Journal of Urban and Environmental Engineering, v.12, n.1, p.77-87

ISSN 1982-3932
doi: 10.4090/juee.2018.v12n1.077087

Journal of Urban and Environmental Engineering

www.journal-uee.org

MULTIVARIATE STATISTICAL- AND GIS-BASED APPROACH DEVELOPED FOR INTEGRATED ENVIRONMENTAL ANALYSIS IN URBAN WATERSHED

Rodrigo Custódio Urban[1]*, Luiz Augusto Manfré[2] and Alexandre Marco da Silva[3]

[1]*College of Environmental Engineering, Pontifical Catholic University of Campinas, Brazil*
[2]*Departament of Transportation, Polytechnic School, University of São Paulo, Brazil*
[3]*Departament of Environmental Engineering, São Paulo State University, Brazil*

Received 29 December 2017; received in revised form 1 June 2018; accepted 1 July 2018

Abstract: This paper aims to assess the environmental quality of a small urban watershed, located at a sub-tropical region highly urbanized in Brazil, using water and soil quality, land cover and terrain characteristics. The proposed methodology was based in physical and chemical features of 40 soil sampling sites, land cover and slope. Principal Components Analysis (PCA) was used to define the best variables to the analysis. The soil quality, land cover and slope data was grouped and categorized in qualitative variables. Multiple Correspondence Analysis (MCA) was applied to cluster the variables. Geographic Information System (GIS) tools were used to build the zoning map. During 12 months water was sampled in two sites in the same river at the watershed. PCA was used again to define water quality and differences between the two sampling sites. Porosity and carbon rate were the principal soil variables to distinguish three different soil zones 1, 2 and 3 represent 15,1%; 9,8% e 75,1% of the area, respectively. Zone 1 present condition that must be conserved to maintain environmental services as water retaining and carbon storage. Related to water quality, the PCA presented differences between dry and wet season. Besides, sampling site 1, located within a vegetation region presents better conditions than sampling site 2, located within urban land cover. The assessment method used multivariate statistics and GIS. The methodology is a useful tool to environmental planning. The replication of this methodology is encouraged, in order to assess its suitability in different conditions, i.e. climate and size.

Keywords: urban soil quality; urban water quality; principal component analysis; multiple correspondence analysis; environmental planning; environmental zoning.

INTRODUCTION

Environmental quality, i.e. air, water and land surface, directly affects human life quality. (Banzhaf *et al.,* 2014; Joseph *et al.,* 2014). Due to the high population density and diverse environmental conditions, urban areas represent the most complex interactions between humanity and environment (Banzhaf *et al.,* 2014).

Soil is an crucial component that influences the environmental quality in urban areas (De Kimpe & Morel, 2000) and descriptions related to soil quality aid the environmental management and guide city planning (Vrščaj *et al.,* 2008). Urban soil and green spaces are fundamental for the water storage in urban environments (Claessens *et al.,* 2014).

However, soil function and quality in urban environments are different from those in agricultural and forest areas, due to different needs for soil use (Vrščaj *et al.*, 2008). Thus, the studies of soil quality in urban areas must be carried out with an unusual perspective, considering its functions for civil construction, urban crops, green areas, waste disposal, rainwater filtration (Effland & Pouyat, 1997; Pedron *et al.,* 2004).

On the other hand, another important feature of urban environmental quality is water. The urban water streams are often studied because of the growing need for water use. Nonetheless, population growth and expansion of economic activities (urban and rural) cause changes in the quality of water resources, impacting on human health and other species (Zeilhofer *et al.,* 2006).

Several studies are conducted to correlate variables of water quality in urban environments and its implications for water sources, recreational activities, consumption, and other uses (Bordallo *et al.,* 2001; Ouyang, 2005; Ouyang *et al.,* 2006), confirming the close relations between water cycle and local land cover, and evidencing that they are essential for the integrated urban water management (Bach *et al.,* 2015).

Projects that aim to investigate relation among the variables regarding urban environmental quality usually demand an array of diverse techniques. Among the techniques used to perform the analysis of correlation between variables, multivariate analysis may be highlighted. This kind of analysis facilitates the interpretation, since it permits the integration of a large amount of data. Exemplifying, the technique of principal component analysis (PCA) is quite used in the interpretation of several environmental data, including soil quality (Silva *et al.,* 2015) and water quality (Parinet *et al.,* 2004; Ma *et al.,* 2010; Aguiar Netto *et al.,* 2013). Through such technique important components are identified and usually they explain part of the variance of the data, reducing the amount of variables into a smaller number of indices and grouping variables with greater correlation between them (Ouyang, 2005; Ouyang *et al.,* 2006; Zimmermann *et al.*, 2008; Moura *et al.*, 2010). Further, the Multiple Correspondence Analysis (MCA) can be applied to aid land use planning (Lavoie *et al.*, 2013).

However, the spatial visualization of the variables enhances the comprehension of urban environmental quality and works and studies involving land use and urban planning depend of this visualization. In the last years, potentialities provided by the continuously expanding Geographic Information Systems (GIS) technology and by the growing availability of digital georeferenced data have been greatly advanced (Borgogno-Mondino *et al.*, 2015), and the GIS become a powerful tool for urban environmental quality analysis (Joseph *et al.*, 2014). The combination of statistical tools and GIS increase the data interpretation possibilities (Plieninger *et al.*, 2013) and the combined use of statistical and GIS techniques need to be better explored. In this sense, this paper aims to evaluate suitability of the analysis of environmental quality in a small urban watershed, considering soil and water quality and terrain characteristics, evaluated under a combined form of statistical and GIS techniques.

MATERIAL AND METHODS

Site descriptionsmatters

The study was conducted in a small-sized and urbanized watershed, located at the central region of Sorocaba City (**Fig. 1**). The municipality of Sorocaba is located in the southeast portion of the State of São Paulo (Brasil). It has an area of 450.4 km^2 and a population of 637.187 inhabitants, being considered that about 99% is urban (IBGE, 2015).

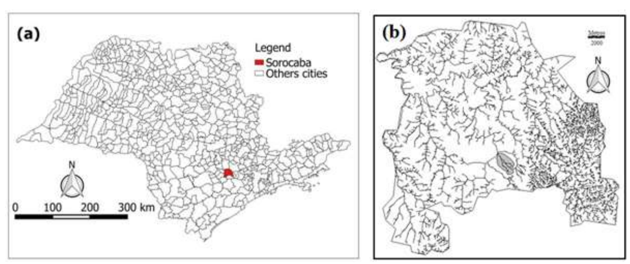

Fig. 1 Sorocaba City location in São Paulo State (a) and the Lavapés watershed (b).

The watershed (river network) is characterized as third order and has an approximate area of 2.88 km². Major land cover class is urbanization (69.5%), followed by fragments of Natural Remnant Vegetation (10.7%), low vegetation (9.5%), uncovered, bare soil (8.2%), and lands used of some agricultural use (2.1%). The lack of riparian vegetation, the dumping of construction wastes at inadequate sites, and the

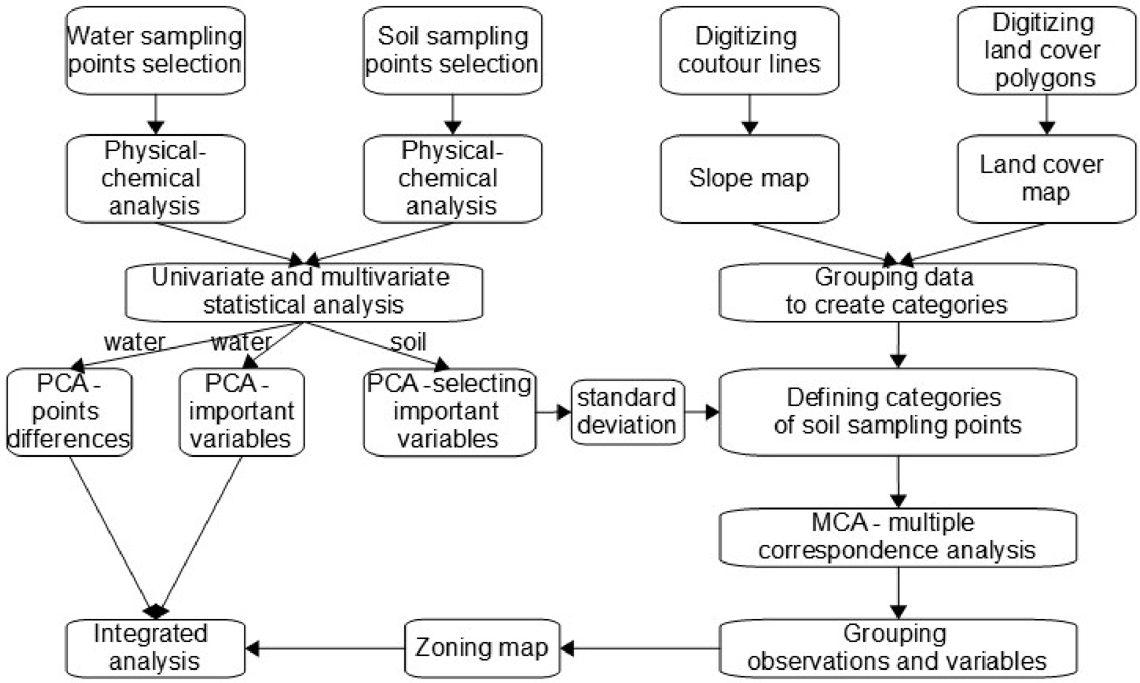

Fig. 2 Integrated environmental analysis methodology flow.

untreated sewage were pointed as the main driving forces of hydrossedimentological disequilibrium in small watershed studied (Silva *et al.*, 2013).

Procedures

The overall strategy of the study is depicted in **Fig. 2**.

Water sample analyses

Water samples and *in situ* data were collected monthly over a year, in 2 sampling sites defined considering the representativeness and accessibility of the sampling sites (Urban *et al.*, 2010). The upper sampling site (site 1) is located at 23°29′31″ S and 47°26′01″ W and the lower sampling site (site 2) is located at 23°30′07″ S and 47°25′23″ W.

Through i*n situ* incursions, we evaluated the dissolved oxygen (henceforward DO) and water temperature. The water sampling followed the procedures described in the Guide for Collecting and Preservation of Water Samples (CETESB, 1987). The samples were delivered to the laboratory for analysis of the variables described in Table 1. The variables were chosen because they are classically used in studies as indicators of water quality and of highly interest for management of water resources.

Table 1 Physicochemical variables quantified and their procedures of measurement

Variable		Method
TS, TVS, TFS		The solids were determined using the evaporation method (APHA 1999).
Temperature		Determined with the use of electronic thermometer with digital display.
EC		Was determined using a benchtop conductivity-meter.
DO		Determined by an electrometric method using an oxygen-saturation-meter.
pH		Determined by the electrometric method. With benchtop pH-meter, according to determinations of APHA (1999).
Chemical parameters (ions)	Alkalinity	Determined by neutralization titration of the acid / base, using sulfuric acid 0.01 mol/L, following the methodology adapted from APHA (1999).
	Hardness Mg and Ca	Determined by titration with reagent kits purchased following methodology adapted from APHA (1999).
	SO_4^{2-} Cl^- K NO_3^- TP	Determined using a spectrophotometer and reagents kit purchased following methodology adapted from APHA (1999).

TS: total solids; TVS: total volatile solids; TFS: total fixed solids; EC: electrical conductivity; DO: dissolved oxygen; Mg: magnesium; Ca: calcium; SO_4^{2-}: sulphate; Cl^-: chlorine; K: potassium; NO_3^-: nitrate; TP: total phosphorus

The data were organized in digital worksheets for analyses of the descriptive statistics and the principal component analysis (PCA). The PCA was selected because is a technique used in physic-chemical analysis of water and enables investigation of the relationships among the variables (Zimmermann *et al.*, 2008; Moura *et al.*, 2010). To better represent the differences between the two sampling sites a PCA was performed for each of them.

Soil sampling campaigns

For the soil sampling campaign, we used a stratified random sampling approach at our study area, considering four existing land cover categories: urban area, pasture, vegetation and bare soil (**Table 2**). In each land cover category, ten soil samples were collected in sampling points randomly determined. In each sampling point one sample (approximately 2 kg) was collected in the superficial soil layer (0-20 cm). The sample was placed into a recipient, which was closed and taken to the laboratory for analyze physical and chemical variables (**Table 3**). Hence, we collected a total of 40 superficial soil samples.

In each sampling point, we also collected another soil sample by means of the cylindrical core method (metallic ring of approximately 254 cm^3), for determining the soil bulk density, following the recommendations of USDA (2008). Some centimeter aside to the sampling point we also analyzed the soil compression using a metallic penetrometer. All the sampling points were georeferenced.

Soil variables were organized and analyzed using descriptive statistics and Principal Components Analysis (PCA). The statistics was calculated with 95% of significance. PCA was used to select the most important soil variables. The selected variables were categorized and transformed in qualitative variables, according to **Table 4**.

Thematic maps

For completing our analysis we also considered the slope map that was elaborated trough the vectorization of topographic maps at 1:10,000 scale, and the generation of Digital Elevation Models (DEM). Slope was divided into the interpretative classes of relief according to the value of declivity (in percentage): very low < 6%; low - 6% to 12%; medium: 12% to 20%; high: 20% to 30% and very high: > 30%.

The land cover map was elaborated by digitizing the polygons features identified in satélite imagery, avaliable in Google Earth©. The land cover classes were defined according to the soil sampling classes (urban areas, grasslands, bare areas and vegetation). The slope and the land cover maps were used to identify to which categories each sampling points belonged to. Thus, it was possible to establish a data bank containing

Table 2 Land cover categories and the respective definitions

Strata – Category of land use	Definition
Urban areas	Wasteland and field of lowland football
Grasslands	Squares and pasture of bovines
Vegetation	Riverbanks and vegetal fragments in areas with greater slopes
Bare areas	Lands prepared for building or for small crops

Table 3 Summary of the methods for soil quality variables considered in this study

Variables	Method
PR	Quantified *in situ* by using a penetrometer of impact (Silva *et al.*, 1998).
pH	H_2O: it was prepared a solution soil x distilled water in the proportion 1:2.5 and, passed 30 minutes, the pH was measured with a bench pHmeter KCl: it was prepared a solution soil × KCl (KCl 1.0 molar) in the proportion 1:2,5 and, passed 30 minutes, the pH was measured with a bench pHmeter (Silva *et al.,* 1998).
W	After collecting, the packaged sample arrived to the laboratory. It was taken an aliquot and the moisture was determined by the mass difference before and after the drying at 80ºC for 48hours
EC	It was measured by using a conductivity-meter after the preparation of a solution soil × distilled water in the proportion 1:1 Oliveira *et al.* (2002).
BD, PD, Por.	The bulk density was quantified by the volumetric ring method. The particle density was quantified by the volumetric balloon method. The porosity was calculated using the soil density data (Vieira, 1988).
Mg, Al, K, Ca, P, C, N, H+Al	Samples sent to the laboratories of University of São Paulo (ESALQ and CENA), Piracicaba, São Paulo

PR: penetrometer resistance (kgf cm^{-2}); W: moisture (%); ED: electric conductivity (μS.cm^{-1}); BD: bulk density (g cm^{-3}); PD: particle density (g cm^{-3}); Por.: porosity (%); Mg: Magnesium (mmol kg^{-1}); Al: aluminum (mmol kg^{-1}); K: potassium (mmol kg^{-1}); Ca: calcium (mmol kg^{-1}); P: Phosphorus (mg kg^{-1}); C: carbon (%); N: nitrogen (%); H+Al: potential acidity (mmol kg^{-1}).

Table 4 Categories thresholds for soil variables transformation from quantitative to qualitative type

Category	Limits of category
Very high	> mean + 2 SD
High	mean + 1 SD : mean + 2 SD
Medium	mean - 1 : mean +1 SD
Low	mean - 2 SD : mean - 1 SD
Very low	< mean – 2 SD

SD: standard deviation

the soil quality from the sampling points (Table 4), relating to the identified categories.

After the data categorizing, a Multiple Correspond Analysis (MCA) was executed, considering it is more appropriated for categorical explanatory variables (Abdi & Valentin, 2007; Abdi & Willians, 2010). Then, the variables and observations were clustered in three categories. According to the categorization and the land cover divisions, three homogeneous zones were defined and a soil zoning map was elaborated for the watershed. The combination of the established zones and the water quality provided an integrated analysis of the environmental characteristics of Lavapés watershed.

RESULTS AND DISCUSSION

Soil quality and zoning map

In order to restrict the variables, we considered factors loadings over 0.7, highlighted in Table 5. The variables may change for other study areas, according to pedogenesis and climatic conditions. The considered variables for the study area were: PD, EC, Porosity, C, N, pH KCl, pH H_2O, K e H+Al. This result shows that particle density is related to land use. However some build residual were find at some sampling sites, and also the superficial soils under high vegetation may contain high levels of organic matter which may affect this attribute. Porosity is very important, as well, being directly related to the land cover.

Table 5 Factor loadings of the four principal components of the soil variables surveyed in the study area, with values higher than 0.7 highlighted according to Kaiser's criteria

Variable	PC1	PC2	PC3	PC4
PR	-0.514	0.163	-0.348	0.598
PD	-0.169	0.035	**0.816**	-0.244
BD	-0.682	0.278	-0.146	-0.449
W	0.584	-0.344	-0.129	-0.483
EC	**0.907**	-0.063	-0.026	0.192
Por.	**0.734**	-0.243	0.529	0.196
C	**0.897**	-0.215	-0.095	0.083
N	**0.892**	-0.219	-0.053	0.113
pH KCl	0.231	**0.906**	0.038	0.015
pH H_2O	0.322	**0.893**	0.035	0.169
P	0.363	0.491	-0.356	-0.219
K	**0.814**	-0.007	0.019	-0.019
Ca	0.608	0.596	-0.115	-0.138
Mg	0.659	0.011	-0.284	-0.313
Al	-0.523	-0.579	-0.244	-0.097
H+Al	0.332	**-0.715**	-0.223	0.078
Eigenvalue	6.200	3.442	1.450	1.156
% variance	38.749	21.511	9.061	7.226
Cumulative %	38.749	60.260	69.321	76.548

PR: penetrometer resistance; PD: particle density; BD: bulk density; W: moisture; EC: electric conductivity; Por.: Porosity; C: carbon; N: nitrogen; P: Phosphorus; K: potassium; Ca: calcium; Mg: Magnesium; Al: aluminum; H+Al: potential acidity

The PCA resulted 16 engeinvalues and four were considered relevant (engeinvalues >1), according to Kaiser Criteria (Kaiser, 1958). These four principal components represent 76.55% of total data variance. **Table 5** presents the factor loadings of each variable.

The highly significant correlation between the two principal components and the values of carbon and nitrogen indicate that the urban soils can contribute to the carbon storage, as also argued by Edmondson *et al.* (2012). The real and potential acidity was also important for the variability at the sampling sites. As well as the electric conductivity represented by the K ions. The threshold values presented in Table 6 were used to establish the transformation between the qualitative and the quantitative data.

Slope and land cover information were also inserted in the analysis. The occurrence percentage of each slope class in the watershed was: very high = 2.7%; high = 7.4%; medium = 28.0%; low = 32.3% and very low = 29.6%. The occurrence of each category of land cover was: vegetation = 10.7%; grassland = 11.6%; bare areas = 8.2% and urban areas = 69.5%. Therefore, the watershed is predominantly urban and low/medium slope. The southeast part of the watershed is steep slope and has high vegetation land use.

The two main factors of MCA are presented in **Fig. 3**. In this sense it was possible to split the data into three groups. The group 1 is composed by soil sampling sites with high values of C, N, K, EC e H+Al, pH and porosity. The group 2 is composed by 10 soil sampling sites, characterized by the low values of C, N, K, EC, pH, porosity and medium values of H+AL and PD. The other soil sampling sites correspond to group 3, with similar characteristics and low data variability. This group has intermediate characteristics for the presented variables, except for PD and H+Al, which is very disperse for PD and H+Al, probably due to the building residual found the sampling sites. In this sense, they are indicators for group.

However the **Fig. 3** presents slope and land cover within some groups, it is not possible to state if they clearly influences on the described variables. Land cover categories are located near to the graphic origin and close to each other. Slope categories are dispersed and not connected. Even that, the land cover map was essential for establishing the map zones (**Fig. 4**). To elaborate this map, the groups of **Fig. 3** were used, highlighting the groups 1 and 2, which had most distinct characteristics. External sites to the watershed were considered for the zoning. Some sites of group 3 were included in group 1 or 2, in order to provide a homogenous zoning (Sites 04, 20, 24 e 35).

Table 6 Categories thresholds for soil variables, highlighted by PCA, transformation from quantitative to qualitative type

Category	PD	EC	Porosity	C	N	pH KCl	pH H_2O	K	H+Al
Very high	> 2.73	> 455.95	> 68.10	> 3.67	>0.29	> 7.53	> 7.21	> 7.30	> 52
High	2.47 - 2.73	329.90 - 455.95	55.85 - 68.10	2.59 - 3.67	0.20 - 0.29	6.88 - 7.53	6.29 - 7.21	5.19 - 7.30	36.42 - 52
Medium	1.96 - 2.47	77.80 - 329.90	31.35 - 55.85	0.43 - 2.59	0.03 - 0.20	5.58 - 6.88	4.46 - 6.29	0.99 - 5.19	5.28 - 36.42
Low	1.70 - 1.96	0.00 - 77.80	19.10 - 31.35	0.00 - 0.43	0.00 - 0.03	4.93 - 5.58	3.54 - 4.46	0.00 - 0.99	0.00 - 5.28
Very low	<1.70	ND	< 19.10	ND	ND	< 4.93	< 3.54	ND	ND

PD: particle density (g cm^{-3}); ED: electric conductivity ($\mu S.cm^{-1}$); Por.: porosity (%); C: carbon (%); N: nitrogen (%); K: potassium (mmol kg^{-1}; H+Al: potential acidity (mmol kg^{-1}); ND: no data

The presence of fragments of remaining natural vegetation predominantly in zone 1 is the main responsible by the definition of most characteristics of such zone, for example: soils with higher superficial porosity, due to the action of vegetation in the soil (Cavernage *et al.*, 1999). The high porosity enhances the capability to retain water no decrease the water flow peaks e the floods effects. These are desirable characteristics for urban soils that are not provided by means of engineering works (Pedron *et al.*, 2004). High values of H+Al may be referred exchangeable aluminum and organic acids, highlighting vegetation areas characteristics (Silva *et al.*, 2008). Therefore, this zone has characteristics that indicate its preservation demand, in order to maintain the environmental quality at study area.

The West and Southwest of the watershed which has a restricted access to soil sampling is highly urbanized and is defined as zone 3. The area share for the watershed is: zone 1 = 15.1%; zone 2 = 9.8% and zone 3 = 75.1%. Table 7 presents the characteristics and indications for each zone.

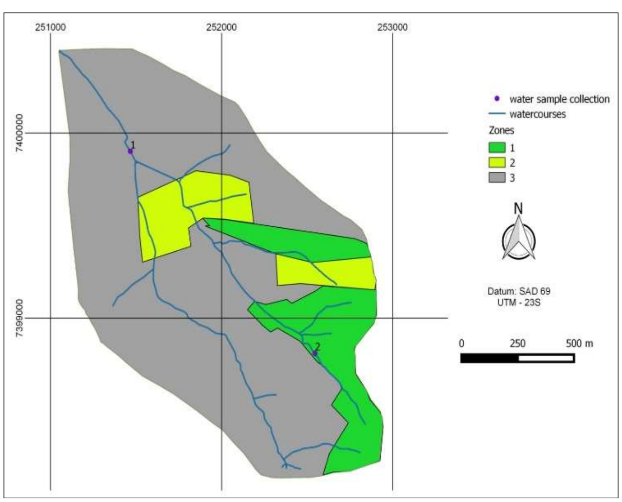

Fig. 4 Lavapés watershed zoning map derived from the land use, soil quality and slope analysis.

Table 7 Features and indications of each zone established for the study area

<table>
<tr><th>Zone</th><th>Main features</th><th>Indications / Importance</th></tr>
<tr><td>1</td><td>Soils slightly acidic, high porosity, high electric conductivity, carbon and nitrogen. Embraces mainly areas covered by fragments of natural remaining vegetation.</td><td>High potential for water retention and storage, as well as carbon sequestration and storage, maintenance of the local environmental quality.</td></tr>
<tr><td>2</td><td>Soils normally compacted, medium-high acid, with low values of electric conductivity and insufficient concentrations of carbon and nitrogen</td><td rowspan="2">Urban settlements with landscaping</td></tr>
<tr><td>3</td><td>Soils with intermediary level of porosity, pH close to neutrality, medium percentages of carbon and nitrogen</td></tr>
</table>

Other evidence of influence of vegetation in the zone 1 is the high carbon value, once this category of land cover leads the soil to concentrate more organic matter and consequently more carbon. Zone 2 has lower percentage of carbon, due to urbanization characteristics. Once this zone presents soil with lower porosity, it is a region indicated for constructions.

Zone 3 has soils that may be associated to urban soils, however it is difficult to categorize this soil type (Pouyat *et al.*, 2007). Considering the diversity of characteristics, with intermediate variable values, it is possible deducing that the soils are composed by diverse materials and has undergone through land cover modifications. Commonly vegetation in these areas is not native and man-induced change soil variables and decrease its quality (Alberti, 2005). However, zone 3 presents better soil quality than zone 2, due to gardens and other uses for urban soils. Therefore, it may be used for landscaping projects and also construction.

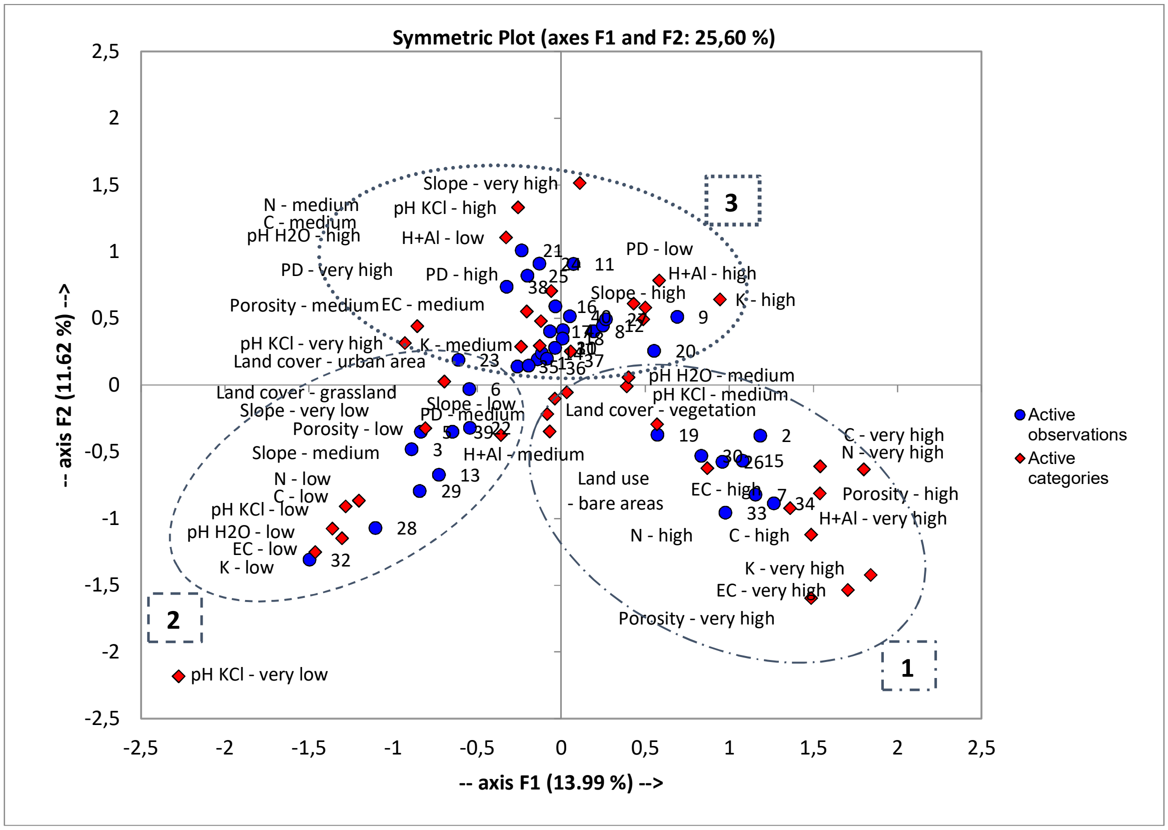

Fig. 3 Soil quality, slope and land use data graphic for the two main MCA factors.

Water quality and integrated analysis

Figure 5 presents the water sampling sites location. Site 1 is within zone 3 and site 2 is within zone 1. Before introducing the water quality analysis, it is important mention that the results do not refer to extreme rainfall events. The survey refers to the minimum quality values obtained in monthly collections. The discussion was focused on the results of studied that considered water quality relatively of surrounding regions of the study area.

The PCA was performed for all variables and both sampling sites. The PCA resulted in 16 eigenvalues, two of them were not considered because they did not represent any percentage of the variance. From the 14 remaining eigenvalues, 5 were considered relevant, according to Kaiser's criterion (Kaiser, 1958). From the selected eigenvalues 5 principal components of data were selected and analyzed, representing 80.72% of the total variance. The top two principal components explained 57.26% of the total variance. The factor loadings for each variable for each major component can be observed in **Table 8**, as well as their eigenvalues and the variance.

Regarding the relationships observed in the PCA, the temperature influences the electrical conductivity and alkalinity and changes the solubility of elements. Solids can be influenced by several variables. For improving the visualization of data, scores and factor loadings obtained from PCA of the two principal components that represent most of the variance (57.26%) were plotted. The interpretation of the graph and its values followed the directions of the work of Zimermman *et al.* (2008) (**Fig. 5**).

The first principal component (PC1) shown some potential to differentiate samples of point 1 and point 2, with some exceptions. Samples of sampling point 1 (numbered 1-12) are mostly on the right side of PC1 along the variables Ca, Cl, hardness, alkalinity, EC, Mg, K, solids, SO_4^{2-}, NO_3^- and TP. Hence it can be assumed that these variables represent samples of the sampling point 1 (zone 3), after receiving loads of diffuse pollutants in your course. The left side of PC1 is constituted predominantly by the samples for the sampling point 2 (numbered 13-24), represented by the variables OD and pH, indicating that characteristics are most influential in this sampling point.

The second principal component (PC2) represents seasonal differences. At the top of the graphic are located the samples for the dry season (April – September). At the bottom are located the wet season samples (October – March). There are 5 groups in the

Table 8 Factorial loads, highlighting the values higher than 0.5 on the five first principal components of the variables related to water quality for the study area

Variable	PC1	PC2	PC3	PC4	PC5
DO	-0.282	0.547	-0.173	-0.177	0.567
TS	0.892	-0.177	0.164	-0.150	-0.181
TFS	0.744	-0.110	-0.325	0.229	-0.132
TVS	0.607	-0.155	0.536	-0.428	-0.140
pH	-0.111	0.543	-0.460	-0.415	-0.407
T	0.532	0.359	-0.225	0.477	-0.299
EC	0.888	0.082	0.065	-0.150	0.080
Hardness	0.829	0.421	-0.220	-0.067	0.153
Ca	0.558	0.668	0.069	-0.188	-0.002
Mg	0.726	-0.023	-0.414	0.088	0.241
Alkalinity	0.840	0.264	-0.014	-0.097	0.105
NO_3^-	0.502	-0.474	-0.533	0.095	-0.014
Cl^-	0.424	0.414	0.484	0.479	-0.030
K	0.613	-0.086	0.365	0.154	0.385
SO_4^{2-}	0.825	-0.341	0.155	-0.124	-0.190
TP	0.383	-0.638	-0.263	-0.149	0.262
Eigenvalue	6.746	2.416	1.685	1.069	1.001
Variance (%)	42.160	15.102	10.534	6.679	6.253
Cumulative variance (%)	42.160	57.261	67.795	74.475	80.728

DO: dissolved oxygen (mg L^{-1}); TS: total solids (mg L^{-1}); TVS: total volatile solids (mg L^{-1}); TFS: total fixed solids (mg L^{-1}); T: temperature (ºC); EC: electrical conductivity (**µS cm^{-1}**); Hardness (mg L^{-1}); Ca: calcium (mg L^{-1}); Mg: magnesium (mg L^{-1}); Alkalinity (mg L^{-1}); NO_3^-: nitrate (mg L^{-1}); Cl^-: chlorine (mg L^{-1}); K: potassium (mg L^{-1}); SO_4^{2-}: sulphate (mg L^{-1}); TP: total phosphorus (mg L^{-1}).

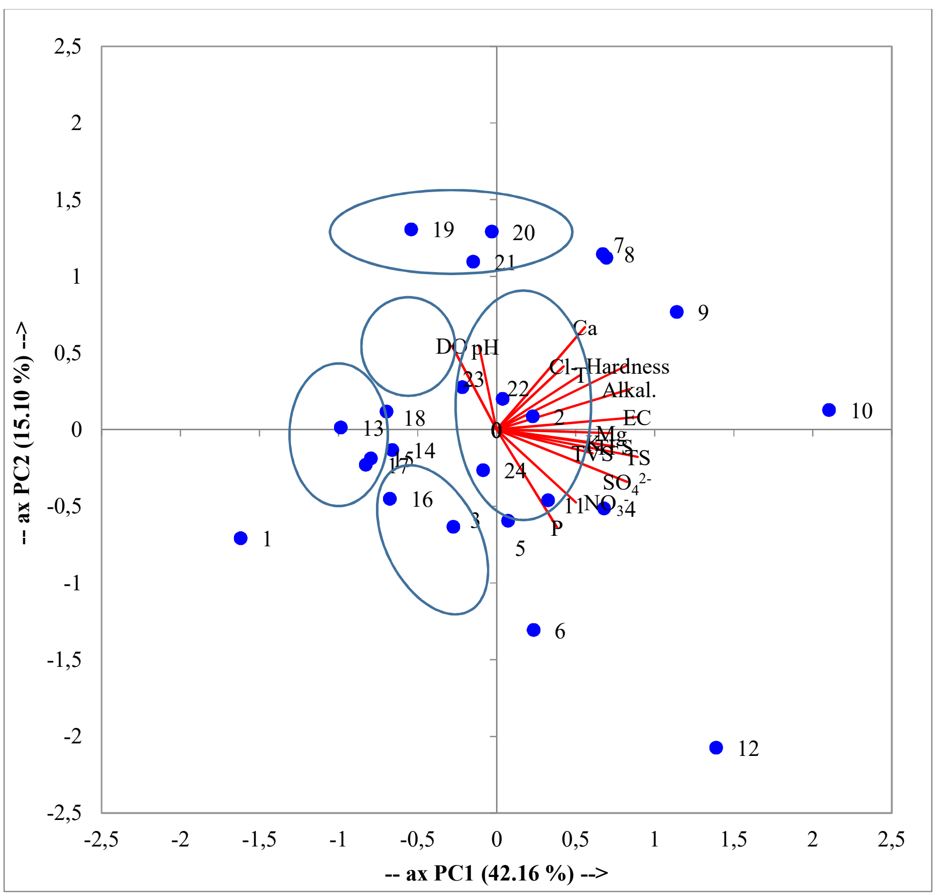

Fig. 5 Graph biplot of scores and factor loadings for the first two principal components of the 2 sampling points in the study area.

biplot graphic of the scores and factor loadings of PC1 and PC2. The leftmost group presents samples 14, 15, 16, 17 and 18, which are those representing the dry season (May-September) at the sampling point 2, indicating similar characteristics. The top group shows samples 7, 8, 19, 20 and 21, which are representative of the rainy season in both sites (September-November in point 2 and October-November in point 1) indicating similarities in water quality during this period. The central group presents samples of December and January at the sampling point 2, suggesting that the OD and pH are the variables that best describe them. The group sited in the left presents the remaining variables, indicating its relationship to water quality. The bottom group presents samples 3, 5 and 6 representatives of the dry period at the sampling point 1.

The sample 12 (March - sampling point 1) deserves attention highlighted by presenting higher values of nitrate and phosphorus, indicating the presence of punctual pollution loads in the stream.

Taking into account the high association between electrical conductivity and the evaluated ions, and also aiming to reduce the variables for individual study of the sampling points (fewer samples) we excluded some variables (ions) from the analysis of the differences among the sampling points. The analysis of both sampling points resulted in 7 eigenvalues, where 1 was disregarded for not presenting relevant data variance. From the 6 remaining eigenvalues, 2 were considered relevant to the sampling point 1 and three relevant to the sampling point 2. The relevance was determined by the Kaiser's criterion (Kaiser, 1958). Selected eigenvalues resulted in their principal components, representing 64.77% of the variance of the data for point 1 and 76.49% of the data variance of point 2. Varimax rotational approach (Kaiser, 1958) was used to maximize simplification of factor loadings to facilitate interpretation. The factor loadings, *eigenvalues* and variances of each component, for points 1 and 2 can be observed in **Table 9**.

The amount of principal components and relevant variables are different among the sample sites. This indicates a differentiation of the interrelationship of quality variables in both sites. However, all the variables are appropriate for the data variance. Soil quality, presented in **Fig. 4** illustrates the difference of the water sampling sites. The first principal component of the sampling point 1 suggesting that increased volatile solids indicate increased organic matter and consequent decrease in dissolved oxygen content. It was observed that the relations presented for point 1 and point 2 are different, and the relationships of volatile solids with dissolved oxygen are present in sampling point 1 and indicate a decrease in water quality from one point to another, confirming the expectations for the presented zones. Average and average thresholds are presented in **Table 10**.

Also in **Table 10**, a significant difference between the DO values was observed. There is a decrease from the point 2 to point 1. DO amount may decrease due to urbanization (Carvalho *et al.*, 2004).

The results of the total, fixed and volatile solids in the stream Lavapés also showed significant differences between point 2 and point 1. Nonetheless, the values in both sampling sites were very high. It is noteworthy that the APHA (1999) indicates that the determinations of fixed and volatile solids not exactly distinguish between organic and inorganic materials, because of the presence of some minerals, as carbonates, chlorides, sulfates, can be volatilized at the same temperatures that organic compounds.

Also according to **Table 10**, the increasing of the values of electrical conductivity between point 1 and point 2 indicate the decline in water quality. The highest values of the stream Lavapés are possibly related to the excessive amount of total solids.

CONCLUSIONS

The proposed methodology for environmental quality assessment is powerful and generated suitable outcome to the objective of the study, being an interesting tool for helping the regional environmental planning. The main strength of this model concerning the potential of be replicable to other watershed in the region and may help the local environmental planning. The three presented zones have different land uses, highlighting the zone 1 which must remain mostly conserved to guarantee or improve the environmental quality indexes for the region. The use of PCA to restrict the soil samples variables was adequate. The replication of this variables set must be validate according to the study area. The application and adaptation of the methodology

Table 10 Comparison of physicochemical variables of water quality between the two sampling points in the stream Lavapés - Sorocaba

Variables	Mean and mean limits with 95% significance	
	Sampling Point 1	Sampling Point 2
DO (mg L^{-1})	2.04 (1.53 - 2.56)	4.29 (2.79 - 5.81)
TS (mg L^{-1})	2242.69 (1885.32 - 2600.07)	1418.97 (1264.34 - 1573.61)
TFS (mg L^{-1})	1535.90 (1273.05 - 1798.74)	1067.18 (953.31 - 1181.04)
TVS (mg L^{-1})	706.79 (424.43 - 989.16)	351.79 (186.36 - 517.23)
pH	7.52 (7.38 - 7.65)	7.46 (7.34 - 7.58)
T (ºC)	24.31 (22.19 - 26.43)	22.62 (20.78 - 24.46)
EC (µS cm^{-1})	300.08 (249.23 - 350.94)	211.75 (189.17 - 234.34)

Table 9 Factor loadings, *eigenvalues* and variance of 2 principal components of the sampling point 1 and three principal components of the sampling point 2 of the studied stream

Variables	Sampling Point1		Sampling Point 2		
	PC1	PC2	PC1	PC2	PC3
DO	**-0.541**	0.437	0.130	**0.798**	-0.117
TS	**0.715**	**0.641**	**0.949**	0.025	-0.068
TFS	0.066	**0.814**	-0.286	0.010	**-0.854**
TVS	**0.856**	0.082	**0.895**	0.014	0.423
pH	**-0.735**	0.039	-0.244	**0.850**	0.141
T	-0.085	**0.695**	0.092	0.085	**-0.812**
EC	**0.584**	**0.592**	0.395	**0.540**	-0.252
Eigenvalue	2.891	1.643	2.227	1.816	1.312
Variance (%)	34.685	30.082	28.912	23.710	23.873
Cumulative variance (%)	34.685	64.767	28.912	52.622	76.495

DO: dissolved oxygen; TS: total solids; TFS: Total fixed solids; TVS: total volatile solids; T: temperature; EC: electric conductivity

in urban areas with different climatic conditions and different sizes is encouraged in order to validate the method in different situations.

Slope did not affect the MCA results. However such environmental factor should be used to aid the zoning map. The analysis of the water sampling sites together did not produce good results for evaluating the degradation of studied region. Despite being made with a smaller number of variables, separate sampling points analysis allowed a better assessment of quality. Sampling points 1 and 2 constitute different scenarios in relation to water quality variables. Sampling site 1 (zone 3) is more impacted than sampling site 2 (zone 1), highlighting the urbanization at Lavapés watershed, confirming the zoning proposition.

REFERENCES

Abdi, H., & Valentin, D. (2007) Multiple Correspondence Analysis. In: N. Salkind (Ed.), *Encyclopedia of Measurement and Statistics* (pp. 652-658). Thousand Oaks, CA: Sage Publications.

Abdi H., & Williams L.J. (2010) Principal Component Analysis. *Wiley Interdisciplinary Reviews: Computational Statistics*, **2**(4),433-459, doi: 10.1002/wics.101

Aguiar Netto, A.O., Garcia, C.A., Alves, J.P.H., Ferreira, R.A., & Gonzaga, M.S. (2013) Pysical and chemical characteristics of water from the hydrographic basin of the Poxim River, Sergipe State, Brazil. *Environmental Monitoring and Assessment*, **185**(5), 4417-4426, doi:10.1007/s10661-012-2880-x

Alberti, M. (2005) The effect of urban patterns on ecosystem function. International Regional Science Review, **28**(2), 168-192, doi: 10.1177/0160017605275160

APHA (American Public Health Association) (1999) *Standard Methods for the examination of water and wastewater*. 20th ed. Washington, DC: APHA.

Bach, P.M., Staalesen, S., McCarthy, D.T, & Deletic, A. (2015) Revisiting land use classification and spatial aggregation for modelling integrated urban water systems. *Landscape and Urban Planning*, **143**, 43-55, doi:10.1016/j.landurbplan.2015.05.012.

Banzhaf, E., de la Barrera, F., Kindler, A., Reyes-Paecke, S., Schlink, U., Welz, J., & Kabisch, S. (2014) A conceptual framework for integrated analysis of environmental quality and quality of life. *Ecological Indicators*, **45**, 664-668, doi:10.1016/j.ecolind.2014.06.002.

Bordalo, A.A., Nilsumranchit, W., & Chalermwat, K. (2001) Water quality and uses of the Bangpakong river (Eastern Thailand). *Water Research,* **35**(15), 3635-3642, doi:10.1016/S0043-1354(01)00079-3

Borgogno-Mondino, E., Fabietti, G., & Ajmone-Marsan, F. (2015) Soil quality and landscape metrics as driving factors in a multi-criteria GIS procedure for peri-urban land use planning. *Urban Forestry & Urban Greening*, **14**(4), 743-750.

Carvalho, C.F, Ferreira, A.L., & Stapelfeldt, F. (2004) Qualidade das águas do ribeirão Ubá - MG. *Revista Escola de Minas*, **57**(3), 165-172. doi: 10.1590/S0370-44672004000300005

Cavernage, A., Moraes, M.L.T., Alves, M.C., Carvalho, M.A.C., Freitas, M.L.M., & Buzetti, S. (1999) Alterações nas propriedades físicas de um LATOSSOLO VERMELHO-ESCURO sob diferentes culturas. *Revista Brasileira de Ciência do Solo*, **23**(4), 997-1003, doi: 10.1590/S0100-06831999000400027

CETESB. (1987) *Guia de coleta e preservação de amostras de água*. 1th ed. São Paulo: CETESB.

Claessens, J., Schram-Bijkerk, D., Dirven-van Breemen, L., Otte, P., & van Wijnen, H. (2014) The soil–water system as basis for a climate proof and healthy urban environment: Opportunities identified in a Dutch case-study. *Science of The Total Environment*, **485-486**, 776-784. doi:10.1016/j.scitotenv.2014.02.120.

De Kimpe, C.R., Morel, J.L. (2000) Urban soil management: a growing concern. *Soil Science*, **165**(1), 31–40.

Edmondson, J.L., Davies, Z.G., McHugh, N., Gaston, K.J., & Leake, J.R. (2012) Organic carbon hidden in urban ecosystems. *Scientific Reports*, **2**, doi:10.1038/srep00963.

Effland, W.R., & Pouyat, R.V. (1997) The genesis, classification, and mapping of soils in urban areas. *Urban Ecosystems*, **1**(4), 217-228, doi: 10.1023/A:1018535813797

IBGE (Instituto Brasileiro de Geografia e Estatística). (2015) *IBGE Cidades*, http://cidades.ibge.gov.br/v3/cidades/municipio/3552205

Joseph, M., Wang, F., & Wang, L. (2014). GIS-based assessment of urban environmental quality in Port-au-Prince, Haiti. Habitat International, 41, 33-40, doi:10.1016/j.habitatint.2013.06.009

Kaiser, H.F. (1958) The varimax criterion for analytic rotation in factor analysis. *Psychometrika*, **23**(3), 187-200, doi:10.1007/BF02289233

Lavoie, R., Lebel, A., Joerin, F., & Rodriguez, M.J. (2013) Integration of groundwater information into decision making for regional planning: A portrait for North America. *Journal of Environmental Management*, **114**, 496-504, doi:10.1016/j.jenvman.2012.10.056

Ma, H., Liu, L., & Chen, T. (2010) Water security assessment in Haihe River Basin using principal component analysis based on

Kendall τ. *Environmental Monitoring and Assessment*, **163**(1), 539–44, doi:10.1007/s10661-009-0856-2

Moura, L.H.A., Boaventura, G.R., & Pinelli, M.P. (2010) The water quality as an indicator of land use and occupation: Gama basin-DF. *Química Nova*, **33**(1), 97-103, doi: 10.1590/S0100-40422010000100018

Oliveira, F.C., Mattiazzo, M.E., Marciano, C.R., & Rossetto, R. (2002) Efeitos de aplicações sucessivas de lodos de esgoto em LATOSSOLO AMARELO distrófico cultivado com cana-de-açúcar: carbono orgânico, condutividade elétrica, pH e CTC. *Revista Brasileira de Ciência do Solo*, **26**(2), 505-519, doi:10.1590/S0100-06832002000200025

Ouyang, Y. (2005) Evaluation of river water quality monitoring stations by principal component analysis. *Water Research*, **39**(12), 2621-2635, doi:10.1016/j.watres.2005.04.024

Ouyang, Y., Nkedi-Kizza, P., Wu, Q.T., Shinde, D., & Huang, C.H. (2006) Assessment of seasonal variations in surface water quality. *Water Research,* **40**(20), 3800-3810, doi:10.1016/j.watres.2006.08.030

Parinet, B., Lhote, A., & Legube, B. (2004) Principal component analysis: an appropriate tool for water quality evaluation and management—application to a tropical lake system. *Ecological Modelling*, **178**(3-4), 295–311, doi:10.1016/j.ecolmodel.2004.03.007

Pedron, F.A., Dalmolin, R.S.D., Azevedo, A.C., & Kaminski, J. (2004) Solos urbanos. Ciência Rural, 34(5) 1647-1653, doi:10.1590/S0103-84782004000500053

Plieninger, T., Dijks, S., Oteros-Rozas, E., & Bieling, C. (2013) Assessing, mapping, and quantifying cultural ecosystem services at community level. *Land Use Policy*, **33**, 118-129, doi:10.1016/j.landusepol.2012.12.013

Pouyat, R.V., Yesilonis, I.D., Russell-Anelli, J., & Neerchal, N.K. (2007). Soil chemical and physical properties that differentiate urban land-use and cover types. *Soil Science of America Journal*, **71**(3), 1010-1019.

Silva, F.C., Eira, P.A., Barreto, W.O., Pérez, D.V., & Silva, C.A. (1998) *Manual de métodos de análises químicas para avaliação da fertilidade do solo*. Rio de Janeiro: Embrapa-CNPS

Silva, E.B., Silva, A.C., Grazziotti, P.H., Farnezi, M.M.M., Ferreira, C.A., Costa, H.A. O., & Horak, I. (2008) Comparação de métodos para estimar a acidez potencial mediante determinação do pH SMP em organossolos da Serra do Espinhaço meridional. *Revista Brasileira de Ciência do Solo,* **32**(5), 2007-2013, doi: 10.1590/S0100-06832008000500022

Silva, A.M., Santos, A.R., Fernandes, R.A., Urban, R.C., Carvalho, R.M., Manfré, L.A., *et al.* (2013) Hydrosedimentological disequilibrium in a small, urbanized watershed. *Acta Limnologica Brasiliensia*, **25**(2), 140-149, doi: 10.1590/S2179-975X2013000200005

Silva, A.M, Manfré, L.A., Urban, R.C., Silva, V.H.O., Manzatto, M.P., & Norton, L.D. (2015) Organic farm does not improve neither soil, or water quality in rural watersheds from southeastern Brazil. *Ecological Indicators*, **48**, 132-146, doi:10.1016/j.ecolind.2014.07.044

Urban, R.C., Silva, A.M., & Manfré, L.A. (2010) Teores e quantidades de alguns elementos no córrego Lavapés, sub-bacia do Rio Sorocaba, São Paulo, Brasil. *Revista Ambiente e Água*, **5**(3), 265-277, doi:10.4136/ambi-agua.168

USDA (United States Department of Agriculture—Natural Resources Conservation Service). (2008) *Soil quality indicators—bulk density.* USDA.

Vieira, L.S. (1988) *Manual da Ciência do Solo*. 1ª ed. São Paulo: Editora Agronômica Ceres.

Vrščaj, B., Poggio, L., & Marsan, F.A. (2008) A method for soil environmental quality evaluation for management and planning in urban areas. *Landscape and Urban Planning*, **88**(2), 81-94, doi:10.1016/j.landurbplan.2008.08.005

Zeilhofer, P., Lima, E.B.N.R., & Lima, G.A.R. (2006) Spatial patterns of water quality in the Cuiabá River Basin, Central Brazil. *Environmental Monitoring and Assessment*, **123**(1), 41-62, doi: 10.1007/s10661-005-9114-4

Zimmermann, C.M., Guimarães, O.M., & Peralta-Zamora P.G. (2008) Evaluation of the quality of the water body of the Tibagi River in the region of Ponta Grossa using principal components analysis (PCA). *Química Nova*, **31**(7), 1727–1732, doi: 10.1590/S0100-40422008000700025