Chemosphere 112 (2014) 114–119

Contents lists available at ScienceDirect

Chemosphere

journal homepage: www.elsevier.com/locate/chemosphere

Chemometric analysis of ecological toxicants in petrochemical and industrial environments

Richard Olawoyin [a],*, Brenden Heidrich [b], Samuel Oyewole [c], Oladapo T. Okareh [d], Charles W. McGlothlin [a]

[a] Oakland University, School of Health Sciences, Rochester, MI 48309, USA
[b] Irradiation Services and Operations, Radiation Science & Engineering Center, The Pennsylvania State University, USA
[c] U.S. Chemical Safety & Hazard Investigation Board, Washington D.C. Metro Area, USA
[d] Division of Environmental Health Sciences, College of Medicine, University of Ibadan, Ibadan, Nigeria

HIGHLIGHTS

- Comprehensive review of chemometrics techniques in environmental assessments.
- Chemometrics methods were used to assess the study area – the Niger Delta.
- Identification of pollution sources and distributions were done using the PCA.
- PCA minimized and orthogonolize the multivariable datasets from samples.
- Accurate results obtained can enhance appropriate mitigating measure designs.

ARTICLE INFO

Article history:
Received 4 February 2014
Received in revised form 14 March 2014
Accepted 23 March 2014

Handling Editor: Tamara S. Galloway

Keywords:
Chemometrics
Toxicants
Principal component analysis
Pollutants
Environment
Variogram

ABSTRACT

The application of chemometrics in the assessment of toxicants, such as heavy metals (HMs) and polycyclic aromatic hydrocarbons (PAHs) potentially derived from petrochemical activities in the microenvironment, is vital in providing safeguards for human health of children and adults residing around petrochemical industrial regions. Several multivariate statistical methods are used in geosciences and environmental protection studies to classify, identify and group prevalent pollutants with regard to exhibited trends. Chemometrics can be applied for toxicant source identification, estimation of contaminants contributions to the toxicity of sites of interest, the assessment of the integral risk index of an area and provision of mitigating measures that limit or eliminate the contaminants identified. In this study, the principal component analysis (PCA) was used for dimensionality reduction of both organic and inorganic substances data in the environment, which are potentially hazardous. The high molecular weight (HMW) PAHs correlated positively with stronger impact on the model than the lower molecular weight (LMW) PAHs, the total petroleum hydrocarbons (TPHs), PAHs and BTEX correlate positively in the F1 vs F2 plot indicating similar source contributions of these pollutants in the environmental material. Cu, Cr, Cd, Fe, Zn and Pb all show positive correlation in the same space indicating similar source of contamination. Analytical processes involving environmental assessment data obtained in the Niger Delta area of Nigeria, confirmed the usefulness of chemometrics for comprehensive ecological evaluation.

1. Introduction

Environmental media are contaminated by pollutants via the deposition of suspended particulate matter (Morillo et al., 2007), contaminated sediments deposition (Zhou et al., 2007) and also through underground water circulation (Duruibe et al., 2007). The contaminants deposited in the soil, sediment, air or water can migrate to the human body through inhalation in their natural state when re-suspended as particulate matters, as soil particles or as adsorbates on dust (Shi et al., 2008).

Several studies have investigated environmental contaminations of soils in and around cities with children's playgrounds (Chen et al., 2005; Vidovic et al., 2005; De Miguel et al., 2006). The major routes of exposure of humans to environmental toxicants are through inhalation, ingestion and dermal contact. Several studies have identified accidental ingestion of soil as the route with the highest tendency of human exposure (especially children) to

---

* Corresponding author. Tel.: +1 248 364 8653.
E-mail address: olawoyin@oakland.edu (R. Olawoyin).

http://dx.doi.org/10.1016/j.chemosphere.2014.03.107
0045-6535/© 2014 Elsevier Ltd. All rights reserved.

contaminants (Ljung et al., 2006; Shi et al., 2008). Volatile toxicants are easily inhaled than ingested but chances of ingesting other toxicants such as lead (Ljung et al., 2006) are higher than being inhaled.

The use of multivariate analytical techniques have gained more steam in recent years but univariate statistical analyses remain useful to a wide range of other environmental scientists. Significant use of chemometrics methods are available, which include source apportionment methods (Zhang et al., 2008; Chen et al., 2009), principal component analysis (PCA) and linear discriminant analysis (LDA) (Seilaff and Einax, 2007). Multivariate analytical techniques give more details on the data structure collected from environmental media than univariate methods since multiple parameters are measured during soil, sediment, water and air quality analysis. In soil analytical studies, a variety of geostatistical tools with a wide range of applications exist (Kanevski et al., 2004).

Attempts have been made by researchers to analyze; polycyclic aromatic hydrocarbons (PAHs) Table 1, polychlorinated napthenes (PCNs) and heavy metals (HMs) contaminations levels in environmental media chemometrics techniques (Bostrom et al., 2002; Mielke et al., 2004; Jarosinska et al., 2006). Although these substances are ubiquitous in the environment (Zhang et al., 2006), but there are concerns when the levels are significantly higher than anticipated background concentrations.

1.1. Pollutants and associated features

1.1.1. Fate and effect

The contamination levels of PAHs, PCNs and HMs are elated in environmental media due to contributions from anthropogenic activities such as petrochemical industrial activities (Morillo et al., 2007). The effects of these toxicants on human health range from mild to severe depending on the concentration and exposure of either an individual toxicant or a mixture of different organic and inorganic toxicants.

2. Methodology

2.1. Principal component analysis (PCA)

The PCA method has improved considerably in recent times, the PCA method has been noted for its applicability as a classic statistical and diagnostic tool (Diamantaras and Kung, 1996), which helps to analyze covariance structures embedded in any multivariate numerical datasets, detect outliers, validate and interpret datasets, and for visualization. Simplified explanation of the PCA method is presented in published studies (Shaw, 2003). However, the specific mathematical manipulations behind the PCA method are presented in other literatures (Jolliffe, 2002; Jackson, 2003). The fundamentals of PCA involves the transformation of a set of multivariate data containing analytical constituents (variables) into a new orthogonal set by allocating total variance to uncorrelated variables (principal components – PCs) using the correlation matrix, whereby the individual variable represent the linear combination of the initial data variables. The PCs are in decreasing order based on factor loading, having the PCs with the largest variance occupying the first PC (PC1) and this follows successively to the PC with the smallest variance (PCn). Eq. (1) shows the mathematical expression for the PC computation with respect to the original variables.

Table 1
PAHs characteristics in sampled media.

Compound	Abbreviation	Benzene ring
Naphthalene	Nap	2
2-Methylnaphthalene	2MNap	2
Acenaphthylene	Acy	2
Acenaphthene	Ace	2
Fluorene	Flu	3
Phenanthrene	Phe	3
Anthracene	Ant	3
Fluoranthene	Flr	4
Pyrene	Pyr	4
Benz[a]anthracene	BaA	4
Chrysene	Chr	4
Benzo[b]fluoranthene	BbF	4
Benzo[k]fluoranthene	BkF	4
Benzo[a]pyrene	BaP	5
Dibenz[a,h]anthracene	DahA	5
Benzo[ghi]perylene	BghiP	6
Indeno(1,2,3-cd)pyrene	InP	6
Benzene, Toluene, Ethylene and Xylene	BTEX	

$$PC1 = a_{11}x_1 + a_{12}x_2 + \ldots + a_{1n}x_n$$

$$PCn = \sum_{j=1}^{n} a_{1j}x_j \tag{1}$$

where a_{1j} is the eigenvectors obtained from the correlation matrix, and x_j is the input variables.

The PCA method is performed sequentially, first by information extraction in the input space (with n-dimensions) to determine the directions of which the input variables x_j display the most substantial variability. The PC coefficients and the eigenvalues ($\lambda_i > 0, i = 1, \ldots, n$) for the correlation matrix ($C = E\{xx^T\}$) with respect to their eigenvectors ($e_i > 0, i = 1, \ldots, n$) make up what is called the loadings are then calculated. This gives a new set of variables that explain the variability in the original dataset; the first PCs retain a greater proportion of the total variance, consequently leading to effective and practical dimensionality reduction exercise. This is an effective means of describing sources, distributions and concentrations of pollutants in environmental materials by displaying multivariate patterns in the original datasets in lower dimensions. However a rational amount of PCs are necessary to be chosen for the avoidance of losing meaning information and poor predictability (Jolliffe, 2002).

3. Principal component analysis (PCA) results

The datasets used in this study include soil, sediment and water quality assessment data collected at different locations in the Niger Delta area of Nigeria (Olawoyin et al., 2012; 2013). The data include physic-chemical, organic and inorganic variables. The PCA was carried out using SAS® version 9.3 and the XLSTAT. Varimax rotation was applied to simplify the factor interpretation by reducing the total number of variables that exhibit high loadings per factor (Everitt and Dunn, 1992). For all the pollutant variables analyzed the correlation matrix was built with equal weights (Chatfield and Collins, 1980).

3.1. Organic substances

The summary statistics for the soil and water PAHs, and sediment physico-chemical valid data (Olawoyin et al., 2013) calculated are presented in Table 2.

The interpretation of PCs in this study comprises of the evaluation of the relationship between the loadings and contamination sources from petrochemical activities. PCs with variables with high loadings depict greater importance from the contamination sources, whereas lower loadings point to lower importance with regards the sources of these contaminations. The PCA in this study suggest a robust solution for the dimensionality reduction of variables based on the score calculation. For the soil PAHs, majority of the pollutants correlation show similar trends in the output space