eurogeo European Journal of Geography

Volume 13, Issue 1, pp. 047 – 068

Article Info
Accepted: 10/03/2022
Corresponding Author: (*) macferre@unicamp.br
DOI: https://doi.org/10.48088/ejg.d.gar.13.1.47.68

Research Article

Regionalization of water quality parameters based on the landscape characteristics of small ungauged basins: a study carried out in south-eastern Brazil

Danilo **Garofalo**[1*] & Marcos **Ferreira**[2]

[1] *Embrapa Environment, Brazil,*
[2] *Institute of Geosciences, UNICAMP, Brazil*

Keywords

Water quality parameters,
Landscape characteristics,
Regionalization,
Small ungagged basins,
Brazil
Spatial Analysis

Abstract

One of the main problems associated with the analysis of water quality parameters (WQPs) in developing countries is the low number of sampling sites in small watersheds. One of the strategies used to solve this problem involves estimating WQPs based on the extrapolation of measurements carried out in other basins using regionalization methods associated with landscape characteristics. The objectives of this study were to evaluate the associations between landscape characteristics and WQPs and to propose a methodology for the regionalization of WQPs based on landscape characteristic data obtained from basins where WQP data were sampled. The study area was the upper Piracicaba-Jaguari River basin located in south-western Brazil. The methodology was based on a survey of 27 environmental variables and 12 water quality parameters in 44 small sub-basins. The non-parametric k-nearest neighbour regression (K-NNR) algorithm was used to estimate the WQP values for the small sub-basins that lacked data. The results showed that the landscape characteristics of the studied sub-basins related to land use and cover significantly influenced the WQPs. The mapped regions showed significant differences among the total dissolved solids, chloride, electric conductivity, pH, salinity, resistivity, dissolved oxygen, and nitrate parameter values.

Highlights:

-Landscape based-approach are efficient for the regionalization of water quality

-Lower water temperatures are associated with larger areas of rainforests

-Lower water temperatures are associated with larger areas of rainforests

-Higher nitrate concentrations are associated with higher areas of pasture

-Land use and cover influence water quality parameters in small river basins

*The publication of the **European Journal of Geography** (EJG) (http://eurogeojournal.eu/) is based on the **European Association of Geographers**' goal to make European Geography a worldwide reference and standard. Thus, the scope of the EJG is to publish original and innovative papers that will substantially improve, in a theoretical, conceptual or empirical way the quality of research, learning, teaching and applying geography, as well as in promoting the significance of geography as a discipline. Submissions are encouraged to have a European dimension. The European Journal of Geography is a peer-reviewed open access journal and is published quarterly.*

1. INTRODUCTION

The growth in population, economic activity, industrialization and agriculture has resulted in an increased demand for water (Qadir et al., 2008). Lakes, reservoirs and rivers are the main sources of continental water for domestic, industrial and irrigation uses. For this reason, it is essential to prevent and monitor water pollution and to have access to reliable information regarding water quality (Varol et al., 2012; Rezaid-Balf et al., 2020).

However, in developing countries such as Brazil, the number of rivers in which water quality parameters (WQPs) are regularly monitored is very small. WQP analyses can be costly depending on the temporal frequency of the observations, the amplitude and accessibility of water bodies and the number of professionals involved in the collection. For this reason, in Brazil, periodic analyses of WQPs are performed on only a few Brazilian water bodies.

Thus, the use of extrapolation and regionalization methodologies for basins without WQP data is fundamental for the water quality management of rivers and lakes. One methodological alternative for estimating WQPs in small sub-basins without in situ measurements involves the use of regionalization models based on landscape data obtained from the sub-basins. The physicochemical and microbiological compositions of river waters are influenced by geomorphological, geological, soil and climatic characteristics as well as by the vegetation cover, land use and land cover of the contributing sub-basins.

The water quality of a water body is very sensitive to the type of landscape that prevails in the watershed. Increasing percentages of agricultural areas is the main factor associated with the degradation of water quality, while increases in forested areas contribute to the improvement of water quality in hydrographic basins (Linlin et al., 2012; Clément et al., 2017). The water quality and pollution in reservoirs and lakes are directly associated with land use (Lee et al., 2009; Li et al., 2016; Pietruszyński and Cieśliński, 2018). The use of integrated spatial approaches based on geographic information systems is an efficient strategy for mapping and monitoring surface and subsurface water pollution (Tcherkezova et al., 2019).

Sub-basin regionalization can be used to spatialize sub-basin clusters with similar water qualities based on datasets composed of WQPs measured in other water bodies and environmental variable values (Larsen et al., 1988; Ravichandran et al., 1996; Pongpetch and Suwanwaree, 2012).

Data extrapolation techniques allow the indirect inference of WQPs from the environmental, land use and land cover characteristics of sub-basins. These methods are important because they make it possible to estimate WQP values for sub-basins without available data using correlations between environmental characteristics and measured WQP data in other sub-basins (Lee et al., 2009; Linlin et al., 2012; Chakraborty et al., 2019).

Learning-based methods such as k-nearest neighbour regression (KNN) have been used to predict and extrapolate the values of water quality variables (Robertson et al., 2003; Dogan et al., 2009; Wang et al., 2010a; Wang et al., 2010b; Mahmoudi et al., 2016; Sattari et al., 2016; Li et al., 2019; Chakraborty et al., 2019). The general goals of these methods are to identify hidden relationships within datasets and build models that reflect the processes that control these systems (Sattari et al., 2016).

In our study, we considered sub-basin groups as landscape units or geosystems characterized by the spatial associations among physical, biological and anthropogenic attributes (Bertrand, 2004). We hypothesized that WQP values are influenced by the landscape characteristics of the contributing sub-basins.

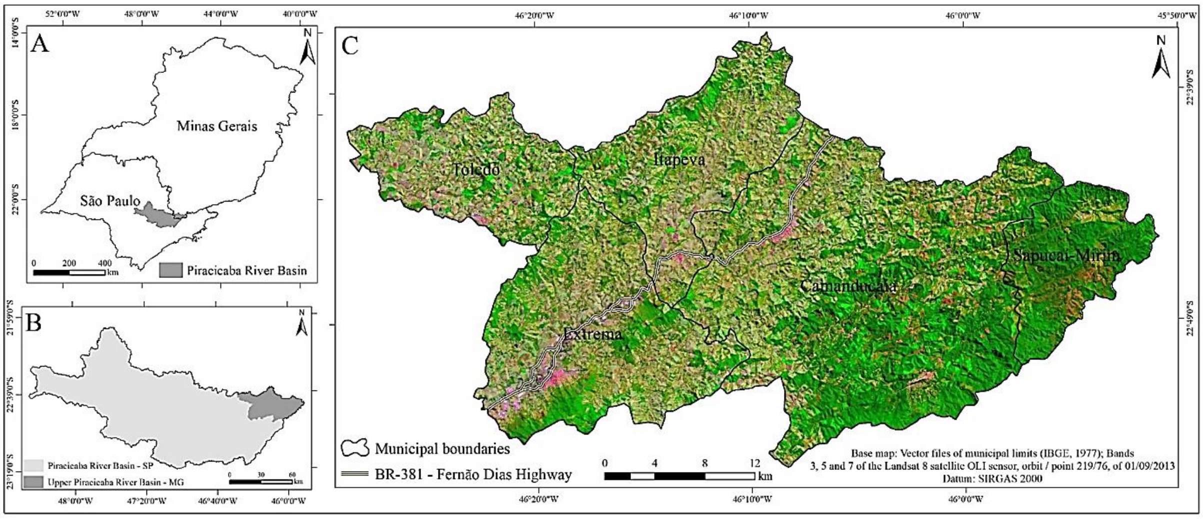
The objectives of this study were to evaluate the associations between landscape characteristics and WQPs and to propose an alternative methodology for the regionalization of WQPs based on environmental sub-basin data and spatial analysis methods for use in sub-basins without on-site monitoring or data collection.

2. STUDY AREA

The study area was the upper Piracicaba-Jaguari River basin (PJ River basin), which has an approximate area of 1,158 km² and is located in the state of Minas Gerais, Brazil, in the Fernão Dias Environmental Protection Area (EPA) (Figure 1).

According to the Köppen climate classification method, the PJ River basin is in the Cwb domain and has a tropical highland climate with dry winters and rainy summers (Sparovek et al., 2007), average temperatures below 22°C in summer and 15°C in winter, and an average annual rainfall total of 1,700 mm. The Piracicaba-Jaguari River basin is an important source of water in south-eastern Brazil and supplies water to part of the metropolitan region of the São Paulo municipality, the most populous metropolis in South America.

Figure 1. Map of the location of the Piracicaba-Jaguari River basin in the south-eastern Brazil.



Base map: Vector files of municipal limits (IBGE, 1977); Bands 3, 5 and 7 of the Landsat 8 satellite OLI sensor, orbit / point 219/76, of 01/09/2013
Datum: SIRGAS 2000

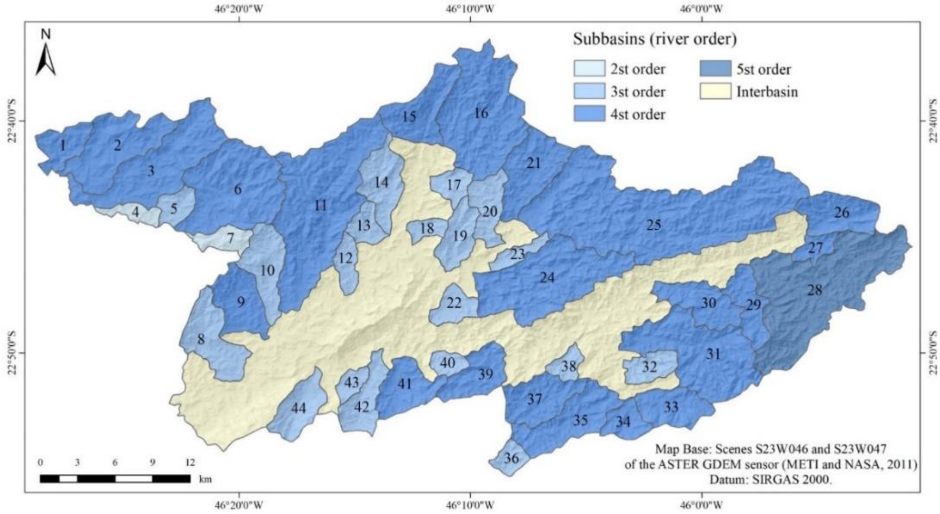
3. MATERIAL AND METHODS

The methodology used in this work is based on the general systems theory (Bertalanffy, 1973; Strahler, 1980). Following this theory, we hypothesized that the WQP values (response variables) of a given hydrographic source are due to the landscape characteristics (explanatory variables) of the source.

3.1 Mapping of the PJ River Sub-Basins

The PJ River basin was divided into sub-basins with catchment areas greater than 2 km² using the Strahler method (Strahler, 1952). The sub-basins that did not meet these specifications were considered inter-basin areas. We used the following materials to delimit the sub-basins: a drainage network map of the Ottocodified Hydrographic Base (ANA, 2015), topographic maps at a 1:50,000 scale and a shaded relief image generated from ASTER GDEM-2 sensor data. Following this procedure, 44 sub-basins were mapped in the PJ River basin (Figure 2).

Figure 2. Map of the sub-basins in the PJ River basin, classified according to the drainage network hierarchy (order).



3.2 Landscape Characteristics Data Collection

The environmental database was composed of 27 variables related to land use and land cover, road density, slope, terrain shape and the morphometry of the sub-basins. The landscape characteristic values were calculated for the 44 sub-basins of the PJ River basin.

3.2.1 Land use and cover

The land use and land cover of the sub-basins were mapped using digital image classification techniques and Landsat-8-OLI multispectral images from 09/01/2013. The multispectral bands were pre-processed using the FLAASH atmospheric correction method (Felde et al., 2003). Then, the multispectral bands with 30-m spatial resolutions were merged with the panchromatic band (at a 15-m resolution).

A land use map was created using object-oriented classification techniques in ENVI 5.4 software (Exelis Visual Information Solutions, 2010) by applying the Feature Extraction and Support Vector Machine (SVM) classifier tools (Vapnik, 1995). The following land use and cover classes were mapped: rainforests, shrub forests, forestry lands, pasture lands, agricultural lands, urban lands, rock outcrops and water bodies. The land use and cover map was validated in the field using the Kappa index (Landis and Koch, 1977). Then, the proportion of the area of each land use and cover class was calculated for all 44 analysed sub-basins.

3.2.2 Road density

Road density mapping in the sub-basins was conducted using topographic maps at a 1:50,000 scale (IBGE, 1977) and an R3G5B7 colour composition multispectral image from the Landsat-8 OLI product. The road density was considered to be the ratio of the total length of roads in each sub-basin to the sub-basin area.

3.2.3 Relief shape

To calculate the terrain shape of each sub-basin, we used the vertical curvature (VN) and horizontal curvature (HN) of the terrain data obtained from the Topodata Project of the National

Institute for Space Research (INPE) (Valeriano et al., 2008). The Topodata digital elevation model was clipped to a vector file of the geographical boundaries of the PJ River basin (IGAM, 2010).

3.2.4 Slope

A slope map was created with ArcGIS 10.5 (ESRI, 2016) and ASTER GDEM2 data (ASTER GDEM, 2014) and represented in percentage units in six classes, according to the methodology used by Embrapa (2006). Then, the average slope values were calculated for the 44 sub-basins.

3.2.5 Morphometric variables of the sub basins

The morphometric variables of the sub-basins were calculated using data from vector-format topographic charts of the upper Piracicaba-Jaguari River basin (ANA, 2015) and ASTER GDEM-2 images. Table 1 shows the equations used to calculate the morphometric variables. The morphometric variable maps were classified into six classes using the Sturges method, and the class intervals were defined by applying the Jenks classification method.

3.3 Water Quality Parameter Data

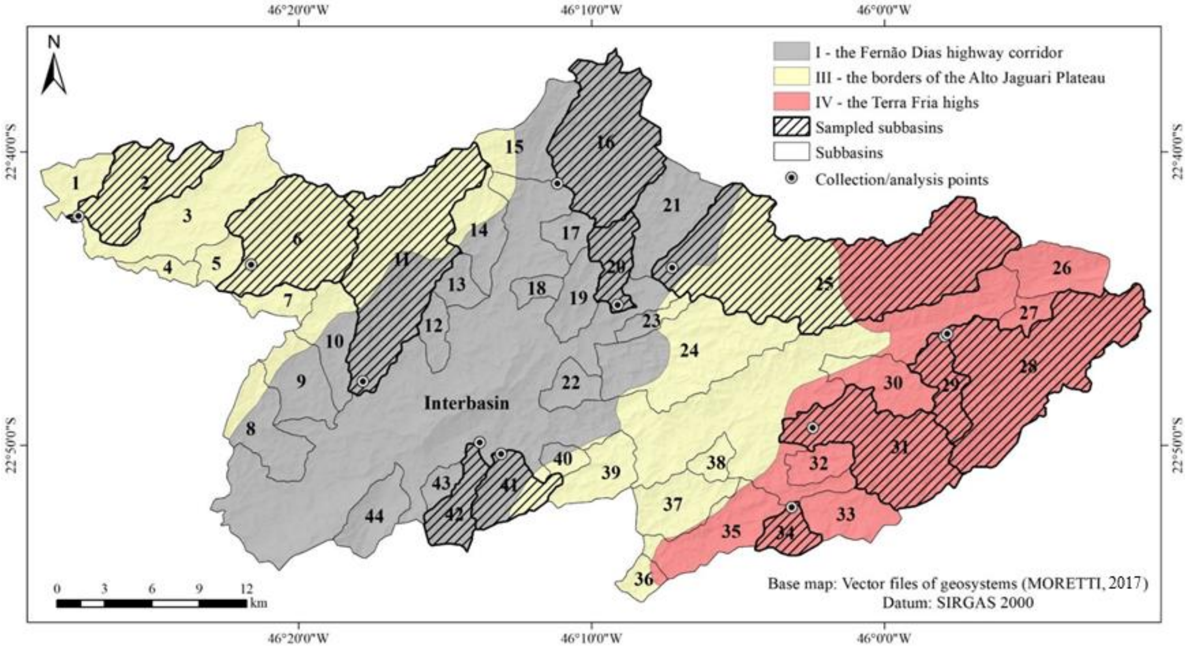
3.3.1 Sampling of WQPs

A sample of 12 sub-basins representing 53.04% of the total area of the PJ River basin was used to collect WQP data (Figure 3). Systematic stratified sampling was used with the objective of selecting representative sub-basins of landscape units in the PJ River basin (Moretti and Ferreira, 2017). The sub-basins for which the WQPs were sampled were selected according to the hierarchy of the sub-basins and the accessibility of the sub-basins to their main channels. Priority was given to the 4th-order sub-basins covering most of each landscape unit area, which had accessible sampling points. The WQP sampling was carried out in August-September (dry season) and December (wet season).

Table 1. Morphometric variables used in the characterization of the sub-basins.

Morphometric variable	**Equation**
Drainage Density (DD)	$DD = \frac{L_t}{A}$ where L_t is the total length of river channels A is the sub basin area
Hydrographic Density (HD)	$HD = \frac{N_t}{A}$ where N_t is the total number of river channels A is the sub basin area
Maintenance Coefficient (MC)	$MC = \frac{1}{Dd.1000}$
Superficial Path Extension (SPE)	$SPE = 0.5Dd.1000$
Topographic Texture (TT)	$Log\ (TT) = 0.219649 + 1.115.\log(Dd)$
Roughness Index (RI)	$RI = \Delta H.DD$

Figure 3. Map of the sampled sub-basins and landscape units of the PJ River basin.



3.3.2 Analysis of the WQP data

The analytical tests employed herein followed the procedures established for physicochemical and microbiological surface water variables and were described in the standard methods for the examination of water and wastewater published by the American Public Health Association (APHA, 1998) and the National Guide for Sample Collection and Preservation published by the National Water Agency (ANA, 2011). The physical, chemical and microbiological parameters used in this study are presented in Table 2.

The physicochemical data collection conducted in the main river channel of each sampled sub-basin was performed in triplicate using the multiparameter device model HI9829 (HANNA, 2013). Sampling was performed at 10-minute intervals at the same point in each river. The equipment calibration followed the manufacturer's technical recommendations (HANNA, 2013).

The microbiological analyses of the total coliforms and *E. coli* were performed in the laboratory by serial dilution with a chromogenic test. An iMPNplateTM-1600 device and chromogenic substrate ONPG-MUG were used to count the microorganisms (CPI, 2012). A Termobac Microprocess portable oven was used to incubate the devices with the samples, and measurements of the variables were taken after 24 hours of incubation at 36 ± 0.5°C. The WQP values were sampled in triplicate in the dry and rainy periods for a total of 72 samples.

3.4 Regionalization of Water Quality Parameters

3.4.1 Regionalization variables

The regionalization of WQPs for all sub-basins, including those without collected data, was carried out using a model to determine the correlations between the landscape characteristics (explanatory variables) and WQP values (response variables). The set of explanatory variables was composed of the data of 27 landscape characteristics, 18 of which corresponded to geomorphological characteristics and 9 of which corresponded to land use and vegetation cover characteristics (Table 3), determined for the population of 44 sub-basins in the PJ River basin.

The dataset referring to the response variables was composed of data comprising twelve WQPs, five of which characterized physical properties, five characterized chemical properties and two characterized microbiological properties (Table 2). The values of these parameters were obtained in 12 of the 44 sub-basins in the dry and wet seasons in triplicate, totalling 72 samples for each water quality parameter. The regionalization procedure was based on the k-nearest neighbour regression (k-NNR) extrapolation model.

Table 2. Physical, chemical and microbiological water quality parameters (WQPs).

	Parameter	Unit
Physical parameters	Temperature (TMP)	°C
	Electric conductivity (ECD)	(μS/cm)
	Resistivity (RES)	(Ω.cm)
	Turbidity (TUR)	(NTU)
	Total dissolved solids (TDS)	(mg/L)
Chemical parameters	pH	-
	Dissolved oxygen (DIO)	(mg/L)
	Nitrate (NIT)	(mg/L)
	Chloride (CHL)	(mg/L)
	Salinity (SAL)	PSU
Microbiological parameters	Total coliforms (TCL)	(most probable number - MPN)
	E. coli (ECO)	(most probable number - MPN)

3.4.2 k-NNR extrapolation model and mapping of the water quality regions

According to the k-NNR model, each object (sub-basin) was represented as a point in an attribute space, and the distance between each pair of points was calculated using the Euclidean distance technique (Faceli, 2011). Each sub-basin was extrapolated through the local interpolation of the sub-basin associated with the nearest neighbours of the training set. The *k-value* is defined by the user, and a small and odd value is recommended (Faceli, 2011).

Extrapolation of the WQPs based on the landscape characteristic values was performed using the entire database, and the weighted averages of the distances among the values were obtained for the three nearest neighbours (k = 3). This step was performed using the Scikit-learn package in the Python language (Rossum and Boer, 1991).

The water quality regions were mapped based on a combination of the extrapolated and measured WQP values and based on the values of the landscape characteristics calculated for the sub-basins. Ward's hierarchical method (Ward, 1963) was used to analyse the homogeneous sub-basin groups. Ward's method is an agglomerative hierarchical grouping method that uses the sum of squares as a criterion to decide which individuals or groups should be merged at each stage of the procedure (Everitt, 2006).

This method is based on the changes in the variations among groups and within groups and is implemented during each grouping step (Mingoti, 2005). The Ward method procedure starts with each object as a single group; then, at each new step in the grouping process, the algorithm calculates the error based on the sum of squares within the group to determine the next two groups to be merged at each step of the algorithm (Ferreira and Hitchcock, 2009).

The number of regions formed during the grouping method was estimated using the silhouette index (Rousseeuw, 1987; Amorim et al., 2015). The silhouette value was calculated using the Euclidean distance to evaluate the grouping of two to six groups. This step was performed in Python (Rossum and Boer, 1991) using the Scikit-learn package.

Table 3. Landscape characteristics used in the *k*-nearest neighbour regression model and the respective variables and units of measurement.

Landscape characteristics	**Variables**
Morphometry	Drainage Density (Dd)
	Hydrographic Density (Dh)
	Maintenance Coefficient (Cm)
	Superficial Path Extension (Eps)
	Topographic Texture (Tt)
	Roughness index (Ir)
Slope	0 to 3 %
	3 to 8 %
	8 to 13 %
	13 to 20 %
	20 to 45 %
	> 45 %
Relief shape (%)	Straight (STR)
	Concave (CCV)
	Convex (CVX)
	Planar (PLA)
	Convergent (CNV)
	Divergent (DIV)
Land use and cover (% of sub basin area)	Agriculture (AGR)
	Pasture (PAS)
	Forestry (FOR)
	Rain forest (RFR)
	Shrub forest (SFR)
	Urban area (UBA)
	Water (WAT)
	Rock outcrop (ROU)
Roads	Road density - ROD (km/km²)

3.4.3 Evaluating the associations between WQPs and landscape characteristics

An analysis of the associations between the WQP values and landscape characteristics related to morphometry, slope, relief shape, land use and land cover and roads (Table 3) was performed by calculating the Pearson's correlation coefficient values. To do this, the free statistical software JASP was used (JASP Team, 2020). Correlation coefficients with significance levels of $p < 0.01$ and $r \geq 0.60$ were considered statistically significant.

4. RESULTS AND DISCUSSION

4.1 Environmental variable data

Table 4 presents the proportions of the areas occupied by different land use and land cover classes. Table 5 presents the proportions of the sub-basins corresponding to each relief shape class and the values of the morphometric parameters. The area proportions corresponding to the slope classes and the road density values are presented in Table 6.

4.2 Sampling of water quality parameters

The results obtained for the dry season indicated that all sampled sub-basins presented good physicochemical conditions according to the water quality standards set by the National Environment Council (Brasil, 2005) (CONAMA Resolution no. 357, March 17, 2005) for Classes I and II in freshwater. With regard to microbiological properties, the CONAMA standards state that the concentration of *E. coli* in freshwater Classes I and II should not exceed the limit of 200 MPN/100 mL. Among the twelve sub-basins sampled, only the main sub-basin river channel, which corresponded to points 1, 9, 10, 11 and 12, was in compliance with this resolution. Table 7 shows the average value of each parameter for the sampled sub-basins in winter.

During the wet season, the sample collections were carried out on December 14, 16 and 18 of 2015. In the five days prior to the sample collection on December 14, the accumulated precipitation values were, on average, 10.42 mm in sub-basin 1, 14.61 mm in sub-basin 2, 17.45 mm in sub-basin 3 and 18.56 mm in sub-basin 4. For the sampling on December 16, the average accumulated rainfall values in sub-basins 5, 6, 7 and 8 were 32.28 mm, 21.37 mm, 12.78 mm and 14.16 mm, respectively.

Table 4. Proportions of areas occupied by land use and land cover classes in the sub-basins of the PJ River basin.

Sub-basins	Land Use and Cover Classes (%)							
	UBA	ROU	SFR	WAT	AGR	PAS	FOR	RFR
1	4.05	4.15	2.96	0.12	14.75	40.84	5.17	27.96
2	4.91	3.48	3.21	0.23	14.73	43.86	5.66	23.92
3	7.45	4.95	5.52	0.11	13.34	39.77	2.72	26.14
4	4.70	2.53	7.64	0.50	21.04	32.58	4.57	26.44
5	13.65	4.75	7.16	0.26	9.91	26.81	11.44	26.02
6	7.94	5.09	3.71	0.19	8.98	34.29	8.08	31.72
7	5.03	5.74	4.63	0.18	5.99	25.90	1.64	50.89
8	5.62	4.65	3.71	0.31	5.82	34.73	5.11	40.06
9	7.27	3.64	4.74	0.08	6.39	26.50	2.78	48.60
10	7.92	4.28	4.20	0.01	5.95	28.63	3.58	45.44
11	6.12	5.04	4.34	0.15	6.92	35.08	3.87	38.47
12	6.73	4.76	8.11	0.02	5.26	47.60	1.75	25.77
13	2.74	4.39	4.49	0.10	4.44	47.94	3.22	32.68
14	5.93	4.21	3.76	0.07	7.80	39.77	3.24	35.23
15	6.45	2.61	4.14	0.05	6.46	22.85	9.04	48.40
16	4.19	5.86	4.90	0.05	6.13	45.71	4.75	28.41
17	5.97	5.25	4.39	0.12	4.62	39.70	2.96	36.98
18	6.69	6.88	5.02	0.23	7.41	49.77	2.51	21.47
19	3.70	6.92	5.09	0.05	3.94	34.58	4.12	41.60
20	5.61	5.88	2.31	0.13	6.55	38.88	1.17	39.47
21	7.22	6.34	4.47	0.21	6.60	40.13	3.62	31.42
22	3.44	4.03	5.84	0.02	9.69	36.95	2.99	37.03

Sub-basins	Land Use and Cover Classes (%)							
	UBA	ROU	SFR	WAT	AGR	PAS	FOR	RFR
23	8.42	10.70	5.14	0.07	8.14	43.40	1.27	22.86
24	3.73	4.77	2.86	0.07	5.92	28.74	5.91	47.99
25	3.64	6.27	4.06	0.07	7.81	26.44	4.49	47.20
26	0.81	4.08	2.36	0.02	3.11	8.33	1.67	79.62
27	0.12	1.38	0.33	0.00	0.51	0.51	14.05	83.10
28	0.26	1.87	2.67	0.00	0.32	1.91	17.04	75.93
29	0.68	3.54	3.11	0.00	2.00	5.80	9.84	75.03
30	3.33	5.83	2.37	0.26	4.70	12.80	6.62	64.08
31	0.67	2.26	0.96	0.02	1.04	1.12	14.78	79.15
32	2.28	4.39	1.76	0.09	1.94	0.99	27.26	61.30
33	5.22	3.11	0.70	0.00	0.88	0.72	4.05	85.32
34	0.37	2.26	0.26	0.00	0.03	0.05	14.01	83.02
35	0.78	1.64	0.71	0.03	1.21	6.56	19.56	69.52
36	3.01	6.24	3.09	0.13	4.83	24.45	6.83	51.42
37	2.21	2.95	2.66	0.04	5.57	12.54	10.00	64.02
38	3.11	2.22	1.12	0.01	0.94	0.33	20.12	72.15
39	5.70	4.49	1.85	0.03	15.31	37.37	5.26	29.98
40	2.05	2.86	0.83	0.03	4.18	31.49	4.35	54.20
41	5.12	4.68	4.66	0.08	6.36	39.58	5.06	34.45
42	4.56	5.42	2.95	0.11	6.85	55.11	3.03	21.96
43	5.91	3.57	4.54	0.00	5.72	51.40	0.98	27.88
44	2.34	3.06	3.10	0.05	4.49	22.27	10.25	54.45

Table 5. Proportions of the sub-basins corresponding to each relief shape class and the values of the morphometric parameters of the sub-basins of the PJ River basin.

Sub-basin	Relief Shape (%)			Morphometric Parameters					
	Concave	Convex	Straight	Dd	Dh	Cm	Eps	Tt	Ir
1	47.21	40.7	12.10	1.87	3.10	536.01	268.00	1.61	768.64
2	46.70	35.12	18.17	1.52	2.33	656.93	328.47	1.52	649.99
3	49.96	36.15	13.87	1.72	3.08	582.05	291.03	1.57	800.62
4	50.9	38.43	10.68	1.35	2.27	742.28	371.14	1.46	179.18
5	51.88	34.44	13.68	1.74	2.83	574.58	287.29	1.58	567.37
6	51.97	36.30	11.73	1.79	2.76	557.61	278.81	1.59	882.33
7	44.91	39.62	15.45	1.72	1.94	582.82	291.41	1.57	789.26
8	46.12	38.91	14.97	1.71	2.81	583.56	291.78	1.57	944.21
9	47.71	37.39	14.90	1.67	1.89	600.56	300.28	1.56	954.11
10	50.78	34.69	14.55	1.90	3.33	527.56	263.78	1.61	1177.11
11	50.12	37.33	12.55	1.89	3.12	530.19	265.10	1.61	1218.42
12	50.8	35.88	13.33	2.34	5.82	426.56	213.28	1.70	862.71
13	53.42	37.27	9.32	1.81	2.72	552.13	276.07	1.59	628.47
14	51.42	38.96	9.61	2.10	3.47	475.92	237.96	1.66	1208.19
15	45.43	39.65	14.91	2.13	3.70	469.25	234.63	1.66	1321.25
16	53.49	37.94	8.57	2.14	3.84	466.88	233.44	1.67	1276.56
17	53.15	34.22	12.62	2.13	3.74	468.75	234.37	1.66	800.00
18	55.54	35.09	9.37	2.14	3.18	466.41	233.21	1.67	870.47
19	54.94	31.43	13.62	1.75	2.45	569.99	285.00	1.58	752.64
20	56.41	36.98	6.62	1.86	3.21	537.97	268.99	1.60	622.71
21	58.63	36.02	5.35	1.83	2.77	546.56	273.28	1.60	578.16
22	53.03	29.73	17.24	1.92	3.12	521.35	260.67	1.62	780.67
23	56.51	33.41	10.08	1.82	2.28	550.27	275.13	1.59	908.65
24	53.5	39.16	7.35	2.38	4.28	419.45	209.72	1.71	1866.74
25	53.39	38.47	8.13	2.38	4.38	420.18	210.09	1.71	2427.52
26	53.33	36.70	9.96	2.95	4.99	339.40	169.70	1.80	1599.90
27	51.29	39.15	9.55	3.44	11.51	290.56	145.28	1.87	1087.54

Sub-basin	Relief Shape (%)			Morphometric Parameters					
	Concave	Convex	Straight	Dd	Dh	Cm	Eps	Tt	Ir
28	54.20	36.35	9.44	3.47	9.70	288.07	144.04	1.88	1926.59
29	50.11	40.32	9.57	3.08	7.81	324.56	162.28	1.82	1451.19
30	54.12	35.06	10.83	2.80	6.48	356.70	178.35	1.78	1239.12
31	53.67	37.99	8.34	2.62	5.41	382.34	191.17	1.75	1553.57
32	51.35	41.88	6.78	2.63	4.67	380.90	190.45	1.75	850.61
33	54.41	34.91	10.68	2.31	3.79	433.51	216.76	1.70	1132.61
34	50.49	32.92	16.59	2.53	4.93	395.52	197.76	1.74	1284.39
35	50.67	38.24	11.09	2.85	6.77	350.45	175.22	1.79	1660.74
36	56.70	36.89	6.42	2.71	4.75	369.03	184.52	1.77	1032.43
37	52.62	40.89	6.51	2.59	6.22	386.08	193.04	1.75	1012.75
38	49.41	39.63	10.96	2.07	3.63	483.11	241.55	1.65	705.85
39	56.03	35.13	8.83	1.99	3.14	503.73	251.87	1.63	1048.17
40	49.32	43.55	7.12	2.58	3.92	387.31	193.65	1.75	1368.43
41	54.03	33.94	12.04	1.73	2.30	578.50	289.25	1.57	1101.12
42	54.78	31.63	13.59	2.12	3.62	471.27	235.64	1.66	1088.55
43	48.74	34.25	17.01	2.09	3.10	478.95	239.48	1.65	937.46
44	50.10	31.04	18.86	1.51	1.65	663.10	331.55	1.51	1048.10

Table 6 - Proportions of the sub-basin areas occupied by slope and road density classes in the sub-basins of the PJ River basin.

Sub-basin	Slope Classes (%)						Road Density (km/km²)
	< 3	3-8	8-13	13-20	20-45	>45	
1	3.36	6.90	11.56	20.99	50.24	6.95	1.58
2	6.09	9.07	12.83	21.45	45.98	4.58	1.87
3	6.26	8.75	12.58	21.41	45.35	5.65	1.91
4	7.58	11.99	15.71	24.87	38.86	1.00	1.61
5	5.85	8.97	12.72	20.88	44.77	6.81	1.88
6	3.37	6.86	10.22	18.12	51.30	10.12	1.59
7	2.92	6.54	10.99	20.37	52.56	6.63	0.68
8	4.22	7.90	11.69	19.46	48.36	8.37	1.75
9	4.62	7.58	11.68	22.32	46.75	7.06	1.36
10	5.75	7.03	10.99	18.78	49.24	8.21	1.51
11	3.86	6.53	10.04	18.48	50.79	10.30	1.75
12	4.15	8.07	12.73	23.69	45.84	5.52	1.85
13	4.68	8.20	10.38	17.03	47.82	11.88	1.96
14	2.69	5.80	8.17	15.29	53.41	14.65	1.96
15	2.47	6.07	8.60	15.85	51.46	15.55	1.94
16	3.34	5.97	8.72	16.29	51.12	14.56	1.55
17	3.40	6.12	8.69	15.41	49.81	16.57	1.77
18	4.23	7.73	10.21	18.14	50.63	9.06	2.04
19	2.80	4.29	7.75	16.09	56.93	12.15	1.29
20	4.42	6.56	10.77	18.70	51.60	7.94	1.78
21	5.09	8.21	11.06	17.30	47.19	11.16	1.60
22	5.89	7.23	11.05	19.35	45.87	10.61	1.02
23	3.38	6.75	9.78	17.53	46.91	15.65	0.69
24	3.51	5.35	7.81	13.96	50.04	19.32	1.07
25	3.85	5.87	8.84	15.67	50.31	15.46	1.07
26	2.34	3.31	5.58	11.77	51.34	25.65	0.64
27	2.58	5.48	8.18	15.95	50.99	16.82	0.77
28	5.24	6.77	9.07	16.23	50.11	12.59	0.45
29	3.38	6.61	9.87	16.19	49.31	14.65	1.29
30	4.55	7.80	11.11	17.97	47.98	10.59	1.24
31	3.80	5.10	7.36	13.30	50.02	20.41	0.66
32	4.05	5.38	7.30	13.10	53.04	17.12	1.78

Sub-basin	**Slope Classes (%)**						**Road Density (km/km²)**
	< 3	**3-8**	**8-13**	**13-20**	**20-45**	**>45**	
33	4.28	6.11	8.72	16.46	49.63	14.80	0.63
34	4.25	6.29	8.65	14.40	50.37	16.04	0.14
35	5.53	6.76	8.81	14.96	48.06	15.88	1.15
36	3.69	4.81	8.34	14.12	51.68	17.37	1.27
37	4.77	4.96	7.24	14.70	54.98	13.35	0.79
38	2.67	5.72	8.07	15.56	49.85	18.12	1.78
39	4.05	6.13	8.14	15.40	48.73	17.56	0.85
40	1.07	2.01	4.60	11.33	54.74	26.25	0.85
41	2.36	4.80	8.11	15.84	49.64	19.26	0.97
42	2.77	4.06	6.97	15.81	54.94	15.46	1.03
43	4.90	10.16	14.03	22.07	38.50	10.34	0.93
44	3.02	4.49	6.37	11.93	52.45	21.74	1.25

Table 7 - Average values of water quality variables in winter by sampled sub-basin.

Sampled point (sub-basin)	**Physical parameters**					**Chemical parameters**					**Microb. parameters**	
	Temp *°C*	**EC** *μS/cm*	**Resist** *Ω.cm*	**Turb** *NTU*	**TDS** *mg/L*	**pH**	**DO** *mg/L*	**NO_3^-** *mg/L*	**Cl^-** *mg/L*	**Sal** *PSU*	**TC** *MPN*	***E. coli*** *MPN*
1 (2)	14.37	33	0.030	8.9	16.6	7.1	6.81	0.27	1.5	0.02	1,600	170
2 (6)	17.28	30	0.033	12.9	14.6	7.0	7.34	0.28	1.13	0.02	1,600	1,600*
3 (11)	15.87	29.3	0.034	8.3	14.6	7.4	7.07	0.26	0.74	0.02	1,600	920*
4 (16)	18.25	52	0.019	4.1	26.0	7.4	6.86	0.64	2.97	0.02	1,600	920*
5 (25)	15	26.3	0.038	23.13	13.0	7.2	6.77	0.30	0.79	0.02	1,600	920*
6 (20)	17.01	44	0.022	2.8	22.0	7.2	6.93	0.54	3.50	0.02	1,600	350*
7 (42)	18.52	76	0.013	3.4	38.0	7.6	7.32	0.70	10.87	0.03	1,600	540*
8 (41)	14.24	37	0.027	5.3	27.3	6.9	6.74	0.21	2.77	0.02	1,600	1,600*
9 (34)	11.63	9	0.111	0.3	4.00	6.5	7.8	0.08	1.67	0.00	220	130
10 (31)	11.97	14	0.071	1.8	7.00	6.9	7.86	0.12	1.47	0.01	350	79
11 (29)	11.44	13	0.076	3.1	7.00	6.9	7.75	0.13	1.23	0.01	1,600	46
12 (28)	11.69	12	0.083	4.4	6.00	6.9	8.53	0.12	1.33	0.01	920	170

* not in accordance with CONAMA Resolution No. 357 of 2005.

The period of five days preceding the sample collection campaign on December 18 recorded average accumulated rainfall values of 51.42 mm in sub-basin 9, 51.66 mm in sub-basin 10, 48.32 mm in sub-basin 11 and 42.86 mm in sub-basin 12. Four of the twelve sampled sub-basins presented nonconformities with the turbidity parameter (Table 6), while the other physicochemical variables were in good condition according to the water quality standards set by CONAMA for freshwater Classes I and II. On average, the physical-chemical variables showed a 66.7% compliance rate. Regarding the microbiological properties of the sampled water, only two sub-basins (9 and 11) showed *E. coli* concentrations below the limit established by the CONAMA Resolution for freshwater in Classes I and II (Table 8).

Table 8 - Average values of WQPs sampled in summer.

Sampled sub basins	Physical parameters					Chemical parameters						Microbi. parameters	
	Tem p °C	EC µS/cm	Resis Ω.cm	Turb. UNT	TDS mg/L	pH	DO mg/L	NO_3^- mg/L	Cl^- mg/L	Sal. PSU	TC MPN	*E. coli* MPN
1	22.27	50	0.02	187.3*	25	6.88	7.31	0.78	3.9	0.02	>1,60	>1,600*
2	23.28	53	0.0189	109.3*	27	6.68	6.49	0.74	3.83	0.02	>1,60	>1,600*
3	21.01	46	0.0213	180*	23	6.98	7.74	0.6	2.9	0.02	>1,60	>1,600*
4	21.03	76	0.0132	34.83	38	6.99	7.53	0.7	3.03	0.03	>1,60	>1,600*
5	20.40	38	0.0263	159.7*	19	6.9	7.44	0.43	1.77	0.02	>1,60	>1,600*
6	21.84	66	0.0152	15	33	7.17	7.17	0.47	2.53	0.03	>1,60	540*
7	23.00	103	0.0097	19.3	52	7.46	5.6	0.49	9.93	0.05	>1,60	920*
8	23.52	83	0.0121	35.9	42	7.04	5.73	0.45	2.73	0.04	>1,60	>1,600*
9	16.39	10	0.1	2.8	5	6.27	7.8	0.21	< 0.6	0.00	1,60	79
10	17.72	15	0.0667	10.1	8	6.7	7.32	0.26	< 0.6	0.01	>1,60	350*
11	18.27	16	0.0625	13.4	8	6.91	7.85	0.27	< 0.6	0.01	1,600	79
12	18.92	14	0.071	74.5	7	6.75	7.95	0.2	< 0.6	0.01	>1,60	1,600*

* not in accordance with CONAMA Resolution No. 357 of 2005.

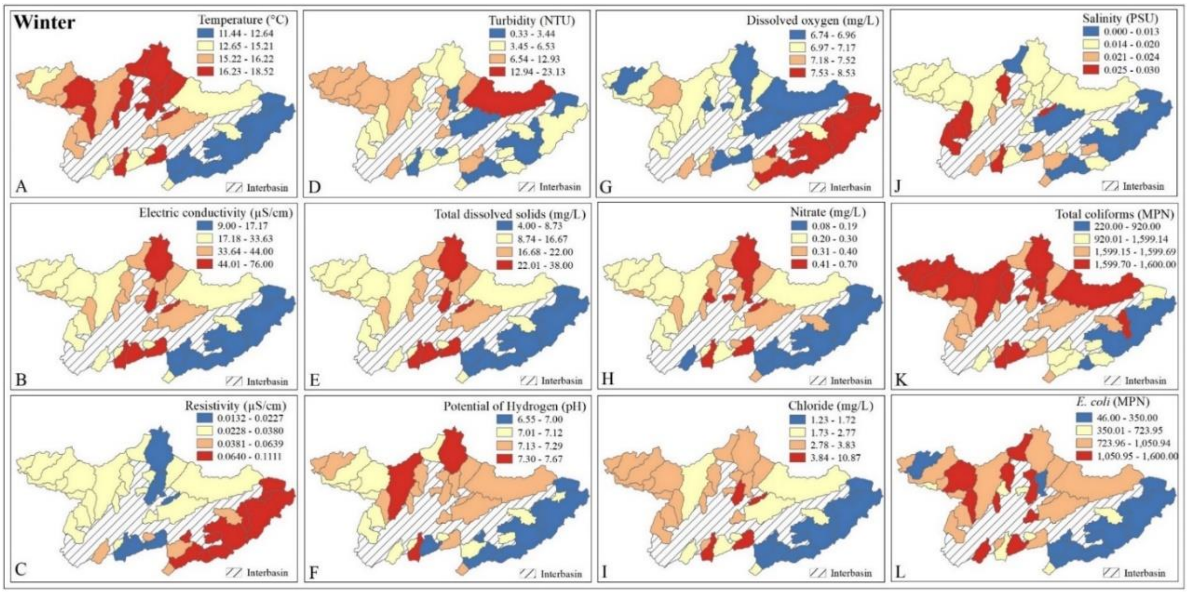
4.2 Extrapolation of the WQPs using the k-NNR Model

Figures 4 and 5 show the extrapolation of the WQPs by the k-NNR model for the 32 sub-basins in the dry and wet seasons, respectively, in addition to the WQP values measured in the main rivers of the 12 sampled sub-basins used to train the model.

4.2.1 Evaluating the associations between WQPs and landscape characteristics

Table 9 presents the Pearson's correlation coefficients between the WQPs and landscape characteristic values. The coefficients that are significant (p<0.01) and higher than r=0.60 are highlighted in bold in the table.

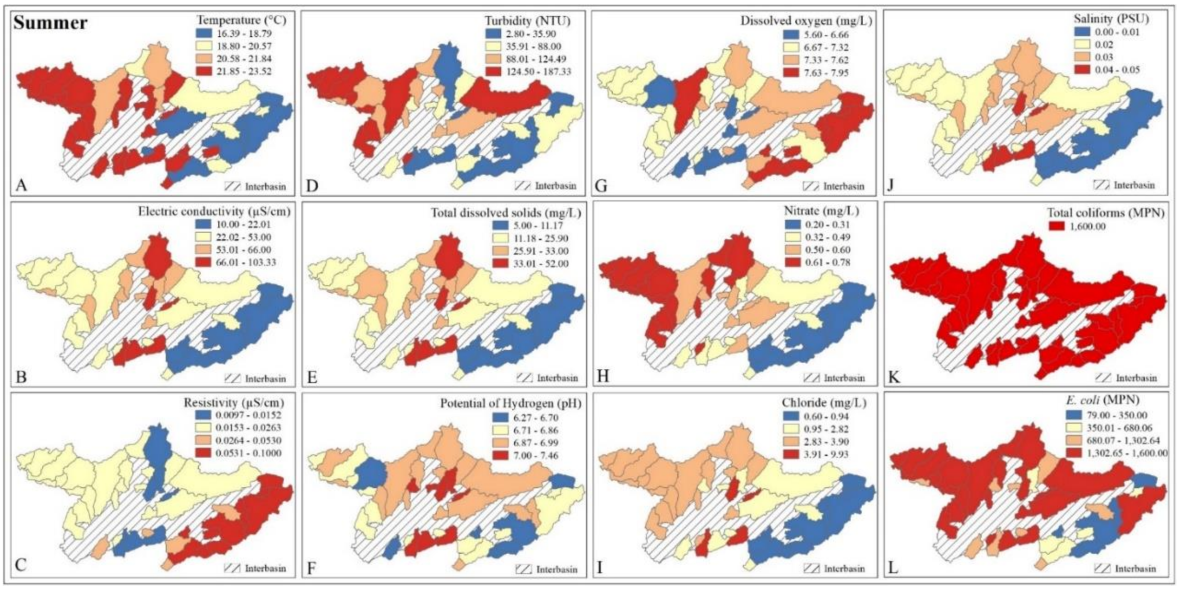
Figure 4 - Maps of the extrapolated concentrations of WQPs in the sub-basins of the PJ River basin during the dry season (winter).



Regarding pH, we observed that the highest values were associated with sub-basins with larger areas of pasture lands (r = 0.77, p <0.01), rainforests (r = 0.73, p <0.01) and rock outcrops (r = 0.60, p <0.01). Lower water temperatures were associated with sub-basins with larger areas of rainforests (r = -0.91, p <0.01) and forestry lands (r = -0.68, p <0.01), higher drainage density values (r = -0.72, p <0.01), and higher topographic texture values (r = -0.69, p <0.01).

Higher turbidity values were associated with sub-basins with smaller surface areas and with slopes above 45% (r = -0.60, p <0.01). Notably, the sub-basins with high dissolved oxygen values were those that had the highest proportions of rainforest areas (r = 0.78, p <0.01) and forestry areas (r = 0.62, p <0.01) and lower proportions of pasture areas (r = -0.78, p <0.01). In addition, dissolved oxygen was higher in sub-basins with higher drainage densities (r = -0.74, p <0.01) and higher hydrographic densities (r = -0.71, p <0.01).

Figure 5. Maps of the extrapolated concentrations of WQPs in the sub-basins of the PJ River basin during the rainy season (summer).



We noted that the highest nitrate concentrations were associated with sub-basins with higher proportions of pasture areas (r = 0.90, p <0.01), urban areas (r = 0.68, p <0.01) and agricultural lands (r = -0.63, p<0.01). On the other hand, sub-basins with higher proportions of rainforests (r = -0.89, p <0.01) and forestry areas (r = -0.69, p <0.01) presented waters with lower nitrate concentrations. This association was also observed in sub-basins with higher drainage densities (r = -0.66, p <0.01), river densities (r = -0.60, p<0.01) and topographic textures (r = -0.62, p <0.01).

Filoso et al. (2003) observed positive correlations between nitrogen concentrations and agricultural and urban land use types. These authors also found a negative correlation between forested areas and the nitrate concentrations in sampled waters. A study by Esteves et al. (2015) confirmed that areas occupied by pasture lands showed high positive correlations with high values of nitrate and EC, especially during the rainy season. Higher chloride values were associated with sub-basins with higher proportions of pasture areas (r = 0.73, p <0.01), while lower chloride values were associated with sub-basins with higher proportions of rainforests (r = -0.72, p <0.01).

We also observed that the highest total dissolved solids values were associated with sub-basins that had higher proportions of pasture areas (r = 0.84, p <0.01) and rock outcrops (r = 0.63, p <0.01). On the other hand, lower values of this parameter were associated with sub-basins with high proportions of rainforests (r = -0.82, p <0.01) and forestry lands (r = -0.62, p <0.01). Analysing the electrical conductivity parameter, we identified that the highest values of this parameter occurred in sub-basins with greater proportions of pasture areas (r = 0.85, p <0.01) and rock outcrops (r = 0.63, p <0.01). Lower electrical conductivity values were associated with sub-basins with higher proportions of rainforests (r = -0.82, p <0.01) and forestry lands (r = -0.62, p <0.01) and with sub-basins with high drainage densities (r = - 0.61, p <0.01).

Table 9 - Pearson's correlation coefficient (*r*) values between the landscape characteristics and water quality parameters of the PJ River basin sub-basins. Statistically significant *r*-values (p < 0.01) are highlighted in bold.

	pH	TMP	TUR	TDS	NIT	CHL	TDS	ECD	RES	SAL	TCL	ECO
UBA	0,44	**0,70***	0,52	-0,54	**0,68***	0,48	0,521	0,521	**-0,66***	0,53	0,39	**0,61**
ROU	**0,60***	0,58	0,12	-0,57	0,56	0,49	**0,632***	**0,631***	**-0,60***	**0,63***	0,34	0,42
SFR	0,46	**0,69***	0,55	-0,53	**0,65***	0,40	0,519	0,520	**-0,68***	0,54	0,36	**0,69***
WAT	0,26	0,40	0,43	-0,28	0,43	0,26	0,236	0,234	-0,41	0,27	0,23	0,26
AGR	0,42	**0,63***	0,53	-0,58	**0,63***	0,49	0,510	0,509	**-0,64***	0,54	0,35	0,52
PAS	**0,77***	**0,91***	0,45	**-0,78***	**0,90***	**0,73***	**0,84***	**0,85***	**-0,91***	**0,85***	0,43	**0,71***
FOR	-0,56	**-0,68***	-0,38	**0,62***	**-0,69***	-0,50	**-0,62***	**-0,61***	**0,69***	**-0,61***	-0,37	-0,52
RFR	**-0,73***	**-0,91***	-0,52	**0,78***	**-0,89***	**-0,72***	**-0,82***	**-0,82***	**0,91***	**-0,83***	-0,45	**-0,74***
ROD	0,39	0,48	0,42	-0,20	0,54	0,17	0,268	0,265	-0,49	0,29	0,48	0,34
DD	-0,33	**-0,72***	-0,46	**0,74***	**-0,66***	-0,54	**-0,60***	**-0,61***	**0,68***	-0,57	-0,31	-0,58
HD	-0,30	**-0,64***	-0,36	**0,71***	**-0,60***	-0,50	-0,570	-0,572	**0,61***	-0,52	-0,27	-0,51
MC	0,26	**0,65***	0,46	**-0,68***	0,58	0,50	0,532	0,534	**-0,62***	0,51	0,29	0,54
SPE	0,26	**0,66***	0,46	**-0,68***	0,58	0,50	0,532	0,534	**-0,62***	0,51	0,29	0,54
TT	-0,30	**-0,69***	-0,46	**0,72***	**-0,62***	-0,53	-0,575	-0,578	**0,65***	-0,55	-0,31	-0,56
IR	-0,17	-0,41	-0,13	0,37	-0,40	-0,33	-0,331	-0,330	0,36	-0,29	-0,29	-0,16
CCV	0,27	0,05	-0,50	-0,16	0,02	0,12	0,275	0,275	-0,08	0,26	-0,05	-0,11
CVX	-0,10	-0,31	0,07	0,47	-0,19	-0,45	-0,434	-0,439	0,23	-0,36	0,10	-0,25
STR	-0,16	0,23	0,39	-0,26	0,17	0,28	0,124	0,128	-0,13	0,07	-0,04	0,32
CNV	0,38	0,08	-0,20	-0,02	0,14	0,06	0,214	0,213	-0,14	0,21	0,25	0,00
DIV	0,14	0,06	0,25	0,15	0,11	-0,07	-0,112	-0,117	-0,10	-0,02	0,01	-0,04
PLA	-0,34	-0,10	-0,13	-0,14	-0,18	0,04	-0,010	-0,005	0,17	-0,10	-0,14	0,04
SLO>45	-0,25	-0,46	**-0,60***	0,19	-0,47	-0,27	-0,191	-0,190	0,42	-0,24	-0,18	-0,35

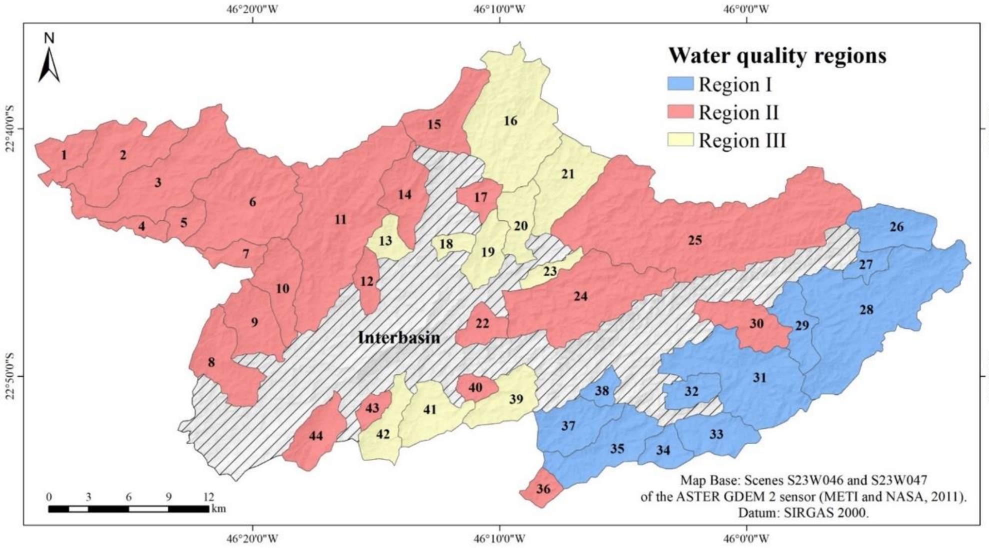
Higher resistivity values were associated with sub-basins with low proportions of pasture areas (r = -0.91, p <0.01), shrub forests (r=-0.68, p<0.01), agricultural lands (r=-0.64, p<0.01) and urban areas (r=-0.66, p<0.01). On the other hand, lower resistivity values were associated with sub-basins with higher proportions of rainforests (r=0.91, p<0.01) and forestry lands (r=0.69, p<0.01) as well as sub-basins with higher drainage densities (r=0.68, p<0.01) and topographic textures (r=0.65, p<0.01).

Higher salinity values were associated with high proportions of pasture lands (r=0.85, p<0.01) and rock outcrops (r=0.63, p<0.01) and with lower proportions of rainforests (r=-0.83, p<0.01) and forestry areas (r=-0.61, p<0.01). E. coli values were positively associated with the proportion of pasture areas (r=0.71, p<0.01) and shrub forests (0.69, p<0.01) and negatively associated with the proportion of rainforests (r=-0.74, p<0.01). The total faecal coliforms did not show a statistically significant association with any landscape characteristic. Studies carried out by Haddad et al. (2010) and Souza et al. (2014) found high positive correlations between E. coli and basins occupied by urban areas and pasture lands.

4.2.2 Regions based on the WQPs and landscape characteristics

According to the silhouette analysis and Ward methods, it was possible to classify the sub-basins of the PJ River basin into three regions based on the WQPs and landscape characteristics of the sub-basins (Figure 6).

Figure 6. Map of the water quality regions of the Piracicaba-Jaguari River basin.



Map Base: Scenes S23W046 and S23W047 of the ASTER GDEM 2 sensor (METI and NASA, 2011). Datum: SIRGAS 2000.

Region I

Region I was located in high-elevation areas, generally above 1,300 metres, formed by hills and mountain reliefs. The largest rainforest areas in the PJ River basin were located in this region. This region comprised the sub-basins with the highest drainage densities in the PJ River basin. Additionally, Region I was composed of the sub-basins with the best water quality among the entire basin. We observed lower temperature, total dissolved solids, chloride, nitrate, salinity and pH values in this region compared to the other regions as well as higher resistivity and dissolved oxygen values (Figure 7).

Region II

The elevations in Region II varied from 900 to 1,300 metres, with a predominance of wide and small hills and large flood plains. In this region, agricultural and livestock activities are well-developed, and the main agricultural crops in this region are corn, potatoes, beans, tomatoes, cassava and cabbage; these agricultural activities make use of pesticides and nitrogen fertilizers. The pasture areas are mainly used for cattle for commercial and family purposes. In both cases, these activities occur extensively without great care for the pastures, which have been degraded in several areas in this region.

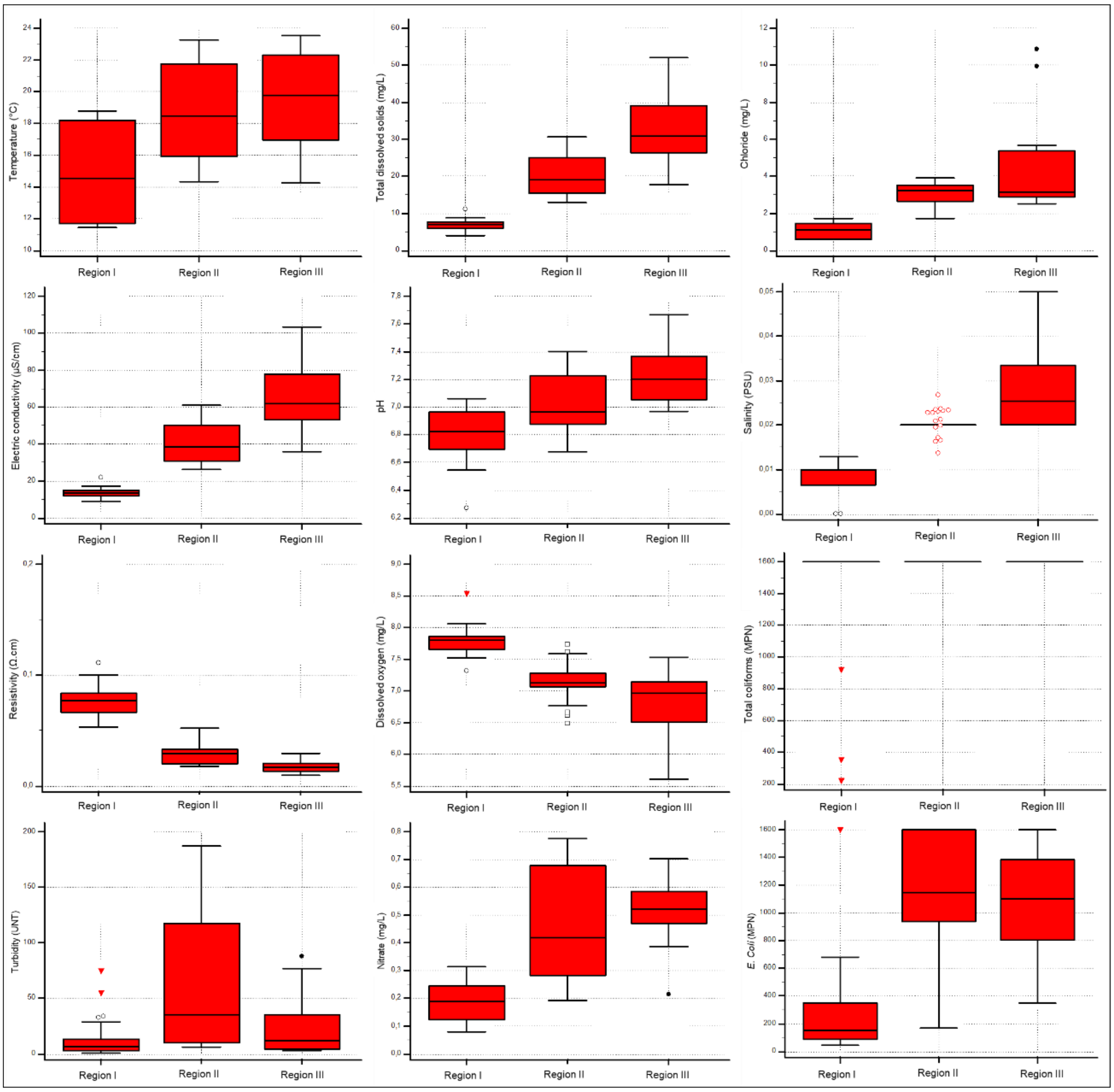
The urban Toledo, Camanducaia and Extrema municipality areas are located in Region II. The sub-basins that comprised Region II presented the highest number of WQPs in non-compliance with the standards established by CONAMA Resolution No. 357. Even when the parameter values were in compliance, they were generally of lower quality than those presented by the sub-basins that formed Region I.

Region III

Region III was located along the Fernão Dias highway corridor and comprised reliefs with elevations predominantly varying from 800 to 900 m. In the sub-basins of Region III, industrial areas, rural neighbourhoods and urban areas are predominant, and these land types are associated with pasture areas. Due to the lower flows of rivers in the dry season, the effluents discharged into rivers presented higher concentrations in this season, contributing to the great variability observed in the EC, TDS, pH, DO, Cl-, salinity and E. coli values (Figure 8). Assessing the effect of land use on lake water quality, Pietruszyński and Cieslinski (2018) related the highest observed concentration of ions to the proximity of farms that generate wastewater.

We noted that there were non-significant differences in the total coliform counts among the three regions ($p = 0.969$), a non-significant turbidity difference between regions I and III ($p = 0.127$), a non-significant E. coli difference between regions II and III ($p = 0.140$), and a non-significant temperature difference between regions II and III ($p = 0.313$). On the other hand, for all other WQPs, significant differences were observed between regions I and II and between regions I and III ($p <0.05$).

Figure 7. Boxplot charts showing the WQPs variabilities in the water quality regions of the PJ River basin.



5. CONCLUSIONS

Landscape characteristics related to land use type and occupation significantly influenced the water quality parameters of the sub-basins in the study area. On the other hand, landscape characteristics related to the shape of the relief and slope did not influence the water quality parameter values.

The geosystemic approach used in this research, which integrated elements of the spatial structure of the landscape with physicochemical and microbiological parameter measurements of the sub-basins, may be used as a complementary strategy for the regionalization of water quality in basins in which no collected data in the streams are available.

The regionalization methodology allowed the extrapolation of water quality variable values collected in the sampled sub-basins to those in the sub-basins without data based on the similarities defined by the land use and land cover types and the environmental parameters. The regions defined according to the proposed methodology showed significant differences among the corresponding total dissolved solids, chloride, electric conductivity, pH, salinity, resistivity, dissolved oxygen, and nitrate parameter values.

However, these regions did not show significant differences among the corresponding E. coli, temperature or turbidity parameter values. We recommend that it is necessary to increase the quantity of samples in future research to assess the significance of the homogeneity of water quality regions.

ACKNOWLEDGEMENTS

The authors thank the São Paulo Research Foundation (FAPESP) for financial aid (Process 13/22185-2).

REFERENCES

Amorim, R. C., and Henning, C. (2015). Recovering the number of clusters in data sets with noise features using feature rescaling factors. *Journal Information Sciences*, 324: 126-145. 126-145. http://doi:10.1016/j.ins.2015.06.039

ANA. Agência Nacional das Águas (2011). *Guia nacional de coleta e preservação de amostras: água, sedimento, comunidades aquáticas e efluentes líquidas*. Companhia Ambiental do Estado de São Paulo, CETESB. https://cetesb.sp.gov.br/wp-content/uploads/2021/10/Guia-nacional-de-coleta-e-preservacao-de-amostras-2012.pdf

ANA. Agência Nacional das Águas. (2015). *Base hidrográfica ottocodificada 1:250.000*. Superintendência de Planejamento de Recursos Hídricos. https://metadados.snirh.gov.br/geonetwork/srv/api/records/0f57c8a0-6a0f-4283-8ce3-114ba904b9fe.

APHA. American Public Health Association (1998). *Standard methods for the examination of water and wastewater*. 20. ed. American Public Health Association, 1220p. http://doi:10.2105/SMWW.2882.219

ASTER GDEM (2014). *Global digital elevation model version 2*. http://gdem.ersdac.jspacesystems.or.jp

Bertalanify, L. von. (1973). *Teoria geral dos sistemas*. Editora Vozes.

Bertrand, G. (2004). Paisagem e geografia física global*. RA´E GA,* v. 8, p.141-152. https://revistas.ufpr.br/raega/article/view/3389/2718

Brasil (2005). *Resolução CONAMA n°357, de 17 de março de 2005. Classificação de águas, doces, salobras e salinas do território nacional*. Diário Oficial da União do Brasil. http://pnqa.ana.gov.br/Publicacao/RESOLUCAO_CONAMA_n_357.pdf

Chakraborty, T., Chakraborty, A. K., and Mansoor, Z. (2019). A hybrid regression model for water quality prediction. *OPSEARCH*, 56: 1167-1178. http://doi:10.1007/s12597-019-00386-z

Clément, F., Ruiz, J., Rodríguez, M. A., Blais, D., and Campeau, S. (2017). Landscape diversity and forest edge density regulate stream water quality in agricultural catchments. *Ecological Indicators*, 72: 627-639. http://doi:10.1016/j.ecolind.2016.09.001

CPI International (2012). *iMPNplateTM-1600 - User Manual and MPN Table*. v.1. CPI.

Li. Y., Degener, J., Gaudreau, M., and Kappas. M. (2016). Adaptive capacity-based water quality resilience transformation and policy implications in rapidly urbanizing landscapes. *Science of The Total Environment*, 569-570:168-178. http://doi:10.1016/j.scitotenv.2016.06.110

Dogan, E., Sengorur, B., and Koklu, R. (2009). Modeling biological oxygen demand of the Melen River in Turkey using an artificial neural network technique. *Journal of Environmental Management,* 90 (2): 1229-1235. http://doi:10.1016/j.jenvman.2008.06.004

EMBRAPA (2006). *Sistema brasileiro de classificação de solos*. 2ª ed., Empresa Brasileira de Pesquisa Agropecuária.

ESRI (2016). *ArcGIS desktop: release 10.5*. Environmental Systems Research Institute.

Everitt, B. S. (2006). *The Cambridge dictionary of statistics*. 3ª ed. Cambridge University Press.

Exelis Visual Information Solutions. 2010. *ENVI v.5.4 manual*. Exelis Visual Information Solutions.

Faceli, K., Lorena, A. C., GAMA, J., and Carvalho, A. C. P. L. F. (2011). *Inteligência artificial: uma abordagem de aprendizado de máquina*. LTC.

Felde, G.W., Anderson, G.P., Cooley, T.W., Matthew, M.W., Adler-golden, S.M., Berk, A., and Lee, J. (2003). Analysis of hyperion data with the FLAASH atmospheric correction algorithm. *In: International Geoscience and Remote Sensing Symposium. Toulouse*. Proceedings. New York: IEEE. http://doi:10.1109/IGARSS.2003.1293688

Ferreira, L., and Hitchcock, D. B. (2009). A comparison of hierarchical methods for clustering functional data. *Communications in Statistics - Simulation and Computation*, 38 (9): 1925-1949. http://doi:10.1080/03610910903168603

Filoso, S., Martinelli, L.A., Williams, M.R. et al. (2003). Land use and nitrogen export in the Piracicaba River basin, Southeast Brazil. Biogeochemistry 65, 275–294. https://doi.org/10.1023/A:1026259929269

Forman, R. T. T., and Alexander. L. E. (1998). Roads and their major ecological effects. *Annual Reviews in Ecology & Systematics*, 29: 207-231. http://doi:10.1146/ANNUREV.ECOLSYS.29.1.207

Haddad, E. A., and Júnior, A. P. M. (2010). Influência antrópica na qualidade da água da bacia hidrográfica do rio São Miguel, carste do alto São Francisco, Minas Gerais. *Geosul,* 25 (49): 79-102. http://doi:10.5007/2177-5230.2010v25n49p79

HANNA instruments® (2013). *Instruction Manual: HI 9829 Multiparameter Meter With available GPS*. http://www.hannainst.com/manuals/HI9829_v0.32_2011_07_01.pdf

IBGE. Instituto Brasileiro de Geografia e Estatística (1977). Cartas topográficas Munhoz. Cambuí. Extrema. Camanducaia. Paraisópolis e Monteiro Lobato. Escala 1:50.000.

IGAM (2010). *Plano estadual de recursos hídricos*. Instituto Mineiro de Gestão das Águas.

JASP Team (2020). *JASP*. Version 0.14.1. University of Amsterdam.

Landis, J.R., Koch, G.G. (1977). The measurement of observer agreement for categorical data. *Biometrics*.; 33(1):159-74. PMID: 843571.

Larsen, D. P., Dudley, D. R., and Hughes, R. M. (1988). A regional approach for assessing attainable surface water quality: An Ohio case study. *J. Soil Water Conservation*, 43: 171-176. https://www.jswconline.org/content/43/2/171

Lee, S., Hwang, S., Lee, S., Hwang, H., and Sung H. (2009). Landscape ecological approach to the relationships of land use patterns in watersheds to water quality characteristics. *Landscape and Urban Planning*, 92: 80-89. http://doi:10.1016/j.landurbplan.2009.02.008

Li, L., Jiang, P., Xu, H., Lin, G., Guo, D., and Wu, H. (2019). Water quality prediction based on recurrent neural network and improved evidence theory: a case study of Qiantang River. China. *Environmental Science and Pollution Research*, 26 (19): 19879-19896. http://doi:10.1007/s11356-019-05116-y

Linlin, X., Liu, R.Z., and Zao, Y.W. (2012). Correlation Analysis of Landscape Pattern and Water Quality in Baiyangdian Watershed. *Procedia Environmental Sciences*, 13: 2188-2196. http://doi:10.1016/j.proenv.2012.01.208

Mahmoudi, N., Orouji, H., and Fallah-Mehdipour, E. (2016). Integration of shuffled frog leaping algorithm and support vector regression for prediction of water quality parameters. *Water Resources Management*, 30 (7): 2195-2211. http://doi:10.1007/s11269-016-1280-3

Mingoti, A. (2005). *Análise de dados através de métodos de estatística multivariada: uma abordagem aplicada*. Editora UFMG.

Moretti, A. I. P., and Ferreira, M. C. (2017). Geossistemas da APA Fernão Dias - MG: mapeamento e caracterização a partir da perspectiva geográfica da paisagem. *XVIII Simpósio Brasileiro de Sensoriamento Remoto,* INPE,. p. 6975-6982.

Pietruszyński, Ł., and Cieśliński, R. (2018). The effects of different land use and hydrological types on water chemistry of young glacial ponds. *Journal of Hydrology*, 564: 605-618. http://doi:10.1016/j.jhydrol.2018.07.049.

Pongpetch, N., Suwanwaree, P. (2012). Spatial water quality assessment and mapping of Lahan Swamp. Chaiyaphum. Thailand. *Procedia Environmental Sciences*, 13: 655-659. http://doi:10.1016/j.proenv.2012.01.057

Qadir, A., Malik, R. N, and Husain, S. Z. (2008). Spatio-temporal variations in water quality of Nullah Aik-tributary of the river Chenab, Pakistan. *Environ Monit Assess*, 140: 43-59. http://doi:10.1007/s10661-007-9846-4

Ravichandran, S., Ramanibai, R., and Pundarikanthan, N.V. (1996). Ecoregions for describing water quality patterns in Tamiraparani basin, south India. *Journal of Hydrology*, 178: 257-276. https://doi.org/10.1016/0022-1694(95)02801-3

Rezaid-Balf, M., Attar, N. F., Mohammadzadeh, A, Murti, M. A., Ahmed, A. N., Fai, C. M., Nabipour, N., Alaghmand, S., and El-Shafie, A. (2020). Physicochemical parameters data assimilation for efficient improvement of water quality index prediction: Comparative assessment of a noise suppression hybridization approach. *Journal of Cleaner Production*, 271. http://doi:10.1016/j.jclepro.2020.122576

Robertson, D. M., and Saad, D. A. (2003). Environmental water-quality zones for streams: a regional classification scheme. *Environmental Management,* 31(5): 581-602. http://doi:10.1007/s00267-002-2955-5

Rossum, G. V., and Boer, J. D. (1991). Interactively testing remote servers using the python programming language. *CWI Quarterly*, 4 (4): 283-304. https://ir.cwi.nl/pub/18204

Rousseeuw, P. J. (1987). Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics*, 20: 53-65. https://doi/10.1016/0377-0427(87)90125-7

Sattari, M. T., Joudi, A. R., and Kusiak, A. (2016). Estimation of water quality parameters with data-driven model. *American Water Works Association*, 108 (4): 232-239. http://doi:10.5942/jawwa.2016.108.0012

Souza, M. M., and Gastaldini, M. C. C. (2014). Water quality assessment in watersheds with different anthropogenic impacts. *Engenharia Sanitária e Ambiental*, 19 (3): 263-274. http://doi:10.1590/S1413-41522014019000001097

Sparovek, G., Van Lier, Q.J., and Neto, D. D. (2007). Computer assisted Köppen climate classification: A case study for Brazil. *International Journal of Climatology*, 27(2): 257-266. http://doi:10.1002/joc.1384

Strahler, A. N. (1980). Systems theory and physical geography. *Physical Geography*, 1(1): 1-27. https://doi.org/10.1080/02723646.1980.10642186

Strahler, A.N. (1952) Hypsometric (area-altitude) analysis of erosional topography. *Geological Society of America Bulletin*, 63, 1117-1141. http://dx.doi.org/10.1130/0016-7606(1952)63[1117:HAAOET]2.0.CO;2

Tcherkezova, E., Stoyanova, V., and Kotsev, T. (2019). A concept of an integrated geodatabase for surface water, soil and groundwater pollution with arsenic in the upper part of Ogosta Valley, northwestern Bulgaria. *European Journal of Geography,* 10(3):6-23, 2019. http://eurogeojournal.eu/articles/1_Tcherkezova_et_al.pdf

Valeriano, M. de M. (2008). *Dados topográficos.* In: Florenzano, T. G. Geomorfologia: conceitos e tecnologias atuais. Oficina de Textos.

Vapnik, V. (1995). *The nature of statistical learning theory*. Springer.

Varol, M., Gökot, B., Bekleyen, A., and Şen, B. (2012). Spatial and temporal variations in surface water quality of the dam reservoirs in the Tigris River basin, Turkey. *Catena,* 92: 11-21. http://doi:10.1016/j.catena.2011.11.013

Wang, H., Ren, Z., and Tang, M. (2010). Design of water quality monitoring based on SVM and its simulation platform by remote sensing. In: Annals of the 3º International Congress on Image and Signal Processing. IEEE, Yantai, China, http://doi:10.1109/CISP.2010.5647474

Wang, X., Ma, L., and Wang, X. (2010). Apply semi-supervised support vector regression for remote sensing water quality retrieving. In: Annals of the Geoscience and Remote Sensing Symposium (IGARSS). IEEE. Honolulu, HI, USA, http://doi:10.1109/IGARSS.2010.5653832

Ward Jr., J. H. (1963) Hierarchical Grouping to Optimize an Objective Function, *Journal of the American Statistical Association*, 58:301, 236-244, http://doi:10.1080/01621459.1963.10500845