

Kriging Geostatistical Methods for Travel Mode Choice: A Spatial Data Analysis to Travel Demand Forecasting

Viviani Antunes Gomes¹, Cira Souza Pitombo¹, Samille Santos Rocha¹, Ana Rita Salgueiro²

¹Department of Transportation Engineering, University of São Paulo, São Carlos, Brazil

²Department of Geology, Federal University of Ceará, Fortaleza, Brazil

Email: viviani_antunes@yahoo.com.br, cira@sc.usp.br, millyblase@gmail.com, geo.ritasalgueiro@gmail.com

Received 15 April 2016; accepted 21 June 2016; published 24 June 2016

Copyright © 2016 by authors and Scientific Research Publishing Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

This paper aims to compare the results of two techniques of Kriging (Ordinary Kriging and Indicator Kriging) that are applied to estimate the Private Motorized (PM) travel mode use (car or motorcycle) in several geographical coordinates of non-sampled values of the concerning variable. The data used was from the Origin/Destination and Public Transportation Opinion Survey, carried out in 2007/2008 at São Carlos (SP, Brazil). The techniques were applied in the region with 110 sample points (households). Initially, Decision Tree was applied to estimate the probability of mode choice in surveyed households, thus determining the numeric variable to be used in Ordinary Kriging. For application of Indicator Kriging it was used the variable “main travel mode” in a discrete manner, where “1” represented the use of PM travel mode and “0” characterized others travel modes. The results obtained by the two spatial estimation techniques were similar (Kriging maps and cross-validation procedure). However, the Indicator Kriging (KI) obtained the highest number of hit rates. In addition, with the KI it was possible to use the variable in its original form, avoiding error propagation. Finally, it was concluded that spatial statistics was thriving in travel demand forecasting issues, giving rise, for the both Kriging methods, to a travel mode choice surface on a confirmatory way.

Keywords

Geostatistics, Kriging, Travel Mode Choice, Spatial Estimation

1. Introduction

Travel mode choice modeling is the process where the means of traveling is estimated. The means of travel is

How to cite this paper: Gomes, V.A., Pitombo, C.S., Rocha, S.S. and Salgueiro, A.R. (2016) Kriging Geostatistical Methods for Travel Mode Choice: A Spatial Data Analysis to Travel Demand Forecasting. *Open Journal of Statistics*, 6, 514-527.

<http://dx.doi.org/10.4236/ojs.2016.63044>

referred to the travel mode, which may be by private automobile, public transportation, walking, etc. Travel mode choice is an important part of travel demand forecasting models that are used for analyses, for instance: 1) Major transportation investment projects since they may attract travelers not only from competing facilities but also competing travel modes; 2) Transit service changes; 3) Pricing policy analyses; 4) Land use planning analyses.

Individuals choose their travel mode considering diverse factors, classified into three groups: 1) Characteristics of the trip maker (car ownership, income, age, etc.); 2) Characteristics of the journey (trip purpose, time of the day, etc.); 3) Characteristics of the transport facility (travel time, monetary costs, comfort, convenience, etc.). Thus, travel behavior involves household and personal characteristics, travel variables and spatially correlated factors, as well as location of households and others destinations, activities distribution in the urban environment [1].

Several studies corroborate the affirmation that travel mode choice issues are strongly related to locations [2] [3]. With the advances of technology, geo-referenced information has become more widely available. Thus, travel demand spatial analysis has been identified as an emerging research area [4].

Bhat and Zhao [5] identify the spatial issues that need to be recognized in travel demand modeling, proposing a multi-level, mixed logit, formulation to address these spatial issues in the context of activity stop generation in the Boston Metropolitan area.

The use of kernel estimation has been already proposed by [6] in association with a logit model in order to assign route choice. This study demonstrates efficient results and the potential of using spatial considerations into travel demand models. Miyamoto *et al.* [7] combine concepts to specify the spatially auto-correlated deterministic and error components within a mixed logit framework.

Páez *et al.* [4] introduce a new indicator of spatial fit that can be applied to the results of discrete choice models. Peer *et al.* [8] apply geographically weighted regression for the approximation of door-to-door travel times in departure time choice models.

Among the techniques of spatial statistics, geostatistics is to be highlighted. Geostatistics enables the development of studies involving spatial autocorrelation, allowing mainly estimating the value of a variable in locations where values are unobserved.

Usually, geostatistics is applied to cases in which spatial continuity is apparent. Despite this limitation, geostatistical modeling has been applied to spatially discrete data for many years [9]. Generally, travel data are spatially discrete. Dealing with this limitation, it is necessary to adapt travel demand data, considering that they are generally discrete variables and have no spatial continuity.

In the literature, however, studies on the application of geostatistics for transportation problems are derived from traffic engineering. Miura [10] presents an approach for predicting car travel time by Kriging. This prediction method is shown to be effective for urban districts with links having changeable travel times owing to congestion. Zou *et al.* [11] propose an improved distance metric called approximate road network distance (ARND) for solving the problem of the invalid spatial covariance function in Kriging caused by the non-Euclidean distance metric.

Following this approach, recent studies have shown that geostatistics is able to estimate transportation demand variables and explain the spatial distribution by maps of Kriging predicted values [12]-[14].

The main aim of this paper is to compare the results of two techniques of Kriging (Ordinary Kriging and Indicator Kriging) that were applied to estimate the Private Motorized (PM) travel mode use (car or motorcycle) in several geographical coordinates of non-sampled values of the concerning variable. The data used was from the Origin/Destination and Public Transportation Opinion Survey, carried out in 2007/2008 at São Carlos (SP, Brazil).

This article is organized in four sections besides this Introduction. Section 2 describes the techniques; Section 3 presents the materials and method. Section 4 exhibits the results, and finally, Section 5 describes the main conclusions.

2. Spatial Analysis Approaches

Geostatistics was developed as an alternative method for exploiting events in which variable values are associated with geographical coordinates. This approach takes account of general spatial statistics because it estimates a continuous surface through a dataset that may be regular or irregularly spatially distributed.

The main point of applying geostatistics follows the idea of characterizing spatial (and/or spatial/temporal) dispersion of an event, assessing uncertainty parameters, concerning its spatial variability, and obtaining a continuous surface estimation. Geostatistics modeling is better defined as the following steps: 1) variography analysis, 2) cross validation, and 3) Kriging.

The primary tool in geostatistical modeling is the semivariogram, which graphically represents a regionalized variable. The semivariogram function is given by Equation (1), where $N(h)$ is the set of all pairwise, and $z(x_i)$ and $z(x_i + h)$ are data values at spatial locations i and $i + h$, respectively.

$$\gamma(h) = \frac{1}{2N} \sum_{i=1}^{n(k)} [Z(x_i) - Z(x_i + h)]^2 \tag{1}$$

Moreover, the representation of an experimental semivariogram requires a further understanding of graphical aspects. Some measures include lag distance and tolerance, cut distance, and angle direction. Another step of variography analysis is to model a theoretical semivariogram based off the experimental one. Of the various theoretical models for adjustments of variogram, the most frequently used are Spherical, Gaussian and Exponential. In this step, the experimental variogram is replaced by a theoretical variogram function, from which is possible to obtain the main parameters for spatial modelling: nugget effect (C_0), Range (a) and Sill ($C + C_0$), see **Figure 1**.

The second step of geostatistics modelling is Cross Validation (CV). CV is a simple way to compare various assumptions either about the model (e.g. type of function to be adjusted, parameters of variograms) or about the data. In the cross validation procedure, each sample value $Z(x_1)$ is removed in turn from the data set and a value $Z^*(x_1)$ at the location is estimated using the remaining $n-1$ samples. The difference between a data value and the estimated value ($Z(x_1) - Z^*(x_1)$) gives an indication of how well the data value fits into the neighborhood of the surrounding data values [15].

Regarding the CV, the geostatistical modeling general method follows the Kriging estimation, which is a linear prediction represented by matricial calculus. The aim of Kriging is to conduct estimates with minimum error and variance (optimizing the model) through the parameters defined in the theoretical semivariogram. The most common univariate Kriging methods are Simple Kriging, Universal Kriging and Ordinary Kriging. In this paper, we compared Indicator Kriging and Ordinary Kriging to estimate PM travel mode use.

Ordinary Kriging is the most widely used Kriging method. Its main goal is to estimate a value at a point of a region, for which the correspondent variogram is known, using data in neighborhood [15].

OK is a method that is often associated with the acronym B.L.U.E. for “best linear unbiased estimator”. OK is linear because its estimates are weighted linear combinations of the available data. It is unbiased since it tries to have the mean residual equal to zero. It is best because it aims at minimizing the variance of errors [16].

For the prediction of the variable Z at a location x_0 , $\{Z(x_0)\}$, the estimator $Z^*(x_0)$ is defined as [9]:

$$Z^*(x_0) = \sum_{i=1}^n \lambda_i Z(x_i) \tag{2}$$

where the λ_i are weights that were calculated considering two constrains: (A) the estimator is not biased; (B) The variance of the estimate is minimal.

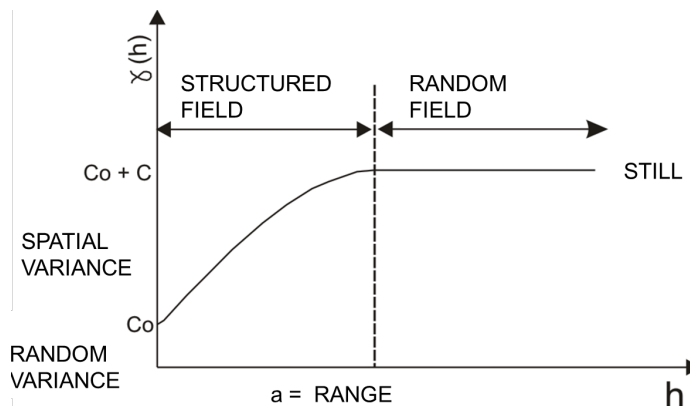


Figure 1. Theoretical variogram (Adapted from [13]).

The variance of the estimate is minimal. These conditions are reached when error is equal zero, in average:

$$E[Z^*(x_0) - Z(x_0)] = 0 \quad (3)$$

Developing the error mathematical expectancy expression, we come to the condition of no bias shown in Equation (4):

$$\sum_{i=1}^n \lambda_i = 1 \quad (4)$$

The variance of estimate error is given by Equation (5):

$$\sigma_E^2 = \text{Var}[Z(x_0) - Z^*(x_0)] \quad (5)$$

The minimization of the estimate error is carried out developing Equation (5), as represented by Equation (6):

$$\sigma_E^2 = E\left[(Z(x_0) - Z^*(x_0))^2\right] - \left(E[Z(x_0) - Z^*(x_0)]\right)^2 \quad (6)$$

Expanding each term on the right side of this equation, one comes to the Equation (7) in terms of the covariance function (C):

$$\sigma_E^2 = C(0) - 2\sum_i \lambda_i C(x_0 - x_i) + \sum_i \sum_j \lambda_i \lambda_j C(x_i - x_j) \quad (7)$$

To minimize the error, it uses the technique of Lagrange multipliers, as shown in Equation (8):

$$L(\lambda_1, \lambda_2, \dots, \lambda_n, \mu) = C(0) - 2\sum_i \lambda_i C(x_0 - x_i) + \sum_i \sum_j \lambda_i \lambda_j C(x_i - x_j) - 2\mu \left(\sum_j \lambda_{j-1}\right) \quad (8)$$

Finally, we obtained the equation system of Ordinary Kriging

$$\begin{cases} \sum_{j=1}^n \lambda_j C(x_i - x_j) - \mu = C(x_0 - x_i) & i = 1, n \\ \sum_{j=1}^n \lambda_j = 1 \end{cases} \quad (9)$$

The Indicator Kriging is an estimation technique with the same basis of the linear Kriging, but it is applied to attributes with non-Gaussian distribution, which are transformed according to a non-linear mapping, codification by means of indication. Applied on a set of numeric sample values, $Z(u = u_a)$, codification by means of indication generates, for a cutting value z_k , a sample set by means of indication $I(u = u_a; z_k)$, of the kind [16]:

$$I(u; z_k) = \begin{cases} 1 & \text{if } Z(u) \leq z_k \\ 0 & \text{if } Z(u) > z_k \end{cases} \quad (10)$$

Codification by means of indication is applied on the whole sample set, creating, for each cutting value, a sample set by means of indication for which sample values are transformed either in 0 or 1. The K cutting values z_k , $k = 1, 2, \dots, K$ are defined according to the number of samples. Definition of a variation computing model depends on the existence of a minimum distribution of zeros (0) and ones (1) within the set of samples codified by means of indication. It can be shown that the best variation computing is obtained for a cutting value equal to the median computed from the sample set [16]. Thus, it is possible to use a single cutting value, equal to or close to the median, in order to generate a single codification by means of indication, known as codification by means of indication by median. Use of a single cutting value reduces computation time, however, this may be restricting if the use of the variation recording model, close to the median, is too different from those obtained by other cutting values.

The set of sample data by means of indication is used to infer values for the random variables by means of indication $I(u; z_k)$, with $u \neq u_a$. The conditional expectancy of the numeric random variable obtained by means of indication $I(u; z_k)$ is calculated by Equation (11).

$$\begin{aligned} E\{I(u; z_k)/(n)\} &= 1 \cdot \text{Prob}\{(u; z_k) = 1/(n)\} + 0 \cdot \text{Prob}\{I(u; z_k) = 0/(n)\} \\ &= 1 \cdot \text{Prob}\{I(u; z_k) = 1/(n)\} = F_s^*(u; z_k/(n)) \end{aligned} \quad (11)$$

Equation (11) presents an extremely important result concerning the inference of the probability distribution

of a random variable. It means that the conditional expectancy of $I(u; z_k)$ yields, for the cutting value $z = z_k$, an estimate of the function of conditioned accumulated distribution, $F^*(u; z_k | (n))$, for numeric attributes. Conditional expectancy $E\{I(u; z_k | (n))\}$ may be estimated by algorithms of Kriging, which employ numeric random variables, codified by means of indication. Kriging by simple indicator is simple linear Kriging applied to a sample set codified by means of indication at the cutting values $z = z_k$, and it has the formulation presented in Equation (12):

$$F_s^* \left[\left((u; z_k) / (n) \right) \right] = \sum_{\alpha=1}^{n(u)} \lambda_{s_\alpha} (u; z_k) i(u_\alpha; z_k) + \left[1 - \sum_{\alpha=1}^{n(u)} \lambda_{s_\alpha} (u; z_k) \right] F_s^* (z_k) \quad (12)$$

where $F_s^* (z_k)$ is the average of the random function of the stationary region, and the weights $\lambda_{s_\alpha} (u; z_k)$ are determined in order to minimize the variance of the estimative error. Considering the sum of the pondering weights equal to 1, a simpler version of Kriging by means of indication is obtained. It is the Kriging by means of ordinary indication, which estimative expression (Equation (13)):

$$F_o^* \left[\left((u; z_k) / (n) \right) \right] = \sum_{\alpha=1}^{n(u)} \lambda_{o_\alpha} (u; z_k) i(u; z_k) \quad (13)$$

The weights $\lambda_{s_\alpha} (u; z_k)$ are obtained by solving the following system of equations of Kriging by means of ordinary indication:

$$\begin{aligned} \sum_{\beta=1}^{n(u)} \lambda_{o_\beta(u; z_k)} C_1 (h_{\alpha\beta}; z_k) + \varphi(u; z_k) = C_1 (h_\alpha; z_k) \quad \forall \alpha = 1, 2, \dots, n(u) \\ \sum_{\beta=1}^{n(u)} \lambda_{o_\beta(u; z_k)} = 1 \end{aligned} \quad (14)$$

where $\varphi(u; z_k)$ is Lagrange's multiplier, $h_{\alpha\beta}$ is the separation vector (which is defined by the positions u_α and u_β), h_α is the vector defined between the positions u_α and u_β , $C_1(h_{\alpha\beta}; z_k)$ is the self-covariance defined by $h_{\alpha\beta}$, and $C_1(h_\alpha; z_k)$ is the self-covariance defined by h_α in $z = z_k$. The self-covariances are determined by the theoretical variation computing model defined by set I for $z = z_k$ [17]. In this paper, the original variable of interest is dummy (0-no use of car/motorcycle; 1-use of car/motorcycle). Thus, in this case there is no cutting value.

3. Material and Methods

3.1. Materials: Study Area, Dataset and Software

3.1.1. Study Area

The study area of this research is the city of São Carlos (São Paulo/Brazil) (Figure 2), with 221,936 inhabitants. 96% of the population lives in urban areas, which covers, approximately, 105 km² of 1137.30 km² of total area [18].

The data used for the development of this study is from the Household interview and Urban Transportation Evaluation Survey applied together in Origin-Destination Survey of 2007/2008 in the city [19]. A set of 5% of households was randomly selected from the municipality water supply database (from which, the respective geographical coordinates were obtained).

3.1.2. Software Packages: Auxiliary Tools

The computing applications used in this paper were the IBM-Statistical Package for the Social Sciences (SPSS) version 22 for DT modeling and the software GeoMS 1.0 for the geostatistical calculation and definition processes of experimental and theoretical semivariograms and Kriging. The software ArcGIS 10.1 was used in order to obtain graphical representations of the results.

3.2. Method

To reach the proposal of this paper, the stages represented in Figure 3, summarized in the subsequent subsections, have been fulfilled.

3.2.1. Data Base Treatment

The database preliminary analysis process led to the elimination of some samples when one, or more, of the following situations occurred: 1) inconsistent or missing data, 2) people who did not travel and 3) households with

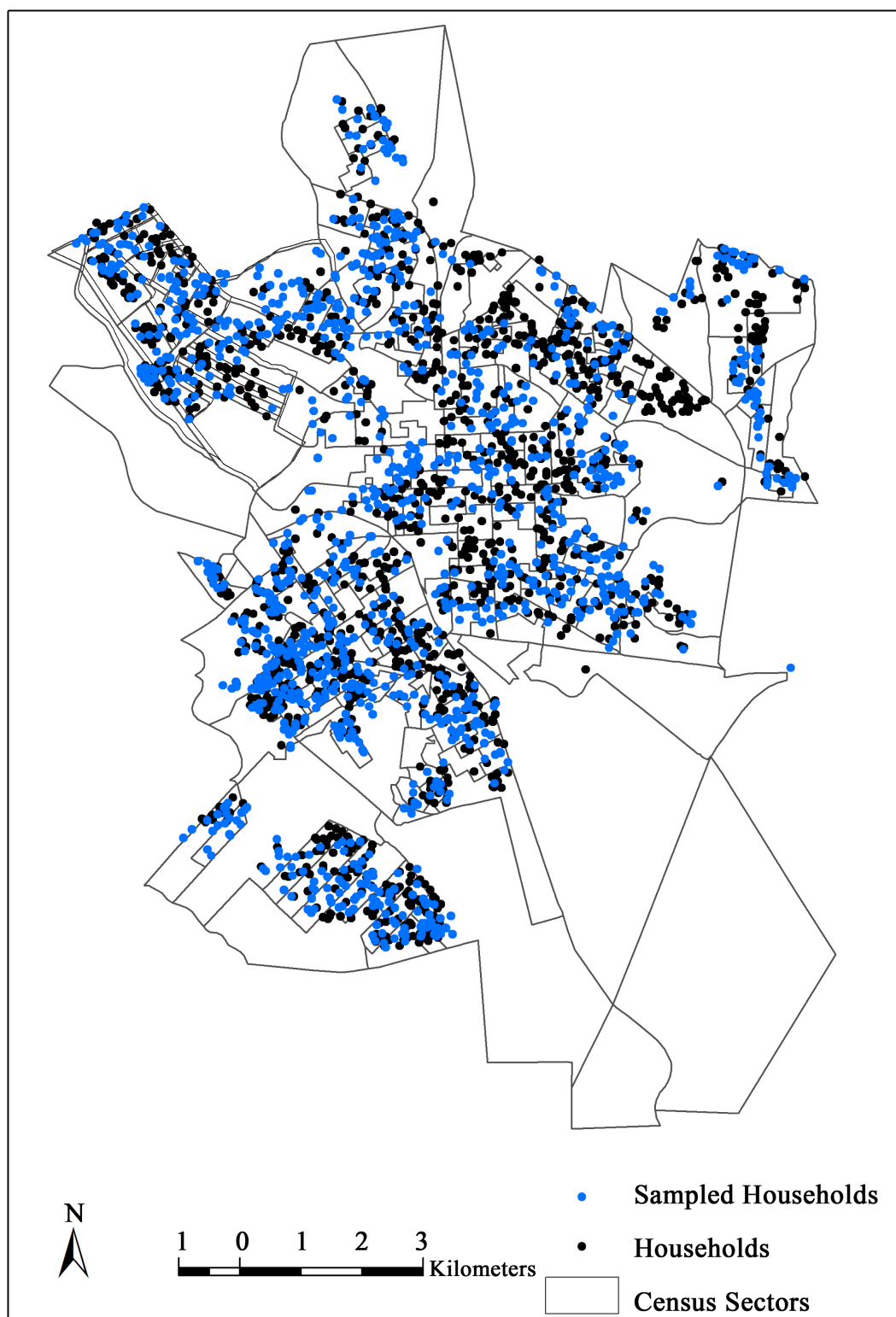


Figure 2. Study Area-São Carlos city. Fonte [19].

repeated geographic coordinates. Thus, the final sample contains 1216 individuals, 22 categorical variables and 4 numeric variables. Such information was then associated with geographic coordinates (latitude and longitude

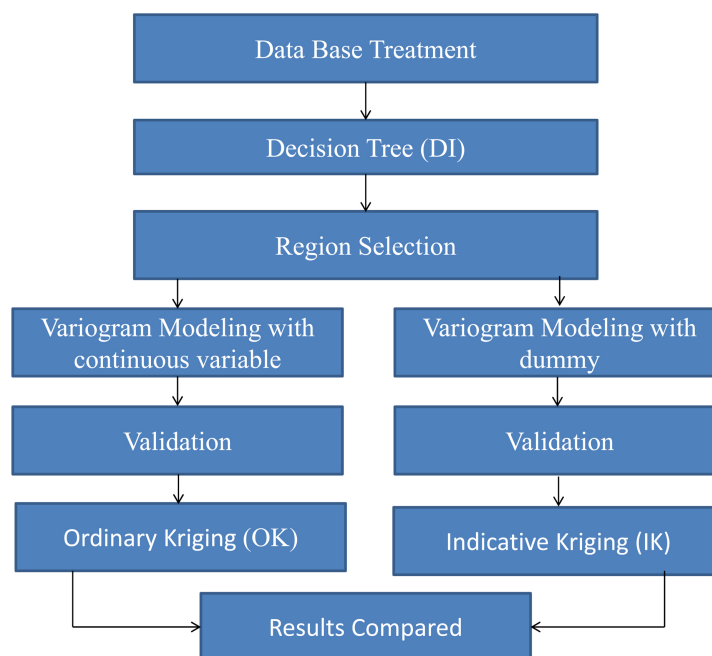


Figure 3. Methodology main steps.

in meters) relative to the households. The main variables in the sample and its description can be found in **Table 1**. In the same table, data used in each technique are presented.

3.2.2. Decision Tree (DT) Application

Decision Tree (DT) analysis, especially the CART algorithm, in this research article, was used to achieve the continuous variable for application of Ordinary Kriging-estimated probabilities of PM travel mode. OK method, used in the next step of this paper, consists of using numeric variables, avoiding the use of the original categorical variable “Main travel mode” for the spatial interpolation of mode choice.

The CART algorithm consists of a sequential binary partitioning of the dataset considering the values of the variables. Tree models are fitted by successively splitting the data to form homogeneous subsets, being the result a hierarchical tree of decision rules useful for prediction or classification [20].

The DT analysis was generated with a sample of 1216 cases, using the CART algorithm and adopting the parameter of a minimum of 25 observations per terminal node. The dependent variable was “Main travel mode”, consisting of three categories (1 transit, 2 car/motorcycle and 3 non-motorized). The independent variables were socioeconomic, travel characteristics and the qualitative measures of transport system, as shown in **Table 1**.

3.2.3. Region Selection

Following the steps of the proposed method, exploratory maps of probabilities of travel mode choice, obtained by DT model, were generated in order to observe spatial patterns. Good estimates for spatial estimation depend, mainly, on spatial structure of the variable to be kriged. From the exploratory spatial analysis, was verified that the variables to be kriged did not present apparent spatial pattern. Thus, it was decided to segregate the city of São Carlos (São Paulo-Brazil) into six regions, considering the criteria of income. Thus, using the Cluster Analysis technique, considering the variable income (categorical, by Minimum Wages) and geographic coordinates of households (latitude and longitude in meters), it were obtained six groups/regions.

A detailed analysis of the probability maps for each of the six regions separately, revealed that Region 2 was the one that presented a more diffusive spatial pattern, ranging from the center to the periphery of the area, as shown in **Figure 4**. This region consists of both low-income households (center of the region) and higher income (peripheral neighborhoods). In addition, Region 2, is a particular one in the city hence it includes the University of São Paulo second campus, an important factor that affects the city dynamics.

Therefore, in order to estimate all the area of Region 2 (*Santa Felicia*), including the locations where mode

Table 1. Main variables and techniques.

Variables	Techniques	Description
Main problem-non motorized mode	DT	(1) risk of running over ; (2) robberies; (3) poor condition of sidewalks; (4) lack of trees
Transit capacity	DT	(1) empty; (2) suitable; (3) crowded; (4) overcrowded
Transit Fleet	DT	(1) very small; (2) small; (3) suitable; (4) upper
Main problem-transit mode	DT	(1) travel time; (2) safety; (3) comfort; (4) itinerary; (5) schedules
Main problem-car mode	DT	(1) traffic jam; (2) lack of parking; (3) high cost
Driver license	DT	(1) yes; (2) no
Gender of household head	DT	(1) male; (2) female
Literacy	DT	(1) complete high school degree or college; (2) incomplete high school degree or less; (3) illiterate
Income	DT	(1) 0 to 2 MW [*] ; (2) 2.1 to 8 MW [*] ; (3) 8.1 to 20 MW [*] ; (4) not answered
Main travel mode	DT	(1) public; (2) private; (3) non-motorized
Motorcycle ownership	DT	number of motorcycles
Car ownership	DT	number of cars
Age	DT	household's head age
Amount of trips	DT	number of trips performed by the household head
latitude and longitude	IK; OK	geographic coordinates of sampled households
Motorized private travel mode	IK	0: Do not use 1: Use
Probabilities of choice of motorized private travel mode	OK	continuous variable from the DT results (between 0 and 1)

MW^{*}: Minimum Wage; DT (Decision TREE); IK (Indicator Kriging); OK (Ordinary Kriging).

choice (PM travel mode choice) was not observed, a database composed by 110 points was submitted to Ordinary Kriging and Indicator Kriging, respectively.

3.2.4. Variograms Modelling: Continuous and Dummy Variables

In modeling the experimental variograms, we used the following two regionalized variables: 1) continuous and 2) dummy. The continuous variable was the probability of choosing PM travel mode, which was then applied in Ordinary Kriging.

Following, the variable “Main travel mode” was discretized in a dummy variable. The value “1” represents the use of PM travel mode and value “0” is for cases of non-use. Subsequently, this variable was used in the Indicator Kriging.

In this study, experimental variograms (constructed based on observed points) for the two variables, were constructed and theoretical models were adjusted with spherical and Gaussian functions, respectively. **Table 2** summarizes the parameters of the theoretical variograms for each regionalized variable (continuous or dummy). **Figure 5(a)** and **Figure 5(b)** illustrates the theoretical variograms in the main direction for the two regionalized variable.

3.2.5. Cross Validation: Continuous and Dummy Variables

In order to assess the accuracy of the model determined in theoretical variograms, parameters such as Mean Error (ME), Mean Normalized Error (MNE) and Variance of Errors (VE), were calculated taking into account the observed and estimated values. In addition, the percentage of correct estimation of PM travel mode was obtained. The performance measures were calculated considering the following equations.

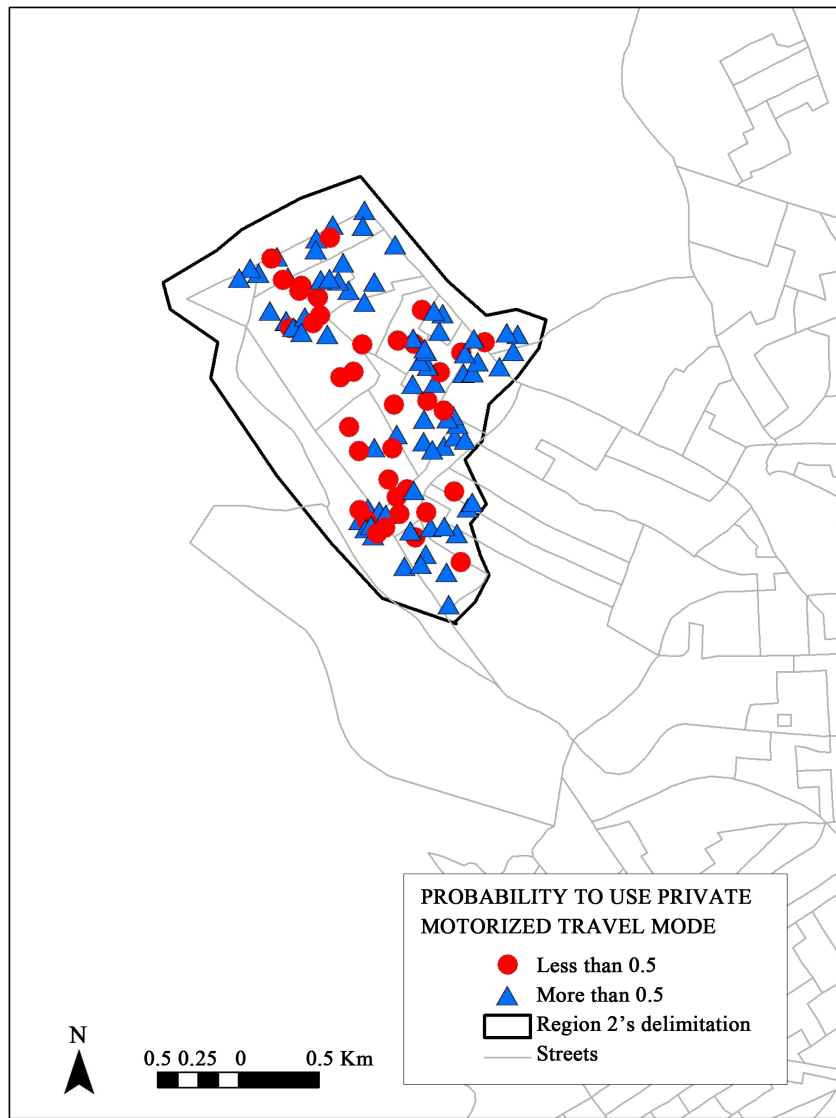


Figure 4. Spatial Distribution of probabilities of motorized private travel mode choice in Region 2.

$$ME = \frac{1}{N} \cdot \sum_{i=1}^N (x_i - y_i) \quad (15)$$

ME = mean error; x_i = estimated value; y_i = observed value.

$$MNE = \frac{1}{N} \cdot \sum_{i=1}^N \frac{(x_i - y_i)}{y_i} \quad (16)$$

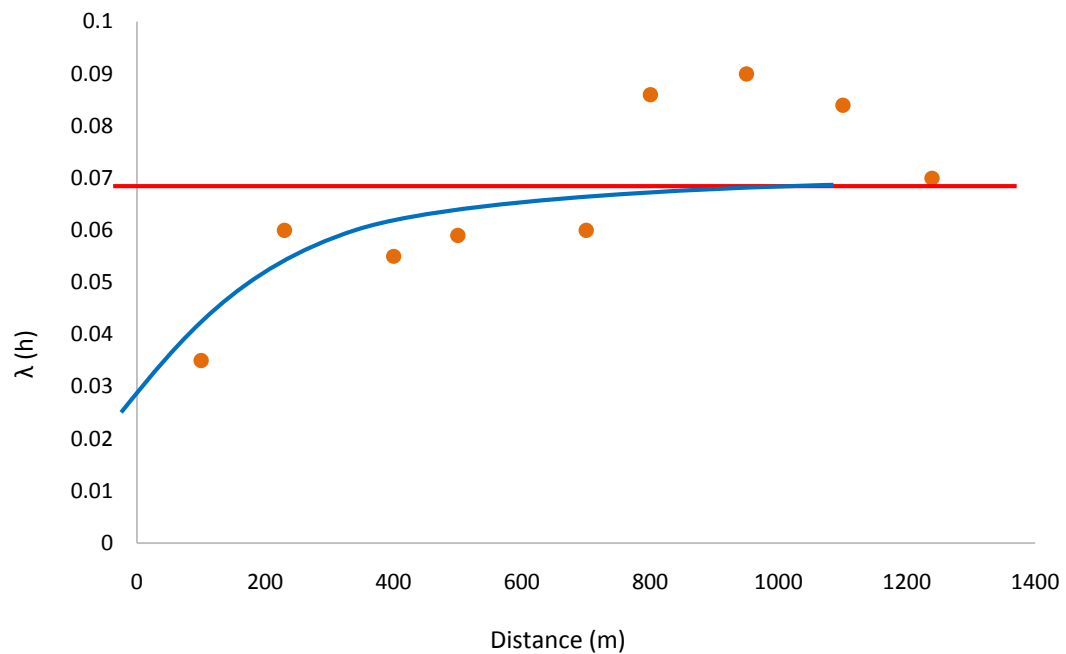
MNE = mean normalized error.

$$var(x - y) = E((x - y) - \mu)^2 \quad (17)$$

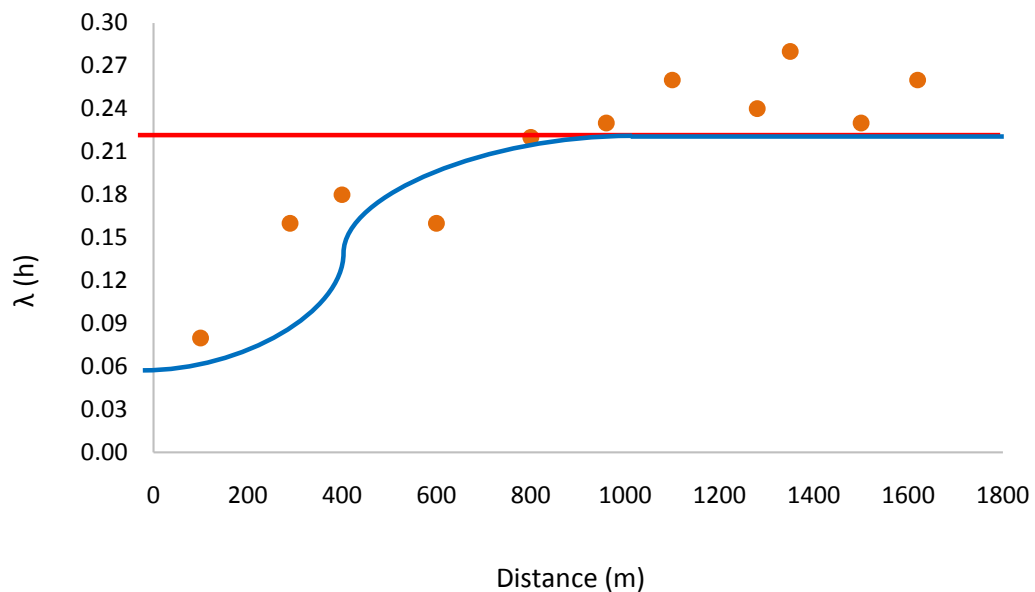
var = Variance of error; $x - y$ = errors; μ = mean of errors.

3.2.6. Kriging: Indicator and Ordinary

For the estimation by OK of the continuous variable and by IK of the dummy variable, a grid of 100×100 meters



(a)



(b)

Figure 5. Theoretical variogram models of axis 30°. (a) Spherical model (continuous variable). (b) Gaussian model (dummy variable).

was established, being its dimensions based on the average distance between households in the study area. Kriging maps were generated by estimation of the two regionalized variables.

4. Results and Discussions

In order to access quality of obtained theoretical variograms, some statistical parameters (ME; MNE and VAR) were calculated and can be observed in [Table 3](#). In addition, the percentage of correct estimation of PM travel mode, considering the estimated and observed values of continuous and dummy variables, was also obtained.

Table 2. Parameters of theoretical variograms for continuous and dummy variables.

Region 2	Continuous variable		Dummy variable	
Axis	30°	-60°	30°	-60°
h (m)	131	132	131	132
Cut distance (m)	100.000	100.000	100.000	100.000
Nugget effect (Co)		0.029		0.070
Range (a)	833 m	363 m	577 m	326 m
Sill (Co + C1)		0.072		0.210
Model		Spherical		Gaussian

Table 3. Cross-validation results.

Region 2	Ordinary Kriging	Indicator Kriging
Mean error	-0.003	-0.001
Normalized mean error	0.6	0.3
Variance of error	0.079	0.104
Hit rate	49%	55%

Upon analysis of cross-validation results, one can see that, despite the low hit rate between observed and estimated values, there were also low values of errors for the both case of Kriging methods. We can also note that the two Kriging methods are very similar. The Indicator Kriging, however, has a higher hit rate, low value of Mean of errors and normalized error.

Important information regarding the accuracy of the theoretical variogram models, as well as the proposed method, that consists in applying two Kriging geostatistical methods to PM travel mode use is listed following: 1) Is to be highlighted that the considered regionalized variables are not natural (variables derive from a previous data treatment and not from direct measures). They were produced by a non-parametric model, the DT model (OK) and from an individual questionnaire (IK); 2) Probably, there are errors associated to survey and data type; 3) The Regionalized Variables are spatially discrete.

Despite the limitations listed for the application of the proposed method, the PM travel mode use was estimated for a 100 × 100 meters grid, for the referred region. Maps generated by Ordinary Kriging estimation and Indicator Kriging are illustrated in **Figure 6(a)** and **Figure 6(b)**.

Observing **Figure 6**, we can note that the Kriging maps from the both Kriging methods are very similar. The results show that this travel mode is more likely to be used in the periphery of the region. Thus, the predisposition for car usage decreases considering the center proximity. Is to be highlighted that the main direction of this variogram (N30E) is clearly translated to the respective map.

The results of both spatial estimation methods are consistent with the reality of the region. Locations with higher probability of car/motorcycle use are exactly those corresponding to neighborhoods of higher income population. Conversely, the center, which is the least probable to use the car, corresponds to low-income households.

The map generated by Indicator Kriging is a surface of probabilities to use PM travel mode (car or motorcycle). Discretizing such values at zero and 1, considering the median of the probabilities, it was obtained a discrete map. The blue color indicates the use of PM travel mode (value 1), while the red color represents non-use. The discrete Kriging map (IK) is shown in **Figure 7**.

5. Conclusions

The results obtained in this study allowed determining the probabilities of PM travel mode (car/motorcycle) use in household locations where travel mode choices were unobserved through two different procedures: 1) joint application of Decision Tree analysis and Ordinary Kriging; and 2) Indicator Kriging.

The resulting maps, obtained from OK and IK, allowed determining that there was a trend in the use of Private

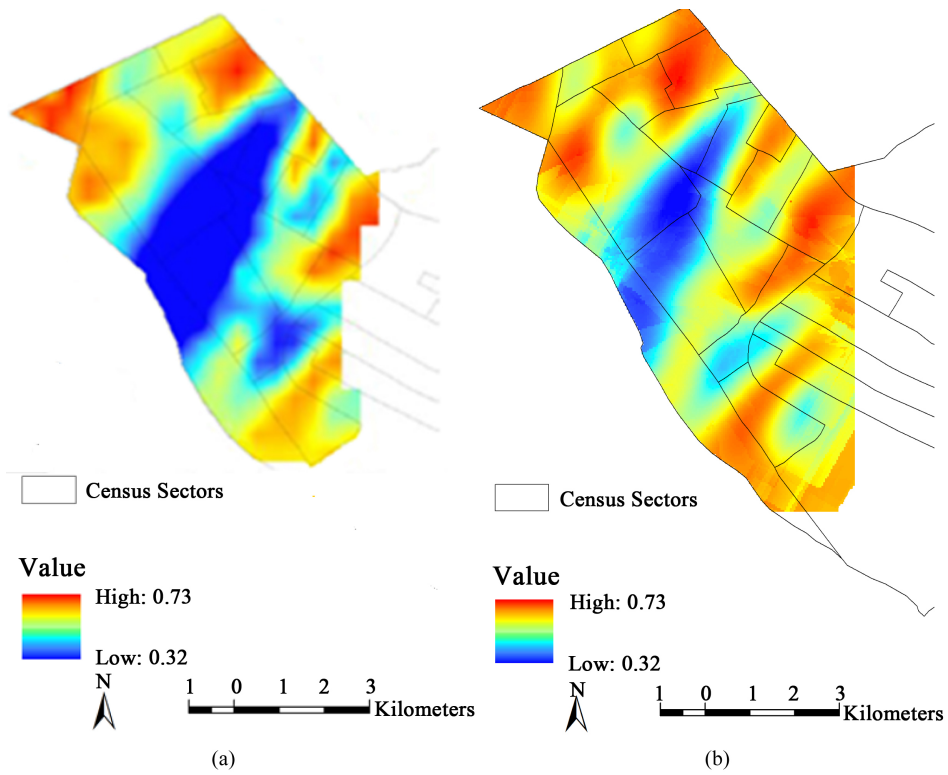


Figure 6. PM travel mode use Surface. (a) Ordinary Kriging. (b) Indicator Kriging.

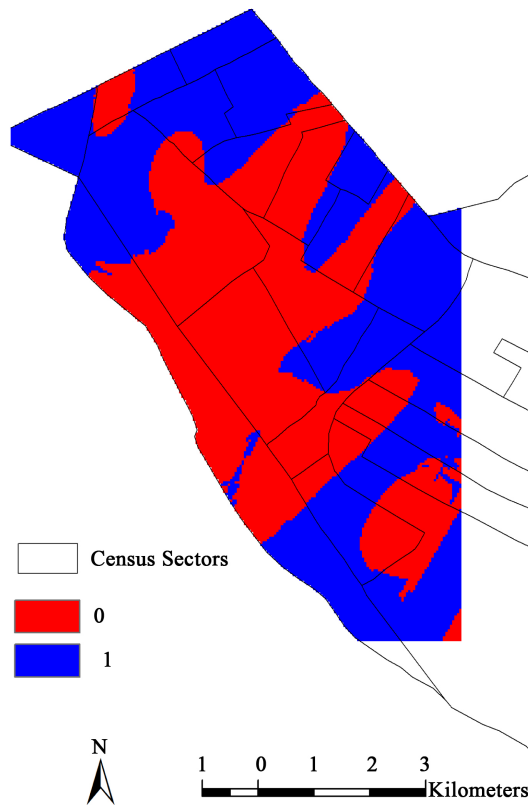


Figure 7. Map of estimation for dummy variable.

Motorized travel mode, which increased from the center to periphery of the selected region. Cross-validation showed good results, for the both approaches, considering mean of errors, normalized mean error and variance of errors.

There was better accuracy in approach by Indicator Kriging, which in addition to higher hit rate and lower mean of errors value, was a univariate procedure.

The application of the Ordinary Kriging approach consisted in a two-step methodology, through the application of DT, which included several covariates. Furthermore, regionalized variable, in this case, was from the previous step. This fact could contribute to error propagation.

Nevertheless, the innovative characteristic of this study should be taken into account. The Kriging geostatistical methods presented are based in unusual techniques in the analysis of travel mode choice. Besides, the results show the success in this combination, which allows a preliminary assessment of spatial particularities of the study area, and, it also emphasizes the necessity of robust/solid information basis, special when dealing with questionnaires.

For future research, we should propose the use of non-spatial traditional techniques for travel mode choice problems (discrete choice models, as instance) for comparison with the performance of spatial analysis based on Kriging approaches.

Acknowledgements

This paper was supported by the *Conselho Nacional de Desenvolvimento Científico e Tecnológico* (CNPQ) and *Fundação de Amparo à Pesquisa no Estado de São Paulo* (FAPESP).

References

- [1] Ortúzar, J.D. and Willumsen, L.G. (2011) *Modelling Transport*. 4th Edition, Wiley, Hoboken. <http://dx.doi.org/10.1002/9781119993308>
- [2] Cervero, R. and Radisch, C. (1996) Pedestrian versus Automobile Oriented Neighborhoods. *Transport Policy*, **3**, 127-141. [http://dx.doi.org/10.1016/0967-070X\(96\)00016-9](http://dx.doi.org/10.1016/0967-070X(96)00016-9)
- [3] Kitamura, R., Mokhtarian, P.L. and Laidet, L. (1997) A Micro-Analysis of Land Use and Travel in Five Neighborhoods in the San Francisco Bay Area. *Transportation*, **24**, 125-158. <http://dx.doi.org/10.1023/A:1017959825565>
- [4] Páez, A., López, F.A., Ruiz, M. and Morency, C. (2013) Development of an Indicator to Assess the Spatial Fit of Discrete Choice Models. *Transportation Research Part B*, **56**, 217-233. <http://dx.doi.org/10.1016/j.trb.2013.08.009>
- [5] Bhat, C. and Zhao, H. (2002) The Spatial Analysis of Activity Stop Generation. *Transportation Research Part B*, **36**, 557-575. [http://dx.doi.org/10.1016/S0191-2615\(01\)00019-4](http://dx.doi.org/10.1016/S0191-2615(01)00019-4)
- [6] Ben-Akiva, M.E., Scott, M.R. and Bekhor, S. (2004) Route Choice Models. *Human Behaviour and Traffic Networks*. Springer, Berlin Heidelberg, 23-45. http://dx.doi.org/10.1007/978-3-662-07809-9_2
- [7] Miyamoto, K., Vichiensan, V., Shimomura, N. and Páez, A. (2004) Discrete Choice Model with Structuralized Spatial Effects for Location Analysis. *Transportation Research Record*, **1898**, 183-190. <http://dx.doi.org/10.3141/1898-22>
- [8] Peer, S., Knockaert, J., Koster, P., Tseng, Y.Y. and Verhoef, E.T. (2013) Door-to-Door Travel Times in RP Departure Time Choice Models: An Approximation Method Using GPS Data. *Transportation Research Part B*, **58**, 134-150. <http://dx.doi.org/10.1016/j.trb.2013.10.006>
- [9] Goovaerts, P. (2009) Medical Geography: A Promising Field of Application for Geostatistics. *Mathematical Geosciences*, **41**, 243-264. <http://dx.doi.org/10.1007/s11004-008-9211-3>
- [10] Miura, H. (2010) A Study of Travel Time Prediction Using Universal Kriging. *TOP*, **18**, 257-270. <http://dx.doi.org/10.1007/s11750-009-0103-6>
- [11] Zou, H., Yue, Y., Li, Q. and Yeh, A.G.O. (2012) An Improved Distance Metric for the Interpolation of Link-Based Traffic Data Using Kriging: A Case Study of a Large-Scale Urban Road Network. *International Journal of Geographical Information Science*, **26**, 667-689. <http://dx.doi.org/10.1080/13658816.2011.609488>
- [12] Pitombo, C.S., Sousa, A.J., Birkin, M. and Quintanilha, J.A. (2010) Comparing Different Spatial Data Analysis to Forecast Trip Generation. *Proceedings of the 12th World Conference on Transport Research Society*, Lisboa, 11-15 July 2010, 1-23.
- [13] Pitombo, C.S., Salgueiro, A.R., Costa, A.S.G. and Isler, C.A. (2015) A Two-Step Method for Mode Choice Estimation with Socioeconomic and Spatial Information. *Spatial Statistics*, **11**, 45-64. <http://dx.doi.org/10.1016/j.spasta.2014.12.002>

-
- [14] Pitombo, C.S., Costa, A.S.G. and Salgueiro, A.R. (2015) Proposal of a Sequential Method for Spatial Interpolation of Mode Choice. *Boletim de Ciências Geodésicas*, **21**, 3. <http://dx.doi.org/10.1590/S1982-21702015000200016>
- [15] Wackernagel, H. (2010) *Multivariate Geostatistics: An Introduction with Applications*. 3rd Edition, Springer-Verlag, Berlin Heidelberg New York.
- [16] Isaaks, E.H. and Srivastava, R.M. (1989) *An Introduction to Applied Geostatistics*. Oxford University Press, Oxford.
- [17] Mendes, R.M. and Lorandi, R. (2002) Engineering Geology Mapping of the Urban Center Area of São José do Rio Preto (Brazil) as an Aid to Urban Planning. *Proceedings of the 9th Congress of the International Association for Engineering Geology and Environment*, **1**, 636-645.
- [18] IBGE (2010) Brazilian Institute of Geography and Statistics. Census Brazilian Population in 2010. <http://www.ibge.gov.br>
- [19] Rodrigues da Silva, A.N. (2008) Preparation of a Travel Database for Assistance of Development Researches in Transportation Planning Area. FAPESP Report, Case No. 04/15843-4. School of Engineering of São Carlos, University of São Paulo, Brazil. (In Portuguese)
- [20] Breiman, L., Friedman, J.H., Olshen, R.A. and Stone, C.J. (1984) *Classification and Regression Trees*. Wadsworth International Group, California.



Scientific Research Publishing

Submit or recommend next manuscript to SCIRP and we will provide best service for you:

Accepting pre-submission inquiries through Email, Facebook, LinkedIn, Twitter, etc
A wide selection of journals (inclusive of 9 subjects, more than 200 journals)
Providing a 24-hour high-quality service
User-friendly online submission system
Fair and swift peer-review system
Efficient typesetting and proofreading procedure
Display of the result of downloads and visits, as well as the number of cited articles
Maximum dissemination of your research work

Submit your manuscript at: <http://papersubmission.scirp.org/>