Article

Data mining using multivariate statistical analysis: The case of heavy metals in sediments of the Msimbazi Creek mangrove wetland

A. Mrutu[1], **G. B. Luilo**[2]
[1]Department of Chemistry, University of Missouri-Columbia, 601 S. College Avenue, Columbia, MO 65211, USA
[2]Department of Chemistry and Chemical Biology, University of New Mexico, USA
E-mail: agnesmrutu@gmail.com

Received 14 August 2013; *Accepted 18 September 2013*; *Published online 1 December 2013*

Abstract

Mangrove wetlands are important biological systems that usually filter out organic and inorganic contaminants from the wastewaters before entering the ocean. Our previous work showed that sediments of the Msimbazi Creek wetland are contaminated with heavy metals and the amounts decreased with increasing depth. However, the hidden relationships between the heavy metals and clay particles were not fully understood based on the numerical data. Therefore this work used the data from literature and the Statistical Package for Social Sciences (SPSS) software to study how significant the relationships are and predict the sources of heavy metals and clays. The results showed that Cd is the only metal that showed insignificant correlations with other heavy metals (with Pb and Zn) while the rest of heavy metals exhibited significant positive correlation (except Pb vs. Ni). Cluster analysis classified the heavy metals based on the concentration and the first 50 cm cores (0-50 cm) had higher heavy metals and % clay than the second 50 cm cores (51-100 cm). The results from the factor analysis suggests that Pb, Cd, Ni, and clay owe their source mostly from anthropogenic activities while Fe, Co, Cr, Zn and sand come from both anthropogenic and natural sources. These results support our previous suggestions that heavy metals and clays found in this wetland have mostly anthropogenic origin. However, we recommend isotopic tracing studies in order to accurately identify the origins of the heavy metals and clays in sediments of Msimbazi Creek mangrove wetland.

Keywords wetland; sediments; heavy metals; correlation; cluster analysis; factor analysis.

Environmental Skeptics and Critics
ISSN 2224-4263
URL: http://www.iaees.org/publications/journals/environsc/online-version.asp
RSS: http://www.iaees.org/publications/journals/environsc/rss.xml
E-mail: environsc@iaees.org
Editor-in-Chief: WenJun Zhang
Publisher: International Academy of Ecology and Environmental Sciences

1 Introduction

Mangrove wetland sediments are important sink organic and inorganic chemical pollutants from point and nonpoint sources (UNEP, 1999). In addition water flowing through the mangroves also brings other materials such sand, silt, clay and organic matter (UNEP, 1999) and these parameters play a great role in the retention and distribution of contaminants. Previous studies have pointed out that waters and sediments of rivers flowing through Dar es Salaam city and the Indian Ocean are contaminated with organic and inorganic pollutants (Mgana and Mahono, 1997; Mohammed, 2002). The surface sediments are reported to have the highest levels of the contaminants (A'khabuhaya and Lodsenius, 1988; Machiwa 1992; DeWolf et al., 2001; Mwevura et al., 2002; Mremi and Machiwa, 2003; Muzuka, 2007).

Our previous study at the Msimbazi Creek mangrove wetland, on the hand, investigated not only the heavy metal levels but also their distribution in the sediment cores and the influencing factors (Mrutu *et al.*, 2013). The factors that were studied include sediment texture (sand, silt, and clay), organic matter (OM), pH, EC, CEC, and redox potential. The descriptive statistics and graphical methods showed that these factors decreased with increased depth while % sand increased with depth except pH which ranged from 6.3 to 7.6 (Mrutu *et al.*, 2013). Since heavy metals levels decreased with increasing depth, it was inferred that all the factors that decreased with depth might have significant influence in the distribution of heavy metals. However, it was not clearly understood how significant the relationships were between heavy metal distribution and the physico-chemical factors. It is generally accepted among scientific community that the concentration of a contaminant or nutrient of anthropogenic origin will decrease gradually with depth in a sediment core (Tam and Wong, 1993). We used this general knowledge to infer that the heavy metals had anthropogenic origin. However, there was a need to perform statistical analysis to determine the significance of the association between heavy metals and physico-chemical factors and determine if the heavy metals owe their sources to anthropogenic activities only or both natural and anthropogenic activities.

Thus, the objective of this paper is to apply multivariate statistical analysis to reveal hidden relationships between heavy metals and also between heavy metals and physic-chemical factors. In this work we report the results from Pearson correlation analysis, cluster analysis and factor analysis using the SPSS software. Factor analysis and correlation analysis are used to suggest the potential source of heavy metals and sediment particles found in the Msimbazi Creek mangrove wetland.

2 Methodology

The study used the raw data from previous work on spatial distribution of heavy metals in sediments of Msimbazi creek mangrove wetland (Mrutu et al., 2013) and the full data is given in Table 1. There were 24 sediment samples and each sample had data on heavy metal contents (Fe, Cd, Cr., Cr, Ni, Zn and Pb), pH, EC, CEC, redox potential (redox), % OM, % sand, % silt, and % clay. The Msimbazi River drains most of the waste from the city to the Msimbazi Creek. The hidden relationships between heavy metals and physico-chemical factors can be revealed by performing multivariate statistical analyses using the Statistical Package for Social Sciences (SPSS) version 19 from SPSS Inc., Chicago. We performed the Pearson correlation analysis, cluster analysis and factor analysis.

The Pearson correlation coefficients (r) measures the degree of associations between two sets of variables and its values range between –1 and +1, which indicates strong negative and positive association respectively, and zero indicates no association (Davis, 1986; Zhang, 2011, 2012) while any r with $p \geq 0.05$ is regarded as an insignificant correlation (sample size for each analysis was 24). The positive correlation between a pair of heavy metals suggest common origin or sources (Ra et al., 2013; Muzuka, 2007; Cynthia et al., 2011). Cluster analysis refers to the bewildering assortment technique designed to perform classification by assigning observations to groups so that each group is more or less homogeneous and distinct from each other (Davis, 1986), The groups that are closely related are displayed in a tree diagram called dendogram. The higher the branches joining at the end point, the more similar (or homogeneous) the samples are (Parks, 1966). The grouping of the parameters from cluster analysis of sediment samples was performed using the Ward Method. Factor analysis is a statistical technique that aims at revealing relationships between variables with smaller variations among them (Harbaugh and Demirmen, 1964).

Table 1 The heavy metals (ppm) and other parameters in sediment samples used in this study.

Sites	#	Depth(cm)	Fe	Cd	Co	Cr	Ni	Zn	Pb	%Sand	%Silt	%Clay	%OM	pH	EC	CEC	Redox
A1	1	0-25	17000	3.6	10.9	50.1	15.2	122.0	76.0	9.1	13.7	77.2	6.0	6.5	7.0	98.2	124.0
	2	26-50	17000	2.1	10.8	51.6	16.9	128.0	70.5	1.0	14.0	85.0	9.4	6.5	11.1	92.4	-14.0
	3	51-75	14000	1.1	9.6	39.2	14.2	121.0	67.0	49.8	21.6	28.7	6.4	6.4	11.1	63.2	-143.0
	4	76-100	11000	1.4	8.7	38.8	11.5	104.0	43.3	53.7	20.8	25.5	5.7	6.5	10.0	45.0	-198.0
C2	5	0-25	21000	3.9	12.5	49.0	15.0	163.0	62.0	12.6	16.3	71.1	5.0	6.2	7.3	88.3	145.0
	6	26-50	22100	3.8	12.0	53.0	11.0	114.0	47.0	19.3	11.7	69.0	3.2	6.5	9.0	83.8	-85.0
	7	51-75	16000	2.7	10.4	44.0	17.0	61.0	41.0	41.2	15.0	43.9	2.4	6.6	9.3	75.0	-167.0
	8	76-100	10400	2.0	8.7	27.0	13.0	50.0	35.0	55.8	13.7	30.0	1.3	6.9	8.9	43.5	-245.0
K2	9	0-25	27000	1.9	17.4	71.2	24.0	173.0	77.7	2.3	47.8	49.9	1.0	7.2	14.0	81.3	12.0
	10	26-50	19000	2.8	13.7	53.4	19.5	171.0	52.4	11.4	31.7	56.9	4.5	7.6	13.7	90.2	-89.0
	11	51-75	20000	1.2	13.8	69.3	20.3	120.0	34.4	3.1	31.4	65.5	3.7	7.5	13.2	53.3	-260.0
	12	76-100	22000	1.4	15.4	52.4	22.2	107.0	27.9	2.0	38.6	59.4	1.1	6.9	15.2	80.5	-378.0
J3	13	0-25	10000	2.9	7.0	38.8	9.4	46.0	38.0	28.8	31.7	39.6	4.2	6.6	8.1	84.2	136.0
	14	26-50	12000	1.7	8.4	42.8	12.1	47.0	43.3	37.8	25.6	36.7	2.5	6.7	8.5	64.2	-190.0
	15	51-75	9000	1.9	6.3	35.1	8.8	38.0	34.6	36.3	20.3	43.3	7.4	6.5	9.0	76.1	-345.0
	16	76-100	7000	1.7	5.0	20.8	6.9	21.0	36.4	50.6	24.4	25.0	6.4	6.5	6.6	43.3	-411.0
J1	17	0-25	18600	2.6	11.9	43.0	11.0	97.0	50.0	15.0	15.0	70.0	5.0	6.5	10.1	89.6	36.0
	18	26-50	16400	2.3	11.3	46.0	11.0	89.2	57.0	49.1	35.1	15.8	6.0	6.1	7.1	48.7	-90.0
	19	51-75	14700	1.4	8.8	43.0	10.0	75.0	55.0	23.3	12.8	63.9	4.2	6.0	9.9	70.7	-245.0
	20	76-100	12800	1.5	10.0	44.0	11.0	69.0	48.0	49.0	13.0	62.0	2.1	6.3	6.3	50.4	-311.0
P1	21	0-25	22000	1.6	14.9	50.6	22.5	119.0	34.9	11.5	43.2	45.3	0.8	6.6	9.0	85.7	90.0
	22	26-50	11000	1.5	8.3	37.4	12.8	93.0	23.3	9.5	41.5	49.0	5.7	6.2	8.5	83.2	-61.0
	23	51-75	5000	0.0	4.1	22.3	7.7	29.0	3.1	43.4	25.2	31.4	2.0	6.4	6.1	46.1	-187.0
	24	76-100	1000	0.6	4.9	13.2	7.1	16.0	7.5	56.2	20.3	23.5	1.1	6.5	5.1	35.7	-254.0

In this study the factor analysis was used for extraction of the variables with the Varimax Kaiser normalization as the rotation method. The analysis reduced the variables by eliminating those that had Measures of Sampling Adequacy less than 0.4 and Measure of Communalities was less than 0.5 (Costello and Osborne, 2005). If any variable has a loading of more than 0.4 in more than one component, it means that there is existence of complex structure for that variable (Costello and Osborne, 2005).

3 Results and Discussion

3.1 Correlation analysis

The depth showed insignificant correlation with pH, % OM, % silt, and EC ($r=-0.010$ to -0.25, $p > 0.05$). On the other hand, redox potential and CEC showed significant negative correlation with depth ($r = -0.91$, $p < 0.01$) and ($r = -0.76$, $p < 0.01$) respectively. This is because the changes in pH, % OM, % silt, and EC with increasing depth were smaller than for redox potential and CEC. The correlation analysis also showed that % sand had a significant positive correlation with depth ($r = +0.61$, $p < 0.01$). These results imply that as depth increases from the surface layer (0 - 25 cm) to the bottom layer (76 - 100 cm) the redox potential, CEC, and % clay decreased while % sand content increased. Electrical conductivity (EC) and CEC were expected to negatively correlate with % sand ($r = -0.59$, $p < 0.01$ and $r = -0.80$, $p < 0.01$) and positively correlate with % clay ($r = +0.33$, $p = 0.12$ and $r = +0.72$, $p < 0.01$) respectively. The negative correlation between % sand and EC or CEC is attributed to less effective surface area available for cation adsorption and electric conductance. Results from a similar study elsewhere concur with the result of this work (James et al. 1982). The significant correlation between CEC and % clay implies that the CEC was higher where the % clay was also higher and vice versa. This is because clay has larger surface area on which metal ion exchanges occur. Thus, % clay in sediment is expected to contribute significantly to adsorption of heavy metals and CEC of the sediments which is in agreement with our previous observations. However, the insignificant correlation between EC and CEC ($r = +0.38$. $p = 0.07$) should not be puzzling since EC is a measure of all electrically charged ions (cations and anions) in sediment solution while CEC refers to cations that can be exchanged with H^+ ions on clay surfaces. It is not expected that all cations in the solution will be available for exchange due factors such as concentration, variable oxidation states of cations and presence of complexing agents. That is why a positive correlation between EC and CEC may not always exist.

Similarly sediment redox potential measures how easy the ions can be reduced or oxidized as defined by the Nernst equation. The Nernst equation shows that the redox potential (E) of sediments depends on the natural logarithm of the ratio of reduced species (gained electrons) to oxidized species (lost electrons) at a particular pH (Delaune and Reddy, 2005). Our analysis showed that the sediment redox potential had negative correlation with % sand ($r = -0.47$, $p < 0.05$) and depth ($r = -0.91$, $p < 0.01$) but redox potential showed positive correlation with CEC ($r = +0.66$, $p < 0.01$), Fe ($r = +0.43$, $p < 0.05$), Cd ($r = +0.58$, $p < 0.01$), Zn ($r = +0.53$, $p < 0.01$), and Pb ($r = +0.46$, $p < 0.05$). These results posed a serious challenge in their interpretation because our study measured the total concentrations (not speciation) of seven heavy metals out of which Zn and Cd are not transition metals (Mrutu et al., 2013). However, redox potential of the sediment is the contribution of all redox reactions that occur in the sediment; and the magnitude of redox potential depends on the stoichiometric ratio of the species reduced and species oxidized in the multi-component sediment matrix.

On the other hand, the positive correlation between redox potential and CEC was expected because CEC accounts for only those cations that are available for exchange or undergoing redox reaction. Organic matter is an assemblage of carbon containing materials that may have carboxylic acid, amine, esters, ketones and aldehydes functional groups (Peuravuori and Pihlaja, 2007; Kanokkantapong et al. 2006; Wershaw et al., 2005). At around neutral pH, it is only the carboxylic acid group may be deprotonated to carboxylate that may provide a site for metal ion coordination. However, the rapid dichromate oxidation method used to determine % OM is destructive (it oxidizes all carbons to CO_2) we could not draw any meaningful inference from its correlation with other parameters.

This work also found that % OM ranged from 1-5 % and a similar study elsewhere recorded only 1.98% (Tam and Wong, 1993). This suggests that many sediment samples have very low % OM and the amounts were too low to exhibit a significant correlation with other parameters. Therefore the retention of heavy metals in this study area not absolutely controlled by one factor rather it could be a combination of the two or more factors that makes metal retention in the sediment a complex phenomenon. Since some heavy metals may exist in sediments not only as ions bound to clays but also as complexes with organic matter/anions or oxides and hydroxides of iron (Tangwira et al., 1992).

Table 2 Pearson correlation (r) with p-values for the sixteen parameters measured in this study (n = 24).

		Depth	Fe	Cd	Co	Cr	Ni	Zn	Pb	% sand	% silt	% clay	% O.M	pH	EC	CEC	redox
Depth		1.00															
Fe	r	-0.53	1.00														
	p	0.007	0.00														
Cd	r	-0.59	0.49	1.00													
	p	0.003	0.02	0.00													
Co	r	-0.43	0.96	0.35	1.00												
	p	0.037	0.00	0.10	0.00												
Cr	r	-0.49	0.91	0.38	0.89	1.00											
	p	0.015	0.00	0.07	0.00	0.00											
Ni	r	-0.31	0.80	0.13	0.89	0.79	1.00										
	p	0.14	0.00	0.54	0.00	0.00	0.00										
Zn	r	-0.52	0.85	0.43	0.84	0.81	0.75	1.00									
	p	0.009	0.00	0.04	0.00	0.00	0.00	0.00									
Pb	r	-0.48	0.61	0.56	0.52	0.61	0.36	0.67	1.00								
	p	0.017	0.00	0.00	0.01	0.00	0.08	0.00	0.00								
% sand	r	0.61	-0.74	-0.34	-0.69	-0.75	-0.68	-0.69	-0.34	1.00							
	p	0.002	0.00	0.10	0.00	0.00	0.00	0.00	0.11	0.00							
% silt	r	-0.25	0.28	-0.27	0.39	0.32	0.51	0.26	-0.15	-0.34	1.00						
	p	0.24	0.19	0.21	0.06	0.13	0.01	0.23	0.50	0.10	0.00						
% clay	r	-0.41	0.59	0.48	0.49	0.60	0.39	0.55	0.44	-0.79	-0.27	1.00					
	p	0.04	0.00	0.02	0.01	0.00	0.06	0.01	0.03	0.00	0.21	0.00					
% OM	r	-0.15	-0.10	0.26	-0.22	-0.02	-0.27	0.13	0.41	-0.07	-0.27	0.18	1.00				
	p	0.48	0.65	0.22	0.29	0.93	0.21	0.56	0.05	0.74	0.20	0.40	0.00				
pH	r	-0.01	0.35	0.00	0.47	0.46	0.61	0.35	0.01	-0.35	0.37	0.11	-0.30	1.00			
	p	0.96	0.10	0.99	0.02	0.03	0.00	0.09	0.97	0.10	0.08	0.61	0.15	0.00			
EC	r	-0.07	0.66	0.00	0.71	0.69	0.76	0.62	0.30	-0.59	0.39	0.33	-0.05	0.67	1.00		
	p	0.75	0.00	0.99	0.00	0.00	0.00	0.00	0.15	0.00	0.06	0.12	0.82	0.00	0.00		
CEC	r	-0.76	0.63	0.65	0.52	0.56	0.47	0.60	0.49	-0.80	0.09	0.72	0.24	0.09	0.38	1.00	
	p	0.00	0.00	0.00	0.01	0.01	0.02	0.00	0.01	0.00	0.67	0.00	0.27	0.67	0.07	0.00	
Redox	r	-0.91	0.43	0.58	0.36	0.36	0.27	0.53	0.46	-0.47	0.12	0.36	0.13	-0.08	-0.04	0.66	1.00
	p	0.00	0.04	0.00	0.09	0.08	0.20	0.01	0.02	0.02	0.56	0.09	0.54	0.72	0.85	0.00	0.00

The data analysis has so far shown that physico-chemical parameters such as % sand, % clay, EC, CEC, redox potential have negative correlation with depth, and positive correlation with % clay (except EC). Since CE, CEC and redox potential are apparently related to the amounts of ions in the sediment matrices, it became imperative to find the correlation between these three parameters and heavy metal levels (Table 2). The results showed that CEC had positive correlation with all heavy metals; EC had insignificant positive correlation with only Cd and Pb; whereas redox potential had insignificant correlation with Co, Cr and Ni only (Table 2).

The correlation analysis between heavy metals can also be used to suggest the potential sources of heavy metals because strong positive correlation generally implies that the heavy metals might owe the same source or sink (Muzuka, 2007; Nyangababo et al., 2005; Singh et al., 2002). Table 2 shows that Fe, Cr, Co, Zn Ni, and Pb (except Pb vs. Ni) showed a significant positive correlation between two metals ($p < 0.01$). This suggests that these metals have a common source and the most probable source is anthropogenic. These metals were most likely drained to the mangrove wetland as wastewater in Msimbazi River flows through the wetland into the Indian Ocean. Empirical data available in literature show that water and sediments of the Msimbazi River are polluted with heavy metals and organic pollutants (Mrutu et al., 2013; Mwevura et al., 2002; DeWolf et al., 2001; DCC, 2004; Machiwa, 1992; A'khabuhaya and Lodsenius, 1988). Our previous study on the distribution of the heavy metals also supports this assertion because the concentration of these metals decreased gradually with increasing sediment depth (Mrutu et al., 2013). Although Cd showed weak positive correlation with Pb and Zn, it also showed insignificant correlation with Fe, Cr, Co, and Ni (Table 2). This finding suggests that Cd, Pb and Zn might be coming from a source different from those for Fe, Cr, Co and Ni or combination of both natural and anthropogenic sources. This also agrees with the results in our previous work in which we found that these metals did not show a sharp exponential decline in concentrations with increasing depth (Mrutu et al., 2013). The common factor analysis employed in the next paragraph is used to unveil the potential sources of the metals, clay and sands.

3.2 Factor analysis

After only three iterations, the rotation converged to two components with Eigenvalue greater than 1 (Table 3). The two components account for 74.9 % of the total variation. Component 1 has strong loading of pH, EC, % sand, Fe, Co, Cr, Ni, and Zn. High concentrations of heavy metals (Fe, Co, Cr, Ni, and Zn) contribute to high EC of the sediments at pH range of 6-8 reported in the previous study (Mrutu et al., 2013). High EC is associated with high metal ions and the pH determines the state of heavy metal ions. At low pH, there is less heavy metal sorption while at high pH sorption of heavy metal ions in sediment is increased (Christensen, 1989). Sand content in the sediments increases with increasing depth from the surface while the EC and heavy metals concentrations decreased (Mrutu et al., 2013). This trend suggests that formerly the area was occupied by sand but due to presence of mangroves fine sand particles, which is mostly quartz (SiO_2), settle down as the water flows through the mangroves.

Table 3 Factor loadings in component 1 and component 2.

Rotated Component Matrix

Latent variables	**Component 1**	**Component 2**
Fe	0.730	0.597
Cd	-0.032	0.821
Co	0.828	0.447
Cr	0.781	0.507
Ni	0.894	0.250
Zn	0.667	0.606
Pb	0.250	0.693
% sand	-0.616	-0.591
% clay	0.326	0.687
pH	0.775	-0.200
EC	0.913	0.034
CEC	0.286	0.829
Redox	-0.019	0.811

Component 2 has strong loading of redox potential, CEC, % clay, Pb, and Cd. This indicates that redox potential, CEC, and % clay are positively correlated and decreased with increasing depth. Redox potential in mangrove sediments is higher in the surface layer and decreases as depth increases from the sediment surface. This is because surface layer is exposed to air especially during low tides. The fact that CEC decreases with increasing depth from the surface can be explained by a decrease in clay with increasing depth from the surface (Appelo and Postma, 1994). These factors may have effect on the distribution of Pb and Cd which have the most stable oxidation states of +2. They may not easily gain electrons (reduction) to free metal or lose electrons (oxidation) to form a stable higher oxidation states at pH 6-8. These ions also may be exchanged on the surface of clay particles which have very high surface area. This suggests that clay significantly affect the distribution of Pb and Cd concentrations in the sediments by sorbing them (Appelo and Postma, 1994). Our previous work also showed that the concentrations of Pb and Cd were higher where there was high % clay and their amounts decreased with decrease in clay content as depth increased (Mrutu et al., 2013).

The factor analysis can also be used to select hidden structures and associations between heavy metals and factors influencing their distribution in the sediment cores (Hunchinson and Sofroniou, 1999). A further look at Table 3 indicates that Fe, Co, Cr, Zn, and % sand have loadings of more than 0.4 in component 1 and component 2. This implies that these variables have complex structures (Hunchinson and Sofroniou, 1999) and therefore Fe, Co, Cr, Zn, and % sand owe their origin to both natural and anthropogenic sources. The most important anthropogenic sources of Fe, Co, Cr, Zn and sand are surface runoffs from the industrial, municipal, and agricultural activities in the city while the natural sources will be the decomposition of leaf litters from mangrove plants and mineral rock formation of the coastal area. On the other hand Cd, Ni, Pb, and % clay had loadings greater than 0.4 only in one component (Table 3) and therefore these metals and clay owe their source mostly to anthropogenic sources (Hunchinson and Sofroniou, 1999). The decrease in concentrations of heavy metals (Fe, Cd, Co, Cr, Ni, and Zn) in sediment profile was also related to the decrease in fine sediment particles (Mrutu et al, 2013). The pH of 6-8 observed in this work favors free exchange of ions between fine clay particles and solutions which also enhance EC of sediments (Christensen, 1989). A pH above 8 will cause most of the heavy metals to precipitate in the solution while a pH below 6 will desorb most of the metal ions adsorbed on the negatively charged site in clay particles.

3.3 Cluster analysis

There were two major clusters shown in the dendogram each of which comprised 12 case numbers that stood for sediment samples analyzed (Fig. 1). The samples were clustered into four groups based on heavy metal concentration at different depths.

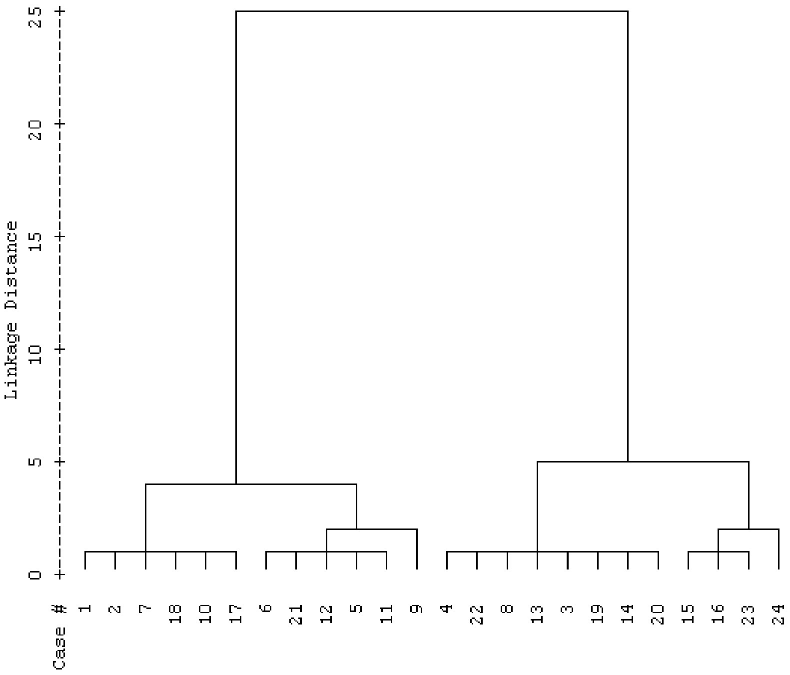

Fig. 1 The dendogram for Hierarchical cluster analysis of the 24 sediment samples.

The first major cluster (on the left) comprised of 12 sediment samples obtained from 0 - 25 cm and 26 - 50 cm depths. Of the 12 samples, nine (75 %) were obtained from the southern bank of the Msimbazi Creek and were divided into three sub-clusters. The first sub-cluster is composed of six samples with case numbers 1, 2, 7, 10, 17 and 18 which stood for sediment samples from 0 - 25 cm and 26 - 50 cm depths at site A1, 51 - 75 cm at site C2, 26 - 50 cm at site K2, 0 - 25 cm and 26 - 50 cm depths at J1 respectively. The first 50 cm cores of sediment recorded high levels of heavy metals due to high silt and clay content (Mrutu *et al*., 2013). Of the six samples in this sub-cluster five came from the 0 - 25 cm and 26 - 50 cores which were expect to have high levels of all parameters except % sand. These samples were characterized by the following features: pH (6.1 to 7.6), EC (7.00 to 13.7 mS/cm), CEC (48.7 to 98.2 meq/100 g), redox potential (-167 to 124 mV), OM (2.4 to 9.4 %), sand (1.00-49.1%), silt (13.7-35.1%) and clay (15.5-77.2%). These samples also had Fe (16,000 to 19,000 ppm), Cd (2.13 to 3.63 ppm), Co (10.4 to 13.7 ppm), Cr (43.0 to 53.4 ppm), Ni (11.0 to 19.5 ppm), Zn (61.0 to 171.0 ppm) and Pb (41.0 to 76.0 ppm).

The second sub-cluster composed of five sediment samples with case numbers 5, 6, 11, 12, and 21 which stood for samples from 0 - 25 cm core at C1, 26 - 50 cm core at C2, 0 - 25 cm core at K2, 51 - 75 cm core at K2, 76 - 100 cm depth at K2 and 0 - 25 cm depth at P1 respectively. These samples were characterized by these features: pH (6.2 to 7.5), EC (9.02 to 15.2 mS/cm), CEC (53.3 to 88.3 meq/100 g), redox potential (-378 to 145 mV), OM contents (0.84 to 4.97 %), sand (2.00 to 19.3 %), silt (11.7 to 43.2 %), clay (45.3 to 71.1 %). Heavy metals levels in these samples were: Fe (20,000 to 22,100 ppm), Cd (1.16 to 3.87 ppm), Co (12.0 - 15.4 ppm), Cr (49.0 to 69.3 ppm), Ni (11.0 to 22.5 ppm), Zn (107.0 to 163.0 ppm) and Pb (27.9 to 62.0 ppm). When this sub-cluster was compared to the first sub-cluster it was found that, on average, it had higher levels of Fe, Co and Cr retained in the first 50 cm cores than the first-sub-cluster. Although the two sub-clusters had similar % clay the second sub-cluster had also higher % silt than the first one. Therefore the observed difference in the heavy metals retained may mostly be attributed to variation in fine sediment texture (clay and silt). Lastly, the case number 9 which stood for sample obtained from site K2 at a depth of 0 - 25 cm. This sample had pH (7.2), EC (14.01 mS/cm), CEC (81.34 meq/100 g), redox potential (12 mV), OM content (1.01 %), sand (2.30 %), silt (47.8 %), clay (49.9 %). The heavy metal levels in this sample were: Fe (27,000 ppm), Cd (1.87 ppm), Co (17.40 ppm), Cr (71.2 ppm), Ni (24.0 ppm), Zn (173.0 ppm) and Pb (77.7 ppm). However, this sub-cluster had higher CEC, % silt, and % clay than the second sub cluster. In general we noted that all sample from 0 - 25 cm depth falling in this first major cluster had positive redox potential which implies that high concentration of reduced species (ionic products/gases) than oxidized species (or elements/ionic reactants) which favor spontaneous redox reaction.

The second major cluster in the dendogram (in the right) consisted of 12 sediment samples of which 75 % were from the northern bank of the Msimbazi creek and fell into three sub-clusters (Fig. 1). The eight samples in the first sub-cluster consisted of case numbers 3, 4, 8, 13, 14, 19, 20 and 22 which represented sediment samples from 51 - 75 cm and 76 - 100 cm cores at A1, 76 - 100 cm core at C2, 51 - 75 cm and 76 - 100 cm cores at J1, and 26 - 50 cm core at P1. These samples had pH (6.0 to 6.9), EC (8.11 to 11.1 mS/cm), redox potential (-311 to +136 mV), %OM content (2.05 to 6.37 %), CEC (43.5 to 84.2 meq/100 g), sand (9.50 to 55.8 %), silt (12.8 to 41.5 %) and clay (25.5 to 63.9 %). Heavy metals levels ranges were: Fe (10,000 to 14,700 ppm), Cd (1.11 to 2.92 ppm), Co (7.0 to 10.0 ppm), Cr (27.0 to 44.0 ppm), Ni (9.42 to 14.2 ppm), Zn (46.0 to 121.0 ppm) and Pb (23.3 to 67.0 ppm).

The second sub-cluster comprised three case numbers 15, 16, 23, and 24, which represented sediment samples from depths of 51 - 75 cm at J3, 76 - 100 cm at J3, 51 - 75 cm and 76-100 cm at P1 respectively. These samples were characterized by pH (6.4 to 6.5), EC (6.05 to 9.00 mS/cm), CEC (43.3 to 76.1 meq/100 g), redox potential (-411 to -187 mV), OM content (2.01 to 7.38 %), sand (36.3 to 50.6 %), silt (20.3 to 25.2 %) and clay (25.0 to 43.3 %). The ranges of heavy metal levels in these samples were: Fe (5,000 to 9,000 ppm), Cd (0.03 to 1.94 ppm), Co (4.09 to 6.30 ppm), Cr (20.8 to 35.1 ppm), Ni (6.91 to 8.79 ppm), Zn (21.0 to 38.0 ppm) and Pb (3.13 to 36.4 ppm). In comparison, the second sub-cluster had lower values for most of the metals measured than the first sub-cluster. This could be attributed to the fact that the percentage of sand was relatively higher in the second sub-cluster than in the first sub-cluster whereas the reverse was true for % silt and % clay. The case number 24, which stood for sediment sample from sampling site P1 at depth of 76-100 cm, was featured with pH (6.5), EC (5.11 mS/cm), CEC (35.69 meq/100 g), redox potential (-254 mV), OM content (1.11 %), sand (56.2 %), silt (20.3 %) and clay (23.5 %). The heavy metals levels in this sub cluster were: Fe (1,000 ppm), Cd (0.57 ppm), Co (4.91 ppm), Cr (13.2 ppm), Ni (7.14 ppm), Zn (16.0 ppm) and Pb (7.48 ppm). The sampling site P1 had the lowest metal levels, % silt and % clay, and the highest % sand of all sediment samples.

The dendogram has successfully classified the samples based on the depth and metal levels. The first major cluster (on the left) consisted of mainly samples from the 0-25 cm and 26-50 cm cores (the first 50 cm layer). Our data showed that this layer had the highest levels of fine particles (% clay and % silt) which have the largest s effective surface area for adsorption or exchange of ions. It also had the highest heavy metal contents because metal ions can easily be bound to the fine particles depend on oxidation states of the ions. It is also expected that redox reaction will occur most frequently in the first layer than the second layer because the first layer is expected to have higher oxygen and metal species to undergo redox reactions. That is why the redox potential in this cluster was positive in the 0-25 cm cores (+12 to +90 mV) and became negative in the 26-50 cm cores (-160 to -14 mV). On the contrary, samples in the second cluster of the dendogram (on the right) consisted of samples mostly from 56-75 cm and 76-100 cm (the second 50 cm layer). This layer was expected to have low dissolved oxygen and heavy metals. It is also expected that less redox reaction will take place in this layer and the cluster analysis revealed that the redox potential in this cluster was mostly negative (-411 to -61 mV) except in 0-25 cm core at J3 where the redox potential was +136 mV.

4 Conclusions

This study demonstrated that the multivariate analysis of sediment data using SPSS is a powerful tool in revealing hidden relationships between parameters. Heavy metals generally showed positive correlation with other heavy metals and that relationship suggests that the metals have mostly anthropogenic origin. We successfully classified the variables using cluster analysis into two clusters that differ in heavy metal content. The heavy metals were higher in the first 50 cm depth than the second 50 cm depth and these results agree with results from our previous study where we found that % clay and heavy metals decreased with increasing depth. Moreover, factor analysis was used to identify potential sources of heavy metals, sand and clay in the sediments. The statistic revealed that Fe, Cr, Co, Zn) and % sand have natural and anthropogenic origins because the variables had loadings of more than 0.4 in both components. On the other hand, the statistic indicated that Cd, Ni, Pb, and % clay had loadings of more than 0.4 only in one component and therefore they are thought to come from only anthropogenic sources. However, the mechanism of metal-clay or metal-organic matter interaction is not known. Therefore, it is recommended that further studies are required to confirm the sources of these materials using a more powerful tool, isotope tracing technique. This technique can not only trace the environmental source of a pollutant but also can provide the contribution of anthropogenic source and natural sources of the materials.

Acknowledgement

We would like to acknowledge the generosity of the Dutch Government Cooperation in Higher Education (MHO) for financing the study. We also extend our appreciation to the University of New Mexico that provided us with free access to the library facilities and SPSS software used for the analysis of the data in this work.

References

Appelo CAJ, Postma D. 1994. Geochemistry, Groundwater and Pollution (2nd edition). A.A. Balkema Publishers, Rotterdam, The Netherlands

Christensen TH. 1989. Cadmium soil Sorption at low concentrations. VII. Correlation with soil parameters. Water, Air, and Soil Pollution, 44: 71-82

Costello AB, Osborne J. 2005. Best practices in exploratory factor analysis: four recommendations for getting the most from your analysis. Practical Assessment Research & Evaluation, 10(7). http://pareonline.net/getvn.asp?v=10&n=7

Cynthia L, Fianko JR, Akiti TT, et al. 2011. Determination of trace elements in the Sakumo wetland sediments. Research Journal of Environmental and Earth Sciences, 3(4): 417-421

Davis JC. 1986. Statistics and Data Analysis in Geology. John Wiley & Sons, New York, USA

DCC (Dar es Salaam City Council). 2004. Dar es Salaam City Profile. Dar es Salaam City Council, United Republic of Tanzania

De Wolf H, Ulomi SA, Backeljau T, et al. 2001. Heavy metal level in the sediments of four Dar es Salaam mangroves. Accumulation in, and effects on the morphology of the Periwinkles, *Littorina scabra* (Mollusca, Gastropoda). Environmental international, 26: 243-249

Delaune RD, Reddy KR. 2005. Redox Potential. In: Encyclopedia of Soils in the Environment (Hillel D, ed). 366-371, Academic Press, USA

Dwivedi SN, Padmakumar KG. 1983. Ecology of Mangrove Swamp Near Juhu Beach, Bombay with Reference to Sewage Pollution. In. Biology and Ecology of Mangroves. Task for vegetation Science Series (Teas HJ, ed). 163-170, Junk Publishers, Lancaster, UK

Harbaugh JW, Demirmen F. 1964. Application to Factor Analysis to Petrologic Variations of Americus Limestone (Lower Permian), Kansas and Oklahoma. Kansas Geological Survey, Special Distribution Publications, 15

Hutcheson GD, Sofroniou N. 1999. The Multivariate Social Scientist: Introductory Statistics Using Generalized Linear Models (1st edition). SAGE Publications Ltd, USA

James CW, Hanks RJ, Jurinak JH. 1982. Modern Irrigated Soils. John Wiley and Sons, New York, USA

Kanokkantapong V, Marhaba TF, Panyapinyophol B, Pavasant P. 2006. FTIR evaluation of functional groups involved in the formation of haloacetic acids during the chlorination of raw water. Journal Hazardous Materials, 136: 188-196

Machiwa JF. 1992. Heavy metal content in coastal sediments of Dar es Salaam, Tanzania. Environment International, 18: 409 - 415

Mgana S, Mahono S. 1997. Land Based Sources and Activities Affecting the Quality and Uses of Marine, Coastal and Associated Freshwater Environment: Tanzania Mainland. Report Presented at the UNEP Regional Workshop Held at the Institute of Marine Sciences, Zanzibar, Tanzania

Mohammed SA. 2002. A Review of water quality and pollution studies in Tanzania. AMBIO, 31(7): 617-620.

Mremi SD, Machiwa JF. 2003. Heavy metal contamination of mangrove sediments and the associated biota in

Dar es Salaam, Tanzania. Tanzania Journal of Science, 29(1): 61-76

Mrutu A, Nkotagu HH, Luilo GB. 2013. Spatial distribution of heavy metals in Msimbazi River mangrove sediments in Dar es Salaam coastal zone, Tanzania. International Journal Environmental Sciences, 3(5): 1641-1655

Mtanga A, Machiwa J. 2007. Assessment of heavy metal pollution in sediment and polychaete worms from the Mzinga Creek and Ras Dege mangrove ecosystems, Dar es Salaam, Tanzania. Western Indian Ocean Journal of Marine Science, 6(2): 125–135

Muzuka ANN. 2007. Distribution of heavy metals in the coastal marine surficial sediments in the Msasani Bay-Dar es Salaam harbour area. Western Indian Ocean Journal of Marine Science, 6(1): 73–83

Mwevura H, Othman OC, Mhehe GL. 2002. Organochlorine pesticideresidues in sediments and biota from the coastal area of Dar es Salaam city, Tanzania. Marine Pollution Bulletin, 45(1-2): 262-267

Nyangababo JT, Henry I, Omutunge E. 2005. Heavy metal contamination in plants, sediments and air precipitation of Katonga, Simiyu and Nyando wetlands of Lake Victoria Basin, East Africa. Bulletin of Environmental Contamination and Toxicology, 75(1): 189-196

Parks JM, 1966. Cluster analysis applied to multivariate geological problems. Journal of Geology, 74(5): 703-715

Peuravuori J, Pihlaja K. 2007. Characterization of Freshwater Humic Matter. In: Handbook of Water Analysis (2nd edition) (Nollet LML, ed). 435-447, CRC Press, Boca Raton, FL, USA

Ra K, Kim E-S, Kim K-Y, Kim J-Y, Lee J-M, et al. 2013. Assessment of heavy metal contamination and its ecological risk in the surface sediments along the coast of Korea. Journal of Coastal Research, 65: 105-110

Singh M, Muller G, Singh IB. 2002. Heavy metals in freshly deposited stream sediments of rivers associated with urbanization of the Ganga plain, India. Water Air and Soil Pollution, 141(1-4): 35-54

Sparks DL. 1996. Methods of Soil Analysis Part 3: Chemical Methods. Soil Science Society of America Book Series No 5. American society of Agronomy-Soil Science Society of America, Madison, Wisconsin, USA

Tagwira F, Oloya T, Nleya GG. 1992. Copper status and distribution in the major Zimbabwean soils. Communications in Soil Science and Plant Analysis, 23(7-8): 659-671

Tam NFY, Wong YS. 1993. Retention of Nutrients and Heavy Metals in Mangrove Sediment Receiving Waste Water of Different Strengths. Environmental Technology, 14: 719-729

Wershaw RL, Leenheer JA, Cox L. 2005. Characterization of Dissolved and Particulate Natural Organic Matter (NOM) in Neversink Reservoir, New York. Scientific Investigations Report 2005-5108, U.S. Geological Survey, USA

Zhang WJ. 2011. Constructing ecological interaction networks by correlation analysis: hints from community sampling. Network Biology, 1(2): 81-98

Zhang WJ. 2012. How to construct the statistic network? An association network of herbaceous plants constructed from field sampling. Network Biology, 2(2): 57-68