STATISTICAL FACTOR ANALYSIS OF HEAVY METAL POLLUTION IN THE CAÍ RIVER BASIN SEDIMENTS (BRAZIL)

Maria Lucia Kolowski Rodrigues, Departamento de Laboratórios, FEPAM, Porto Alegre, RS, Brazil,
Maria Tereza Raya-Rodríguez, Departamento de Ecologia, UFRGS, Porto Alegre, RS, Brazil

ABSTRACT

This study tested the use of factor analysis to obtain quality indexes representing the grade of heavy metals pollution in the sediments of Caí River basin (Brazil). The information was summarized in three principal components, which were associated to different potential sources - natural contribution of basalts; metallurgical industries and tanneries. A chromatic classification of the indexes allowed to locate the sampling stations on the basin map, according to the sediment pollution status.

RÉSUMÉ

Cette étude montre l'usage de l'analyse factorielle pour déterminer les indices de qualité représentatifs de la pollution par les métaux lourds dans les sédiments du bassin de la Rivière Caí (Brésil). Les informations se résumées en trois principaux facteurs, qui sont associés aux différentes sources - contribution naturelle des basaltes; industrie métallurgique et tanneries. Pour présenter les résultats, les indices obtenus ont été classifiés avec différentes couleurs et les stations d'échantillonnage ont été représentés en fonction de la pollution des sédiments sur la carte du bassin.

1. INTRODUCTION

The analysis of multivariate environmental data, resulting from heavy metals monitoring in sediment samples, requires statistical methodologies that enable the recognition of tendencies and the preparation of reliable diagnoses. The characterization of fluvial sediments by the assessment of multiple variables allows the detection of correlations, positive or negative, that can indicate potential sources of contribution, natural or anthropic.

Factor analysis is a multivariate statistical method often used to prepare quality indexes. In this technique, the initial set of variables is substituted by a smaller group of factors or hypothetic variables, which preserve as much information contained in the original variables as possible (Fachel 1976). Factor analysis identifies and quantifies basic standards of variation in the data set, and allows the construction of an index that accounts for the variability with a smaller, simpler number of vectors than those obtained using the original data (Lohani and Mustapha 1982). The basic stages of factor analysis involve preparing the correlations matrix between variables, the extraction of factors and a possible rotation, seeking a final solution with simpler factors that are easier to interpret. At the beginning of the process, all the variables are equally important. The degree and significance of the correlation of each variable with the others will determine its importance in the final composition of the quality index and its capacity to express the environmental situation. Applying this technique, it is possible to evaluate the existence of associations between the variables, to determine the degree of relevance of each one in the environment diagnosis, and to compare the sampling units (spatial and temporal measures resulting from data collection), over time and space, detecting the existence of trends (Fachel 1976).

Preparing a quality index involves the combination of a group of correlated variables on a common scale, obtaining a single numerical value that can synthetically express the environmental quality of each sampling unit. This group should contain the most significant variables of the database, in a manner such that the resulting indexes can describe the whole situation and reflect changes in a sensitive, representative manner. Some information may be lost in the simplification process but, for adequately projected indexes, this loss is minimal and such that it does not cause distortion or misinterpretation of results (SDD 1976, Kishi and Lucca 1991).

Various examples of applying factor analysis to obtain water quality indexes are found in the literature (Shoji et al. 1966, Lohani and Mustapha 1982, Haase et al. 1989, Haase and Possoli 1993). The final product of the work is almost always a map showing the water quality, and representing spatial and seasonal variations (SDD 1976). Some studies recommend testing the use of factor analysis to obtain quality indexes for other environmental compartments, such as sediments, that are long term indicators.

Some care must be taken as regards the limitations of the technique, when the results are interpreted. The indexes are specific to the data set from which they were calculated, and it is not possible to perform a comparison with factor scores obtained using other databases. Since they summarize information, they are generic, and cannot be applied to decisions that require more detailed, precise knowledge. Although imperfect, the indexes represent a simple, understandable means of communication between technicians, public at large, and government authorities, about the environment quality and sites that have been more heavily polluted (Kishi and Lucca 1991).

2. METHODOLOGY

2.1 Study Area

The Caí River, located in the Rio Grande do Sul State in south Brazil, belongs to the Guaíba Hydrographic Basin. It extends for 260 km and is geographically characterized by the transition between the Brazilian Plateau (1000 m altitude) and the Central Depression, almost at the sea level. The total catchment area of the Caí Basin is about 5000 km^2, and the local geology consists mainly of Serra Geral Formation rocks, in which basalts that are naturally rich in metallic elements predominate.

On the banks of the Caí River there are major urban and industrial concentrations. A major part of the total pollution load discharged into this river comes from the industrial activity, the most important being petrochemicals, metals, tanneries, food and textiles, which are potential sources of heavy metals. The region has also concentrated agricultural activities, such as viticulture, vegetable and fruit production, in which fungicides and inputs containing these elements are applied (FEPAM/GTZ 1997).

2.2 Collection and Analysis of Sediment Samples

As described by Rodrigues (1997), between July 1994 and July 1995, four sampling expeditions were performed at seven stations along the Caí River and four at the mouth of important tributaries. In the Caí River, the sampling stations were identified by the letters CA followed by the distance in kilometers from the mouth. For comparison, sediment was collected at a site above the areas impacted by domestic and industrial effluents.

The selected elements cadmium (Cd), copper (Cu), lead (Pb), nickel (Ni), chromium (Cr), vanadium (V), zinc (Zn), aluminum (Al), iron (Fe), manganese (Mn) and titanium (Ti) were analyzed in the silt-clay fraction, on dry weight, using inductively coupled plasma optical emission spectrometry (ICP/OES). To extract the total metals content, the samples were digested with nitric and hydrofluoridric acids, in PTFE closed vessels, in an oven at 150°C for 4 hours, and then dried on a hot plate. For potentially mobile metals, a partial extraction, involving pre-digestion with hydrogen peroxide and hot extraction with HCl 0.3 M, was performed (Malo 1977).

The samples duplicates were processed with ultrapurified water and p.a. grade reagents, previously tested as to metals content. In order to evaluate the accuracy of the results, the reference material CANMET STSD-4 was analyzed concurrently, with recoveries between 95 and 105%, except for chromium (75%). Total aluminum was analyzed by X-ray fluorescence only in the third sampling expedition. Organic matter was also determined, the samples being calcinated at 450°C until reaching a constant weight. Grain size was analyzed by sieving and sedimentation techniques.

2.3 Statistical Analysis of Data

In order to obtain quality indexes that would be representative of each sampling unit, a data base with 22 columns (variables) and 44 lines (sampling units) was set up. Total Al was ignored, in order to avoid gaps in the data matrix. On the purpose to rank the sampling units as to affinity to the group of studied variables, the data were processed with the software Statistica for Windows 4.3TM, by Factor Analysis, using the Principal Components option. Initially a Pearson correlations matrix was determined, defining the relevant universe of the variables. Considering this matrix, the possibility of reducing the data was explored, by building a set of new variables, the common factors, based on the interrelationships of the original data (Haase and Possoli 1993).

The correlation of variables with the extracted factors resulted in the factor loadings matrix, in which the first component tends to be more general, representing the most important common part of the variables analyzed. This factor is the best summary of the linear relationship exhibited by the data. The second principal component is independent of the first one (orthogonal), considering only the residual variance not included in the first factor, and so successively for the other axes. In this study it was decided to retain three factors for interpretation, accounting for approximately 67% of the total variance. Since the configuration of the factor structure is not unique, an attempt at rotation was made, with a view to obtaining simpler factors that could be interpreted and identified with the nature of the observed variables (Fachel 1976). In this case, rotation did not improve the interpretation of factors, and the solution without rotation was adopted. Based on the factor loadings matrix, the factor coefficients (weight of each variable in the index composition) were determined and used to calculate the factor scores or quality indexes (QI), given by the Equation 1:

$$QI = c_1 z_1 + c_2 z_2 + \dots + c_n z_n \qquad [1]$$

QI = quality index (factor score) obtained for the sample unit;
c_n = factor coefficient for the variable *n*
z_n = standardized variable

All the variables participate in the index composition, although some have lower coefficients, since they are less important to describe the environment to be studied. The variable that has the highest correlation with the others is most relevant to the final index.

Since the quality index has a distribution with a mean equal to zero and variance equal to 1, it is expected that 99% of the values calculated will be between -3 and +3. The extreme values, on the one hand, represent poor environmental conditions, while on the other a better quality occurs.

In order to improve the visualization of factor analysis results, bi- and three-dimensional dispersion graphs were prepared, ranking the sampling units in a comparison space, which displays spatial and temporal variations of the three obtained quality indexes.

For the purpose of obtaining a simplified instrument to disseminate the results, the sampling stations were mapped according to the degree of sediments pollution. Initially, for each one of the three components, the sampling units with a score below the inferior quartile received a weight of 3, which corresponded to a worse environmental quality. Indexes above the superior quartile were considered as having better environmental quality, and received a weight of 1. Values between inferior and superior quartile were classified as being of intermediate quality, with multiplication factor of 2. Next, the number of times each sampling station was classified in a given category was looked at, multiplying the value obtained by the weight attributed and adding up the three results. According to this criterion, and considering the existence of four results for each sampling site, the smallest number of points possible in each local, concerning one component, was equal to 4 (the best environmental quality). The maximum number, equivalent to the worst environmental quality, was 12. Considering this 8-point range of variation, three class intervals were arbitrated :

a) $4 \leq x < 6$ units, corresponding to categories of better quality, and represented by the colour white;

b) $6 \leq x \leq 10$ units, equivalent to an intermediary quality, which received a dot as symbol;

c) $10 < x \leq 12$, equivalent to worse sediment quality and represented by the black colour.

As a final result of the score, it was possible to classify each sampling station according to the chromatic scale, which could be represented on a map summing up the results of monitoring.

3. RESULTS AND DISCUSSION

The analyses results are summarized in Table 1. Table 2 presents the factor loadings matrix, where one sees the decomposition of the information into three principal components, characterized by different variables, which may be associated with potential sources of contribution.

On the first axis (Figs. 1 and 2), which expressed 30.6% of the information contained in the data, the variables total and partial Fe, total Ni, total and partial Mn, total Cr, total and partial Ti, total and partial V were better correlated. Most of them can be associated with the basaltic rocks that predominate in the Caí Basin, and exert the greatest influence on soil and sediment formation. This connection suggested the contribution of surface runoff and erosive processes that are most important in Forromeco and Cadeia Streams flood plain, in the area of the plateau slopes.

On the second axis (Fig. 2), that accounted for 28.3% of the total variance, the variables partial Cd, total and partial Cu, total and partial Pb, partial Ni, total and partial Zn were the most important. The good correlation of these parameters indicated a potential contribution of Caxias do Sul industrial complex, where metallurgy, electroplating and textiles predominate. Another possible contribution would be from

Table 1. Heavy Metals and Organic Matter Contents (Dry Weight, Fraction < 63 μm) in the Sediments from Caí River Basin (RS, Brazil), Average of 11 Sampling Locations in the Period from July 1994 to July 1995.

Variable	n	Mean	P/T (%)	Minimum	Maximum
Cd T (mg/kg)	44	1.10	44.3	0.445	1.73
Cd P (mg/kg)	44	0.487		< 0.220	0.811
Cu T (mg/kg)	44	102	47.3	48.1	158
Cu P (mg/kg)	44	48.2		16.0	82.5
Pb T (mg/kg)	44	34.8	58.3	21.2	49.2
Pb P (mg/kg)	44	20.3		11.9	36.0
Fe T (mg/kg)	44	53457	30.6	2659	72276
Fe P (mg/kg)	44	16333		9946	23355
Ni T (mg/kg)	44	50.2	27.3	7.68	71.2
Ni P (mg/kg)	44	13.7		1.28	26.8
Zn T (mg/kg)	44	136	61.5	67.4	352
Zn P (mg/kg)	44	83.7		24.2	321
Mn T (mg/kg)	44	1080	88.0	432	1704
Mn P (mg/kg)	44	950		389	1479
Cr T (mg/kg)	44	117	46.8	14.1	305
Cr P (mg/kg)	44	54.8		2,31	223
Al T (mg/kg)	11	89155	8.3	84300	106800
Al P (mg/kg)	44	7415		4966	9633
Ti T (mg/kg)	44	10429	0.8	8196	14071
Ti P (mg/kg)	44	78.9		32.1	142
V T (mg/kg)	44	230	21.6	101	356
V P (mg/kg)	44	49.6		19.2	78.9
Organic Matter (%)	44	14.8		10.7	18.5

Note: T=total content; P=potentially mobile content

Table 2. Factor Loadings Matrix.

Variable	Factor 1	Factor 2	Factor 3
Cd T	0.366	-0.428	0.028
Cd P	0.140	-0.526	-0.323
Cu T	0.390	-0.836	0.107
Cu P	0.267	-0.863	0.028
Pb T	-0.084	-0.618	-0.265
Pb P	-0.168	-0.847	-0.169
Fe T	0.702	-0.030	0.446
Fe P	0.650	0.386	-0.169
Ni T	0.898	-0.269	0.146
Ni P	0.385	-0.849	-0.058
Zn T	0.038	-0.880	0.064
Zn P	-0.028	-0.920	-0.110
Mn T	0.809	0.072	0.128
Mn P	0.801	0.163	0.104
Cr T	0.749	0.089	-0.488
Cr P	0.527	0.087	-0.718
Al P	0.271	0.099	-0.327
Ti T	0.691	0.311	0.246
Ti P	0.538	0.389	-0.159
V T	0.877	0.167	0.219
V P	0.749	0.291	-0.225
Organic Matter	-0.370	0.481	-0.419
Explained variance	6.728	6.216	1.733
Total proportion of variance (%)	30.6	28.3	7.9

Note: T=total content; P=potentially mobile content

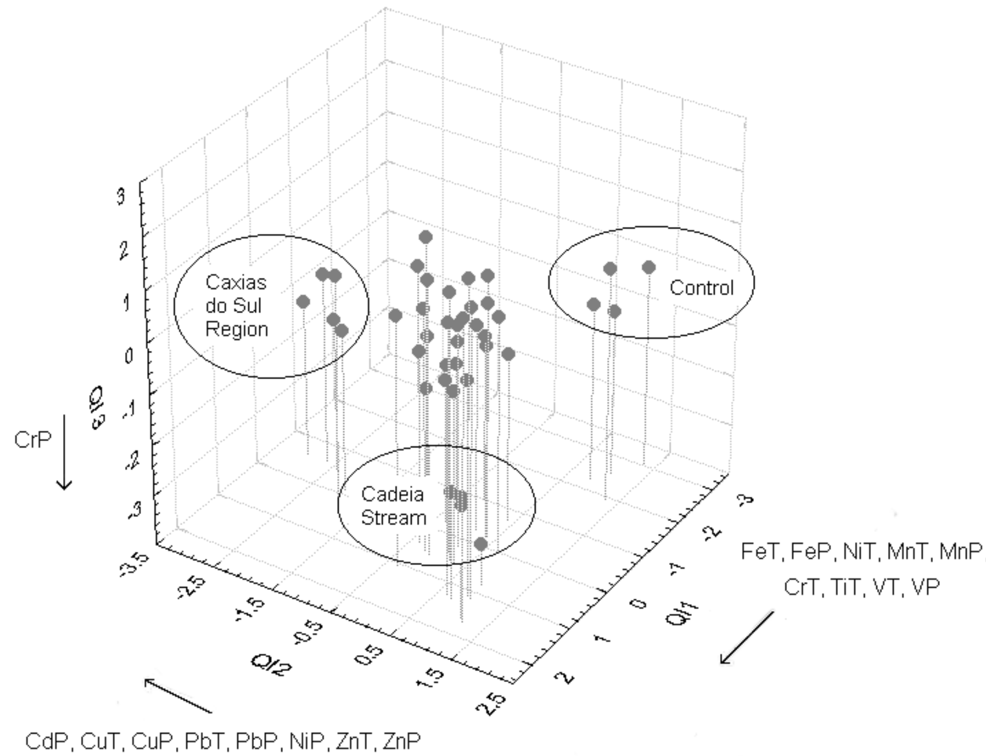

Figure 1. Dispersion of the sampling units according to the factor scores (quality indexes) on the three principal components. The variables most correlated with the components are indicated (T = total metal content; P = partial or potentially mobile content). The arrow direction indicates ascending concentrations of the elements in the sediments.

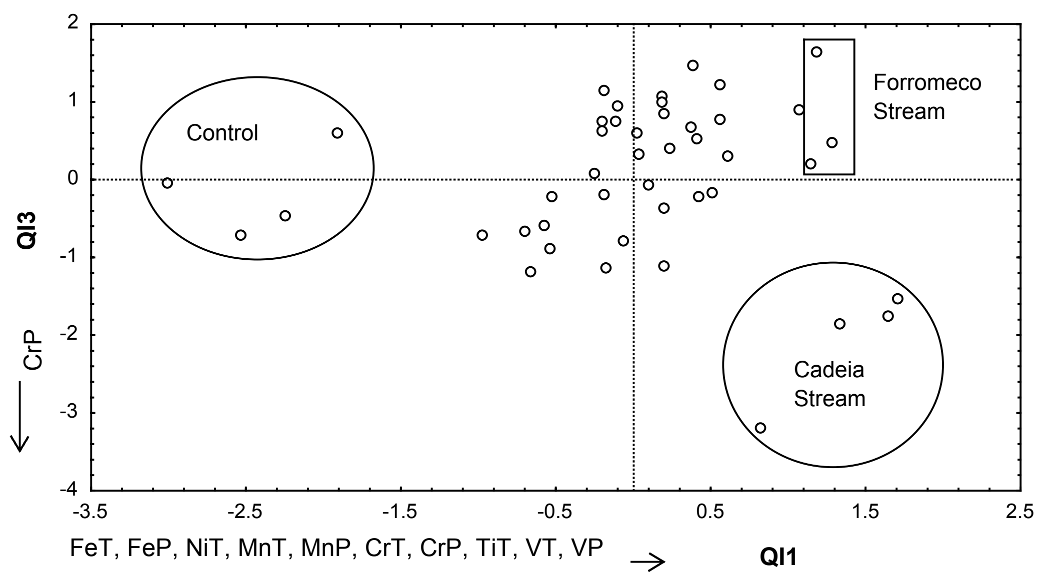

Figure 2. Dispersion diagram of the sampling units according to the factor scores (quality indexes) on first and third principal components. The variables most correlated with the components are indicated (T = total metal content; P = partial or potentially mobile content). The arrow direction indicates ascending concentrations of the elements in the sediments.

vineyards located in this region, where copper compounds-based fungicides are used. The sites characterized by a low environmental quality as regards this index were the mouth of Pinhal and Belo Streams and the point CA 136 in the Caí River.

The third axis (Figs. 1 and 2), which represented 7.9% of the total variance, is better correlated with the variable partial Cr. The intermediate and positive correlation of this variable with organic matter suggested the contribution of tanneries, located in the region drained by the Cadeia Stream, which is outstanding in this dimension.

The representation of the quality indexes on three- and bi-dimensional graphs (Figs. 1 and 2) enabled the observation of spatial tendencies. No temporal trends were detected. In the area under influence of the petrochemical complex, the sampling units presented a small dispersion around the scores average, including the upstream sites CA 24.1 and CA 050. Therefore, no detectable influence of this complex on the contents of the elements analyzed in the sediment was observed during the monitoring period. The sediment samples collected in the control point (CA 210) presented the lower concentrations of the studied metals, and consequently showed the best environmental quality.

As to the categorization of the indexes to be displayed on the basin map, the result of the calculation of the quartiles, values considered as delimiting the initial classification, is indicated in Tab. 3. Meanwhile, the final result of the chromatic categorization is shown in Fig. 3.

Table 3. Values of the quartiles used to classify the indexes

	Inferior	Superior
Quality index 1 (QI1)	-0.206	0.554
Quality index 2 (QI2)	-0.345	0.667
Quality index 3 (QI3)	-0.680	0.754

4. SUMMARY AND CONCLUSIONS

The use of statistical factor analysis to obtain sediment quality indexes enabled the segregation of areas impacted by different contributions of heavy metals, from natural and anthropic sources. The information was summarized into three main components, which were associated with different potential sources of contribution. In the first, where the Forromeco and Cadeia Streams were highlighted, the best correlated variables indicated a natural contribution of the local geology, where basalts predominate, and the effect of surface runoff on unvegetated soils. The second component, stressing typical elements of metallurgical, electroplating and textile industries effluents, indicated a potential influence of the industrial complex in the region of Caxias do Sul, and the use of copper-based fungicides in the vineyards concentrated in this area. The third component, in which the amount of potentially mobile Cr was relevant, showed the impact of tanneries on the Cadeia Stream sediments. The control point presented the best quality in relation to the investigated elements. During the monitoring period, no detectable influence of the petrochemical complex on the contents of the elements analyzed in the sediment was observed. However, the final stretch of the Caí River, under influence of this complex, is a preferential site for fine particles settlement and may represent a reservoir for trace elements, which could undergo remobilization to the water column. This section of the river deserves, therefore, great attention in terms of monitoring and preservation of the environmental quality. The graphic representation of the factor analysis results enabled a good visualization of spatial tendencies, while the chromatic categorization of the indexes allowed to identificate the sampling stations on a map, according to the relative degree to which the sediments were affected by heavy metals pollution.

5. REFERENCES

Fachel, J.M.G. 1976. Análise fatorial. Dissertation, Instituto de Matemática e Estatística da Universidade de São Paulo, São Paulo, Brasil, 81 p.

FEPAM/GTZ. 1997. Levantamento dos principais usos do solo e da água na bacia hidrográfica do rio Caí, Vol.1, Projeto de Cooperação Técnica Brasil / Alemanha, Porto Alegre, Brasil, 138 p.

Haase, J., Krieger, J.A., and Possoli, S. 1989. Estudo da viabilidade de uso da técnica de análise fatorial como instrumento na interpretação da qualidade da água da bacia hidrográfica do Guaíba, RS, Brasil. Ciência e Cultura, Vol. 41, pp. 576-582.

Haase, J., and Possoli, S. 1993. Estudo da utilização da técnica de análise fatorial na elaboração de um índice de qualidade da água: comparação entre dois regimes hidrológicos diferentes. Acta Limnológica Brasiliense, Vol. 6, pp. 245-255.

Kishi, R.T., and Lucca, S.J. 1991. Índices de qualidade da água aplicados à Lagoa Negra/RS. Proc. of the XVI Congresso Brasileiro de Engenharia Sanitária e Ambiental, Porto Alegre, Brasil, pp. 120-131.

Lohani , B.N. and Mustapha, N. 1982. Indices for water quality assessment in rivers: a case study of the Linggy River in Malaysia. Water Supply & Management, Vol. 6, pp. 545-555.

Malo, B.A. 1977. Partial extraction of metals from aquatic sediments. Environmental Science & Technology, Vol. 11, pp. 277-282.

Rodrigues, M.L.K. 1997. Diagnóstico da poluição por elementos-traço no sedimento da bacia hidrográfica do Rio Caí, RS. Dissertation, Instituto de Biociências da Universidade Federal do Rio Grande do Sul, Porto Alegre, Brasil, 124 p.

SDD - Scotish Development Department. 1976. Development of a water quality index. Applied Research & Development, ARD.3. Report, Scotish Development Department, Edinburgh, 62 p.

Shoji, H., Yamamoto, T., and Nakamura, T. 1966. Factor analysis on stream pollution of the Yodo River system. Air & Water Pollution, Vol. 10, pp. 291-299.

ACKNOWLEDGEMENTS

To FINEP/PADCT for the financial support, to Dr. Jandyra Fachel (UFRGS) for the revision of the statistical analysis, and to the Laboratory of FEPAM for helping in the sediment sampling and heavy metals analysis.

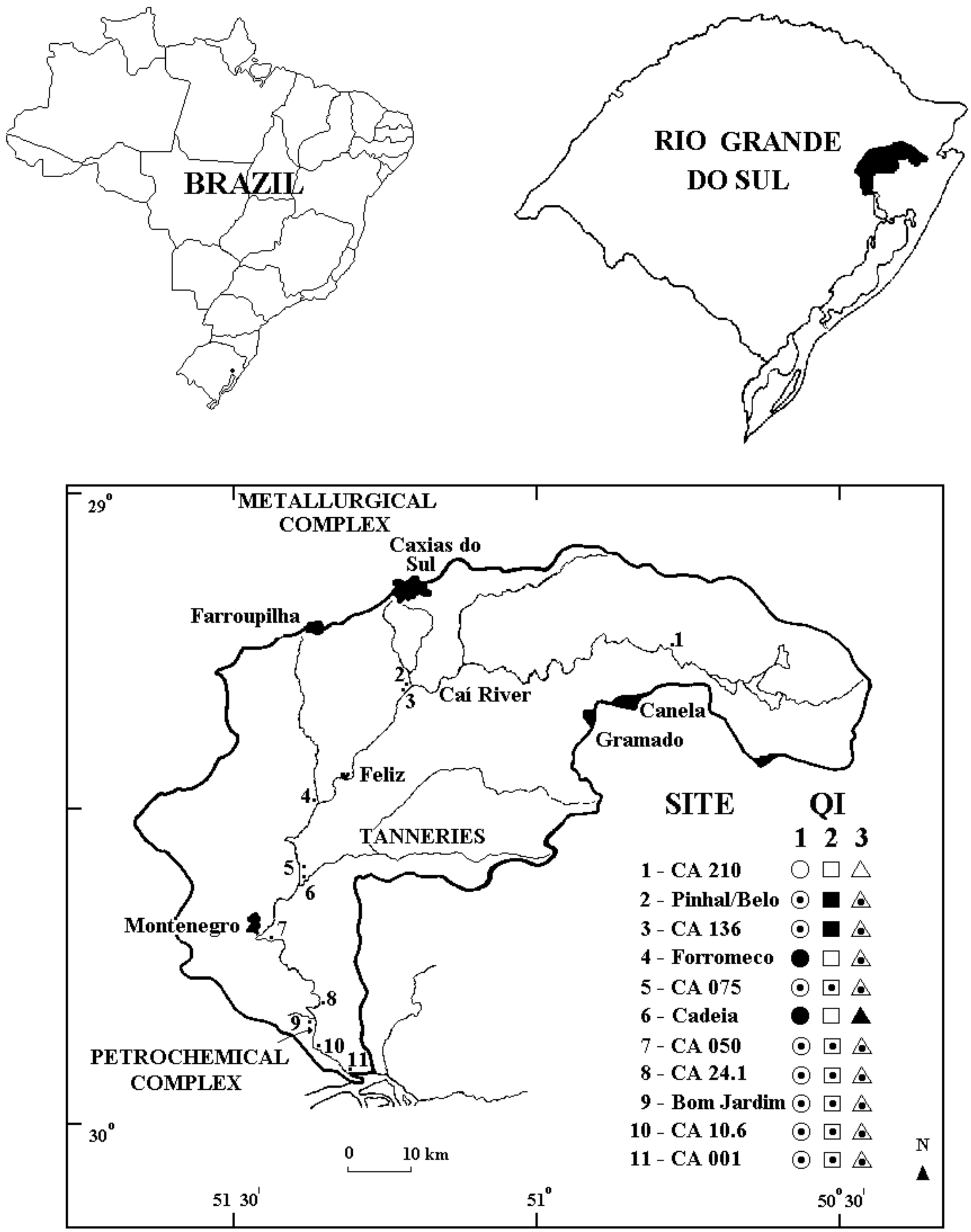

Figure 3. Location and classification of sediment sampling stations in the Caí River basin (RS, Brazil), according to the heavy metals pollution status. Quality Index 1 (circle) was associated to a natural contribution of basalts; Quality Index 2 (square) with metallurgical, galvanoplastic and textile industries, besides copper compounds used in viticulture; and Quality Index 3 (triangle) with the pollution by tanneries. The white colour represents better environmental quality, black corresponds to worse quality, and the dotted symbol to an intermediary sediment pollution by the investigated elements.