Prediction of Car Parking Occupancy in Urban Areas Using Geostatistics

Thales Cervi[1], Luiz Melo Jr.[1], Myriam Delgado[1], Semida Silveira[2], Ricardo Lüders[1]

[1]CPGEI – Universidade Tecnológica Federal do Paraná (UTFPR) – Curitiba – Brazil

[2]Systems Engineering Program, Cornell University Ithaca, NY 14853 – USA

{tcervi@alunos.,luizledo,myriam,luders@}utfpr.edu.br, ssemida@hotmail.com

Abstract. *Public on-street car parking is an important shared resource of a city infrastructure with a significant impact on traffic. This paper proposes a geostatistical model aimed to predict parking occupancy rates for different periods of the day. In the study case, the occupancy representation considers the georeferenced position of spots for a particular area of Los Angeles (USA). Different models are compared and their parameters are estimated using the available dataset of the parking area. The final model is chosen to generate a kriging map that helps to understand and predict the occupancy rates. The end goal is to open doors for modeling and predicting urban phenomenons with Geostatistics to help with planning public parking policies in high density urban areas.*

1. Introduction

On-street Parking is an important city resource not only for mediating goods deliveries but also for demanding a vast use of land [Inci 2015]. It can represent an important issue for traffic and greenhouse emissions due to the activity of looking for a parking space by driving a car. According to [Shoup 2006], on average, 30% of cars in traffic are cruising for parking in a study made with congested downtown areas from around the world. Moreover, they spend from 3.5 to 14min (8.1min on average) in this activity. Some studies show that between 9% and 56% of the traffic is cruising for parking [Zhu et al. 2020] and it has a negative effect on the travel time of regular traffic.

There is a growing trend of making data publicly available and consuming it in the context of the Information Systems designed for integrated cities [Bernardini et al. 2017]. From the urban mobility perspective, data driven approaches are made available by IoT technologies [Al-Turjman and Malekloo 2019], [Lin et al. 2017]. They support drivers and parking service providers with information about the status of a parking system (e.g. occupancy, number of vehicles entering and leaving the parking lot). Some works deal with Intelligent Parking Reservation (IPR) systems which gather the provided information and help drivers to reduce the time necessary for finding a spot, parking the vehicle, and paying the fee. However, few of them address the forecast of parking space [Caicedo et al. 2012].

Our work proposes a geostatistical approach to predict the occupancy of on-street parking spaces at different hours of the day using historical parking records in particular urban areas. A spatiotemporal model considers temporal data and spatial interdependence of parking spots due to poles of attraction (services that usually drive the demand for parking like public services or shopping malls, for instance). To the best of our knowledge, it is the first time geostatistics is applied to predict parking occupancy.

2. Related Work

Several works in the literature consider two dimensions for predicting parking places in an integrated parking system to support users in deciding when and where to find parking spots. The first dimension usually deals with the prediction of occupancy from historical data while the second computes the actual parking demand from traffic sensors or simulated data.

A methodology proposed by [Caicedo et al. 2012] predicts real-time parking space in IPR architectures. It aggregates simulated drivers' preferences and parking availability. A real-time availability forecast (RAF) algorithm combines current and historical information. The RAF algorithm is a mix of real and simulated information. The results are compared with a one-by-one approach which simulates the process of searching and finding a parking place for each vehicle. There are no relevant differences when those results are compared to the aggregated approach adopted by RAF which is less time-consuming.

[Vlahogianni et al. 2016], considers two modules: i) a real-time occupancy time-series prediction based on multilayer perceptron networks, and ii) a static approach for estimating the probability of finding a parking space. The first module provides a short-term prediction by using historical data while the second is based on parametric hazard-based modeling which deals with estimating the time necessary to restore normal condition after an incident. The results are generated from data of sensors located in the city of Santander, Spain, and show that the duration of free parking space follows a Weibull distribution. A comparative analysis of other Machine Learning (ML) techniques for the same scenario in Santander can be found in [Awan et al. 2020]. Some works deal with parking forecast by making predictions of parking demand with neural networks, using incoming sensor data, and a probability model of parking space utilization, based on historical data. Basically, they approach the problem as one-dimensional time series forecast. [Camero et al. 2019] propose a new deep learning technique based on Recurrent Neural Networks (RNN) and evolutionary algorithms. It is compared with other ML applied to predict the occupancy of parking lots in Birmingham, UK. The results show similar or better outcomes than other predictors like polynomial fitting, Fourier series, k-means clustering, and time series [Stolfi et al. 2017].

Multivariate autoregressive models that take into account both temporal and spatial correlations of parking availability are investigated in [Rajabioun and Ioannou 2015]. The experiments consider real-time parking data in the areas of San Francisco and Los Angeles. They argue that probability distribution based models are not accurate enough to capture changes in on-street parking due to multi-dimensional dependencies of parking data. They also consider spatial correlations of parking data by observing that changes in one parking facility affect nearby facilities. They propose an autoregressive model by considering both temporal and spatial correlations simultaneously.

The discussed related works use simulation with temporal data. Differently, this paper applies geostatistics to address on-street parking supposing that spots are usually affected by georeferenced points of interest. Although [Rajabioun and Ioannou 2015] addresses the spatial dimension, it is indirectly considered as a correlation of parking data measures on nearby facilities. The georeferenced information is not explicitly included in the model as we propose in our geostatistical model.

3. Methodology

The present paper proposes a geostatistical method of modeling and predicting the on-street public parking occupancy rate phenomenon. Each public park spot is georeferenced, whereas its occupancy rate in a given period of time is a random variable. The geostatistical inference builds a probability distribution of parking occupancy according to the geographical coordinates of parking spots.

3.1. Geostatistical Model

Georeferenced random variables can be represented by a multivariate distribution. In this work, we use the Gaussian multivariate distribution whose covariance matrix takes into account distances between points in space associated with georeferenced random variables. Nearby points have samples with great similarity. The degree of similarity is controlled by spatial correlation functions (e.g exponential, Gaussian). The statistical model is then given by $Y(x_i) = \mu + S(x_i) + Z(x_i), \quad i = 1, \dots, n$ where $Y(x_i)$ is the variable of interest associated to a particular point x_i in geographical space, μ models the average level of Y_i, $S(x_i)$ models the correlation between nearby points of x_i, and $Z(x_i)$ models the uncertainty in the measure of $Y(x_i)$.

The following assumptions are considered:

1. $S(x_i)$ is a weakly stochastic process with $E[S(x_i)] = 0$, $Var(S(x_i)) = \sigma^2$, and $\rho(h_{ij}) = Corr\{S(x_i), S(x_j)\}$ for the euclidean distance $h_{ij} = ||x_i - x_j||$.
2. $Z(x_i)$ is normally distributed with $N(0, \tau^2)$.

According to [Ribeiro and Diggle 2007], **Y** is associated to a multivariate Gaussian normal distribution given by (1) in matrix form:

$$\mathbf{Y} \sim NM(\mu\mathbf{1}, \sigma^2\mathbf{R} + \tau^2\mathbf{I}) \tag{1}$$

such that **1** is an n-dimensional vector of 1's, **R** is a square matrix of order n with elements $\rho(h_{ij})$, and **I** is the identity matrix of order n. Examples of Exponential and Gaussian correlation functions $\rho(h_{ij})$ are given by $\rho(h_{ij}) = \exp(-h_{ij}/\phi)$ and $\rho(h_{ij}) = \exp\left(-(h_{ij}/\phi)^2\right)$, respectively.

The parameter ϕ is called range and controls the distance from which the variables of interest have same variance, i.e., the structured variability in space is no longer observed. Correlation functions control how the correlation decreases between two points in space (abruptly or smoothly). The growth of semivariance caused by the decrease of correlation can be evaluated by a semivariogram.

The semivariogram is characterized by the semivariance $\gamma(h)$ between two measurements of Y spaced h from each other according to (2).

$$\gamma(h) = \frac{1}{2n}\sum_{i=1}^{n}[Y(x_i) - Y(x_j)]^2 \tag{2}$$

The statistical model expects that the variance increases as the distance between the two points increases in any direction (isotropy). The nugget effect $\gamma(0) = \tau^2$ is due to point properties. Moreover, after some distance $h = \phi$, the semivariance converges to $\gamma(\phi) = \sigma^2$, i.e., the range ϕ is a threshold that divides the structured field of correlated samples from the random field of independent samples in space.

The semivariogram is a descriptive analysis technique to verify the spatial structure of measurements. If the statistical model expectations are found for a particular dataset, the variable **Y** of interest is modeled by (1) which is used for prediction as well. The prediction $\hat{T}$ is obtained from theorems of the conditional Gaussian distribution $T|Y$ according to $E[T|Y] = \hat{T} = \mu + \sigma^2 \mathbf{r}'(\sigma^2 R + \tau^2 \mathbf{I})^{-1}(Y - \mu\mathbf{1})$. Matrix **r** correlates the measured data Y with the predicted point T which generates a data interpolation called kriging. The loss function $LOSS(\theta)$ given by (3) measures the quality of models using different variogram parameters $\theta = (\phi, \tau^2, \sigma^2)$,

$$LOSS(\theta) = \sum_k n_k(\hat{\gamma_k} - \gamma_k(\theta))^2 \tag{3}$$

where $\hat{\gamma_k}$ and $\gamma_k(\theta)$ are values of the empirical and theoretical variograms, respectively, weighted by n_k values of each bin k, according to [Ribeiro and Diggle 2001]. The best model which minimizes $LOSS(\theta)$ is chosen among different correlation functions. The parameter θ of the best model is estimated by maximum likelihood estimation from (1).

3.2. Dataset

The dataset used for the experiments is publicly provided by Los Angeles (USA) Department of Transports [LADOT 2022] with information on the city public on-street parking. It contains historical occupancy data about several parking spots, each one identified by a unique ID and a set of information including its geographical position, latitude and longitude coordinates later converted to UTM.

There is also a set of binary events representing each ID occupancy state on a given timestamp, such that 1 represents spot taken, and 0 spot free on that given time. From those events, it is possible to extract the parking spot occupancy as a binary function of time. This work considers a set of 11 different time periods of the day obtained by dividing working hours from 08:00 am to 07:00 pm into time windows of one hour size. For each time window, the probability of occupancy can be calculated as the average of the reported occupancy rate for each week day. Taking the average over several different days helps smooth the heterogeneous measurements of the dataset. The reported occupancy is the fraction of that time window in which the spot occupancy binary function is 1.

The work focuses on a subarea of Los Angeles city to reduce the amount of data to be processed. Also, a data cleaning procedure ensures that only parking spots with more transactions than average are selected, removing outliers with low number of collected transactions. After the initial data cleaning, 690 on-street parking spots, with an average number of transactions around 6000 each, remain for the chosen time frame. The dataset considered in this work contains data from January to October 2021, excluding weekends and holidays.

4. Results

This section presents the results of the geostatistical analysis on the Los Angeles (USA) public parking available data. The same procedure was repeated for each time window.

The results have been generated using *geoR* [Ribeiro and Diggle 2001], a geostatistics library of the software package R. The addressed data has been loaded into a

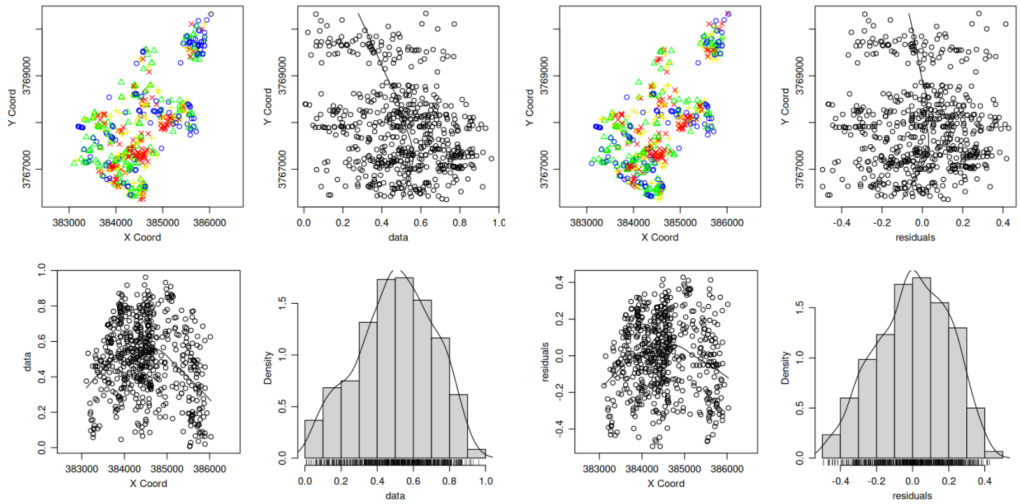

Figure 1. Exploratory analysis of parking data (09:00-10:00 am). The four plots on the left: i) distribution of parking spots of LA (upper left); ii) occupancy rate data against Y coordinates (upper right); iii) occupancy rate data against X coordinates (lower left); iv) histograms of the occupancy rate (lower right). The four plots on the right: similar meaning with bias filtered in coordinate Y.

specific library object of type *geodata* to facilitate the geostatistical analysis. The procedure starts with an exploratory analysis, see results in Figure 1. Color plots show on-street public parking spots with their respective UTM positions. The data is divided into quartiles following the default *geoR* color code with blue representing the lowest occupancy rate and red the higher one. The occupancy rate data is then plotted against axes X and Y (B&W plots of parking spots in Figure 1). These plots help to identify any uni-directional bias that might be affecting the data. Histograms of the occupancy rate are also plotted to understand and verify the normality of data distributions. The plot of occupancy rate data against Y coordinates in Figure 1 shows a slight linear data bias in direction Y, an effect observed in all time windows.

The semivariogram is then generated and checked against four correlation functions: Exponential, Mattern, Spherical and Gaussian, as shown in Figure 2. The best correlation function is chosen by computing the loss value according to (3) for each model, with parameters adjusted by the maximum likelihood estimation function *variofit* of *geoR*.

Figure 3 shows the loss of each model with different correlation functions for each period of the day relative to the worst model (bars are proportional to the loss difference between the considered model and the worst one). It is possible to observe that there is no better correlation function for all time windows, although the Exponential function has the best value (i.e. lowest relative loss value) in 8 of 11 time windows, coming tightly close to the best value on 2 of the remaining 3 windows.

After choosing the exponential correlation function, the estimated model parameters $\theta = (\phi, \tau^2, \sigma^2)$ are computed for each period of the day as shown in Table 1. The range ϕ allows a better understanding of how far a parking spot occupancy can affect a different spot occupancy during the day, as shown in Figure 4.

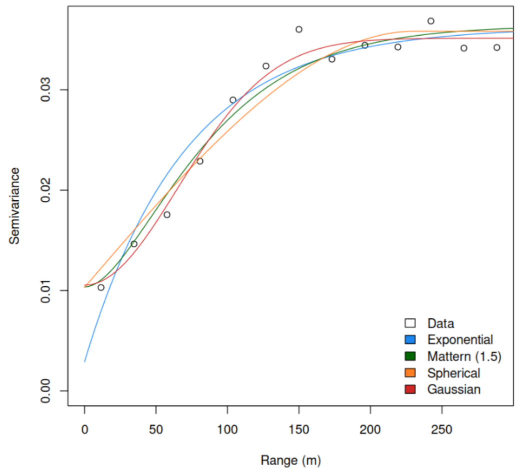

Figure 2. Semivariogram for the data from 09:00-10:00 am and for the four models with different correlation functions

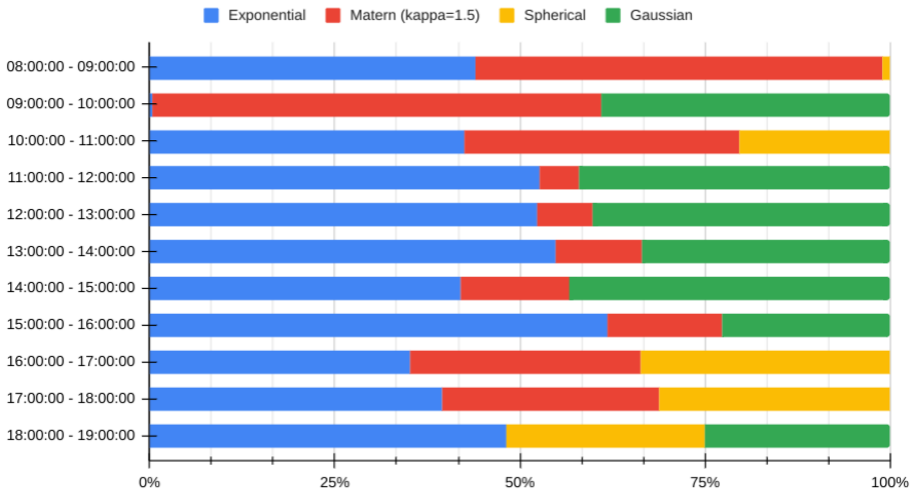

Figure 3. Loss of each model with different correlation functions for each period of the day. Bars are proportional to the normalized difference to the worst model (big bars are better)

Table 1. Estimated model parameters for each time period of the day.

Time period	σ^2	τ^2	ϕ	Time period	σ^2	τ^2	ϕ
08:00-09:00	0.0305	0.0026	51.94	14:00-15:00	0.0190	0.0103	94.56
09:00-10:00	0.0347	0.0037	75.47	15:00-16:00	0.0306	0.0083	75.29
10:00-11:00	0.0305	0.0071	96.85	16:00-17:00	0.0475	0.0059	70.57
11:00-12:00	0.0231	0.0099	107.7	17:00-18:00	0.0511	0.0045	81.03
12:00-13:00	0.0201	0.0105	108.7	18:00-19:00	0.0527	0.0052	87.22
13:00-14:00	0.0179	0.0109	105.2				

The kriging map of the considered area is presented in Figure 5. It is computed with the *krige.conv* function of *geoR* using the chosen model for 09:00-10:00 am. A simulation of 1000 predictions is made in a grid of points that are 75 m distant from each other, according to the range of 75.74 m obtained for this time window. When the

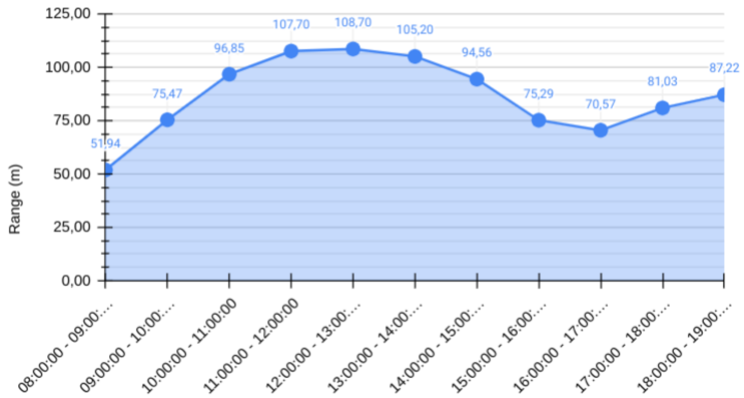

Figure 4. Parking occupancy spatial structure range (ϕ) during the day

map is overlapped with the Los Angeles city map, it is possible to understand the parking occupancy as heat map and visualize where the occupancy rate is low (green area) or high (red area). The kriging map is an important result of geostatistical analysis. It allows the prediction of unknown occupancy rate of parking spots with coordinates known inside the modeled area. The selected kriging area overlapped with the city map corresponds to the so called Los Angeles Fashion District. This area contains a high concentration of fashion stores, which may explain the high parking occupancy in this area.

Figure 5. Kriging map of parking occupancy between 09:00-10:00 am for LA area

5. Conclusion

This work presented a geostatistical model for predicting occupancy of on-street car parking facilities in the city of Los Angeles. Four correlation functions were evaluated with the Exponential function giving the best results. A total of 11 periods of the day were considered from 08:00 am to 07:00 pm with different estimated parameters for each one. The range of parking occupancy presented peak values around 12:00 pm. It probably means that crowded areas extend the spatial structure correlations around 100 m at peak hours, which is compatible with a circular area of 200 m (or one block on a street). The obtained kriging map has shown an occupancy partially constrained by the topology of streets. The results have limitations such as the small number of spots with available data, and poorly distributed parking spots. The issues could be further investigated in the future, not least using data from different cities for the development of a robust model. We expect that the proposed approach could be used for optimizing the delivery of goods in urban areas with restricted parking hours and places, reinforcing the trend of using historical parking data analysis to guide public policies and decisions related to parking fee adjustment.

Acknowledgments

This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Code 001, by the National Council for Scientific and Technological Development – CNPq (grants 314699/2020-1 and 439226/2018-0), and partially supported by the Smart City Concepts in Curitiba Project.

References

Al-Turjman, F. and Malekloo, A. (2019). Smart parking in IoT-enabled cities: A survey. *Sustainable Cities and Society*, 49.

Awan, F. M., Saleem, Y., Minerva, R., and Crespi, N. (2020). A comparative analysis of machine/deep learning models for parking space availability prediction. *Sensors (Switzerland)*, 20.

Bernardini, F., Viterbo, J., Vianna, D., Martins, C., Medeiros, A., Meza, E., Moratori, P., and Bastos, C. (2017). General features of smart city approaches from information systems perspective and its challenges. In Boscarioli, C., Araujo, R., and Maciel, R., editors, *I GranDSI-BR: Grand Research Challenges in Information Systems in Brazil 2016-2026*, pages 25–40. Brazilian Computer Society (SBC).

Caicedo, F., Blazquez, C., and Miranda, P. (2012). Prediction of parking space availability in real time. *Expert Systems with Applications*, 39.

Camero, A., Toutouh, J., Stolfi, D. H., and Alba, E. (2019). Evolutionary deep learning for car park occupancy prediction in smart cities. volume 11353 LNCS.

Inci, E. (2015). A review of the economics of parking. *Economics of Transportation*, 4.

LADOT (2022). Los Angeles city open data: Parking meter occupancy.

Lin, T., Rivano, H., and Mouel, F. L. (2017). A survey of smart parking solutions. *IEEE Transactions on Intelligent Transportation Systems*, 18.

Rajabioun, T. and Ioannou, P. (2015). On-street and off-street parking availability prediction using multivariate spatiotemporal models. *IEEE Transactions on Intelligent Transportation Systems*, 16.

Ribeiro, P. J. and Diggle, P. J. (2001). geoR: A package for geostatistical analysis. *R-NEWS*, 1(2).

Ribeiro, P. J. and Diggle, P. J. (2007). *Model-based Geostatistics*. Hardcover, 1 edition.

Shoup, D. C. (2006). Cruising for parking. *Transport policy*, 13(6):479–486.

Stolfi, D. H., Alba, E., and Yao, X. (2017). Predicting car park occupancy rates in smart cities. volume 10268 LNCS.

Vlahogianni, E. I., Kepaptsoglou, K., Tsetsos, V., and Karlaftis, M. G. (2016). A real-time parking prediction system for smart cities. *Journal of Intelligent Transportation Systems: Technology, Planning, and Operations*, 20.

Zhu, Y., Ye, X., Chen, J., Yan, X., and Wang, T. (2020). Impact of cruising for parking on travel time of traffic flow. *Sustainability (Switzerland)*, 12.