MODELING OF THE NEED FOR PARKING SPACE IN THE DISTRICTS OF MOSCOW METROPOLIS BY USING MULTIVARIATE METHODS

Roman Sidorchuk*, Sergey Vladimirovich Mkhitaryan, Irina Ivanovna Skorobogatykh,
Anastasia Alekseevna Stukalova, Anastasia Vladimirovna Lukina
Plekhanov Russian University of Economics, Russian Federation

The growth of metropolis cities and consequently the number of vehicles cruising within their boundaries create a permanent problem of dissatisfaction with the amount of parking space and its over-occupancy. The results of continuous observation of parking lots in Moscow and data on registered cars in the city districts was the initial basis for this study. The data was processed by IBM SPSS Statistics 20 statistical program to obtain descriptive statistics indicators of parking space in Moscow, the analysis of cause-and-effect relations and subsequent multivariate modeling using regression analysis; log it regression; discriminant analysis; "classification trees" (decision tree). The results clearly show the possibility of applying the methods of multivariate statistics, log it regression and "classification trees". Both models allow for using the explanatory variables "proportion of parking lots with violations" and "number of parking spaces in the street and road network" to analyze the impact on parking lot occupancy. Also, the descriptive statistics analysis revealed that when the number and proportion of parking lots with violations are 2 times higher on average in the districts with over-occupied parking lots versus the districts where the parking lot occupancy is not so high, and the number of paid parking lots is over 10 times less. The increase in the proportion of parking spaces with violations ranging from 0 to 0.2% entails a sharp increase in parking space occupancy (up to 90%), while a further increase in the proportion of parking spaces with violations does not entail a significant increase in the parking occupancy.

Key words: parking, parking space modeling, parking policy, metropolis, smart technologies

INTRODUCTION

Mass urban migration of people will entail, per some forecasts, to the fact that "by 2040 cities will be populated by 70% of the world's population" [20]. Such population growth in the cities will lead to significant transport problems, one of which is the growth of traffic and the parking space occupancy. The solution of this problem requires proper planning and parking policy management.

This article presents a study on modeling the availability of parking spaces in the metropolis, as exemplified by the metropolitan city of Moscow. This study is aimed at finding tools for modeling the parking space occupancy s for the subsequent creation of technological smart solutions within the "Smart City" strategy. An important issue for the development of municipal parking policy in the metropolis is related to the provision of various districts with parking spaces. A significant challenge here is to develop a model aimed to determine the demand for parking spaces in municipal districts. The importance of solving this problem is based on the fact that the problem of parking space arrangement is becoming critical in municipal planning [17, 23], in the face of population growth in major Metropolitan areas [2, 7 8].

The metropolis of Moscow is not an exception here, as the number of cars per 1,000 inhabitants is equal to that in the world's largest metropolises [19].

In this paper, we present a part of the results of the analysis of the study concerning parking lots in Moscow that was conducted from December 26, 2018 throughout June 30, 2019. The outcomes of field studies of parking spaces available in the metropolis of Moscow, performed by 285 professors, experts and students of Plekhanov Russian University of Economics were the basis of the initial data used for the analysis. An observation method was used to obtain the data [16].

The literature review revealed certain gaps in the analysis of the factors influencing the parking space occupancy in the metropolitan urban areas. Therefore, the objectives of this study were:

- To obtain descriptive statistics of the indicators of parking space available in Moscow;
- To identify the factors of influence on the basis of the analysis of cause-and-effect relations between the explanatory indicators of parking space and parking space occupancy.
- To analyze and verify the effectiveness of predictive models for determining the parking occupancy.

Based on the established objective of the study, the following hypotheses are proposed:

H1. The dependence of the parking occupancy indicator on predictors is nonlinear;

H2. The models of parking space provision in the metropolis can be formulated on the basis of the revealed cause-and-effect relations between the explanatory indicators of parking space and parking space occupancy by applying the methods of multivariate modeling.

*professor_sidorchuk@mail.ru

Roman Sidorchuk, et al. - Modeling of the need for parking space in the districts of Moscow metropolis by using multivariate methods

iipp

The data array was analyzed in this study using the IBM SPSS Statistics 20 software package. The calculations of total absolute and averaged relative indicators for each district of Moscow, as well as the calculation of additional indicators were carried out based on the obtained data array by using consolidated tables. Occupancy (%) as the proportion of the number of parked cars and the number of parking spaces was calculated on the basis of initial data for each parking lot. Then, with the help of correlation analysis and scattering patterns, causal relations between the explanatory indicators of parking space and parking lot occupancy were determined. The parking lot occupancy models were developed on the basis of the obtained data using statistical methods of multivariate modeling: regression analysis; logit regression; discriminant analysis; classification trees.

The obtained results showed that the models, where the predicted indicator is the parking lot over-occupancy in the city district were of the greatest interest for the parking occupancy modeling in terms of quality and efficiency. These are the logit regression model and the model of classification trees.

The novelty of the study is stipulated, firstly, by the revealed cause-and-effect relations between the occupancy of parking spaces, over-occupancy of parking lots and parking spaces with violations; secondly, by the hypothesis about the nonlinear nature of the relation between the indicators and the parking lot occupancy that was set and confirmed in the course of the study; third, by the outcomes of multivariate modeling, which revealed the most appropriate methods of the parking lot occupancy modeling.

We assume that the city authorities can contribute to the creation of smart solutions for the parking policy implementation on the basis of the proposed models, both in the metropolis, as a whole, and in particular urban districts.

LITERATURE REVIEW

Two main directions in the parking space modeling can be identified in the reviewed literature, namely: simulation based on queuing theory and modeling by using statistical methods.

Queuing theory based modeling

The literature sources suggest the use of the queuing theory to model various parameters of parking methods and parking spaces [9, 29]. The modeling and management of transport behavior of drivers based on the queuing theory with the use of smart parking systems is an important issue being discussed in the literature. This will reduce unstructured parking and parking violations, and also will increase the parking space occupancy efficiency in the permitted parking lots [6, 22, 24, 30, 34].

Queuing theory studies and models the processes provoking the occurrence of the units to perform certain type of services and the servicing of these units [11]. The studied objects may include production processes, supply processes, transport, trade, stocks. For example, the authors demonstrate the application of the queuing theory to model the distribution of elevators in the underground parking [35]. The random nature of the main parameters of order queuing systems is a common feature of all problems of the queuing theory:

- The number of units received for service and the time interval between their receipt are random variables;
- The duration of each request service is random [12].

The queuing system is usually divided into two parts:

- Serviced subsystem (population, customers, and drivers) - where the demand (need) for servicing arises;
- Servicing subsystem, which is designed to cover the incoming servicing demands.

The literature presents the studies comprising an attempt to model a navigation system to predict parking occupancy and availability of parking spaces, based on queuing theory [15]. In some studies, this problem is localized for the part of urban space. For example, it pertains to the business part of the city [32]. Other researchers propose a method for determining the number of the entry terminals at the entrances to the closed parking lot (garage), also based on queuing theory [13].

Modeling based on statistical methods

Approaches to modeling based on statistical methods are more widely presented in the literature. Various statistical methods and sets of indicators can be used to model parking space occupancy [5]. The literature sources provide examples of modeling the user behavior and the time required to park a car [14]. This study deploys two types of multiple linear regression models using SPSS software. Based on the analysis of models, the most appropriate explanatory variables affecting the demand for parking are identified having regard to the socio-economic behavior of the user. The authors proved that the most influential explanatory variables for this model are the travel distance and the income per month. The cumulative model study identifies the most appropriate explanatory variable to evaluate the parking demand and, according to the researchers, can be used to determine the number of parking spaces that will be provided to meet the parking demand.

In some works [3], parking modeling serves as the basis for finding "Smart City" solutions, where flow management, space availability management and transport system development is based on simulation and modeling proceeding from formalization using the Discrete Event System Specification. Other researchers base their models on formulation of forecasts using cumulative data on the availability of parking spaces [26].

Zheng et al. [35] proposed an interesting approach of combining statistical data with the results of cluster anal-

ysis of "parking anomalies" to solve the problem with modeling of parking space availability. Some researchers analyze the parking lots operated by private operators using a limited number of indicators: Parking capacity, parking fee and access period [31]. Other researchers reduce the variety of indicators of parking spaces to the cost and time spent on search for available parking space. "From the view of microeconomics, the chosen car park is determined by the value of utility. Motorists will always choose the car park with the maximal utility value, which is related to the consumed time and the cost" [35].

A number of studies notes that the different cities will most likely face the same situation with parking spaces, i.e. between 8 and 18 on weekdays – high demand for parking. Researchers can develop predictive models for such urban areas (e.g. office buildings, restaurants, residential areas, etc.) that possess parking occupancy data [10].

Summing up the review, it can be stated that the literature contains a gap in the use of cause-and-effect relations between the explanatory indicators of parking space and parking space occupancy for the subsequent modeling of the provision of parking spaces in the metropolis. In addition, the authors focus mainly on either a separate parking lot or a parking lot in a part of the city, without analyzing the parking lots in different districts of the city in their totality.

METHODS

Field study and descriptive statistics

In this paper, we use data on parking lots in Moscow obtained in the period of field study and data collection by observation from December 26, 2018 throughout June 30, 2019. Two hundred and eighty faculty members, employees and students of Plekhanov Russian University of Economics participated in the observation [4]. The method of observation corresponded to the approaches described in the literature [16]. An array of data describing the various indicators of parking lots in Moscow was obtained as a result. For the subsequent processing, the received data on observations of parking space of Moscow was assumed in this paper including: a list of parking spaces with indication of city district for each parking lot, parking ring reference, number of parking spaces, number of parked cars, number of cars parked with violations, position relative to the street and road network (roadside, SRN), parking tariff, method of placing (parking or storage).

The data with its distribution by districts of Moscow and the "ring references" were used for processing. The districts of Moscow include the following municipal administrative territories: 125 municipal districts, 19 settlements and 2 city districts. In our study, we used data on the districts, which had been a part of Moscow before the administrative expansion in 2012 [1].

The ring reference is an indicator that determines the position of the parking lot relative to the ring roads dividing the city. The ring reference is used by the municipal authorities to set basic parking tariffs. There are 4 rings: Boulevard Ring (BR), Garden Ring (GR), Third Transport Ring (TTR) and Moscow Ring Road (MRR) [18]. The ring reference zones are shown in Table 1.

Table 1: Indicators of parking space in Moscow by ring reference

No.	Indicator
1	Internal territory between TTR and MRR
2	Internal territory of BR
3	Internal territory between GR and TTR
4	Internal territory between BR and GR
5	Outside MRR

The analysis of tariffs was based on the principle of formation and not their cost values:

- Common, when the cost of parking does not depend on its duration;
- Progressive, when the cost of parking increases based on its duration [25].

In addition, the Department for Transport and Road Infrastructure Development of the Moscow City Government provided data on the number of the cars registered in the districts of Moscow. However, data on the number of cars registered in the districts of Moscow are not available in the public domain.

Parking occupancy (%) as the proportion of the number of parked cars and the number of parking spaces was calculated on the basis of initial data for each parking space.

The preliminary data processing consisted in the calculation by using consolidated tables of total absolute and averaged relative indicators for each district of Moscow, as well as the calculation of additional indicators. As a result, the following indicators should have been obtained for each district (Table 2).

Modeling

Modeling of parking occupancy was carried out using a variety of statistical methods of multivariate modeling:

- Regression analysis;
- Logit regression;
- Discriminant analysis;
- "Classification trees".

These modeling tools are multivariate statistical methods and allow consider the impact of a set of indicators on the dependent variable [5].

Multiple regression analysis allows to evaluate the relation between several explanatory variables (also called

regressors or predictors) and the dependent variable. Regression analysis provides a mathematical description of a particular type of dependencies. The regression analysis results in a model represented by the regression equation. The measure of the model adequacy to the initial data (model quality) is the measure of agreement – determination factor – R^2 and the residual characteristics.

Discriminant analysis and logit regression are predictive models, in which the resulting measure (response) is categorical and the explanatory variables are measured in an interval (quantitative) scale. Discriminant analysis interprets the data area as consisting of separate sets, each of which is characterized by variables with multivariate normal distribution.

Logit regression evaluates the probability of occurrence of a certain event with an increase in the set of factors affecting it, this function is not linear, but S-shaped.

Classification trees are a heuristic method of examining many explanatory variables to determine such combinations of categories that yield the highest percentage under the desired response condition. If successful, the resulting tree shows which explanatory variable is most closely related to the target variable.

RESULTS

Primary data analysis allows the solution of two problems:

- To obtain descriptive statistics of the indicators of parking space available in Moscow;
- To identify the cause-and-effect relations between the explanatory indicators of parking space and parking space occupancy.

Descriptive statistics of parking space indicators are presented in Table 3. The minimum, maximum, and average values of indicators for Moscow districts were obtained. Parking occupancy varies from 42% to 97%, whereas the average value of 81% is significantly closer to the maximum value; therefore, the number of districts with parking occupancy above the average value exceeds the number of districts with less occupancy. Parking occupancy exceeds 90% in 50 districts.

The number of cars registered in the district varies from 1.92 thousand cars to 90.91 thousand cars, while the average value is 31.53 thousand cars; in a similar manner, the number of districts with the number of registered cars below the average value exceeds the number of districts with a large number of registered cars. The number of registered cars ranges from 2 up to 40 thousand in more than 60 districts.

The number of parking and storage spaces, thousand spaces is found to vary from 2.04 and 0 thousand spaces up to 95.44 and 15.87 thousand spaces respectively, whereas the average value is 14.46 thousand parking spaces and 3.60 thousand storage spaces Similarly, the number of districts with the number of parking and storage spaces below the average value is found to surpass the number of districts with a large number of parking and storage spaces. In more than 70 districts, parking spaces amount to 13-27 thousand, and storage spaces up to 4 thousand.

The share of parking lots with violations varies from 0 up to 4.5%, with the average value being 0.8%. Similarly, the number of districts with a share of parking lots with violations below the average value surpasses the number of districts with a large share of parking lots with violations. In 50 districts, the share of parking lots with violations is less than 0.33%.

Table 2: Indicators of parking space in Moscow by districts

No.	Indicator, units.	Notation in equations
1	Parking space occupancy (average for the district), %	Y
2	Number of cars registered in the district, thousand pcs.	x_1
3	Parking, thousand spaces	x_2
4	Storage, thousand spaces	x_3
5	Parking spaces with violations, t/s	x_4
6	Share of parking lots with violations, (5)/(2)/1000	x_5
7	Common (charge tariff), for parking	x_6
8	Progressive (charge tariff), for parking	x_7
9	Number of paid parking lots, (7)+(8)	x_8
10	Number of parking lots between the Boulevard Ring and the Garden Ring	x_9
11	Number of parking lots between the Garden Ring and TTR	x_{10}
12	Number of parking lots between TTR and MRR	x_{11}
13	Number of parking lots outside MRR	x_{12}
14	Number of parking lots in SRN	x_{13}

The number of paid parking lots ranges from 0 up to 3.238, the average value is 63 parking lots, the number of districts with the number of paid parking lots below the average value exceeds the number of districts with a large number of paid parking lots. Over 110 districts have up to 200 paid parking lots.

The standard deviation shows the average deviation from the mean values, the variation coefficient is the standard deviation in fractions of the average; therefore, the variation coefficient should be used for the comparative assessment of the variability degree of indicators between the districts (Figure 1).
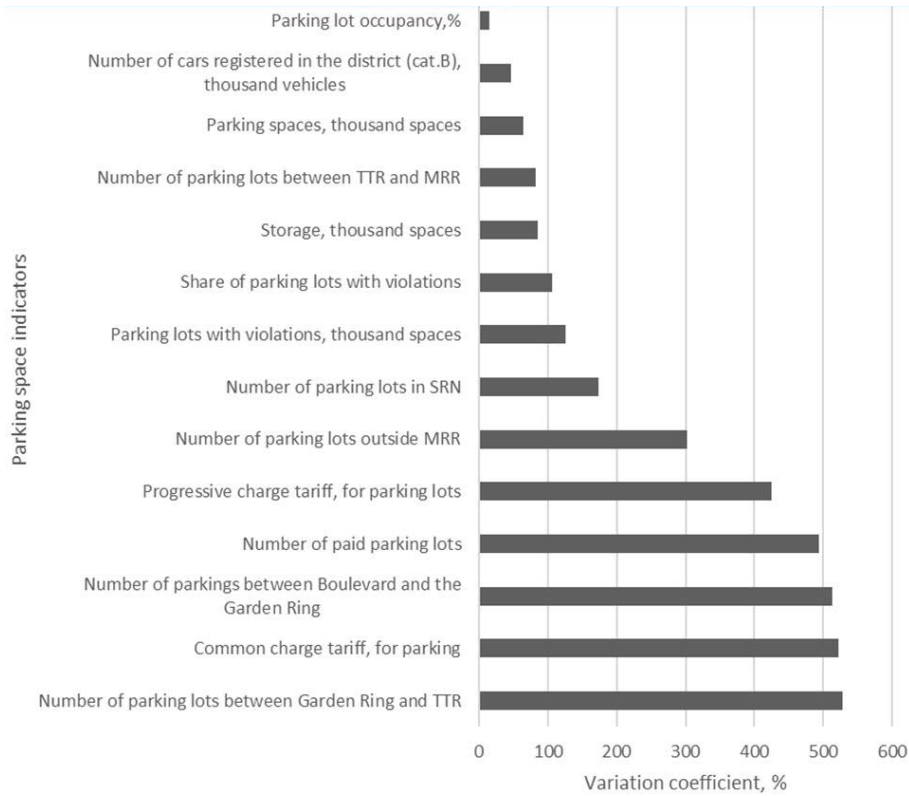

Figure 1: Variation coefficients of parking space indicators

Table 3: Descriptive statistics of parking space indicators in Moscow

No.	Indicator	Minimum	Maximum	Average	Mean square Deviation	Variation coefficient, %
1	Occupancy of parking lots, %	42	97	81	13	15
2	Number of cars registered in the district, thousand cars	1.92	90.91	31.53	14.50	46
3	Parking, thousand spaces	2.04	95.44	22.25	14.46	65
4	Storage, thousand spaces	0	15.87	4.20	3.60	86
5	Parking lots with violations, t/l	0	1.390	177.75	222.79	125
6	Share of parking lots with violations	0	0.045	0.008	0.009	106
7	Common (charge tariff), for a parking lot	0	3.204	58.65	306.25	522
8	Progressive (charge tariff), for a parking lot	0	120	4.05	17.213	425
9	Number of paid parking lots	0	3.238	63	310	494
10	Number of parking lots between the Boulevard Ring and the Garden Ring	0	675	19	99	513
11	Number of parking lots between the Garden Ring and TTR	0	3.025	59	314	527
12	Number of parking lots between TTR and MRR	0	1.870	377	311	83
13	Number of parking lots outside MRR	0	958	43	130	302
14	Number of parking lots in SRN	0	3.198	174	301	173

The highest variability is found with the following indicators:

- Number of parking lots between the Garden Ring and TTR;
- Common charge tariff for parking lots;
- Number of parking lots between the Boulevard and the Garden Ring;
- Number of paid parking lots.
- The smallest variability:
- Parking, thousand spaces;
- Number of cars registered in the district, thousand cars
- Parking lot occupancy, %.

Moreover, the lowest variation coefficient is the indicator of parking lot occupancy (15%).

An additional occupancy rate of parking lots in the district (k) was introduced based on the occupancy rate of parking lots (y) to determine the average occupancy rate of parking lots according to the formula.

$$k = \begin{cases} 1, \text{if } y \geq 85\% - \text{parking lots are over} - \text{occupied} \\ 0, \text{if } y < 85\% - \text{parking lots are not over} - \text{occupied} \end{cases} \quad (1)$$

The share of districts with over-occupied parking lots amounted to 52%.

The cause-and-effect relations between the explanatory indicators of parking space and parking occupancy were assessed using the correlation analysis and the scatter patterns. The correlation analysis revealed the presence of a poor statistical linear relation between the parking occupancy, parking lots with violations and the share of parking lots with violations (pair correlation factors were r=0.37 and r=0.42, respectively). There was no statistically significant correlation with any other indicators found. In this regard, a

repeated correlation analysis with squared values of variables was carried out, based on the assumption of the parabolic nature of the influence of indicators on the parking lot occupancy. As a result, a statistically significant relation between the parking lot occupancy and the following parameters was revealed:

- Parking, thousand spaces (*r=0.47*);
- Storage, thousand spaces (*r=0.48*);
- Parking lots with violations, t/s (*r=0.45*);
- Number of parking lots between TTR and MRR (*r=0.30*);
- Number of parking lots outside MRR (*r=0.32*).

To clarify the nature of the most significant relation between the parking lot occupancy and the share of parking lots with violations, a scatter pattern was created (Figure 2).

According to the diagram, a moderate increase in the proportion of parking lots with violations from 0 to 0.2% entails a sharp increase in parking lot occupancy (up to 90%), further increase in the proportion of parking lots
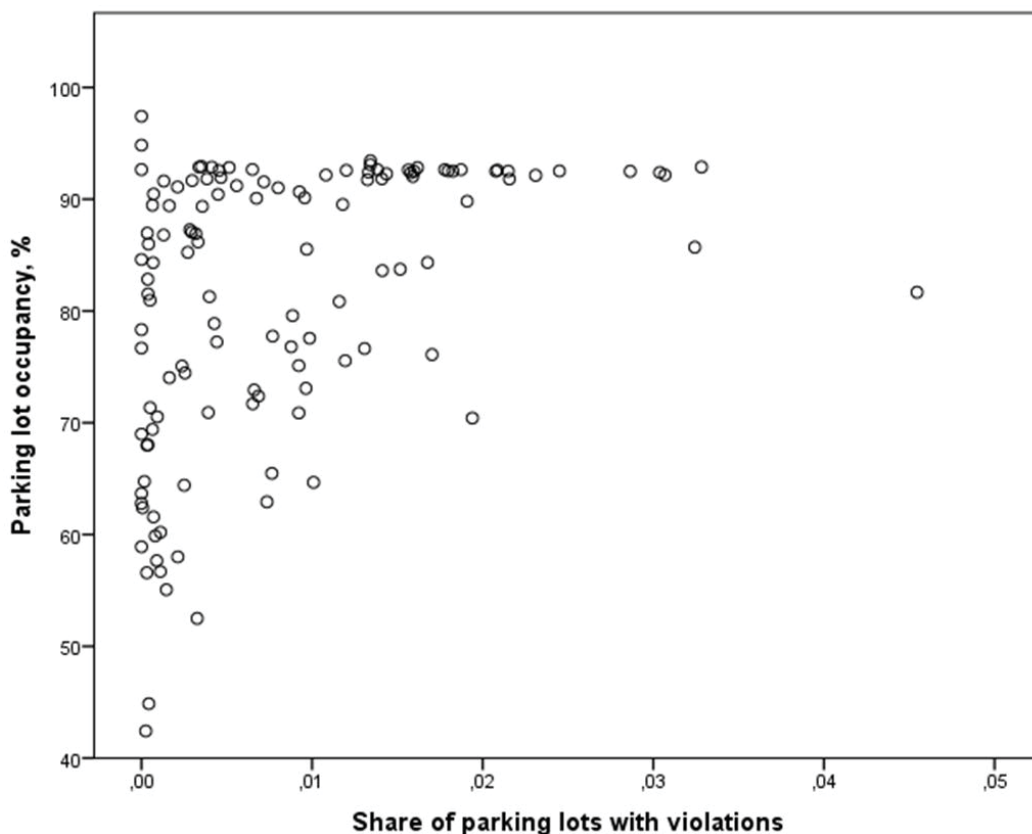

Figure 2: Scatter pattern characterizing the relation between the parking lot occupancy and the share of parking lots with violations

Table 4: Average values of parking space indicators for parking lots with different levels of occupancy

Average values of indicators	Indicator of parking occupancy in the district	
	Not over-occupied parking lots	Over-occupied parking lots
Share of parking lots with violations	0.0054	0.0110
Parking lots with violations, t/s	114	235
Number of paid parking lots	120	11
Number of parking lots between the Boulevard Ring and the Garden Ring	41	0
Number of parking lots in SRN	244	110

with violations does not entail a significant increase in the parking lot occupancy.

A comparison of average values was used to assess the relation between the indicator of the parking lot occupancy in the district and other factors; the statistical significance of differences in the mean values was assessed using the t-criterion. The differences were found between the average values for the following indicators at the significance level of less than 5% (Table 4).

Thus, in the districts with over-occupied parking lots, the number and the share of parking lots with violations are averagely more than 2 times higher than with not over-occupied, and the number of paid parking lots is over 10 times less. It is noteworthy that there are no over-occupied parking lots between the Boulevard and the Garden Ring. The number of parking lots in SRN in the districts with over-occupied parking lots is averagely more than 2 times less than in the not over-occupied districts.

Modeling of parking occupancy

Model 1-Multiple Linear Regression. Multiple regression analysis was used to develop a model of parking occupancy dependence (y) on other indicators. As a result of the preliminary study, the nonlinear nature of the relation of some variables with the parking lot occupancy was revealed, which was taken into account at model development. Initially, the following values were included as explanatory variables:

x_1 is the number of cars registered in the district, thousand cars;

x_2 is the parking, thousand spaces;

x_3 is storage, thousand spaces;

x_5 is the share of parking lots with violations;

x_8 is the number of paid parking lots;

x_9 is the number of parking lots between the Boulevard and the Garden Rings;

x_{10} is the number of parking lots between the Garden ring and TTR;

x_{11} is the number of parking lots between TTR and MRR;

x_{12} is the number of parking lots outside MRR;

x_{13} is the number of parking lots in SRN.

A model including only significant variables at a significance level of less than 5% (2) was obtained after the regression analysis.

This model explains only 40% of the input data variance

$$\hat{y} = 78.261 + 1446.659x_5 - 27997.961x_5^2 + 0.051x_8 + 0.014x_{11} + 0.014x_{12} - 0.088x_{13} + 1.03 \cdot 10^{-5}x_{13}^2 \qquad (2)$$

(based on the determination factor value), since the parking occupancy (y) dependence is affected by the parameters other than those included in the model. However, the model may be interpreted:

- Increment of the parking occupancy caused by change in the share of parking lots with violations at unchanged other parameters of the model is not linear and is shown in Figure 3, the extremum falls on 2.6% share of parking lots with violations, apparently before 2.6% the direct impact of parking lots with violations on the parking lot occupancy occurs due to the lack of parking spaces, and from 2.6% the opposite effect may occur due to the high cost of parking, which forces the drivers to look for the ways of avoiding payment;

- Increase in the number of paid parking lots x_8 by 100 units is caused by an average 5% increase in parking occupancy, which might be explained by the usual location of paid parking lots in places with the greatest demand for parking;

- Increase in the number of parking lots between TTR and MRR (x_{11}) and outside MRR (x_{12}) by 100 will lead to an average increase in the parking occupancy by 1.4% for each location, apparently, this is caused by the fact that bedroom communities are concentrated in these locations;

- Increment of parking occupancy caused by the change in the number of parking lots in SRN (x_{13}) with unchanged other parameters of the model is nonlinear and shown in Figure 4, the inverse relation may be explained by the fact that parking lots in SRN are more convenient for parking, visible and accessible for the drivers than the parking lots outside SRN.
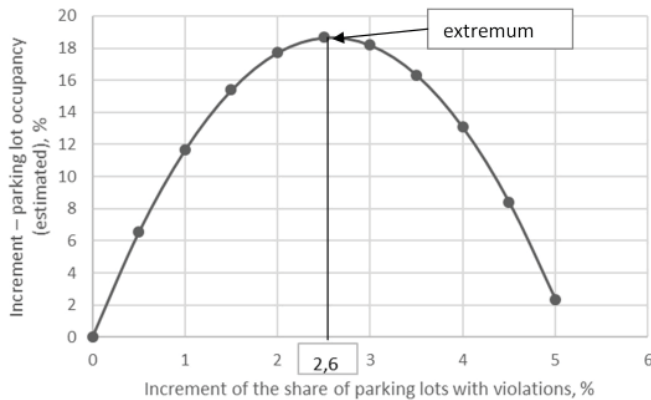
Figure 3. Dependence of the increment of the estimated parking occupancy on the increment of the share of parking lots with violations, %
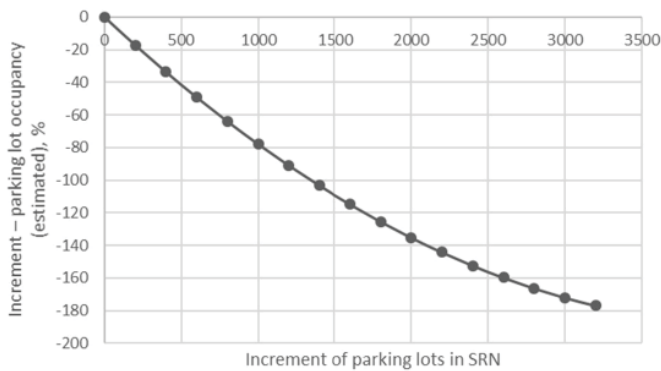

Figure 4: Dependence of the increment of the estimated parking occupancy on the increment of parking lots in SRN

Model 2 - Logit Regression. The multiple logit regression allowing to develop a model for calculating the probability of a certain event (p) based on a set of factors (z) affecting it was applied to obtain a model with a higher quality (having better predictive abilities than the 1st model). The probability is described as an S-curve.

The indicator of parking occupancy in the district (k) calculated by formula 1 was used as a dependent variable.

Initially, the following values were included as explanatory variables:

x_1 is the number of cars registered in the district (B category), thousand cars;

x_2 is the parking, thousand spaces;

x_3 is storage, thousand spaces;

x_5 is the share of parking lots with violations;

x_8 is the number of paid parking lots;

x_9 is the number of parking lots between the Boulevard and the Garden Rings;

x_{10} is the number of parking lots between the Garden ring and TTR;

x_{11} is the number of parking lots between TTR and MRR;

x_{12} is the number of parking lots outside MRR;

x_{13} is the number of parking lots in SRN.

A model including only significant variables at a significance level of less than 5% (3) was obtained after the regression logit analysis.

$$p = \frac{e^z}{1 + e^z} \qquad (3)$$

Where p is the probability of parking over-occupancy

$$z = -0.751 + 334.874x_5 - 6541.646x_5^2 + 4.33 \cdot 10^{-5}x_5^2 + 6.57 \cdot 10^{-6}x_{11}^2 - -7.96 \cdot 10^{-5}x_{13}^2$$

This model is nonlinear for all explanatory variables and explains only 52% of the initial data variance (based on Nigelkerk R^2 value), the proportion of the correctly predicted values of the indicator of parking occupancy in the district (k): "Not over-occupied parking lots" - 80% and "Over-occupied parking lots" - 80%, which characterizes the satisfactory quality of the model.

Interpretation of the model is as follows:

- Increment of the indicator z caused by change in the share of parking lots with violations at unchanged other parameters of the model is not linear and, similarly as in model 1, the extremum falls on 2.6% share of parking lots with violations (Figure 5); apparently before 2.6% the direct impact of parking lots with violations on the probability of parking over-occupancy (p) may be explained by the lack of parking spaces, and from 2.6% the opposite effect may occur due to the high cost of parking, which forces the drivers to look for the ways of avoiding payment;

- Increase in the number of paid parking lots x_8 by 100 units is caused by an average increase z by 0.43 and increases the probability of parking over-occupancy (p), which might be explained by the usual location of paid parking lots in places with the greatest demand for parking;

- Increase in the number of parking lots between TTR and MRR (x_{11}) by 100 will result in an average increase of z by 0.06 and increases the probability of parking over-occupancy (p), apparently it is caused by the high concentration of bedroom communities in this location;

- Increase in the number of parking lots in SRN (x_{13}) by 100 will lead to an average decrease of z by 0.8 and reduces the probability of parking over-occupancy (p), the inverse relation may be explained by the fact that parking lots in SRN are more convenient for parking, visible and accessible for drivers than the parking lots outside SRN.

$$z = -0.751 + 334.874x_5 - 6541.646$$

It is easy to notice the similar results of interpretation of the 1st and 2nd models, which confirms the validity of the conclusions:

- Multidirectional influence of the share of cars parked with violations on the parking lot occupancy;

- Close direct connection between the number of paid parking lots and the parking lot occupancy;
- Direct impact of the number of parking lots outside TTR on the parking lot occupancy;
- Reverse influence of the number of parking lots in SRN on the parking lot occupancy;
- Presence of significant additional factors affecting the parking lot occupancy, not included in the model.

Model 3. Stepwise discriminant analysis was used as an alternative to the logit regression model, to develop a predictive model to determine the category a district will belong to, based on the average parking lot occupancy (k) – irrespective of the fact, whether the parking lots are over-occupied or not.

The indicator of parking occupancy in the district (k) calculated by formula 1 was used as a dependent variable.

Initially, the following values were included as explanatory variables:

x_1 is the number of cars registered in the district, thousand cars;

x_2 is the parking, thousand spaces;

x_3 is storage, thousand spaces;

x_5 is the share of parking lots with violations;

x_8 is the number of paid parking lots;

x_9 is the number of parking lots between the Boulevard and the Garden Rings;

x_{10} is the number of parking lots between the Garden ring and TTR;

x_{11} is the number of parking lots between TTR and MRR;

x_{12} is the number of parking lots outside MRR;

x_{13} is the number of parking lots in SRN.

After carrying out the discriminant analysis, the model consisting of 2 discriminant Fisher functions (for 2 categories of parking occupancy degree) was obtained: not over-occupied, over-occupied), which includes only significant variables at a significance level of less than 5% (4, 5).
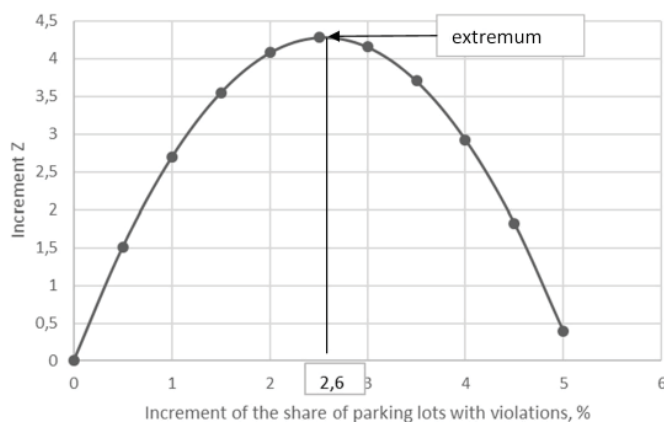

Figure 5: Dependence of the increment of logit model indicator z on the increment of the share of parking lots with violations

$$F_{not\ over-occupied} = -2.356 + 92.598x_5 - 2658.428x_5^2 + 0.017x_{13} - 4.844 \cdot 10^{-6}x_{13}^2 \tag{4}$$

$$F_{over-occupied} = -2.266 + 423.669x_5 - 9437.800x_5^2 + 0.005x_{13} - 1.588 \cdot 10^{-6}x_{13}^2 \tag{5}$$

This model is nonlinear with all explanatory variables. The measure of stress, Wilks' lambda, is 0.67; the canonical correlation factor is 0.58. The share of correctly predicted indicator values of parking occupancy in the district (k): "Not over-occupied parking lots" - 75% and "Over-occupied parking lots" - 75%, which characterizes the satisfactory quality of the model. These indicators are 5% lower than the similar values of the logit model, but the model obtained as a result of the discriminant analysis includes 2, rather than 4 explanatory variables.

The practical application of the model is based on substitution of values of explanatory variables in Fisher's discriminant functions, and depending on the fact the value of which function is greater, the district falls into the category by the degree of parking lot occupancy.

Model 4. The fourth model is developed with the help of "classification trees" and allows to characterize the districts on the basis of a set of indicators by the average degree of parking lot occupancy (k): over-occupied or not.

The indicator of parking occupancy in the district (k), calculated by formula 1, was used as a target (dependent) variable.

Initially, the following values were included as explanatory variables:

x_1 is the number of cars registered in the district (B category), thousand cars;

x_2 is the parking, thousand spaces;

x_3 is storage, thousand spaces;

x_5 is the share of parking lots with violations;

x_8 is the number of paid parking lots;

x_9 is the number of parking lots between the Boulevard and the Garden Rings;

x_{10} is the number of parking lots between the Garden ring and TTR;

x_{11} is the number of parking lots between TTR and MRR;

x_{12} is the number of parking lots outside MRR;

x_{13} is the number of parking lots in SRN.

The CHAID method based on Pearson's Chi-squared test and allowing to identify explanatory variables that have the greatest impact on the target variable change was used to construct the tree; Bonferroni adjustment was not included, the number of observations in terminal nodes was set to at least 10, in parent nodes to 20.

The classification results are presented in Tables 5 and 6.

Table 5: "Classification Tree" Gains of the districts by the degree of parking lot occupancy. target category: not over-occupied parking lots

Node No.	Node		Gain		Response rate, %	Index, %
	Number of observations	Share of observations, %	Number of observations	Share of observations, %		
9	17	13.7	16	27.1	94.1	197.8
8	26	21.0	17	28.8	65.4	137.4
3	25	20.2	14	23.7	56.0	117.7
7	19	15.3	8	13.6	42.1	88.5
4	19	15.3	4	6.8	21.1	44.2
5	18	14.5	0	0.0	0.0	0.0

In the gain tables, the nodes of the "classification tree" are arranged in descending order of the observation share corresponding to the target category (response). The first column contains the node number in the "classification tree", the second and the third column contain the number and the share of observations in the node (in our case, Moscow districts), the next two columns contain the number and the share of observations in the node corresponding to the target audience (gain) across all nodes. Response share is the share of observations that correspond to the target audience in the node. Index is the ratio of the response share for the node and for the entire sample (the latter is represented in the root (zero) node).

Thus, the nodes with an index significantly greater than 100% are of interest for each target category.

Let us consider the most significant nodes for each target category.

Target Category: Not over-occupied parking lots. The following nodes are of interest:

No. 9 includes 17 districts (13.7% of the total number of districts) of which 16 parking lots are not over-occupied (94.1%), which is 1.98 times higher than the average in the sample; characteristics: Parking lots in SRN – over 202; the share of parking lots with violations – no more than 1.18%;

No. 8 includes 26 districts (21% of the total number of districts) of which 17 parking lots are not over-occupied (65.4%), which is 1.37 times higher than the average in the sample; characteristics: Parking lots in SRN - from 130 to 202, the share of parking lots with violations – more than 1.18%.

Target Category: Over-occupied parking lots. The following nodes are of interest:

No. 5 includes 18 districts (14.5% of the total number of districts) of which all are over-occupied (100%), which is 1.91 times higher than the sample average; characteristics: Parking lots in SRN – less than 130, the share of parking lots with violations – more than 1.18%;

No. 4 includes 19 districts (15.3% of the total number of districts) of which 15 parking lots are over-occupied (23.1%), which is 1.51 times higher than the sample average; characteristics: Parking lots in SRN – less than 130, the share of parking lots with violations – from 0.327% to 1.18%.

DISCUSSION AND CONCLUSIONS

First, it is necessary to explain the selection of the modeling methodology. The literature review demonstrates that there are two approaches to modeling based on the queueing theory (QT) and the multivariate statistical methods. Having analyzed the literature, we believe that QT with consideration of the methodology, is applicable to an individual parking lot, as it was described in the literature [11, 12]. In addition, with certain assumptions, QT may be used for clusters of parking lots or local urban

Table 6: "Classification Tree" gains of the districts by the degree of parking lot occupancy target category: over-occupied parking lots

Node No.	Node		Gain		Response rate, %	Index, %
	Number of observations	Share of observations, %	Number of observations	Share of observations, %		
5	18	14.5	18	27.7	100.0	190.8
4	19	15.3	15	23.1	78.9	150.6
7	19	15.3	11	16.9	57.9	110.4
3	25	20.2	11	16.9	44.0	83.9
8	26	21.0	9	13.8	34.6	66.0
9	17	13.7	1	1.5	5.9	11.2

spaces, as in the work [32], which are considered as a multi-channel queuing system in the event when drivers in the process of searching for a parking space consider the parking lots included in the cluster as acceptable parking places.

First, it is necessary to explain the selection of the modeling methodology. The literature review demonstrates that there are two approaches to modeling based on the queueing theory (QT) and the multivariate statistical methods. Having analyzed the literature, we believe that QT with consideration of the methodology is applicable to an individual parking lot, as it was described in the literature [11, 12]. In addition, with certain assumptions, QT may be used only for clusters of individual parking lots or local urban spaces, as seen in the work Yang et al. [32]. This assumptions dictating necessity of adapting of QT to multi-channel queuing systems in the event that drivers in the process of searching for a parking space consider the parking lots included in the cluster as acceptable places to park. The QT can be used (as it is described in the literature and from our observations) for individual parking lots only and for urban spaces. With some assumptions QT can be transferred to the multi-channel queuing systems for the individual parking lots, when drivers in the process of searching for a parking space consider the parking lots included in the cluster as acceptable places to park.

In cases where the whole system of urban parking lots is considered, we believe it is more correct to use the methodology of statistical methods of multivariate modeling, as proposed in the literature [5]. This approach makes it possible to develop models based on sample data that describe the influence of a set of factors on the dependent variable or on the probability of a certain event. The multivariate statistical methods allow to assess the adequacy of models and include statistically significant factors in the model. Such tasks were fundamental for this study.

In cases where the whole system of urban parking lots is considered, we believe it is better to use the methodology of multivariate modeling, as proposed in the literature [5]. This approach enables developing the models based on sample data that describes the influence of a set of factors on the dependent variable or on the probability of a certain event. The multivariate statistical methods allow assessing the adequacy of models and include statistically significant factors in the model.

In this paper, we do not describe the procedure for collecting initial data based on the observation of parking lots in Moscow. Various methods and models were used in the processing of the study findings.

In this paper, we do not describe the procedure for collecting initial data based on the observation of parking lots in Moscow. Various methods and models were used in the processing of the study findings.

The dependence of parking occupancy indicator on the predictors was verified to confirm hypothesis 1. The cause-and-effect relations between the explanatory indicators of parking space and parking occupancy were assessed using the correlation analysis and the scatter patterns.

The dependence of parking occupancy indicators on the predictors was verified to confirm hypothesis 1. The cause-and-effect relations between the explanatory indicators of parking spaces and parking occupancy were assessed using the correlation analysis and the scatter patterns. The correlation analysis revealed the presence of a poor statistical linear relation between parking occupancy and

parking lots with violations (the pair correlation factor was $r=0.37$). There was no statistically significant correlation with any other indicators found. A hypothesis related to the non-linear nature of the relation between the indicators and parking occupancy occurred. A repeated correlation analysis was performed and scatter patterns were created to verify this hypothesis.

The correlation analysis revealed the presence of a poor statistical linear relation between the parking occupancy, parking lots with violations and the share of parking lots with violations (pair correlation factors were r=0.37 and r=0.42, respectively). There was no statistically significant correlation with any other indicators found. A hypothesis related to the nonlinear nature of the relation between the indicators and the parking occupancy occurred. A repeated correlation analysis was performed and scatter patterns were created to verify this hypothesis.

The repeated correlation analysis was carried out with squared values of variables, based on the assumption of the parabolic nature of the influence of indicators on the parking lot occupancy. The hypothesis was confirmed, the dependence of the indicator of parking occupancy on predictors was nonlinear and there was a statistically significant relation between the parking occupancy and the parameters:

- Parking, thousand spaces ($r=0.47$);
- Storage, thousand spaces ($r=0.48$);
- Parking lots with violations, t/s ($r=0.45$);
- Number of parking lots between TTR and MRR (r=0.30);
- Number of parking lots outside MRR ($r=0.32$).

In addition, during the initial analysis of the collected data, descriptive statistics were calculated for each district. In contrast to the study conducted by Ionita et al. [10], the cost of fuel was not considered, but the parameter of cars parked with violations was included. In addition, we used data on all parking lots, including the "unorganized/informal" ones. The analysis revealed an interesting pattern, according to which the number and proportion of parking violations were 2 times higher on average in the districts with over-occupied parking lots versus the districts where parking occupancy was not so high, and the number of paid parking lots was over 10

Table 7: Comparative evaluation of models for determining the demand for parking spaces

No.	Model name	Predicted value	Explanatory indicators	Indicators of model quality / efficiency	Model advantages	Model disadvantages
1	Multiple linear regression	Degree of parking lot occupancy in the district, %	1. Share of parking lots with violations 2. Number of paid parking lots 3. Number of parking lots between TTR and MRR 4. Number of parking lots outside MRR 5. Number of parking lots in SRN	Explains 40% of the variance in the initial data	Able to predict the quantitative indicator of parking occupancy	Low quality indicator of the model
2	Logit regression	Probability of parking over-occupancy in the district	1. Share of parking lots with violations 2. Number of paid parking lots 3. Number of parking lots between TTR and MRR 4. Number of parking lots in SRN	Explains 52% of the variance of the initial data, the share of correctly predicted values of the parking occupancy indicator in the district (80%)	The highest quality indicator among the models considered	The probability of parking over-occupancy is considered rather than, the occupancy degree
3	Discriminant analysis	The fact of parking over-occupancy	1. Share of parking lots with violations 2. Number of parking lots in SRN	Explains 33% of the variance of the initial data, the share of correctly predicted values of the parking occupancy indicator in the district (75%)	The simplest model among those considered	Low quality indicator of the model
4	Classification trees 1	The fact of parking over-occupancy	1. Share of parking lots with violations 2. Number of parking lots in SRN	The categories of districts with deviations from the average parking over-occupancy indicator up to 91-98% are revealed.	Enabling to develop rules for determining districts with similar indicators of parking occupancy	The probability of parking over-occupancy is considered rather than, the occupancy degree
5	Classification trees 2	Degree of parking lot occupancy in the district, %	1. Share of parking lots with violations 2. Number of parking spaces 3. Number of registered passenger cars (B category) 4. Number of parking lots in SRN	The categories of districts with deviations from the average indicator of parking occupancy up to 14-20% are revealed.	Enabling to develop rules for determining districts with similar indicators of parking occupancy	Low efficiency indicators of the model

times less. In addition, according to the diagram, the increase in the proportion of parking spaces with violations from 0 to 0.2% entails a sharp increase in parking lot occupancy (up to 90%), while a further increase in the proportion of parking lots with violations does not entail a significant increase in the parking occupancy.

Four approaches using approaches of multivariate statistics were analyzed to verify the hypothesis of the applicability of multivariate modeling to create a model of parking adequacy in the metropolis on the basis of the cause-and-effect relations between the explanatory indicators of parking spaces and parking occupancy (y). The results of the comparative evaluation of the model are presented in Table 7.

Based on the comparative evaluation of the models, the following conclusions can be drawn:

1. All models have the following explanatory variables that affect parking occupancy:

- The share of parking lots with violations; and
- The number of parking lots in SRN.

2. The most interesting models in terms of quality and efficiency are Nos. 2 and 4 (Table. 7), in which the predicted indicator presents the fact of parking lot over-occupancy in the district:

- Logit regression; and
- "Classification tree".

Thus, we have identified two models that give similar results, but the probability of parking over-occupancy will be calculated for each district in the logit regression; in case of "classification trees", it will be correlated with a certain group (node), which represents the total value of the probability of parking over-occupancy by districts. The obtained result confirms hypothesis 2.

The obtained models can be used in the "Smart City" systems [27, 28].

At present, several cities in the world have begun to monitor the parking area to provide the drivers with information about the free space availability. But as the researchers note "however, this approach is limited by the high costs of sensors that need to be installed throughout the city" [10]. However, such sensor system is not sufficiently represented in various districts and parking lots even in very developed cities. In addition, the sensor-based system does not allow for prediction of the parking space development in the city, and operates only with the available data. Our approach enables to integrate the models of parking occupancy dependence on other indicators into smart technologies and use it irrespectively of sensor availability.

We consider further study in terms of integrating the model with the data flow received from city video cameras, in order to take into account the factor of a motor vehicle temporary parking in the particular district based on the time it stayed in the city. We suggest focusing on the data specifying the vehicle license plate number and the time interval when it was registered by the cameras in different parts of the district or the city as a whole. We intend to be guided by the vehicle shield number and the time slot of its registering between the cameras installed in various parts of the district or the city in general.

ACKNOWLEDGEMENTS

REFERENCES

1. Administrative Territories and Districts of Moscow http://mosopen.ru/regions, accessed on 2019-08-19.

2. Almeida, J.C., Moreira, A., Moreira, L., Arbilla, G. (2008). Primary Emission Ratios Obtained from the Monitoring of Criteria Pollutants in Rebouças Tunnel, Rio De Janeiro, Brazil. Periódico Tchê Química, vol. 5, no. 9, 13-18.

3. Dominici, A., De Gentili, E., Capocchi, L., & Santucci, J. F. (2018). Smart Parking: Integration and Data Management by Modeling and Simulation Using Connected Objects According to the DEVS Formalism. Proceedings of the 2018 4th International Conference on Universal Village (UV), p.1-4.

4. DTRID Contract No. 190. (2018) The Study of the Parking Spaces for Moscow Citizens and Recommendations for Their Effective Use. Department for Transport and Road Infrastructure Development of the Moscow City Government (DTRID). Research project, Reg. No. Research, Development and Technological Work (RDTW) AAAAAH-A19-119062190046-5.

5. Dubrov A.M., Mhitarjan V.S., Troshin L.I. (2011) Multivariate Statistic Methods. Finance and Statistics, Moscow.

6. Fedorenko, R.V., Zaychikova, N.A., Abramov, D.V., Vlasova, O.I. (2016). Nash equilibrium design in the interaction model of entities in the customs service system. Mathematics Education, vol. 11, no. 7, p. 2732-2744.

7. Fickling D., He, E. (2018) The Future of Transport is the Future of Cities. Bloomberg. https://www.bloomberg.com/opinion/articles/2018-11-06/polluted-megacities-push-transport-s-future-toward-the-rails, accessed on 2019-03-10.

8. Fithra, H., Sirojuzilam, H., Saleh, S. M., Saputra, J. (2019). Road Network Connectivity and Freight Transportation for Supporting the Development of the Northern Zone of Aceh. Journal of Southwest Jiaotong University, vol. 54, no. 3.

9. Gong, Q., Midlam-Mohler, S., Serra, E., Marano, V., & Rizzoni, G. (2013). PEV Charging Control for a Parking Lot Based on Queuing Theory. 2013 American Control Conference, p. 1124-1129.

10. Ionita, A., Pomp, A., Cochez, M., Meisen, T., & Decker, S. (2018). Where to Park?: Predicting Free Parking Spots in Unmonitored City Areas. Proceedings of the 8th International Conference on Web Intelligence, Mining and Semantics, p. 22.

11. Khinchin, A.Ya. (1963). Studies on Math Queuing Theory (Vol. 236). Fizmatgiz, Moscow.

12. Kleinrock, L. (1975). Queuing Systems, Volume I: Theory.

13. Krpan, L., Maršanić, R., Milković, M. (2017). A Model of the Dimensioning of the Number of Service Places at Parking Lot Entrances by Using the Queuing Theory. Tehnički vjesnik, vol. 24, no. 1, 231-238.

14. Latif, M., Singla, S., Vadav, V. (2019). Modeling Off-Street Parking Based on User's Behavior Using SPSS Software, International Journal of Innovative Technology and Exploring Engineering, vol. 8, no. 8, 2360-2372.

15. Ma, J., Clausing, E., & Liu, Y. (2017). Smart On-Street Parking System to Predict Parking Occupancy and Provide a Routing Strategy Using Cloud-Based Analytics. SAE Technical Paper, no. 2017-01-0087.

16. Malhotra, N.K., Birks, D. F. (2007). Marketing Research: An Applied Orientation. Pearson Education, New York.

17. Mingardo, G., van Wee, B., Rye, T. (2015). Urban Parking Policy in Europe: A Conceptualization of Past and Possible Future Trends. Transportation Research Part A: Policy and Practice, vol. 74, 268-281.

18. Moscow Transport Rings. The History. https://moscowchronology.ru/transport_rings.html, accessed on 2019-09-19.

19. Moscow Transport System Development. http://transport.mos.ru/common/upload/public/Инфоцентр/транспортная%20система%20Москвы_rus.pdf, accessed on 2019-03-10..

20. Mouchili, M.N., Aljawarneh, S., Tchouati, W. (2018). Smart City Data Analysis. Proceedings of the First International Conference on Data Science, E-learning and Information Systems, p. 33.

21. Muramatsu, T., Oguchi, T. (2016). Proposal and Application of Parking Area Performance Measurement Methodology. Transportation Research Procedia, vol. 15, 628-639.

22. Pandey, D., Hanchate, S. (2018) Navigation-Based Intelligent Parking Management System Using Queuing theory and IOT. 2018 Second International Conference on Green Computing and Internet of Things (ICGCIoT).

23. Paula, M. (2006). Inimigo Invisivel: Metais Pesados E A Saude Humana. Periódico Tchê Química, vol. 3, no. 6, 37-44.

24. Pel, A.J., Chaniotakis, E. (2017). Stochastic User Equilibrium Traffic Assignment with Equilibrated Parking Search Routes. Transportation Research Part B: Methodological, vol. 101, 123-139.

25. Progressive Tariff for Parking Will Help Avoid Parking Lot Over-Occupancy in the Capital Downtown (2015). https://marino.mos.ru/presscenter/news/detail/2025650.html, accessed on 2019-08-19.

26. Richter, F., Di Martino, S., Mattfeld, D.C. (2014). Temporal and Spatial Clustering for a Parking Prediction Service. Proceedings of the 2014 IEEE 26th International Conference on Tools with Artificial Intelligence, p. 278-282.

27. Sidorchuk, R.R., Murtuzalieva, T.V., Skorobogatykh I.I., Musatov, B.I. (2018). Marketing Management System of Continuous Control for Indicators of Expectations of Consumers and Use of Big Data. Annals of Marketing MBA, vol. 2.

28. Sidorchuk, R., Musatova, Z., Mkhitaryan, S., Nevostruev, P., Komleva, N. (2016). Smart Technologies in Public Transport and Their Perception by the Youth Audience. Indian Journal of Science and Technology, vol. 9, no. 42, 10-17485.

29. Sobolev, A.N. (2008) The Classification and Identification of Economic Models and Types of Transportations with Associated Transport Services: Author's Abstract ... by the Candidate of Economic Science: 08.00.05 Far Eastern State Transport University]. Khabarovsk.

30. Thomas, D., Kovoor, B.C. (2018). A Genetic Algorithm Approach to Autonomous Smart Vehicle Parking System. Procedia Computer Science, vol. 125, 68-76.

31. Thong, P.H., Soh, A.C., Jaafar, H., Ishak, A.J. (2013). Real-Time Monitoring System for Parking Space Management Services. Proceedings of the 2013 IEEE Conference on Systems, Process & Control (ICSPC).

32. Yang, Z., Gao, Ch., Chen, D. (2016). Optimization of Parking Supply in Central Business District Based on the Relationship Between Car Travelling and Car Parking. Systems Engineering Theory and Practice, vol. 8, no. 19, 2091-2010.

33. Zhang, J., Zheng, W. (2016). Research on Stochastic Behavior of Traffic flow. Journal of Southwest Jiaotong University, vol. 51, no. 3.

34. Zhang, M.J., Zhang, Z.B., Du, X.L. (2017). Research on Number of Car Elevators for Cylindrical Underground Garage Based on Queuing Theory. China J. Highw. Transp, vol. 30, pp. 133-139.

35. Zheng, Y., Rajasegarar, S., Leckie, C., Palaniswami, M. (2014, April). Smart Car Parking: Temporal Clustering and Anomaly Detection in Urban Car Parking. Proceedings of the 2014 IEEE Ninth International Conference on Intelligent Sensors, Sensor Networks and Information Processing (ISSNIP), p. 1-6.

36. Zhenyu, M. E. I., Xiang, Y., Jun, Chen, Jun, & Wei, W. (2010). Optimizing Model of Curb Parking Pricing Based on Parking Choice Behavior. Journal of Transportation Systems Engineering and Information Technology, vol. 10, no. 1, 99-104.