reveals a more detailed description of the polluted coastal region.

![Figure 2 Hierarchical dendrogram (Ward’s method of linkage) for 39 sampling sites (sites 1–10 – Lake Beloslavsko; sites 11–21 – Lake Varnensko; sites 22–28 Varna gulf near to lake Varnensko; sites 29–35 – Varna Gulf near to coast; sites 36–39 – Bourgas gulf)](https://iwaponline.com/wst/article-pdf/46/8/45/426010/45.pdf)
The sites in the Bourgas gulf (36–39) represent an independent group (I) of heavily polluted area (oil refinery). They are definitely separated from all other sites and this is due to the enhanced indicator concentrations. The next well-formed group (III) comprises sites from the moderately contaminated lake buffer zones (sites 1–28), which indicates sites from the two lakes and Varna gulf sites located near to Lake Varnensko. The third group (II) indicates the intermediate level of pollution (higher than the buffer zone contamination but less than the Bourgas gulf area) of the sites originating mainly from the Varna gulf area (sites 29–35).

The factor loading matrix is listed in Table 1. Four factors describe almost 90% of the total variance of the system. The first one contains dominantly copper, manganese and iron and could be conditionally named “natural” since these elements are typical major constituents of Black sea coastal sediments. The second factor includes zinc, cadmium and chromium, the third – lead and arsenic and the fourth – nickel and cobalt. The last three factors reflect the typical anthropogenic influences of heavy metals from various sources such as chemical and glass production plants, oil refineries, steel-works and melting plants. The detected pollution pattern indicates in a semiquantitative way the emission sources. In Figure 4 the 3-D loading plot (PC1 vs. PC2 vs. PC3) is presented.

The second part of the study deals with trend and principal components analysis of the water quality data from Kamchia River and its tributaries. As already mentioned above nine sampling sites along the Kamchia river basin were considered. The results from the trend analysis are summarized in Table 2.

**Table 1** Factor loadings (Varimax normalized; marked loadings are higher than 0.7) for four principal components

<table>
<thead>
<tr>
<th>Element</th>
<th>PC1</th>
<th>PC2</th>
<th>PC3</th>
<th>PC4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cu</td>
<td>0.95</td>
<td>0.08</td>
<td>-0.09</td>
<td>0.04</td>
</tr>
<tr>
<td>Pb</td>
<td>0.04</td>
<td>-0.04</td>
<td>-0.88</td>
<td>-0.04</td>
</tr>
<tr>
<td>Mn</td>
<td>0.96</td>
<td>0.04</td>
<td>0.16</td>
<td>0.17</td>
</tr>
<tr>
<td>Zn</td>
<td>0.27</td>
<td>0.90</td>
<td>0.04</td>
<td>0.11</td>
</tr>
<tr>
<td>Co</td>
<td>0.31</td>
<td>-0.04</td>
<td>0.17</td>
<td>0.89</td>
</tr>
<tr>
<td>Cd</td>
<td>-0.15</td>
<td>0.88</td>
<td>-0.26</td>
<td>-0.05</td>
</tr>
<tr>
<td>Cr</td>
<td>-0.13</td>
<td>0.82</td>
<td>0.05</td>
<td>0.27</td>
</tr>
<tr>
<td>Fe</td>
<td>0.95</td>
<td>-0.14</td>
<td>0.07</td>
<td>0.11</td>
</tr>
<tr>
<td>Ni</td>
<td>-0.01</td>
<td>0.16</td>
<td>-0.04</td>
<td>0.81</td>
</tr>
<tr>
<td>As</td>
<td>-0.14</td>
<td>0.16</td>
<td>-0.86</td>
<td>-0.06</td>
</tr>
<tr>
<td>% Expl.var</td>
<td>29.5</td>
<td>27.1</td>
<td>16.5</td>
<td>15.8</td>
</tr>
</tbody>
</table>
In this table only qualitative data from the linear regression analysis are given for all sampling sites and all indicators. The trends of decrease (D) and increase (I) for each chemical component are given for each site. The statistically significant trends are indicated by s, the non-significant are indicated as ns. The lack of any trend is shown with the abbreviation no tr.

### Table 2  Summarized results from trend analysis of Kamchia river monitoring data

<table>
<thead>
<tr>
<th>Indic.</th>
<th>61u</th>
<th>62r</th>
<th>63u</th>
<th>66co</th>
<th>67u</th>
<th>242str</th>
<th>371u</th>
<th>373be</th>
<th>374aft</th>
</tr>
</thead>
<tbody>
<tr>
<td>pH</td>
<td>I-s</td>
<td>I-s</td>
<td>I-s</td>
<td>I-s</td>
<td>I-s</td>
<td>I-ns</td>
<td>I-ns</td>
<td>I-ns</td>
<td>I-ns</td>
</tr>
<tr>
<td>Diss.m.</td>
<td>D-ns</td>
<td>D-ns</td>
<td>D-ns</td>
<td>I-ns</td>
<td>I-ns</td>
<td>D-ns</td>
<td>I-ns</td>
<td>D-ns</td>
<td>D-ns</td>
</tr>
<tr>
<td>Susp.s.</td>
<td>D-ns</td>
<td>D-ns</td>
<td>D-ns</td>
<td>D-ns</td>
<td>D-ns</td>
<td>D-ns</td>
<td>I-ns</td>
<td>D-ns</td>
<td>D-ns</td>
</tr>
<tr>
<td>Diss.O₂</td>
<td>D-ns</td>
<td>I-ns</td>
<td>I-ns</td>
<td>I-ns</td>
<td>I-ns</td>
<td>D-ns</td>
<td>I-ns</td>
<td>D-ns</td>
<td>D-ns</td>
</tr>
<tr>
<td>Ac.</td>
<td>D-ns</td>
<td>D-ns</td>
<td>D-ns</td>
<td>D-ns</td>
<td>D-ns</td>
<td>D-ns</td>
<td>I-ns</td>
<td>D-ns</td>
<td>D-ns</td>
</tr>
<tr>
<td>NH₄</td>
<td>D-ns</td>
<td>D-ns</td>
<td>D-ns</td>
<td>D-ns</td>
<td>D-ns</td>
<td>D-ns</td>
<td>I-ns</td>
<td>D-ns</td>
<td>D-ns</td>
</tr>
<tr>
<td>NO₂</td>
<td>No tr</td>
<td>I-ns</td>
<td>I-ns</td>
<td>I-ns</td>
<td>I-ns</td>
<td>I-ns</td>
<td>I-ns</td>
<td>I-ns</td>
<td>I-ns</td>
</tr>
<tr>
<td>NO₃</td>
<td>D-ns</td>
<td>D-ns</td>
<td>D-ns</td>
<td>D-ns</td>
<td>D-ns</td>
<td>D-ns</td>
<td>I-ns</td>
<td>D-ns</td>
<td>D-ns</td>
</tr>
<tr>
<td>PO₄</td>
<td>D-ns</td>
<td>D-ns</td>
<td>D-ns</td>
<td>D-ns</td>
<td>D-ns</td>
<td>D-ns</td>
<td>I-ns</td>
<td>D-ns</td>
<td>D-ns</td>
</tr>
<tr>
<td>Cl</td>
<td>D-ns</td>
<td>D-ns</td>
<td>D-ns</td>
<td>D-ns</td>
<td>D-ns</td>
<td>D-ns</td>
<td>I-ns</td>
<td>D-ns</td>
<td>D-ns</td>
</tr>
<tr>
<td>SO₄</td>
<td>D-ns</td>
<td>D-ns</td>
<td>D-ns</td>
<td>D-ns</td>
<td>D-ns</td>
<td>D-ns</td>
<td>I-ns</td>
<td>D-ns</td>
<td>D-ns</td>
</tr>
<tr>
<td>Fe</td>
<td>D-ns</td>
<td>D-ns</td>
<td>D-ns</td>
<td>D-ns</td>
<td>D-ns</td>
<td>D-ns</td>
<td>I-ns</td>
<td>D-ns</td>
<td>D-ns</td>
</tr>
<tr>
<td>Mn</td>
<td>D-ns</td>
<td>D-ns</td>
<td>D-ns</td>
<td>D-ns</td>
<td>D-ns</td>
<td>D-ns</td>
<td>I-ns</td>
<td>D-ns</td>
<td>D-ns</td>
</tr>
<tr>
<td>Ca</td>
<td>I-s</td>
<td>D-ns</td>
<td>D-ns</td>
<td>D-ns</td>
<td>D-ns</td>
<td>D-ns</td>
<td>I-ns</td>
<td>D-ns</td>
<td>D-ns</td>
</tr>
<tr>
<td>Mg</td>
<td>I-ns</td>
<td>D-ns</td>
<td>D-ns</td>
<td>D-ns</td>
<td>D-ns</td>
<td>D-ns</td>
<td>I-ns</td>
<td>D-ns</td>
<td>D-ns</td>
</tr>
<tr>
<td>Hard.</td>
<td>D-ns</td>
<td>D-ns</td>
<td>D-ns</td>
<td>D-ns</td>
<td>D-ns</td>
<td>D-ns</td>
<td>I-ns</td>
<td>D-ns</td>
<td>D-ns</td>
</tr>
<tr>
<td>BPK</td>
<td>D-ns</td>
<td>D-ns</td>
<td>D-ns</td>
<td>D-ns</td>
<td>D-ns</td>
<td>D-ns</td>
<td>I-ns</td>
<td>D-ns</td>
<td>D-ns</td>
</tr>
<tr>
<td>Koli</td>
<td>D-ns</td>
<td>D-ns</td>
<td>D-ns</td>
<td>D-ns</td>
<td>D-ns</td>
<td>D-ns</td>
<td>I-ns</td>
<td>D-ns</td>
<td>D-ns</td>
</tr>
</tbody>
</table>

Note: D – decreasing trend; be – before the dam; No tr – no trend at all; str – along the main stream site; I – increasing trend; aft – after the dam; s – significant; c – coastal site; u – urban site; ns – non-significant; r – rural site; c – coastal site

In this table only qualitative data from the linear regression analysis are given for all sampling sites and all indicators. The trends of decrease (D) and increase (I) for each chemical component are given for each site. The statistically significant trends are indicated by s, the non-significant are indicated as ns. The lack of any trend is shown with the abbreviation no tr.

Several important conclusions could be derived from the trend study. At all sites the pH value has a definite increasing trend, which is an indication for decrease of the water acidity. The BOD₅ and acidity values are also decreasing at all sites (the exception is site 371 with a statistically non-significant increasing trend). One could easily find dominantly decreasing trends in the content of dissolved matter and suspended solids, ammonium, nitrate, phosphate and water hardness. This is an indication for improvement of the overall water quality.

Figure 4  PCA loading 3D plot (PC1 vs. PC2 vs. PC3) for 10 chemical components
It is quite important to note that in many cases the data to be treated for trend analysis contain typical outliers (extreme maximum or minimum values), which hinder the proper determination of a real trend. For instance, very often calcium and magnesium contents are almost constant for the whole period of observation and only single outliers could influence one or other course of (usually statistically non-significant) trend. That is why a recommendable procedure in determining a trend is a careful check of the raw data distribution.

If one takes as a special site the coastal site 66, it could be expected that its component trends are a result of overlapping of tendencies from the main river stream on one hand (sites 61, 62, 63) and, on the other, from the influences of the tributaries like Luda Kamchia (sites 67, 242, 371, 373 and 374). For pH it is obvious that the “coastal” trend (significant trend of increase, i.e. decrease of acidity) is really a confirmation of the overall pH increase at all other sites under consideration. A quite similar situation is observed for those indicators, which indicate statistically significant trends at the coastal site: BOD₅, ammonium, nitrate and hardness. But, it is also seen, that the main stream is more important in formation of the “coastal pattern” at least for these indicators (pH, dissolved oxygen, nitrite, BOD₅, acidity, suspended solids, chloride, ammonium, nitrate, hardness). A special exception is sulfate with a decreasing trend (both statistically significant and non-significant) at all sites except for the coastal one where a turnover is observed – increasing (non-significant) trend. Actually, there are two outliers in the raw data (1981 and 1990), which influence the trend.

The biological indicator does not indicate a significant trend, in most of the cases its values are almost constant.

An interesting idea for the next step in the investigation of Kamchia River basin is to calculate the role of water inflows for the formation of the coastal pattern.

The performance of PCA of the Kamchia river data set (all sites and all monitoring periods) suggested the formation of 5 latent factors responsible for the data structure and possible natural and anthropogenic sources in the region:

- **PC1**: $pH$, dissolved oxygen, nitrite – “acidity” factor
- **PC2**: BOD₅, acidity, ammonium, suspended solids, phosphate – “turbidity” factor
- **PC3**: dissolved matter, chloride, calcium, magnesium – “solubility” factor
- **PC4**: sulfate, nitrate, hardness – “salt” factor
- **PC5**: iron, manganese – “anthropogenic” factor.

These PC models confirm, in principle, ideas about the role of the different indicators in the formation of surface water quality.

**Conclusion**

The application of different multivariate statistical methods proves to be an efficient tool in achieving better understanding of the state of the environment. It is recommended to combine various approaches to gain better information on the system of interest or to try to predict its future trends.

Several important conclusions could be derived from the present environmetric study.

- Three zones of the Black Sea marine environment with different levels of pollution are separated by interpretation of the sediment analysis data by environmetric methods;
- Four latent factors are extracted from the sediment data by PCA, one of them indicating natural effects and the other three – anthropogenic influences from oil refinery, cement and chemical plants;
- The long-term water quality of the Kamchia river estuary is analysed and trends for 18 indicators of the surface water are determined with the summarising conclusion for improvement of the water quality;
- Five latent factors determine the structure of the river water data set reflecting the
natural (turbidity, solubility) and anthropogenic (acidity, metal content, dissolved salts) influences on the water quality.

References


*Standard Methods for the Examination of Water and Wastewater* (1997). APHA, AWWA, WPCF.