

A Chemometric Study on the Accumulation of Heavy Metals Along the Mogi Guaçu River Basin

Maria R. C. da Silva^a, Káthia M. Honório^a, Janete Brigante^b, Evaldo L.G. Espíndola^b,
Eny M. Vieira^a, Maria T. P. Gambardella^a and Albérico B. F. da Silva^{*,a}

^a Instituto de Química de São Carlos, Universidade de São Paulo, CP 780, 13560-970 São Carlos - SP, Brazil

^b Departamento de Hidráulica e Saneamento, Escola de Engenharia de São Carlos, Universidade de São Paulo, São Carlos - SP, Brazil

Um estudo sobre os impactos ambientais causados por atividades agrícolas e industriais, ao longo da bacia do rio Mogi Guaçu, foi realizado neste trabalho. A concentração de alguns metais (Fe, Cd, Cu, Cr, Zn, Ni, Pb e Mn) foi determinada em quinze estações de coleta de sedimentos e em dois períodos diferentes: chuvoso (Março/2001) e seco (Junho/2001). Os métodos de reconhecimento de padrões PCA e HCA foram utilizados para agrupar as amostras coletadas em São Paulo e em Minas Gerais e obter informações relevantes sobre as possíveis fontes de contaminação do rio Mogi Guaçu de acordo com o período de coleta das amostras. As análises quimiométricas realizadas, utilizando as amostras coletadas em Março/2001 e em Junho/2001, indicaram que somente a concentração de Cd, Cu, Cr e Pb discrimina as amostras coletadas em São Paulo das amostras coletadas em Minas Gerais. A influência do período de coleta também foi analisada e os resultados obtidos indicaram que a concentração de quatro metais (Cd, Cr, Zn e Mn) foi importante para discriminar as amostras coletadas em Março/2001 das amostras coletadas em Junho/2001. De forma geral, a concentração dos metais estudados se apresenta em altos níveis em todas as estações de coleta e este fato caracteriza um ambiente com alto grau de contaminação causada, principalmente, por atividades antropogênicas.

A study on the environmental impacts caused by agricultural and industrial activities along the Mogi Guaçu river basin was carried out. Concentrations of some metals (Fe, Cd, Cu, Cr, Zn, Ni, Pb and Mn) were determined for fifteen collection stations in two different seasons: rainy (March/2001) and dry (June/2001). Pattern recognition methods (Principal Component Analysis and Hierarchical Cluster Analysis) were used to group samples collected in São Paulo and Minas Gerais and obtain relevant information on possible polluting sources of the Mogi Guaçu river according to the collection season of the samples. A chemometric analyses performed using samples collected in March/2001 and June/2001 indicated that only the concentrations of Cd, Cu, Cr and Pb discriminate samples collected in São Paulo from those collected in Minas Gerais. The influence of the collection season was also analysed and the results indicated that four metals (Cd, Cr, Zn and Mn) were important to discriminate samples collected in March/2001 (rainy season) from those collected in June/2001 (dry season). We verified that the concentrations of the metals studied are at high levels in all collection stations and this fact characterizes an environment with a high degree of contamination caused mainly by anthropogenic activities.

Keywords: Mogi Guaçu river, sediments, metal concentration, principal component analysis, hierarchical cluster analysis

Introduction

Aquatic ecosystems cover about 3/4 of the earth surface. Most of this water (about 97%) is salty and it is found in the oceans. The evaporated water forms clouds in the atmosphere and it precipitates like rain in the terrestrial

surface constituting fresh water. About 77% of the fresh water in the world is found in the polar ice caps and icebergs and 23% is underground. Only a small percentage of this water is found in rivers and lakes.¹ However, rivers and lakes represent only a small proportion of fresh water on earth (0.004% and 0.33% respectively). Fresh water has a fundamental role in civilization development (drinking water source, way of transportation, electric power source

* e-mail: alberico@iqsc.usp.br



Figure 1. Location of the Mogi Guaçu river basin in São Paulo and Minas Gerais states and the indication of the collection stations.

etc.). With the increase in human population, the natural landscape was being modified, for example: forests were deforested, agricultural monocultures arose, land was removed for flood prevention, coastal areas were landed and big drainage systems were built. Besides, great water volumes are being used by industries, public government and agricultural activities. The results of these activities have an impact on the physical, chemical and biological structure of the rivers, lakes and dams and, consequently, there is a decrease in the useful potential of these systems.²

In this work, a study on the environmental effects caused by man in the Mogi Guaçu river basin (Figure 1) is carried out with the goal of providing subsidies for the treatment of contaminated areas by heavy metals such as lead, cadmium and mercury. Our main purpose here is to find out some relationships between the concentration of some metals and the kind of activity performed by man along the Mogi Guaçu river basin in the São Paulo and Minas Gerais regions.

The Mogi Guaçu river (Big Snake, in tupi-guarani dialect) has its nascent located in Bom Repouso (Minas Gerais - MG) with average altitude of 1650m. In São Paulo, the Mogi Guaçu river goes through strongly industrialized and urbanized regions as well as large extensions of

farmable lands.³ The environmental study of the Mogi Guaçu river basin in São Paulo can be facilitated by analysing the main polluting sources of each one of the locations that are related, direct or indirectly, to the collection stations adopted in this work. Some agricultural and industrial activities along the Mogi Guaçu river are summarized in Table 1. The more expressive agricultural activity along the São Paulo route of the Mogi Guaçu is sugar cane. In Bom Repouso (Minas Gerais), 368 nascent of the Mogi Guaçu river are found and the predominant crops in this area are potato and strawberries. This kind of cultures, allied to economic aspects, has been advancing over the region without any preoccupation with the environmental impacts.

Table 1. Main polluting agricultural and industrial activities of the Mogi Guaçu river in São Paulo (SP) and Minas Gerais (MG) states

Activity	Municipal district (State)
Non-metallic minerals	Mogi Guaçu (SP)
Metallurgic	Mogi Guaçu (SP); Itapira (SP)
Cellulose and paper	Mogi Guaçu (SP); Itapira (SP)
Wood and furniture	Itapira (SP)
Textile and footwear	Mogi Guaçu (SP)
Leather factory	Luiz Antônio (SP); Mogi Guaçu (SP)
Sugar factory	Pradópolis (SP); Itapira (SP)
Potato and strawberry fields	Bom Repouso (MG)

The Mogi Guaçu river basin has some industries with polluting potential, as presented in Table 1. Metals, eliminated by these industries, can accumulate in lakes and rivers (in sediments) changing the environmental conditions. The chemical action of heavy metals has attracted great environmental interest due to the fact these metals do not present biodegradable character. This fact determines the presence of heavy metals in the global biogeochemical cycles and these metals can remain accumulated in aquatic life at considerably high levels.

The heavy metal term is widely recognized and used for a group of metals that are associated with pollution and toxicity such as lead, cadmium, mercury, arsenic and uranium. In this group (heavy metals) are also included other metals but they are considered biologically essential (in low concentration), for example: copper, manganese, selenium and zinc.⁴ The term "heavy metal" does not necessarily imply "toxic metal", as many metals are considered indispensable nutrients to plants and human beings at low concentration. Several metallic ions are essential for live organisms and other ions (such as sodium, calcium, potassium, manganese, iron, cobalt, molybdenum, copper and zinc) have fundamental importance for man.

However, the ingestion of metals is considered toxic in spite of some metals be essential for live organism.

The presence of heavy metals in the aquatic environment, in high concentration, causes mortality of fish and other animals. In soils, heavy metals can be retained by different mechanisms, mostly when the soil is rich in organic matter and has a pH greater than seven (what reduces the leaching of the soil and the entrance of metals into the water). In this kind of environment, the heavy metals drastically reduce the fertility and the development of plants.^{5,6} Also, the presence of heavy metals in the human organism, from the food chain, can provoke countless diseases due to the cumulative effect and can cause death.⁷

Actually, our main purpose in this work is to find out some relationships between the concentration of heavy metals and the kind of activity performed by man along the Mogi Guaçu river in the São Paulo and Minas Gerais regions. The discrimination performed by chemometric methods becomes important since different activities are developed in these two regions (industrial activities in São Paulo and agricultural activities in Minas Gerais) and based on the concentration of the metals studied our results can be used to assess the origin (São Paulo or Minas Gerais) of new collected samples.

Material and Methods

Experimental data set

The choice of the collection locations took into account the use and occupation of the soil (agricultural, industrial and extracting activities, urban agglomerates and permanent preservation areas). Five regions of the Mogi Guaçu river were selected and 15 collection points were established (which can be identified in Figure 1).

For each region studied, a number of collection stations (enough to allow a characterization of the alterations in the biogeochemical processes along the Mogi Guaçu river) was established based on the degree of human interference. Regarding the sampling frequency, two collections were accomplished during March/2001 (rainy season) and June/2001 (dry season).

The sediment collection for analysis was made by using collectors of "core and Eckamn's dredge".⁸ After collection, the samples were submitted to a pretreatment, *i.e.*, a procedure of wet digestion was carried out. Afterwards, the material was analysed using the atomic absorption technique (Hitachi model 28100) obtaining the concentration of the metals studied. The metals selected for this work were: iron (Fe), cadmium (Cd), copper (Cu), chromium (Cr), zinc (Zn), nickel (Ni), lead (Pb) and

manganese (Mn). The concentration of each metal studied was obtained in triplicate, and the mean value of them is presented in Table 2.

Chemometric analysis

In this work we employed exploratory data techniques for obtaining relevant information about the data set. The main goal of these techniques is to reduce the data set and obtain possible relationships between the variables (concentration of the metals) and the kind of activity performed along the Mogi Guaçu river. The two exploratory data techniques used in this work were: Principal Component Analysis (PCA) and Hierarchical Cluster Analysis (HCA).

Principal Component Analysis (PCA) is a multivariate statistical technique and its central idea is grouping correlated variables, generating a new set of variables called Principal Components (PCs), onto which the data set is projected. These PCs are built as linear combination of the original variables and have the important property of being completely uncorrelated. In PCA analysis, the position of the samples in the new coordinate system (PCs) is represented by the score matrix, while the loadings matrix gives the importance (weight) of the original variables on the PCs. The first new axis, PC1, is chosen in such a direction that maximizes the variance along that axis. The second axis must be chosen orthogonal to the first one and in the direction to describe as much variance left as possible and so on.^{9,10}

Hierarchical Cluster Analysis (HCA) examines the distances between the samples in a data set and this information is represented as a two-dimensional plot called dendrogram. The HCA method is an excellent tool for preliminary data analysis and it is useful to analyse data sets for expected or unexpected clusters, including the presence of outliers. It is advisable to analyse the dendrogram in conjunction with PCA, since they give complementary information in different forms. In HCA each point forms, initially, an only cluster and then the similarity matrix is analyzed. The most similar points are grouped forming one cluster and the process is repeated until all the points belong to an only cluster.¹¹ In this work we used the incremental technique for grouping the samples studied.¹²

For the chemometric analysis in this work, the variables (metal concentration) were autoscaled so that they could be compared to each other in the same scale. In the autoscaling, each element of the data matrix is mean-centered and scaled to a variance of one. The exploratory data analyses (PCA and HCA) were performed by using the computational package Pirouette 2.0.¹³ In our PCA

Table 2. Metal concentration (mg kg⁻¹) in the sediment samples collected in March and June/2001

Location	Fe	Cd	Cu	Cr	Zn	Ni	Pb	Mn
March 2001								
MP ₁	40,500	3.0	32.0	62.5	38.4	23.1	34.2	707.3
MP ₂	78,500	1.3	18.0	68.5	100.0	29.8	26.3	633.9
MP ₃	26,570	3.7	49.0	92.5	54.8	39.3	50.2	801.3
MP ₄	22,480	2.75	20.0	43.0	15.6	19.8	33.0	113.5
MP ₅	35,250	3.3	20.0	42.5	17.4	21.0	36.7	120.3
MP ₆	28,000	4.0	14.5	68.0	57.0	29.7	49.5	767.3
MP ₇	14,500	3.75	16.0	46.5	25.4	20.8	38.6	142.4
MP ₈	29,000	4.0	6.0	86.5	39.1	33.5	53.0	344.5
MP ₉	37,500	4.5	10.0	77.0	50.0	33.7	54.0	260.4
MP ₁₀	46,500	4.0	7.5	75.0	53.5	36.9	53.3	278.2
MP ₁₁	10,100	5.0	16.0	68.5	50.0	36.4	63.1	137.8
MP ₁₂	20,750	4.6	10.0	29.0	50.0	40.0	73.3	332.3
MP ₁₃	17,500	4.1	10.0	35.0	57.9	21.7	45.0	705.0
MP ₁₄	90,250	3.4	21.5	75.0	12.6	18.1	32.0	838.0
MP ₁₅	39,000	4.7	13.0	71.5	40.8	38.0	73.0	332.3
June 2001								
JP ₁	59,750	1.2	8.5	51.5	50.0	32.0	21.7	130.0
JP ₂	45,750	3.3	18.5	13.5	48.5	24.5	42.6	140.0
JP ₃	12,870	1.8	15.0	79.0	57.5	35.0	44.5	337.3
JP ₄	10,000	1.9	9.5	49.5	50.0	21.05	42.2	260.0
JP ₅	10,000	2.0	9.0	74.0	62.5	34.0	39.7	162.5
JP ₆	44,070	2.0	6.5	65.0	62.5	31.8	39.0	135.0
JP ₇	15,870	1.5	5.5	32.0	50.0	26.5	27.0	250.0
JP ₈	32,500	2.65	38.0	20.5	50.0	24.1	50.5	510.0
JP ₉	17,000	1.95	25.0	28.0	50.0	28.5	35.8	500.0
JP ₁₀	97,000	1.75	35.0	34.5	50.0	22.3	32.4	215.0
JP ₁₁	93,000	2.85	5.0	44.5	46.7	22.2	52.0	825.0
JP ₁₂	31,000	3.3	4.0	19.5	54.6	34.6	54.2	750.0
JP ₁₃	30,500	2.25	10.0	37.5	75.0	28.1	39.7	350.0
JP ₁₄	83,320	2.3	6.0	27.5	75.0	27.7	40.5	750.0
JP ₁₅	41,500	2.7	25.0	42.5	50.0	31.8	52.2	450.0

and HCA analyses, we defined the samples collected in March/2001 as MP_i and the samples collected in June/2001 as JP_i. The index *i* corresponds to areas where the samples were collected (from 1 to 15 in Figure 1).

Results and Discussion

Initially, the results obtained for each collection season (March and June/2001) were analysed individually and, subsequently, the two seasons were compared with the goal to find the relationships between the collection seasons and the metal distribution.

March 2001

From the data set presented in Table 2, we carried out the PCA analyses and the results obtained show that only four metals (Cd, Cu, Cr and Pb) are important to discriminate samples from São Paulo and Minas Gerais regions. The first three principal components describe

96.25% of the overall variance. The first PC explains 51.29% of the variance and the second PC describes 29.86%. These two PCs correspond to 81.15% of the overall variance, *i.e.*, PC1 and PC2 give most of the information on the systems studied.

The main result obtained with PCA is showed by the score plot (the spatial distribution of the samples in the new coordinate system - PC1 x PC2) displayed in Figure 2. The loading values of the most important variables in PC1 and PC2 are expressed by the equations:

$$PC1 = 0.644 \text{ Cd} - 0.410 \text{ Cu} - 0.052 \text{ Cr} + 0.644 \text{ Pb} \quad (1)$$

$$PC2 = 0.183 \text{ Cd} + 0.509 \text{ Cu} + 0.815 \text{ Cr} + 0.197 \text{ Pb} \quad (2)$$

From Figure 2 we can see that PC1 discriminates samples from São Paulo (labeled MP1 to MP8 in Figure 2) and Minas Gerais (labeled MP9 to MP15 in Figure 2). Analysing Figure 2 and equation (1) one can see that samples collected in São Paulo present negative values

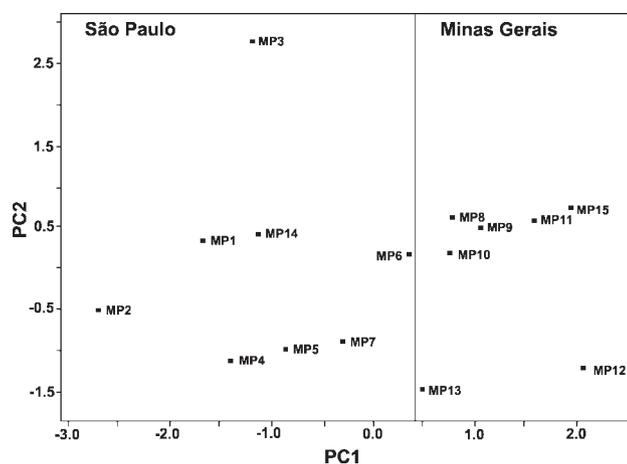


Figure 2. Score plot for samples collected in March/2001.

for PC1, *i.e.*, on average these samples have higher concentrations of Cu than the ones from Minas Gerais (see Table 3). According to the characteristics presented by these samples, we can conclude that the high levels of Cr can be attributed to the effluents originated from sugar, leather and paper factories, which are the main activities present at São Paulo route. The samples collected in Minas Gerais have positive values for PC1 and this means high concentrations of Cd and Pb. This indicates indiscriminate use of pesticides and fertilizers on the potato and strawberry fields in Minas Gerais region. Also, it is interesting to notice that the sample MP3 present a different behavior when compared with the other São Paulo samples, since it presents a very high concentration of Cr due to the presence of a sugar factory near this collection station (MP3).

The two exceptions in the PCA classification are the samples MP8 and MP14. The sample MP8 was collected in São Paulo region, but it was considered as a sample from Minas Gerais region. This occurred due to two factors: (i) the proximity of the sample MP8 to Minas Gerais region, and (ii) the collection season (rainy), for which the transport of solid particles, nutrients and agricultural

chemicals (pesticides and fertilizers) occurs, from this collection station, characterizing this sample as a sample from Minas Gerais region. The sample MP14 was collected in the Minas Gerais region, but it was considered as a sample from São Paulo region. This can be explained due to the agricultural activities developed near this collection station (MP14), leading to the increase of Cu (originated from pesticides used on the potato and strawberry fields in Minas Gerais region).

After the PCA analysis, we carried out the HCA analysis with the goal of grouping the similar samples. Figure 3 illustrates the dendrogram obtained with the HCA analysis. The vertical lines in Figure 3 represent the samples studied and the horizontal lines the similarity values between pairs of samples, a sample and a group of samples and sample groups. From Figure 3 we can see that two clusters are formed: (i) the first cluster is mainly formed by the samples collected

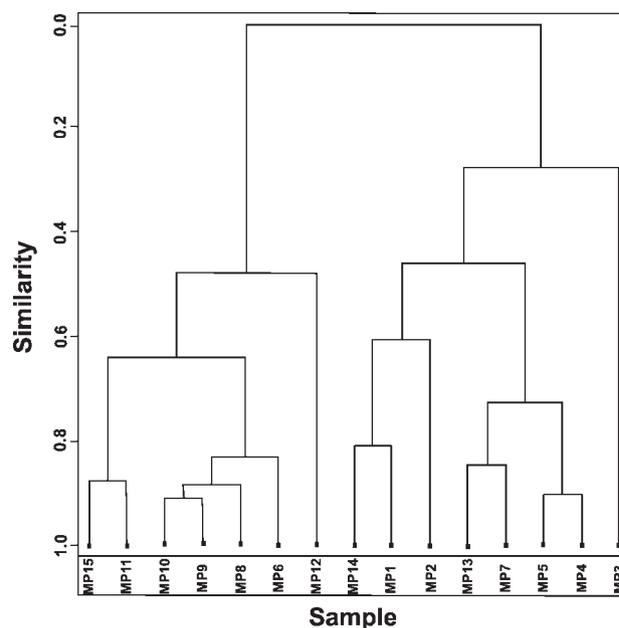


Figure 3. Dendrogram obtained for samples collected in March/2001.

Table 3. Average values of metal concentrations (mg kg^{-1}) and standard deviations (SD) for the samples collected in March and June/2001

Region	Fe	Cd	Cu	Cr	Zn	Ni	Pb	Mn
March 2001								
São Paulo	34,350.00	3.225	21.937	63.750	43.462	27.125	40.187	453.812
SD (SP)	19,465.16	0.902	13.100	19.146	27.606	7.071	9.608	305.118
Minas Gerais	37,371.43	4.328	12.571	61.571	44.971	32.114	56.242	412.000
SD (MG)	26,761.35	0.535	4.782	20.462	15.172	8.617	14.981	256.937
June 2001								
São Paulo	28,851.25	2.043	13.812	48.125	53.875	28.618	38.400	240.600
SD (SP)	19,346.01	0.659	10.697	24.328	5.987	5.215	9.461	132.015
Minas Gerais	56,188.57	2.442	15.714	33.428	57.328	27.885	43.828	548.571
SD (MG)	33,671.53	0.540	12.406	8.956	12.289	4.553	8.828	231.061

in São Paulo (MP1, MP2, MP3, MP4, MP5 and MP7), and (ii) the second group is mainly formed by the samples collected in Minas Gerais (MP9, MP10, MP11, MP12 and MP15). Two samples collected in Minas Gerais region (MP13 and MP14) were incorrectly grouped with the samples from São Paulo region: the sample MP13 has similar concentration of Pb to the concentration presented by the samples from São Paulo (low concentration of Pb due to a small use of pesticides), and the sample MP14 presented similar levels of Cd, Cu and Pb when compared with the samples from São Paulo, *i.e.*, low concentrations of Cd and Pb and high level of Cu. Also, two samples collected in São Paulo region (MP6 and MP8) were incorrectly grouped as samples from Minas Gerais region. In the collection region of the MP6 sample there are many sugar cane crops and, consequently, indiscriminate use of pesticides, characterizing this sample as originated from Minas Gerais.¹⁴ The sample MP8 was classified as a sample from Minas Gerais due to its localization, since it is localized in the border region between São Paulo and Minas Gerais states.

In fact, when we compare Figures 2 and 3 we can also see the anomalies detected by PCA and HCA and commented in the previous two paragraphs, *i.e.*, the two samples (MP8 and MP14) that were classified incorrectly by PCA were also classified incorrectly by HCA, and the two samples (MP6 and MP13) that were classified incorrectly by HCA were classified correctly by PCA.

The discrimination obtained by PCA and HCA methods is important as different activities are developed in these two regions (industrial activities in São Paulo region and agricultural activities in Minas Gerais). Thus, based on the concentrations of Cd, Cu, Cr and Pb, our PCA and HCA results could be used to assess the origin (São Paulo or Minas Gerais) of new collected samples.

June 2001

From the data set obtained for June/2001 samples (see Table 2), we also carried out PCA and HCA analyses. The PCA results obtained are similar to the ones for samples of March/2001, *i.e.*, just the concentration of four metals (Cd, Cu, Cr and Pb) was important to group the São Paulo and Minas Gerais samples. The PCA results showed that the first three principal components describe 98.27% of the overall variance. The first PC explains 52.24% of the variance and the second PC describes 27.84%. These two PCs (PC1 and PC2) correspond to 80.08% of the overall variance.

Figure 4 shows the score plot (PC1 x PC2) and the loading values for the most important variables (Cd, Cu, Cr and Pb) are represented by equations (3) and (4):

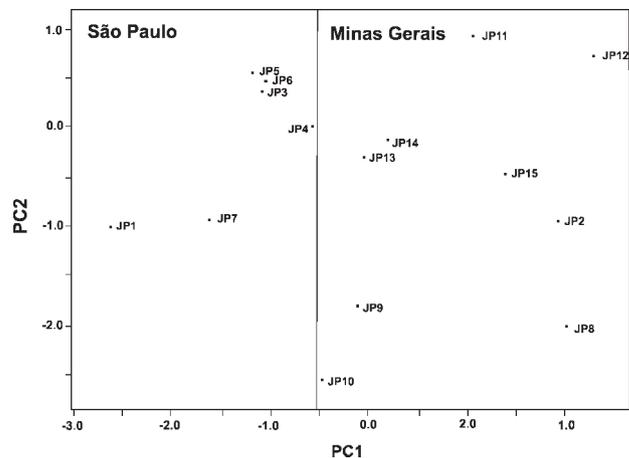


Figure 4. Score plot for samples collected in June/2001.

$$PC1 = 0.657 \text{ Cd} + 0.209 \text{ Cu} - 0.443 \text{ Cr} + 0.573 \text{ Pb} \quad (3)$$

$$PC2 = 0.225 \text{ Cd} - 0.762 \text{ Cu} + 0.471 \text{ Cr} + 0.383 \text{ Pb} \quad (4)$$

From Figure 4 we can see that samples collected in São Paulo region in the dry season (labeled JP1 to JP8 in Figure 4) have negative values for PC1, while samples collected in Minas Gerais in the dry season (labeled JP9 to JP15 in Figure 4) have positive values for PC1.

From Figure 4 and equation (3) we can see that samples collected in São Paulo present high concentration of Cr due to its high negative coefficient in the loading equation (see equation (3)). The high concentration of Cr in these collection stations (JP1 to JP8) is explained by the presence of several highly polluting activities such as sugar, leather and metallurgy factories, which contribute to the increase of the Cr concentration in these samples.

The samples collected in Minas Gerais have positive values for PC1 and from equation (3) we can see that the metals Cd, Cu and Pb positively contribute for PC1, *i.e.*, these samples have high concentrations of Cd, Cu and Pb. The contamination of these samples are mainly originated from pesticides and fertilizers used on the potato and strawberry fields which are the main agricultural activities in these regions. For example, the presence of high levels of Cu in Minas Gerais samples is the result of the use of large amount of fungicide on the potato and strawberry fields in Minas Gerais, since copper sulphate is largely used as fungicide.

The exceptions in this classification are the samples JP2 and JP8. The sample JP2 was incorrectly classified since it was collected in São Paulo, but it was considered as a sample from Minas Gerais. The region where the sample JP2 was collected is characterized by sugar cane crops (and a large amount of pesticides is used in these crops) which have

high concentrations of Cd, Cu and Pb and, therefore, the sample JP2 (from São Paulo) was classified as a sample from Minas Gerais. The sample JP8 was also classified as a sample from Minas Gerais due to its localization, *i.e.*, this collection station (JP8) is localized in the border region between São Paulo and Minas Gerais states.

The HCA results obtained for the samples collected in June/2001 are summarized in Figure 5. From Figure 5 we can see two different groups: (i) samples collected in São Paulo (JP1 and JP3 to JP7) and two samples collected in Minas Gerais (JP13 and JP14). These two samples (from Minas Gerais) were incorrectly grouped to the cluster containing the São Paulo samples due to their low concentration of Pb which is similar to the samples from São Paulo. This occurs because the JP13 and JP14 samples were collected in nascent regions and, consequently, they are not contaminated with pesticides that contain Pb; (ii) samples collected in Minas Gerais (JP9 to JP12 and JP15) and two samples collected in São Paulo (JP2 and JP8). The region where the sample JP2 was collected presents sugar cane crops and, consequently, pesticides are used in this region and this characterizes the sample JP2 as from Minas Gerais. The sample JP8 was also classified as a sample from Minas Gerais due to its localization, *i.e.*, this collection point (JP8) is located in the border region between São Paulo and Minas Gerais states.

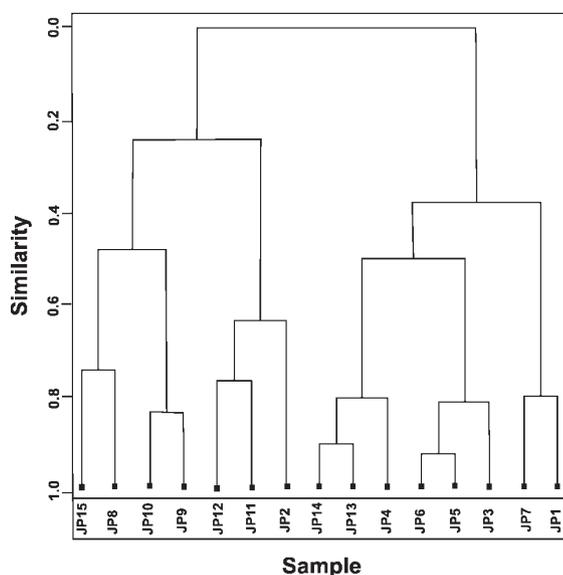


Figure 5. Dendrogram obtained for samples collected in June/2001.

Two seasons (March and June/2001)

The aim of analysing the March and June/2001 seasons together is to find out some relationship between the concentration of the metals (arising from agricultural or

industrial activities) and the sample collection season (rainy and dry).

From the data set presented in Table 2 we carried out PCA and HCA analyses. The PCA results showed that four metals (Cd, Cr, Zn and Mn) were important to discriminate the samples from March/2001 and June/2001. The first three principal components describe 86.96% of the overall variance. The first PC explains 35.71% of the variance while the second PC describes 29.04%. These two PCs correspond to 64.75% of the overall variance. The score plot obtained is presented in Figure 6 and the loading values are presented in equations (5) and (6).

$$PC1 = 0.680 \text{ Cd} + 0.373 \text{ Cr} - 0.619 \text{ Zn} - 0.123 \text{ Mn} \quad (5)$$

$$PC2 = 0.233 \text{ Cd} + 0.485 \text{ Cr} + 0.401 \text{ Zn} + 0.742 \text{ Mn} \quad (6)$$

From Figure 6 we can see the good discrimination between the samples collected in March/2001 and June/2001. Also from Figure 6 it is possible to see that the samples collected in March/2001 (rainy season) have positive values for PC1 while the samples collected in June/2001 (dry season) have negative values for PC1.

According to Figure 6 and equation (5), the samples collected in the rainy season (March/2001) present high concentrations of Cd and Cr due to their positive coefficient in the loading equation (see equation (5)). The high concentrations of Cd and Cr in these collection stations (MP1 to MP15 in Figure 6), generally originated from industrial activities, are a result of the rainy season which is responsible for the homogeneous concentration of the metals in all collection stations, since the effects caused by industrial activities (carried out in São Paulo) are extended to the other collection stations from Minas Gerais.

The samples collected in June/2001 (dry season) present negative values of PC1 and this means high concentrations

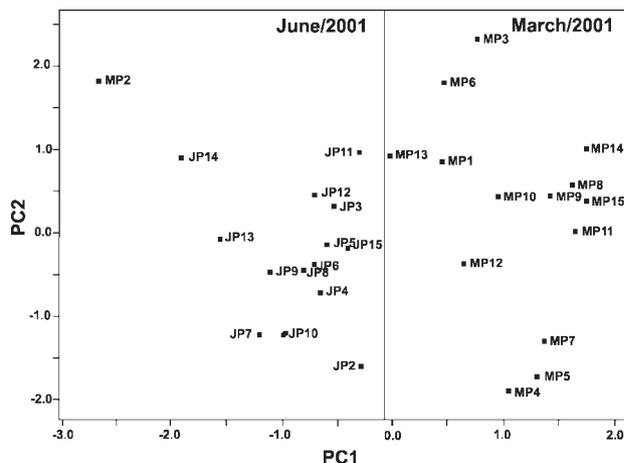


Figure 6. Score plot for samples collected in March and June/2001.

of Zn and Mn (see Figure 6 and equation (5)). We can attribute the high concentration of Zn to the indiscriminate use of pesticides and fertilizers in all samples collected, and the high concentration of Mn is originated from the soil composition and some chemical products used in several industrial processes have this metal (Mn) in their composition. The only exception is the sample MP2 which was collected in March/2001 (rainy season), but it was classified as a sample collected in June/2001 (dry season). This occurs due to the presence of sugar cane crop in this region which is responsible for the high concentration of Zn, indicating the use of pesticides in this crop (sugar cane). Therefore, we can see that it is possible to characterize a sample collected in the rainy season (March/2001) using the concentrations of Cd and Cr (originated from industrial activities) while samples collected in June/2001 (dry season) can be characterized by the concentrations of Zn (originated from pesticides and fertilizers) and Mn (from soil composition and industrial processes), indicating a high level of contamination in these regions caused by anthropogenic activities.

The HCA analysis was used to group the samples collected in March and June/2001 and analyse the similarity among the variables studied (metal concentration). Figure 7 shows the dendrogram obtained for the variables and we can see that there are two distinct clusters: a group with high similarity among Mn, Fe and Cu and another cluster formed by Pb, Cd, Ni, Cr and Zn. An explanation on this grouping needs to take into account that metals originated from anthropogenic contamination are associated with organic matter present in the thin fraction of the sediments, adsorbed on Fe/Mn hydrous oxides or precipitated as

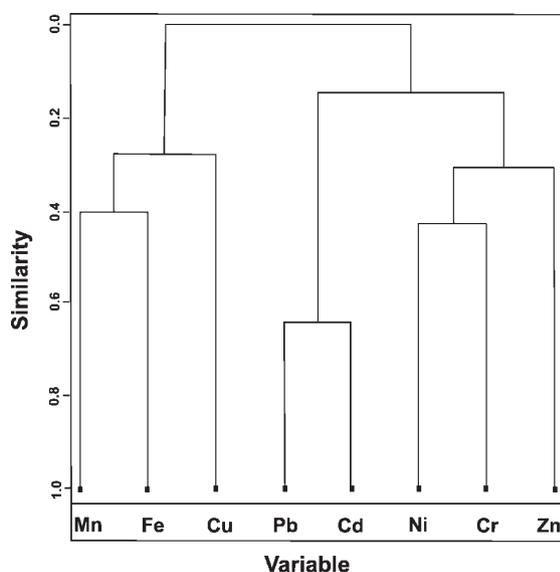


Figure 7. Dendrogram obtained for the variables (metal concentration) used in this work.

hydroxides, sulphides and carbonates.¹⁵ So, we can group these metals according to their similar characteristics, *i.e.*, the bioavailability of Mn and Fe in the water depends on their redox potential and the adsorption of Cu in the sediments varies according to the presence of iron and manganese oxides. The second group (Pb, Cd, Ni, Cr and Zn) is characterized by the metal adsorption in the sediments, *i.e.*, the presence of thin fractions such as clay helps the adsorption of the metals in the sediments.

Figure 8 illustrates the dendrogram obtained for the samples collected in March and June/2001. From Figure 8 we can see that one cluster is formed by samples collected in the rainy season (Group A, samples MP1 to MP15) and the second cluster (Group B) is formed by the samples collected in the dry season (samples JP1 to JP15). The sample JP2 was incorrectly grouped in Group A (samples collected in March/2001) due to its high concentration of Cd in the dry season (which is similar to samples collected in March/2001). The samples MP1, MP2 and MP13 were also incorrectly grouped in Group B (samples collected in June/2001) due to the similar concentrations of Cd, Cu and Pb when compared to Cd, Cu and Pb concentrations presented in the samples collected in June/2001. Based on these results, we can conclude that only the concentration of four metals (Cd, Cr, Zn and Mn) can be used to group samples from São Paulo and Minas Gerais regions with a high level of reliability.

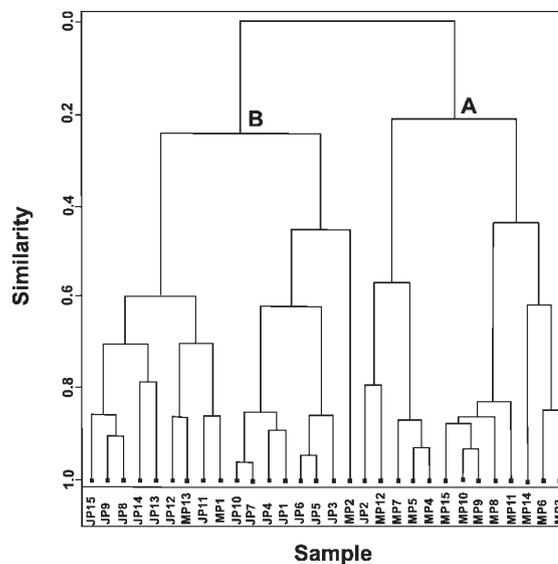


Figure 8. Dendrogram obtained for samples collected in March and June/2001.

Conclusions

A multivariate data analysis was performed in this work (PCA and HCA). The aim was to discriminate samples collected from São Paulo and Minas Gerais regions and

obtain information on the main contributions for the environmental impacts along the Mogi Guaçu river basin.

From the PCA results, for samples collected in March/2001, we could see that PC1 can discriminate samples from São Paulo and Minas Gerais only based on the concentration of four metals (Cd, Cu, Cr and Pb), *i.e.*, the samples collected in São Paulo have high concentrations of Cu, which can be attributed to the effluents originated from sugar, leather and paper factories besides the pesticides applied on the sugar cane. Samples collected in Minas Gerais have high concentrations of Cd and Pb and this indicates indiscriminate use of pesticides and fertilizers on the potato and strawberry fields in Minas Gerais region. The HCA analysis showed similar results to those obtained with PCA, *i.e.*, samples collected in São Paulo region were correctly grouped in one cluster and another cluster was formed by samples collected in Minas Gerais.

For samples collected in June/2001, we verified a good discrimination between samples from São Paulo and Minas Gerais according to the concentration of only four metals (Cd, Cu, Cr and Pb). The PCA results indicated that samples collected in São Paulo, in the dry season, present high concentration of Cr due to the presence of several highly polluting activities such as sugar, leather and metallurgy factories which contribute to the increase of the concentration of Cr in these samples. Samples collected in Minas Gerais have high concentrations of Cd, Cu and Pb indicating that these metals are mainly originated from pesticides and fertilizers used on the potato and strawberry fields (the main agricultural activities performed in these collection stations). The HCA results also indicated a good grouping between the samples from São Paulo and Minas Gerais only using four metals (Cd, Cu, Cr and Pb) in the analysis. Therefore, the discrimination performed by PCA and HCA is important, as different activities are developed in these two regions (industrial activities in São Paulo and agricultural activities in Minas Gerais) and based on the concentrations of Cd, Cu, Cr and Pb our PCA and HCA results can be used to assess the origin (São Paulo or Minas Gerais) of new collected samples.

The influence of the collection season (rainy and dry) was also analysed with the aim to determine the main contamination sources in these two seasons. The PCA and HCA results indicated that it is possible to characterize a sample collected in the rainy season (March/2001) using the concentrations of Cd and Cr (originated from industrial activities). The samples collected in June/2001 (dry season) can be characterized by the concentration of Zn (originated from pesticides and fertilizers) and Mn (from soil composition and industrial processes), indicating a high

level of contamination in these samples caused by anthropogenic activities.

The results obtained in this work by using the PCA and HCA methods also showed that chemometric methods can be useful to identify the location of a sample collected in the two regions studied in this work (São Paulo and Minas Gerais) as well as to obtain information on the kind of polluting sources (agricultural and industrial activities).

Acknowledgments

The authors would like to thank CAPES, CNPq, FAPESP and FINEP (Brazilian agencies) for the financial support.

References

1. Salbu, B.; Steinnes, E.; *Trace Elements in Natural Waters*, CRC Press: Boca Raton, 1995.
2. Lloyd, R.; *Pollution and Freshwater Fish*, Blackwell: New York, 1992.
3. CETESB, *Relatório de Qualidade das Águas Interiores do Estado de São Paulo*, Editora CETESB: São Paulo, 2003, (http://www.cetesb.sp.gov.br/Agua/agua_geral.asp), Accessed in March, 2004.
4. Russel, J. B.; *Química Geral*, McGraw-Hill: São Paulo, 1994.
5. Lacerda, D. L.; Solomons, W.; *Mercury from Gold and Silver Mining: A Chemical Time Bomb*, Springer-Verlag: Berlin, 1998.
6. Jardim, W. F.; *Cienc. Hoje* **1988**, 7, 78.
7. Chapman, O. M.; Allen, H. E.; Godfredsen, K.; Zraggen; M. N.; *Environ. Sci. Technol.* **1996**, 30, 448.
8. Mudroch, A.; McKnight, S. D.; *CRC Handbook of Techniques for Aquatic Sediments Sampling*, CRC Press: Boca Raton, 1991.
9. Jolliffe, I. T.; *Principal Component Analysis*, Springer-Verlag: New York, 1986.
10. Dillon, W. R.; Goldstein, M.; *Multivariate Analysis*, John Wiley & Sons: New York, 1984.
11. Kowalski, B. R.; Bender, C. F.; *J. Am. Chem. Soc.* **1972**, 94, 5632.
12. Everitt, B. S.; *Cluster Analysis*, Arnold: London, 1993.
13. *Pirouette Multivariate Data Analysis for IBM PC Systems*, Version 2.0, Infometrix: Seattle, WA, 1996.
14. Brigante, J.; Espíndola, E. L. G.; *Limnologia Fluvial: Um Estudo no Rio Mogi-Guaçu*, Rima: São Carlos, 2003.
15. Soares, H. M. V. M.; Boaventura, R. A. R.; Machado, A. A. S. C.; Esteves da Silva, J. C. G.; *Environ. Pollut.* **1999**, 105, 311.

Received: November 19, 2004

Published on the web: August 30, 2005

FAPESP helped in meeting the publication costs of this article.