Scientific Journal of Silesian University of Technology. Series Transport
Zeszyty Naukowe Politechniki Śląskiej. Seria Transport

Volume 123 **2024**

p-ISSN: 0209-3324

e-ISSN: 2450-1549

DOI: https://doi.org/10.20858/sjsutst.2024.123.9

Journal homepage: http://sjsutst.polsl.pl

Article citation information:
Matyja, T., Stanik, Z., Włodkowski, K. A method of generating customer requests in a car rental simulation model. *Scientific Journal of Silesian University of Technology. Series Transport*. 2024, **123**, 191-208. ISSN: 0209-3324. DOI: https://doi.org/10.20858/sjsutst.2024.123.9.

Tomasz MATYJA[1], Zbigniew STANIK[2], Krzysztof WŁODKOWSKI[3]

A METHOD OF GENERATING CUSTOMER REQUESTS IN A CAR RENTAL SIMULATION MODEL

Summary. Optimization of business processes and policies in a car rental company is an ongoing topic. Especially that each of these companies operates in slightly different local conditions and external environments. Testing new optimization algorithms is best done with simulation methods. A comprehensive rental simulation model can be built, for example, using the SimEvents library of the Matlab/Simulink environment. The paper focuses on the problem of preparing a sequence of customer requests in the short-term car rental system, necessary to carry out simulations in the SimeEvents environment. It was assumed that these data may come from the real world, or they may also be artificial data. A method of generating artificial sequences of customer requests and the structure of input data necessary to carry out this process have been proposed. The use of machine learning to build models that transform real data in such a way that it can be randomized was also tested.

Keywords: short-term car rental, client's requests, machine learning, SimEvents

[1] Faculty of Transport and Aviation Engineering, The Silesian University of Technology, Krasińskiego 8 Street, 40-019 Katowice, Poland. Email: tomasz.matyja@polsl.pl. ORCID: https://orcid.org/0000-0001-6364-619X
[2] Faculty of Transport and Aviation Engineering, The Silesian University of Technology, Krasińskiego 8 Street, 40-019 Katowice, Poland. Email: zbigniew.stanik@polsl.pl. ORCID: https://orcid.org/0000-0003-1965-4090
[3] MM Cars Renta Sp. z o.o., Lotnisko 81 Street, 40-271 Katowice. Email: k.wlodkowski@mmservicelease.pl. ORCID: https://orcid.org/0009-0001-0063-3233

1. INTRODUCTION

Car rental companies can operate according to different business models [14]. Most often, a registered customer orders the service via the company's website or the application on the phone. The vehicle is released at one of the company's stations (offices), or, if there is such an option, the vehicle is delivered to the indicated address. The return of the vehicle can take place at the same station (two-way, round trip), at another station (one-way trip), or at the address indicated. In the case of car sharing, it is frequently possible to leave the vehicle anywhere (free-floating car sharing) [6].

In connection with the activities of car rental companies, there are many interesting logistical and decision-making problems. A significant issue is making decisions about accepting or rejecting vehicle bookings to ensure maximum use of the fleet and customer satisfaction. The most popular and frequently researched concern in the literature is the problem of optimal relocation of vehicles [14]. Vehicle relocation aims to balance the demand and supply of cars in a specific location. Relocation can be based on the operator, then it is carried out by the rental staff. Another way is relocation based on users, encouraged to do so by proposed discounts or clients, who are appropriately selected at the stage of the decision to accept a reservation. Another problem is maintaining the size of the vehicle fleet that is optimal for the current needs. Supplementing it by purchasing new cars from manufacturers (“fleeting”) or resale of cars after a certain period of use (“de-fleeting”) [8].

Simulation research allows for the search for new algorithms for the functioning of the vehicle rental system in order to increase the efficiency of processes [12, 13]. The problem of relocation is most often considered as a flow optimization issue. Among other things, a mathematical programming-oriented approach and a simple linear model based on integer flow variables [2], time expanded network (time-space network) [4, 5], Petri nets [3] are used. Models based on event simulations [9] are suitable for simulating the processes taking place in a car rental company.

In the future, the authors plan to develop a simulation model representing the functioning of car rental companies, using event-driven simulation technology. This model would enable testing of various sales policies, algorithms optimizing customer and vehicle flows, algorithms for making short-term decisions, e.g., in the field of vehicle relocation, as well as other processes occurring in this type of enterprise.

The article focuses on the issue of developing a rental customer request generator that could be implemented in a rental model built using the SimEvents library of the Matlab/Simulink environment [11]. Such a generator is one of the most important elements of the car rental modelling system. At the stage of designing the generator, it is necessary to formulate basic assumptions regarding the functioning of the rental company and to decide on the data structure describing a typical customer request.

Simulations always require input. The correctness of the results depends on their quality, and therefore, such data should be objective and reliable. Real-world data is the most desirable, but it is usually difficult to obtain. Real data in some cases may be biased. The disadvantage of real data is also that they are static (deterministic), i.e., they do not change. Statistical analysis of real data allows detecting trends and estimators of parameters necessary to generate artificial data, e.g., parameters of probability distributions. Based on the statistical analysis of real data, an artificial data generator can be built. Such data will be similar to real data, but at the same time, will not be identical to it. Artificial data can also be random. Data artificially generated based solely on arbitrary assumptions and estimates, without a thorough analysis of real data, may be unreliable.

In the work, an analysis of real-world data obtained from a company dealing with short and medium-term vehicle rental was carried out. On this basis, a scalable method for generating artificial data was proposed. In addition, the possibility of using machine learning to generate such data was tested. In accordance with the idea suggested in paper [1], the following were used: Gaussian mixture model to generate the times of customer notification, start of vehicle rental, vehicle return; a model of a binary classification tree to generate vehicle delivery and return stations.

2. CUSTOMER REQUEST GENERATOR

2.1. The general idea of a vehicle rental simulation model

It is assumed that the company deals in the short-term rental of passenger vehicles (RAC) for a period of 1 to 30 days. The minimum rental time is one day. This type of car rental is more interesting than medium-term rental (MTR) because there is a higher turnover of cars and the likelihood of relocation is increased. The presented model of the customer request generator can be easily scaled and adapted to the needs of simulating car rental by the hour in car sharing systems.

The company has N stationary stations (offices) where clients can rent and return a vehicle, located in a specific area. The customer can rent a vehicle at one of the rental points and then return it at any point (two-way or one-way trip mode is allowed). If necessary, it is also possible to simulate cases where the vehicle is delivered or picked up at any location specified by the customer (free-floating car sharing). Such a virtual station may have a number equal to zero. The location of all stations as well as locations indicated by the customer can be described using GPS coordinates.

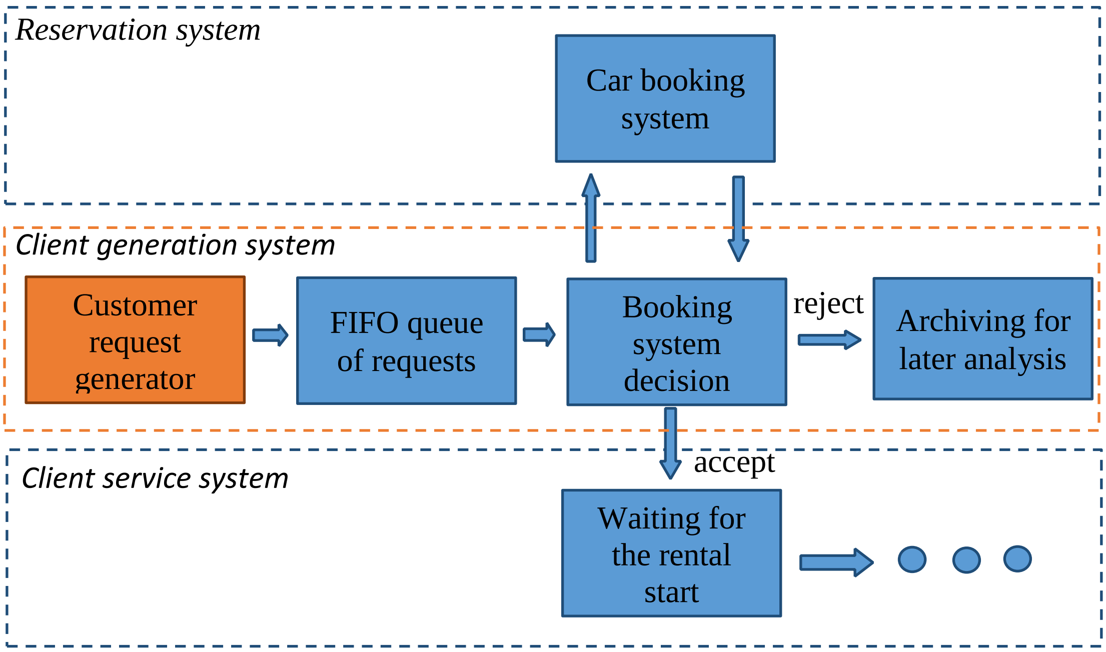


Fig. 1. Simplified schematic diagram of the rental model

One day was assumed as the basic unit of time in the model. A day does not have to be interpreted as the equivalent of 24 hours, it is rather a part of a day - a working day, corresponding to the opening hours of the rental stations. The occurrence of each event during the day is described by a fraction of a day. This allows timing the customer requests and other events that may occur during the simulation.

The schematic diagram of the car rental model is shown in Fig.1. The subject of interest is the subsystem that generates customer requests. An important role in the model will also be played by the booking subsystem and the customer service subsystem, which must also include other subsystems, e.g., a subsystem simulating the use of a car by the customer.

When booking a vehicle, the customer specifies the start and end date of the rental, the starting, and ending station and chooses the class of the vehicle. Optionally, it also provides location coordinates when picking up or returning a vehicle outside stationary stations.

Car rental companies offer their customers vehicles in various categories of size and equipment. They usually use SIPP codes (Standard Interline Passenger Procedure) so customers can easily compare different models of vehicles. SIPP or ACRISS (Association of Car Rental Industry Systems Standards) car classification is a 4-letter code, in which: the 1st character denotes the vehicle category (size, luxury factor); the 2nd character defines the vehicle chassis type; the 3rd character defines the transmission and drive; the 4th character defines the fuel and whether the car has air conditioning or not. From the customers' perspective, the most important thing is the size of the car, i.e., how many passengers and how much luggage can be accommodated. The European fleet is divided into the following categories: mini, economy, compact, intermediate, standard, full-size, premium, luxury. In each of them, an additional elite subcategory can be distinguished (e.g., compact elite). The term elite has been selected by ACRISS to identify a category of vehicle that is superior to another of equal body size. There is one more category: special. In total, there will be 17 categories of car sizes [15]. Companies can freely limit the range of vehicles available to customers, or extend it to a selection of other car parameters, including even specific brands and models [10]. For this reason, a variable parameter defining the number of available vehicle types should be included in the simulation.

2.1. Generator data structures

The SimEvents library has been equipped with the ability to dynamically generate events based on assumed probability distributions, for example, exponential distribution appropriate for customer requests. According to the authors, this method may not be sufficient if it has considered the volatility of demand for services over time and the existing trends. For this reason, it was decided to use the simplest solution consisting in preparing a sequence of customer's requests in advance, before starting the simulation. This will enable the simulation to incorporate both real and artificially generated data.

Based on the general idea of the car rental model described above, the minimum set of data describing the application of one customer in the form of a vector can be determined:

$$(\tau_r, \tau_s, \tau_e, S, E, C) \tag{1}$$

where: τ_r – request time, τ_s – rent start time, τ_e – rent end time, S – start rent office index, E – end rent office index, C – rent car type.

If the vehicle is to be delivered to the customer at the address provided, the data should be supplemented with information about the location of the pick-up point (when S=0) and/or return of the vehicle (when E=0). These can be, for example, GPS coordinates in the format $(\varphi_S, \lambda_S, \varphi_E, \lambda_E)$, which can be used to determine the distances to stationary rental points.

If historical data are available, there are no obstacles to determining a sequence of customer requests on their basis, which can be implemented in a model made in the SimEvents environment.

If the data is to be generated artificially, it was assumed that the most important factor is the rental start time, and that it must be generated first. T_s Because of this, it will be possible to take into account the phenomena of volatility and seasonality in the demand for vehicle rental. Assuming a single probability distribution for the entire simulated rental company operation period would obscure and flatten these effects.

Then the request time is generated. It was assumed that the interval between the request and the start of the rental will have an exponential distribution with a mean $\bar{I}_{rs}$. This will be a truncated distribution due to constraints on real values $I_{rs_{min}} \leq I_{rs} \leq I_{rs_{max}}$. The rental end time is generated similarly ($\bar{I}_{se}$ – mean of rental duration, $I_{se_{min}} \leq I_{se} \leq I_{se_{max}}$).

The selection of the rental starting point is made randomly based on an arbitrarily assumed probability distribution:

$$pdf_s(T_s) = [p_i(T_s)] \quad i = 1, \dots, N; \ \sum_{i=1}^{N} p_i = 1 \tag{2}$$

In the case of generating the end station, we have to deal with conditional probability. To describe the density of this probability, a square matrix was arbitrarily selected. The elements of this matrix determine the probability of choosing the end station number j provided that the starting station was:

$$pdf_e(T_e) = \left[p_{ij}(T_e)\right] \quad i,j = 1, \dots, N; \ p_{ij} = P\left(E_j | S_i\right) \quad \sum_{j=1}^{N} p_{ij} = 1 \tag{3}$$

It was assumed that the pdf_s vector and the pdf_e matrix may change during the simulation. Selections of the appropriate version of the vector and matrix are made by converting the simulation time into predetermined periods T_s and T_e. If the customer has the option of indicating any pick-up location or return of the vehicle location, the dimensions of the vector and the probability density matrix should be increased by 1.

Exactly on the same principle, a vector describing the distribution of the probability density of the vehicle class selection pdf_C should be adopted. Its length depends on the available classes, numbered from 1 to M.

Generating location GPS coordinates can be solved in different ways. The easiest way is to define the boundary ranges of coordinates in the area of operation of the car rental company and determine the GPS coordinates by two random numbers from a uniform distribution. Another possibility is to arbitrarily create a two-dimensional Gaussian mixture model with the components concentrated around the locations of stationary stations, and randomize locations using this distribution.

3. REAL-WORLD DATABASE ANALYSIS

Real-world data were obtained from a company that conducts both short-term rental (RAC) and medium-term rental (MTR), lasting from one month to two years. For this reason, these data are not fully representative of the simulation model being prepared. However, some regularities can be observed on their basis, which were then used to prepare the artificial data generator.

The actual data were properly processed and filtered. Datetime fields were converted to the time unit assumed in the simulation - one day. The data were sorted chronologically according to the start time of the reservation. In a few cases, records with missing information were rejected or corrected.

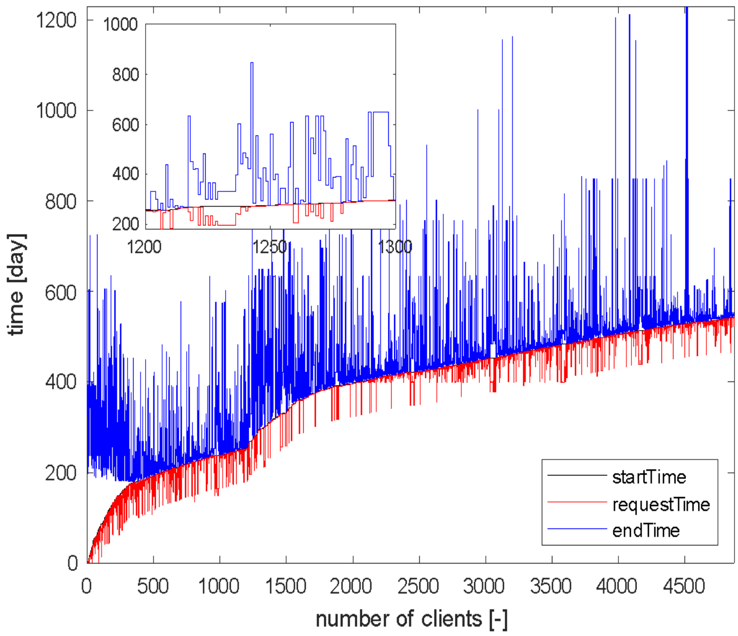


Fig. 2. Schedule of customer requests, start, and end of rental

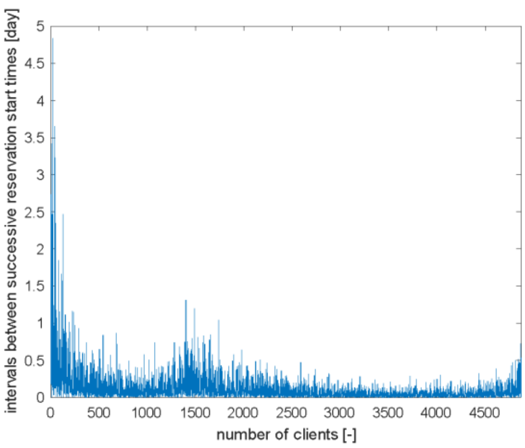


Fig. 3. Intervals between successive rental start times

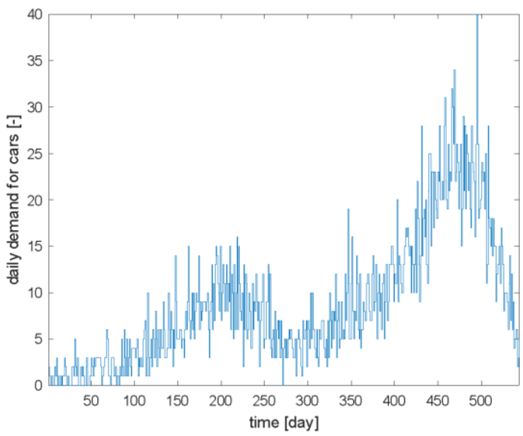


Fig. 4. Daily demand for vehicles based on customer requests

Fig. 2 shows the time distribution of customer requests, rental start and end times. According to the assumptions of the model, the basis is the booking start time line (black). Around this line are the booking request time (red line) and booking end time (blue line). The selected fragment of the graph is shown enlarged. It is visible that the company allows to book cars even several months in advance. In the initial phase, most of the notifications concerned medium-term rentals. Later, both short-term and medium-term rentals were serviced. Intervals between successive rental start events are shown in Fig. 3. At the beginning, they are even several days long, then these times shorten significantly below the value of 0.5 of the contractual working day. The daily demand for vehicles shows seasonality with a clear upward trend (Fig. 4). Refusals to accept vehicle reservations accounted for slightly more than 6% of all applications.

Interestingly, customers most often chose Monday as the vehicle release day and Friday as the return day (Fig. 5). For short-term rentals, the most likely rental duration was 5 days. The company rents a lot of vehicles on a medium-term basis (Fig. 6). The company provided several stations where the customer can rent or return a car.

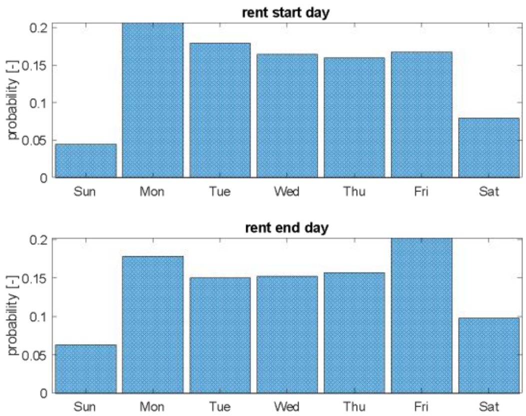


Fig. 5. Vehicle rental and return probabilities by day of the week

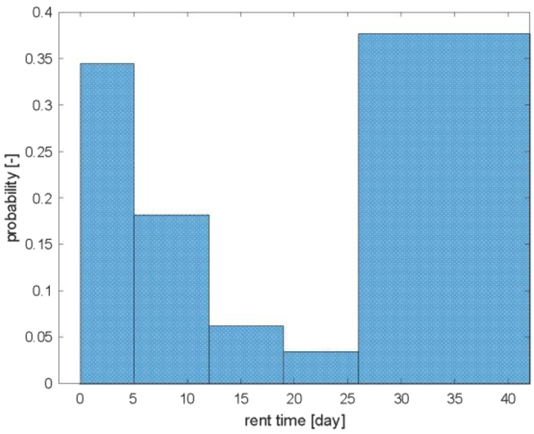


Fig. 6. The probabilities of the duration of car rent

Fig. 7 shows how the number of vehicles rented in individual company offices was distributed, and similarly the number of vehicles returned. The matrix of conditional probability distributions of returned cars is graphically presented in Fig. 8. Only a small percentage of customers dropped off the vehicle at a different office than they picked it up. For this reason, the probability values on the diagonal are dominant.

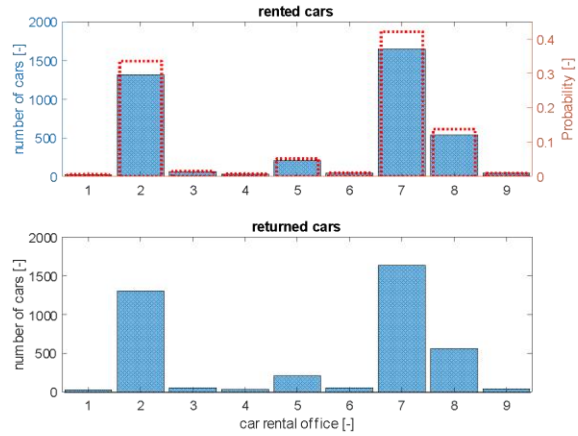


Fig. 7. Number of vehicles borrowed (probability distribution) and returned by rental office number

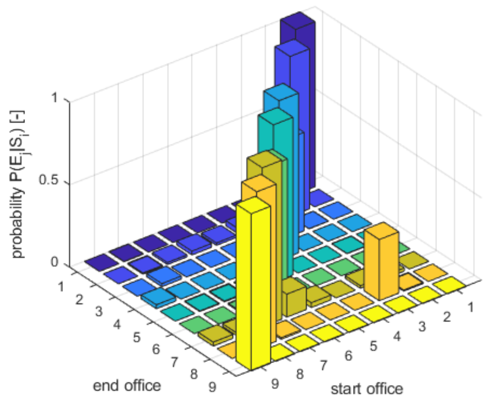


Fig. 8. Probability distributions of returning the car to a specific office provided it was rented from this or another office

The last analysed issue was the structure of the fleet of cars chosen by customers. The company uses the ACRISS classification. Based on the classification code, the histogram shown in Fig. 9 was created. Then, the percentage share in the rental of individual types of cars was calculated. The largest group are vehicles in the compact class (Fig. 10).

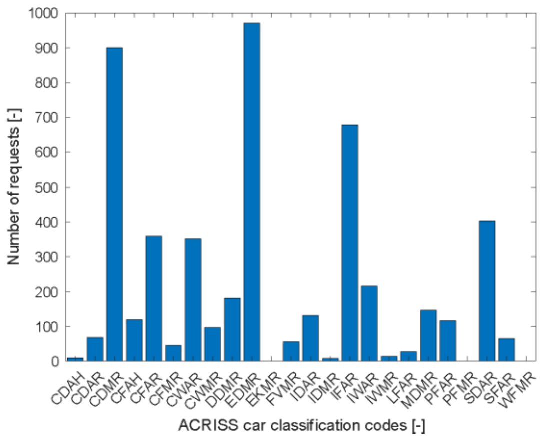


Fig. 9. Number of vehicles by classification code

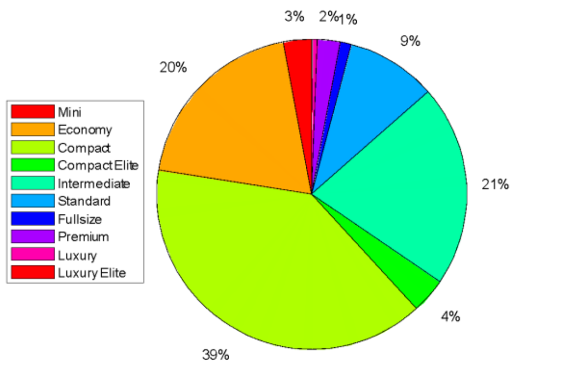


Fig. 10 Percentage share of individual types of vehicles

4. ARTIFICIAL DATA GENERATOR

Since the obtained real data turned out to be insufficient, an attempt was made to generate artificial data, assuming that the vehicles are rented for a short time. The starting day was January 1, 2023 (important due to the later calculation of the days of the week). The time horizon was set at two years (730 days). Efforts were made to reproduce the regularities noticed during the analysis of real data as best as possible.

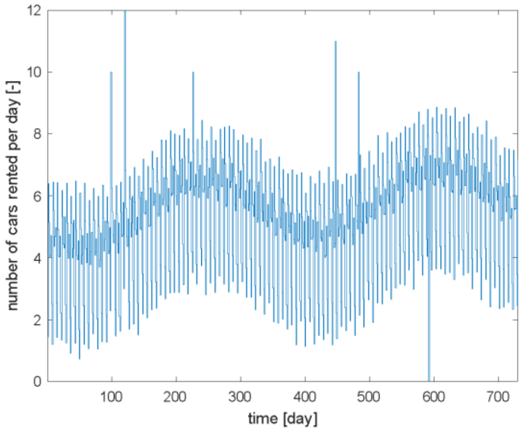


Fig. 11. Generated number of cars per day

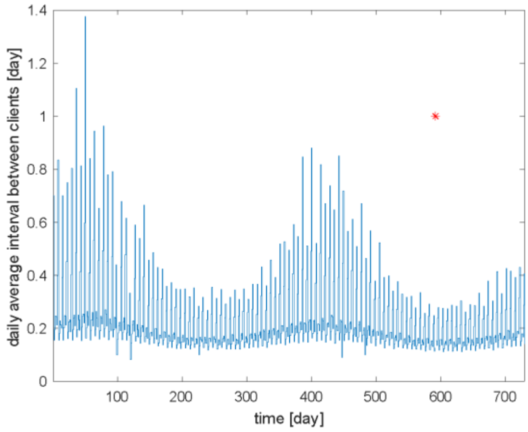


Fig. 12. Daily averaged interval between clients

As mentioned earlier, it is best to begin generating the start of the rental times. It can be difficult to directly determine the sequence that describes rental start times. Therefore, first, the number of vehicles to be rented each day was generated. The appropriate data series were obtained by adding together four time series: trend, weekend cycle, seasonality, and noise.

It was assumed that the trend will be modeled by a polynomial of the type $c_0 + c_1 t + c_2 t^2 + \cdots$. The coefficient c_0 determines the average number of vehicles. Fractional values are allowed. The weekend cycle is represented by a 7-element vector (Sunday is the first day), the values of which determine the increases or decreases in the number of vehicles in relation to the average value, depending on the day of the week. The seasonality series is obtained using harmonic functions, giving the amplitudes of the number of vehicles and the periods and phase shifts in days. The noise series is defined by random numbers between zero and one multiplied by an assumed constant noise level.

Finally, the data series obtained is overwritten, at selected points of the timeline, with arbitrarily adopted values - holiday impacts. If negative values occur in the series as a result of the addition operations, they are reset to zero.

The results of generating the number of vehicles for each day are shown in Fig. 11. The reciprocal of the number of vehicles allows determining the average daily interval between customers (Fig. 12). If the number of vehicles is zero, then the interval is infinite. This case is marked with a red asterisk in the figure.

In the next step of the method, for each day, based on the average daily interval between customers, customer requests are generated using exponential distribution. The first interval obtained in this way is randomly shortened - the event moves closer to the beginning of the day. Subsequent intervals are generated until the sum of all of them exceeds the time of one day. The last interval is discarded because it corresponds to an event that would occur on the following day. In this way, a sequence of intervals between successive customer notifications is created, which allows the generation of rental start events throughout the expected time of the simulation, i.e., up to the 730th day. Intervals generated in this way are shown in Fig. 13.

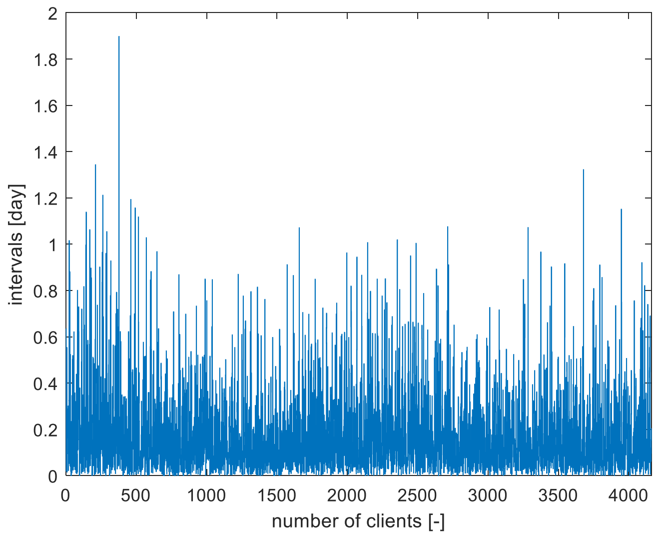


Fig. 13. Intervals between successive rental starts

The cumulative sum of the intervals determines the occurrence of successive rental start events on the time axis. By subtracting the randomly generated intervals between the start and request, the request time is obtained. Similarly, by adding the length of the rental time, the rental end time was calculated. It was assumed that the mean value of the intervals between the request and the start of the rental is $\bar{I}_{rs} = 7$ days, and the randomly generated interval must meet the constraint of $0.5 \leq I_{rs} \leq 14$ days. Values going beyond the accepted range are removed and the random procedure itself is repeated. Similarly, the average rental period $\bar{I}_{se} = 5$ days with the restriction $1 \leq I_{se} \leq 30$. The results of such operations are shown in Fig. 14.

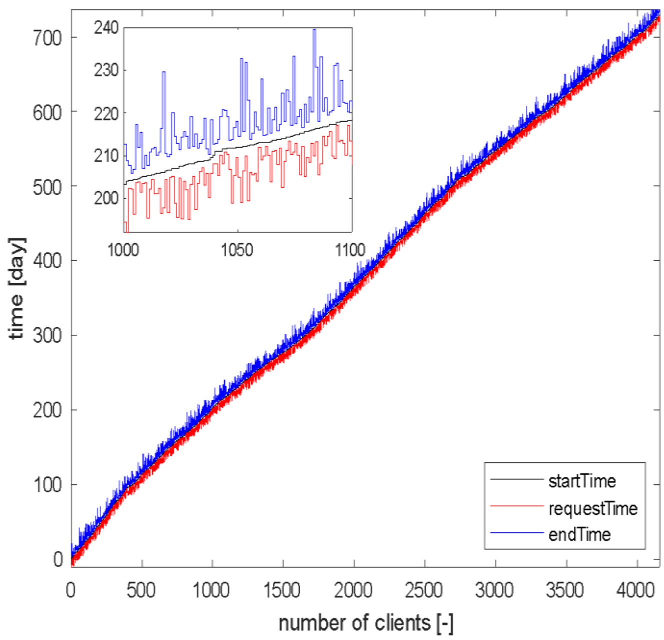


Fig. 14. The time of the events: request, start and end of the rental

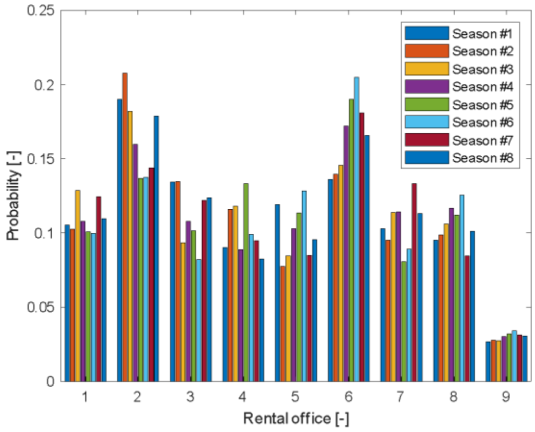


Fig. 15. Pdf for start station

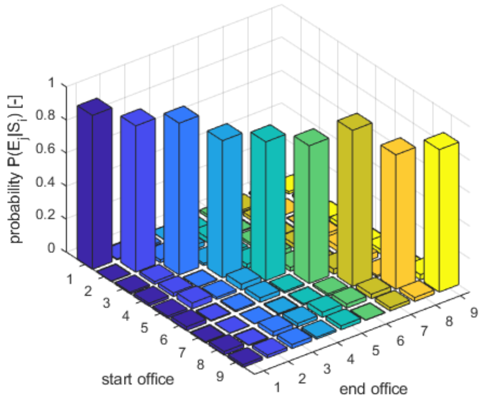


Fig. 16. Pdf matrix for end station

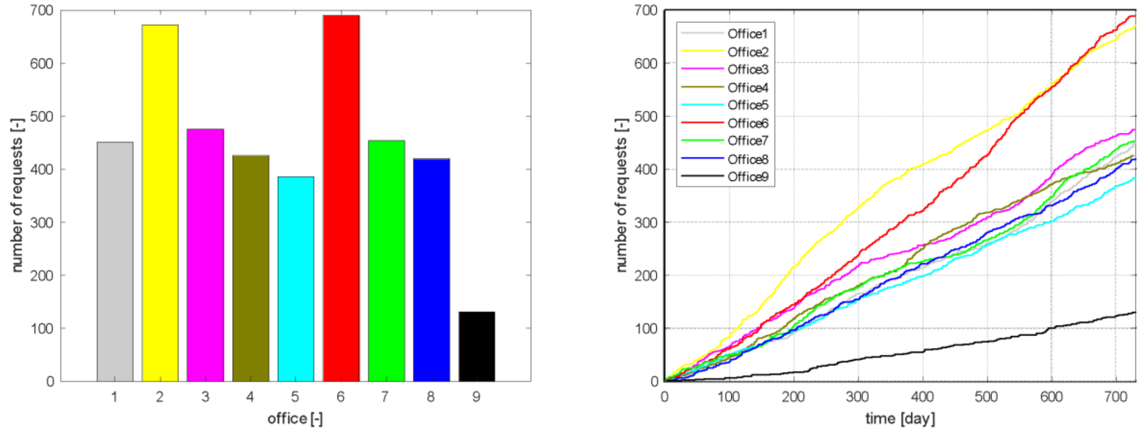


Fig. 17. The results of random selection the starting station;
on the left – cumulated; on the right – distribution over time

Generating further customer request data (the place of rental start and rental end) requires arbitrary determination of the density probability for the starting point and the matrix of conditional probabilities for the rental end point. In the case of the starting point, it was assumed that the probability values change every 90 days (Fig. 15). In the case of the rental end point, one density probability matrix graphically shown in Fig. 16 was used. It was assumed that a two-way rental was the most likely. The results of the starting point randomization are shown in Fig. 17. The results of the randomization of the rental termination points based on the probability density (Fig. 16) are shown in Fig. 18. The bars in Fig. 16 and Fig. 18 are not identical, but very similar.

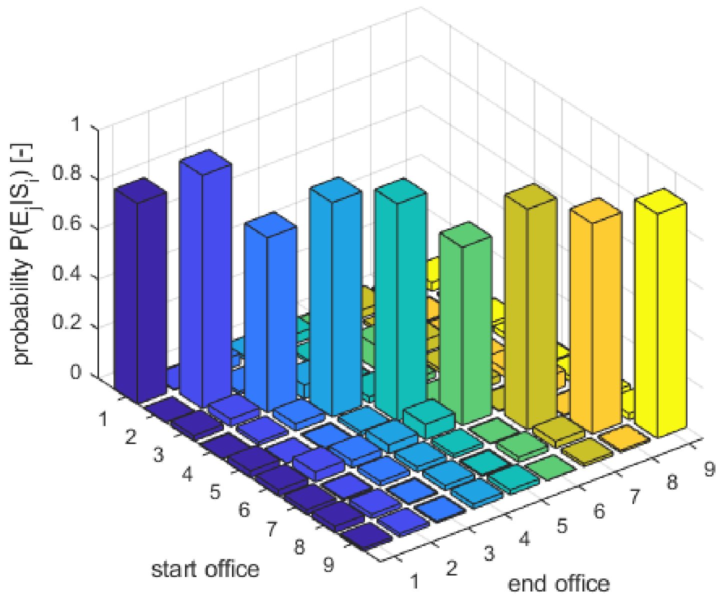


Fig. 18. The results of random selection the end station (transformed to probability density)

Generating the type of vehicle selected by the customer during the booking is done in the same way as above, using an arbitrarily selected probability distribution.

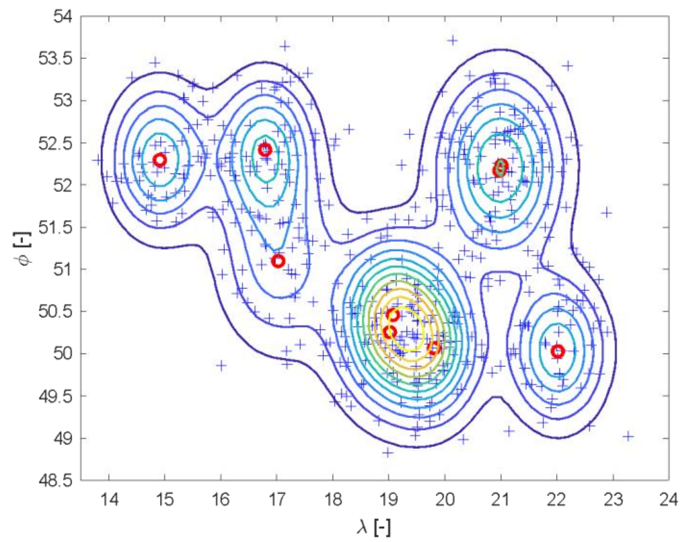


Fig. 19. Results of random selection of the locations outside stationary stations

In order to generate artificial locations indicated by customers, a nine-component bivariate Gaussian mixture distribution with equal mixing proportions was prepared. The mean values of the distributions were the GPS coordinates of stationary car rental stations. It was assumed that the radius served by the station with numbers 2,6,7,8 was 0.5 degrees of latitude (just over 50 km), while for the remaining stations it was only 0.3 degrees. This was taken into account by specifying the values of the elements of the covariance matrix, which was diagonal.

Fig. 19 shows stationary station placement (red circles), the probability distributions function of location selection, and randomly generated locations. This method of generating locations has an advantage over randomization from uniform distributions because it considers customer behaviour. Customers are aware that their order will be rejected if they choose a location too far from existing stationary stations.

5. GAUSSIAN MIXTURE MODEL

In the search for tools for prediction and possible forecasting of data for the car rental simulation model, mathematical models dedicated to time series and machine learning models were tested. Analyzing real data, e.g., daily demand for vehicles based on customer request (Fig. 4) or adequate artificial data (Fig. 11), it can be seen that these are time series having such features as trend and seasonality. Unfortunately, it was not possible to obtain satisfactory results using SARIMA (Seasonal Auto Regressive Integrated Moving Average) econometric models [7]. The residual errors of the tested models were too large for both real and artificial data.

Based on work [1], a Gaussian mixture model (GMM) was used to predict times (τ_r, τ_s, τ_e), trained with the use of artificial and real data. Various variants of input variables were tested, both in the form of event times and their intervals. The best results were obtained in the case of a model with three variables in the form of event times and 12 components of the mixture. However, it should be emphasized that the BIC (Bayes Information Criterion) and the AIC (Akaike Information Criterion) quality indices of the models with fewer components (e.g., 8) were only slightly worse than the model with 12 components.

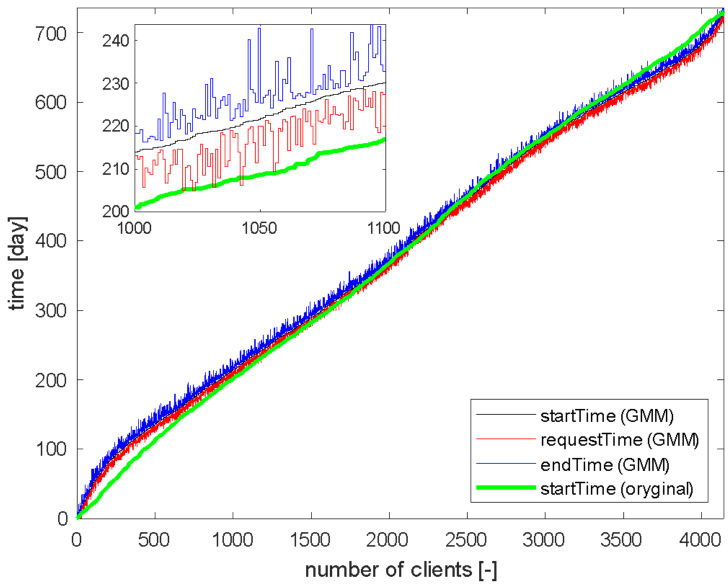


Fig. 20. Event times from GMM

On the basis of the trained GMM model with the use of artificial data, values of time events were then randomly generated. Not all values generated in this way met the imposed constraints $(0 < \tau_s \leq 730) \wedge (\tau_r \leq \tau_s - 0.5) \wedge (\tau_e \geq \tau_s + 1)$. Such invalid data had to be filtered out and the random choice repeated. Another disadvantage of this method of generation is that the data is not chronological and requires sorting. Another disadvantage is that the length of the generated data must match the length of the input data.

The generation results are shown in Fig. 20. For comparison, the original time course of the rental start times has also been plotted. It is apparent that the GMM data have slightly different forms than the original data. This may adversely affect the preservation of trend and seasonality features present in the original data. The consequences of selecting a shorter or longer data sequence than the original data are explained in Fig. 21. Then the number of events in the analysed period changes. At the same time, the GMM is quite stable even though the generated data sequences are slightly different (Fig. 22). This feature can be used to test the sensitivity of the simulation model to input data.

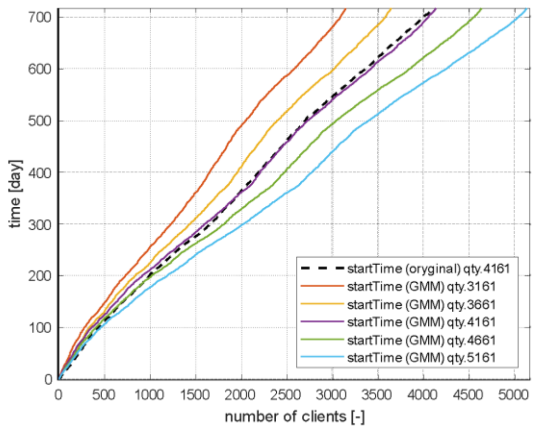


Fig. 21. Influence of the data series length

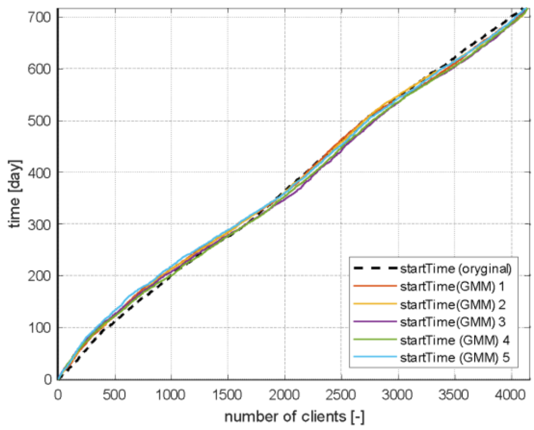


Fig. 22. Several series of data randomly selected from the GMM

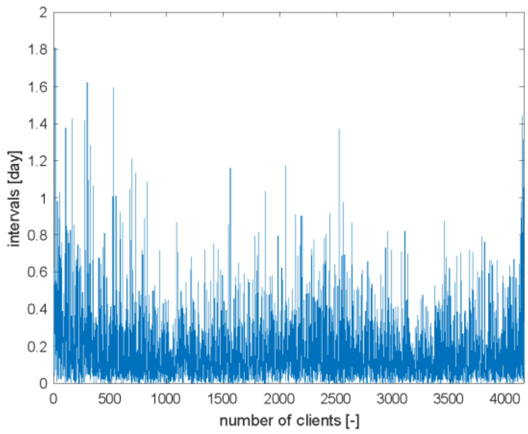


Fig. 23. Intervals between customers

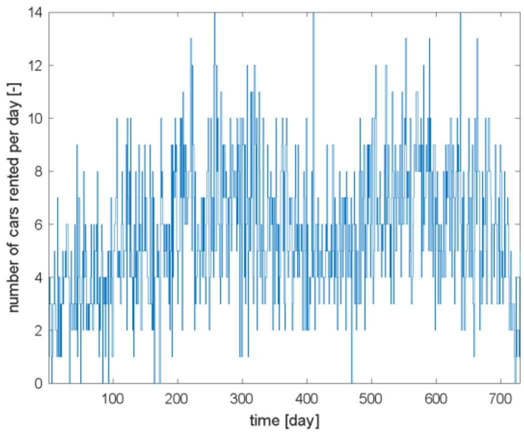


Fig. 24. Vehicles per day

Based on the start times generated from the GMM, graphs of intervals between customers (Fig. 23) and the number of vehicles each day (Fig. 24) were reconstructed. To evaluate the GMM data, it would be necessary to compare these plots with their counterparts for artificial data (Fig. 13) and (Fig. 11). There is a similarity in the values and time course of the functions, but they are generally different data.

The procedure was repeated for real-world data. The results are presented in Fig. 25 and Fig. 26 (comparison of event times) and Fig. 27 (daily demand for vehicles). Based on the comparison of Fig. 25 with Fig. 2, it can be concluded that the structure of booking times (more - early bookings) and rental termination times (more - longer rental times) has changed significantly. In real-world data, there were both very long and very short periods at the same time. The GMM model clearly averaged them out.

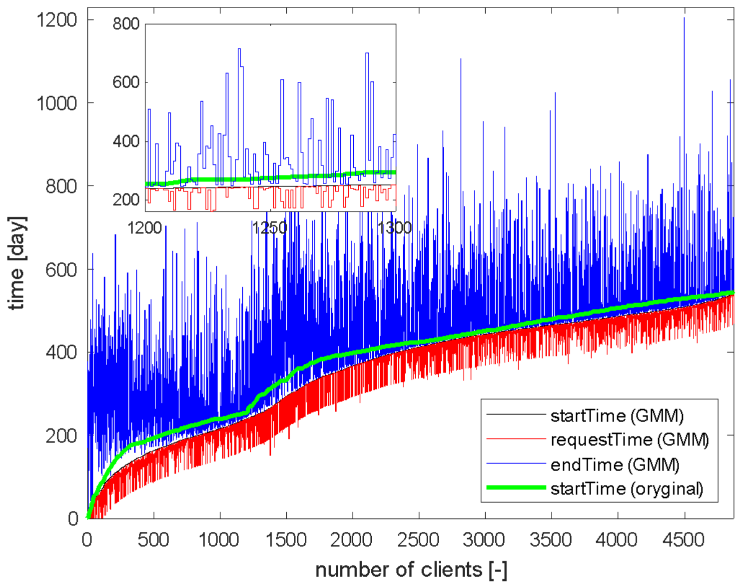


Fig. 25. GMM trained with real-world data

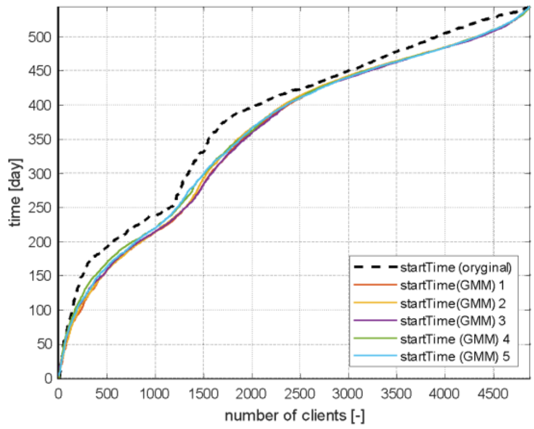


Fig. 26. Time course of several series of random data

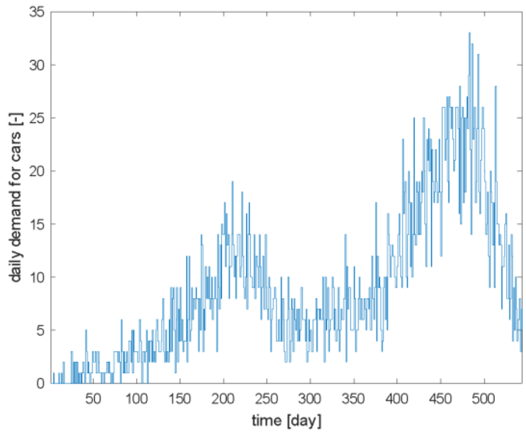
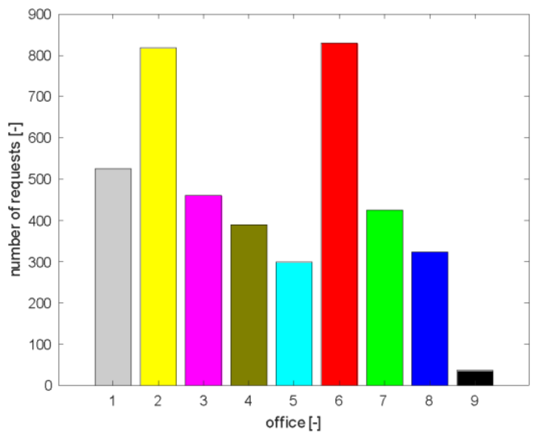


Fig. 27. Daily demand for vehicles

6. CLASSIFICATION TREES

Based on artificially generated data (τ_s, S, E), an attempt was made to train two binary classification decision trees, which could be used in the future to predict the starting and final point of renting a vehicle.

In the case of the first tree, the input variable (also known as predictor) was the rental start time τ_s, while the output (response) was the booking start point S. The prediction results using the trained booking start point tree are shown in Fig.28. It can be compared with the simulation results shown in Fig. 17. It can be seen that the general trends have been preserved. There are still two leading rental points with numbers 2 and 6 and the least frequently chosen point with number 9. At the same time, the number of inquiries for the leading points has been slightly overestimated, while for point 9 it has definitely been underestimated. The time histories of individual office queries also differ from those obtained directly from artificially generated data.



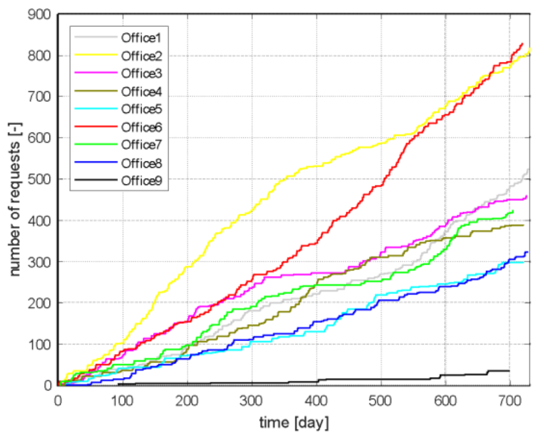
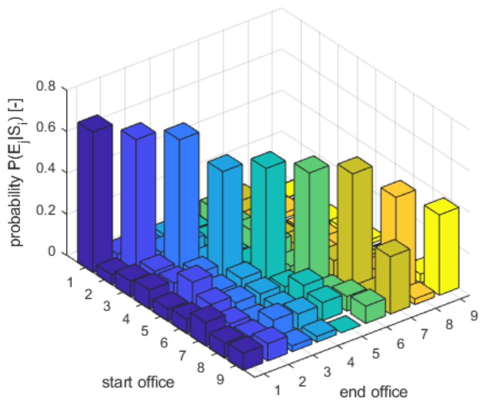


Fig. 28. Rental starting point prediction results using a trained classification tree;
on the left – the number of inquiries depending on the office number;
on the right – the number of queries spread over time



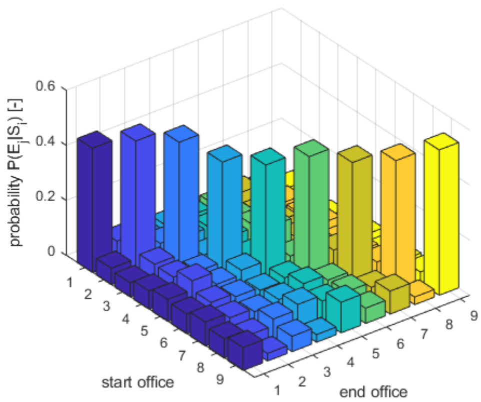


Fig. 29. Rental endpoint prediction results;
on the left – only one input variable (S); on the right – two input variables (τ_s, S)

For the second tree, the output variable was the booking endpoint. Two variants were tested. The first one considers only one input variable: start time. In the second, there were two input variables: reservation start time and starting point. It appears that the second variant better reflects the fact that the probability of choosing the end point is conditioned by the choice of the starting point. The prediction results are shown in Fig. 29. They should be compared with the simulation results shown in Fig. 16, where it was assumed that the dominant rental method is the two-way variant (the values on the diagonal of the probability density matrix are then dominant). The analysis of the data obtained from the prediction showed that the probability of one-way rental increased at the expense of the two-way probability. Interestingly, the case with one input variable (start time) gives better results because the probability values on the diagonal are higher (Fig. 29 left).

7. CONCLUSIONS

The method of preparing a sequence of requests from car rental customers proposed in the work can be used both in the case of having a real database and in the case of generating artificial data. The sequence of requests will be a set of necessary information to generate events in the model made in the SimEvents environment.

The obtained real-world data are not representative for short-term rental. However, their analysis allowed to detect certain regularities and observe seasonality and trends. For example, the specific days of the week when the rental starts and ends, or the specific length of the rental. While developing the method of generating artificial data, efforts were made to reproduce the most characteristic features of real data. The proposed method of generating artificial data is easy to modify and to scale it to the case of renting per hour. Unfortunately, artificially generated data is usually too regular. In the developed method, it can be tuned by increasing the noise level.

An alternative to the proposed method of generating artificial data may be the use of machine learning models: trained Gaussian mixture models and trained binary classification trees. However, the tests of these models for the available data showed some disadvantages, discussed in the text of the article. Briefly summarizing the results of testing machine learning models, it can be said that they can distort trends and cyclicality existing in the input data. This was particularly evident in the case of GMM training with real-world data. In the real data, there were bookings with very short or very long advance times. The GMM matching algorithms averaged these values, and consequently bookings with very short advance times disappeared. The averaged advance time for all bookings has increased. The same was true for the duration of the rental. Initially, short 5-day rentals and long rentals, almost two years long, dominated. After the transformation to GMM, short-term rentals almost disappeared.

Despite some disadvantages, machine learning models can be a source of fresh artificial random data obtained on the basis of other data.

References

1. Brendel Alfred Benedikt, Christian Rockenkamm, Lutz M. Kolbe. 2017. „Generating Rental Data for Car Sharing Relocation Simulations on the Example of Station-Based One-Way Car Sharing”. In: *Proceedings of the 50th Hawaii International Conference on System Sciences*: 1554-1563. Hilton Waikoloa Village, Hawaii, USA, January 4-7, 2017. ISBN: 978-0-9981331-0-2.
2. Carlier Aur´elien, Alix Munier Kordon, Witold Klaudel. 2014. „Optimization of a oneway carsharing system with relocation operations”. In: *10th International Conference on Modelling, Optimization and SIMulation MOSIM 2014.* Nov 2014, Nancy, France. hal-01294548.
3. Clemente M., M.P. Fanti, A.M. Mangini, W. Ukovich. 2013. "The vehicle relocation problem in car sharing systems: Modelling and simulation in a Petri net framework". In: *Application and Theory of Petri Nets and Concurrency. PETRI NETS 2013. Lecture Notes in Computer Science* 7927: 250-269. Edited by J.M. Colom, J. Desel. Springer, Berlin, Heidelberg. DOI: 10.1007/978-3-642-38697-8_14.
4. Conejero J. Alberto, Cristina Jordán, Esther Sanabria-Codesal. 2014. „An Iterative Algorithm for the Management of an Electric Car-Rental Service”. *Journal of Applied Mathematics* 2014 (Article ID 483734): 1-11. DOI: 10.1155/2014/483734.

5. Fink A., T. Reiners. 2006. „Modelling and solving the short-term car rental logistics problem" *Transportation Research Part E: Logistics and Transportation Review* 42(4): 272-292. DOI: 10.1016/j.tre.2004.10.003.
6. Jorge D., G. Correia. 2013. „Carsharing systems demand estimation and defined operations: A literature review". *European Journal of Transport and Infrastructure Research* 13(3): 201-220. DOI: 10.18757/ejtir.2013.13.3.2999.
7. Milenković Miloš, Libor Švadlenka, Vlastimil Melichar, Nebojša Bojović, Zoran Avramović. 2018. „SARIMA modelling approach for railway passenger flow forecasting". *Transport* 33(5): 1113-1120. ISSN: 1648-4142. DOI: 10.3846/16484142.2016.1139623.
8. Nair R., E. Miller-Hooks, 2011. „Fleet management for vehicle sharing operations". *Transportation Science* 45(4): 524-540. ISSN: 0041-1655. DOI: 10.1287/trsc.1100.0347.
9. Repoux Martin, Burak Boyacı, Nikolas Geroliminis. 2015. „Simulation and optimization of one-way car-sharing systems with variant relocation policies". In: *94th Annual Meeting of the Transportation Research Board At: Washington D.C.*
10. Turoń K., A. Kubik, F. Chen. 2022. „What Car for Car-Sharing? Conventional, Electric, Hybrid or Hydrogen Fleet? Analysis of the Vehicle Selection Criteria for Car-Sharing Systems". *Energies* 15: 4344.
11. SimEvents User's Guide. 2022. MathWorks Inc.
12. Wang H., R. Cheu, D.H. Lee. 2010. „Dynamic relocating vehicle resources using a microscopic traffic simulation model for carsharing services". In: *Computational Science and Optimization (CSO), 2010 Third International Joint Conference* 2010(1): 108-111. DOI: 10.1109/CSO.2010.98.
13. Wang Yuxuan, Feng Huixia. 2020. „Optimization and Simulation of Carsharing under the Internet of Things". *Hindawi Mathematical Problems in Engineering* 2020 (Article ID 4873048): 1-8. DOI: 10.1155/2020/4873048.
14. Yang Y., W. Jin, X. Hao. 2008. „Car rental logistics problem: A review of literature". In: *Service Operations and Logistics, and Informatics, 2008. IEEE/SOLI 2008. IEEE International Conference on Volume 2.* DOI: 10.1109/SOLI.2008.4683014.
15. ACRISS. „Car Codes". Available at: https://www.acriss.org.

Received 02.12.2023; accepted in revised form 12.03.2024