REDUCING WAGONS ACCUMULATION TIME IN CLASSIFICATION YARDS BY GENETIC ALGORITHM

Sanjin Milinković*, Rajko Karličić*, Slavko Vesković*, Miloš Ivić*, Ivan Belošević*
* University of Belgrade, Faculty of Transport and Traffic Engineering, Belgrade, Serbia
s.milinkovic@sf.bg.ac.rs, rajkokarlicic@yahoo.com, veskos@sf.bg.ac.rs, m.ivic@sf.bg.ac.rs, i.belosevic@sf.bg.ac.rs

***Abstract* – The process of railcars accumulation is one of the most important operations carried out at marshalling yards as it is the part of train forming process. Accumulation parameter is used for the evaluation as the indicator that measures the efficiency of the processes in the classification yard. It is used to determine the average collection train time per average number of wagons. This paper presents a model for reducing of collection times by genetic algorithm. Model uses genetic algorithm to search for the initial railcar groups from which to start the collection process and thus minimize the accumulation parameter. The model was tested for the Belgrade marshalling yard to calculate the optimal schedule of collection process for direct trains.**

I. INTRODUCTION

Marshalling yards are complex railway stations where freight trains are disassembled and assembled or rearranged in order to create trains according to wagon flows. Trains operations in marshalling yard are planned according to the schedule of the arriving trains. These tactical plans must include details on arriving trains, lists of wagons currently in the yard and other parameters (including type and characteristics of the shunting operation, repairing of the wagons, shunting of part-load shipments and supplying of the refrigerated railcars (wagons) with ice. Trains consist of wagons with different routes of referral, so they can be disassembled in marshalling yard and then assembled to form new trains consisting of wagons with same referral routes. This procedure typically performed in marshalling yards reduces the transport time and increases efficiency.

The operations that are necessary in order to disassemble and assemble trains are performed in receiving yard, in the zone of gravity (hump) and in the classification yard [1]. Classification yard consist of many tracks where wagons from disassembled trains are queued until the collection process is finished. The new train will be formed when the numbers of accumulated wagons reach the predefined criterion for certain direction. The process of accumulation is performed simultaneously with other operations and like all operations the goal is forming trains [2]. In the process of accumulation of wagons it is necessary to take into account all the wagons that are at a given time at the station by routes (tracks in classification yard). The aim of the accumulation is to collect a certain number of wagons and then to form and dispatch the train. The accumulation time is dependent on the number of wagons present in the classification yard, on the average number of wagons per train for each observed route and on the schedule and composition of the arriving trains. Accumulation parameter is mostly used to evaluate the efficiency of the classification yard performance, as well to comparison, and as an indicator of the process of accumulation it depends on the wagons group that has initiated accumulation. Processes in the classification yard are organized according to the limitations from the applied technology, and by the train parameters for the each departing route of referral [1, 3]. The accumulation of wagons by routes is the longest part of the train forming process, so by minimizing the accumulation parameter we are reducing the costs of the transport. The model for reducing the accumulation time should be able to determine the initial group of wagons for each route of referral so that the total collection times in the classification yards are minimal. We have tested the genetic algorithm method for the optimization of the accumulation time.

Genetic algorithms (GA) are evolutionary computation technique based on a similarity with the processes of selection and evolution in nature. Most of the methods that are called genetic algorithms have the following common elements: population of chromosomes, criterion (fitness) function, genetic algorithm operators and random mutation [4].

Genetic algorithms in transport are most often used for solving scheduling and routing problems [5, 6]. In railways problems, genetic algorithms are used for dispatching of train operations, for locomotive assignment problem, for solving railway traffic control conflicts etc. [7, 8, and 9].

The first step in the application of the genetic algorithms method is to represent a defined problem by a string of genes which is called chromosome. After chromosomal representation it is necessary to generate random population and evaluate the fitness of each chromosome in the population. Next step is to create a new population by repeating the application of genetic operators (selection, crossover and mutation). The newly created population is used for further work of the algorithm until a stopping criterion is fulfilled.

Goal of this paper is to examine the possibility of applying genetic algorithms as a decision support model to determine the best solution for initial wagons in the process of accumulation. We have created a hybrid model consisting of GA model (in Matlab [10]) that is connected to the spreadsheet software for input and output data.

II. THE PROBLEM DESCRIPTION

The process of accumulation of the wagons is the process of forming trains [2]. This process is done in marshalling yards. The duration of the process of accumulation depends on the total number of wagons for a specific referral as well as the number of wagons in the train that will be dispatched. Qualitative factors influencing the process of accumulation wagons are technology in the yard and the pattern of wagon arrivals.

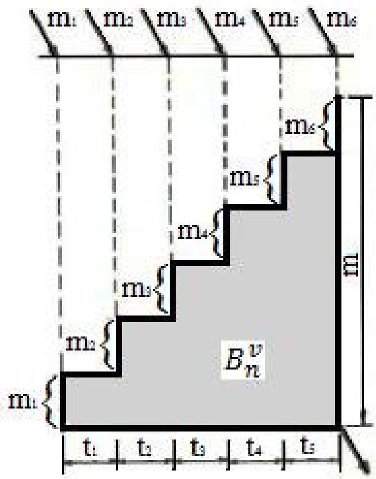

Figure 1. Calculation of train accumulation time

Accumulation parameter (Fig.1) is calculated by:

$$c = \frac{B_n}{m}[h], \text{ for } B_n = \sum t_{gri}\, m_{gri}\ [wagon \cdot h],$$

Where:

B_n – Number of wagon hours of accumulation for the collection of wagons for *i* route of referral,
t_{gr} – Time of group of wagons waiting during the collection process (except for the last, ending group) in hours,
m_{gr} – Number of wagons in a group.

Calculation of the accumulation time (Fig. 1) is by following equation:

$$B_n = m_1(t_1 + t_2 + t_3 + t_4 + t_5) + m_2(t_2 + t_3 + t_4 + t_5) + \ldots + m_5 t_5 + m_0 t_0\ [wagon \cdot h]\ .$$

In order to reduce the time of accumulation it is necessary to perform the following operations: larger groups of wagons should be at the end of accumulation process; and, the largest intervals between group arrivals should be between collected trains.

The process of accumulation should be observed from the moment of wagons arrival at the yard. Operations that are done simultaneously after the arrival are operations at the receiving yard and up to the final operations of forming trains.

Accumulation process is organized by train dispatcher who can prioritize some trains, for example trains in which there are groups of wagons that completes the collection of the train. The complex systems of marshalling yards can be analysed by simulation modelling in order to include all processes [11, 12]. In this model we observe only direct trains, i.e. trains that are formed for departure to the next station and will not change the composition on the route. The procedure for forming these types of trains is called single-stage sorting. Trains formed by a multi-stage sorting method contains wagons that needed to be humped again in order to form the trains that have groups of railcars sorted by the stations along the route [13, 14].

The aim of the model is to mark railcars initial groups for each direction in order to minimize accumulation parameter for the entire station (for all tracks forming direct trains). Model was tested on the Belgrade marshalling yard. To minimize the accumulation parameter it is necessary to determine from which group of wagons the accumulation process starts. The object of accumulation process is to create trains, in such a way that wagons are accumulated for train compositions of individual referral according to the plan of forming. Trains arrive at the station at the time that is defined in MS Excel spreadsheet input data table for each direction of referral. During accumulation process, groups of wagons that are pending consists of a certain number of wagons. Railcars are accumulated to a given number of wagons in the train and when accumulated to sufficient number of wagons for the formation of a train, train is starting the dispatching procedure. Number of wagons in the train varies for direct routes of referral and its value changes depending on the characteristics of the route and tracks. Input data for the model stored in the Excel tables are: arrival times of groups at accumulation process; and the number of wagons in groups. Excel tables are also used to calculate accumulation parameter for the opening sequence of the initial accumulation groups per direction (or track, as the direction is presented by one track). Output data, accumulation parameter for each combination of initial groups of railcars, is calculated from within the Excel thru a series of calculations. For example, accumulation parameter is calculated for opening sequence of initial groups where all twelve direct routes of referral are given by initial groups defined by *i*-th incoming group on *j*-th track. The input data for GA is also defined in Excel file for a chromosome that shows from which groups the accumulation process starts and at the same time indicates the direction of referral of direct trains.

The Matlab genetic algorithms toolbox (*gatool*) [10] were used in order to calculate the minimum accumulation parameter for the entire station. Matlab uses the data defined in *MS Excel* and by using *GA tools* instantly changes the groups from which accumulation begins. The "cooperation" of *MS Excel* and *GA Toolbox* is in exchanging the data where *MS Excel* is used as a fitness function with predefined analytical routines for calculating the accumulation parameter. Model implemented in *GA Toolbox* use data received from *MS Excel* to find minimum accumulation parameter by examining from which group accumulation should begin. This is achieved by importing the results of the fitness function (accumulation parameter *c*) from MS Excel to the GA model. Fitness function is a variable defined by a complex algorithm. Consequently, the value of the fitness function will change when altering the chromosomes or sequence of initial groups by directions. The GA model examines from which group accumulation should begin in order to find minimum accumulation parameter. In the next step, the GA model imports the results of the fitness function from the Excel fitness function as results of the new sequence generated by the GA model. Thus, the GA model starts the genetic algorithms optimization process by loading the value of fitness function where the accumulation of wagons starts from the *i*-th group for all 12 routes of referral. Then, it changes the value of the chromosome (by changing the number of groups from which accumulation begins) in fitness function and, after the recalculation of the fitness function, the result is imported as a new input data for the GA model. This procedure is performed until the required stopping criterion is fulfilled, i.e. until the minimum value of accumulation parameter is found.

III. IMPLEMENTATION OF THE MODEL

Belgrade marshalling yard (Fig. 2) was used for the implementation of the model as the biggest station in Serbia. Station is semi-automatic, gravitational single-sided marshalling yard with consecutive arrangement of receiving (14 tracks) and classification-dispatching (48 tracks) yards. According to the technology, it is gravitational marshalling yard with double-track hump. Belgrade marshalling yard dispatches direct and pick-up trains, but in this paper we have tested only direct trains or single-stage sorting. Track numbers in classification yard (routes of referral) are labelled by numbers from 1 to 12 and they include directions to: Pozega, Prijepolje, Nis, Presevo, Dimitrovgrad, Subotica and Sid.

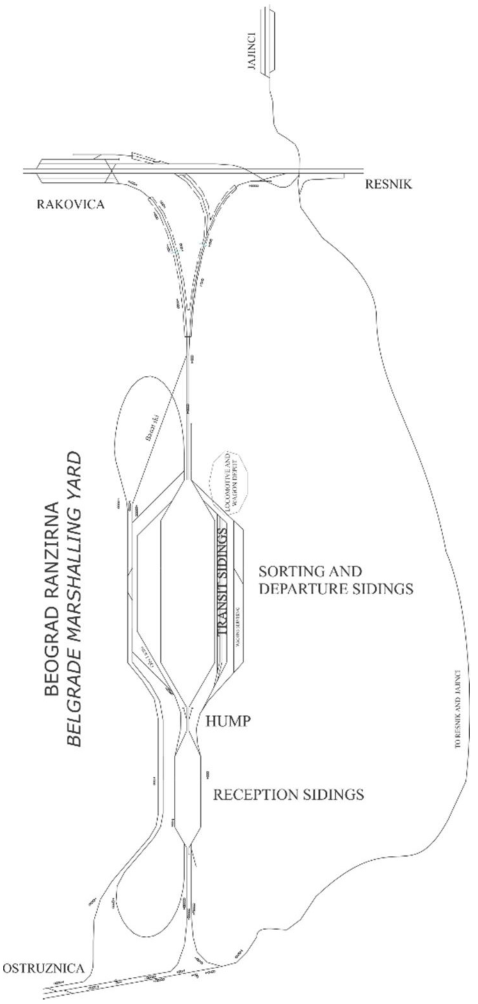

Figure 2. Belgrade Marshalling yard

Input data table is constructed in order to define arrival time of the groups to accumulation process as well as the number of wagons in a group that comes to the accumulation for all of routes of referral. Input data is stochastic and defined by the final reports from the Belgrade marshalling yard. Data on the arrival times, number of railcars, directions, and dispatching times extracted from the official reports are used to define stochastic models. For example, number of wagons in each group varied from one to eight (eight is defined as a limit due to technical limitations of the hump), and uniform distribution was used to define the number of wagons in groups for each route. The initial assumption is that all the processes of the accumulation and then dispatching of trains are finished within 24 hours.

Arrival time of groups at accumulation was generated as a result of two theoretical distributions: Erlang distribution of the second order is commonly used for generating the interval between train arrivals at the station [1], and exponential distribution is used to define the number and directions of wagons in inbound trains. Next step is to calculate the time that wagons spend at the accumulation process. After completion of the process of accumulation it is possible to calculate the parameters of the accumulation of individual and the total accumulation parameter of the entire station. Total accumulation parameter will be dependent on weighted parameters for all 12 routes of referral:

$$c = c_{ni1}\left(\frac{N_{di1}}{\sum N_{di}}\right) + c_{ni2}\left(\frac{N_{di2}}{\sum N_{di}}\right) + \cdots + c_{ni12}\left(\frac{N_{di12}}{\sum N_{di}}\right),$$

where:

c - Accumulation parameter for the station,

$c_{ni1} \dots c_{ni12}$ - Accumulation parameters by routes of referral,

$N_{di1} \dots N_{di12}$ - The sum of all wagons coming to the accumulation.

Input data includes the chromosome that is represented as a string consisting of as many digits as there are referral routes.

In order to start the *gatool* in Matlab it is necessary to define the fitness function, variables, the appearance of chromosome (defined in MS Excel), as well as all the other parameters to provide the best solution.

The initial appearance of chromosome can be changed in each generation, depending on time constraints or depending on the objective function. Changes in chromosome will be adjusted in GA Toolbox.

Fitness function *c* (accumulation parameter for the Belgrade marshalling yard) is defined in Matlab as an external parameter and retrieved from the Excel for each iteration. Excel file calculates the accumulation parameter *c* for a given input *X*, where *X* is defined as string of initial groups for accumulation.

In addition to using Genetic Algorithm Optimization Tool it is possible to run genetic algorithm function directly from a command line. Starting of genetic algorithm function from the command line was performed in the following way:

1. *Xmin* and *Xmax* are defined to determine from which group the accumulation starts. The lower limit for *X* is

equal to 1, and the upper ranges to the number of groups per routes of referral;
2. After the first step it is defined that *X* takes integer values (*X* represents the group from which the accumulation begins);
3. Next, optimization options structure (*opts*) is created with defined population size, number of generations, items related to the selection, crossover and mutation. A structure options is passed to the optimization function later on;
4. When the optimization options structure is created, genetic algorithm function is initiated;
5. After reaching stopping criterion, final solution is obtained i.e. accumulation parameter of Belgrade marshalling yard.

Variable *X* takes integer values, so there are certain restrictions when setting up optimization options structure. Some of the more important restrictions are:

- Only *doubleVector* population type.
- No custom creation function (*CreationFcn* option), crossover function (*CrossoverFcn* option), mutation function (*MutationFcn* option), or initial scores (*InitialScores* option).
- Genetic algorithm uses only the binary tournament selection function (*SelectionFcn* option), and overrides any other setting.
- No hybrid function. Genetic algorithm overrides any setting of the *HybridFcn* option.
- Genetic algorithm ignores *ParetoFraction*, *InitialPenalty* and *PenaltyFactor* options.

The genetic algorithm attempts to minimize a penalty function, not the fitness function. The penalty function includes a term for infeasibility. This penalty function is combined with binary tournament selection to select individuals for subsequent generations. The penalty function value of a member of a population is:

- If the member is feasible, the penalty function is the fitness function.
- If the member is infeasible, the penalty function is the maximum fitness function among feasible members of the population, plus a sum of the constraint violations of the (infeasible) point.

Given that the fitness function is the same as the penalty function when there is defined area of feasible solutions and in the model bounds for *X* are defined, in the remainder of this paper will continue to be used the term fitness function.

IV. RESULTS

In order to obtain the final solution testing was performed for a population size of 12 individuals, 20 and 24 individuals. These values were chosen based on the recommendation that the population of scale *n* and *2n* is optimal for a specific problem, where *n* is the length of the chromosome [4]. Since we defined the length of chromosome as 12 strings, population size from 12 to 24 was selected. Further, because the testing with a population size of *n* and *2n* only applies to some of the problems, testing was extended for the population size of 50, 100 and 200 individuals. Using a larger number of individuals in the population extends the time searching for solutions but with a larger population it is possible to produce better results.

Along with the change in population size, number of generations was changed and also defined as stopping criterion. Testing was carried out for 24, 50 and 100 generations.

Table I shows results for the case of 12, 20, 24, 50, 100 and 200 individuals in the population and 24, 50 and 100 generations. For a certain values of accumulation parameter the algorithm has interrupted its work. The reason for termination is that average change in the penalty fitness value is less than *options.TolFun* (tolerance on the constraint violation) and constraint violation is less than *options.TolCon* (termination tolerance on the function value). Standard value for *TolFun* and *TolCon* is *1.0000e-6* and should not be lower than 1.0000 e-14. According to the tests the lowest value of the accumulation parameter is 8.8578 hours and it is a case when there are 100 individuals in the population and 50 generations. For the next two cases (100 generations and 100 generations with set *TolFun* and *TolCon* to 1.0000-12), the lowest value of the accumulation parameter is repeated, i.e. the algorithm interrupts operation before completing all 100 generations.

TABLE I. VALUES OF ACCUMULATION PARAMETER FOR DIFFERENT CASES

Population	Generation			
	24	50	100	100*
12	9.59	9.59	9.59	9.59
20	9.71	9.34	9.34	9.34
24	9.21	9.15	9.15	9.15
50	9.20	8.92	8.92	8.92
100	8.96	8.86	8.86	8.86
200	9.33	9.33	9.33	9.33

100* - 100 with *TolFun* and *TolCon*

Fig. 3. and Table I shows changes in fitness function by generations. The number of individuals in a population is 100 and the number of generations 50 and in this case the value of accumulation parameter is 8.8578 hours. This represents the minimal accumulation parameter for Belgrade marshalling yard.

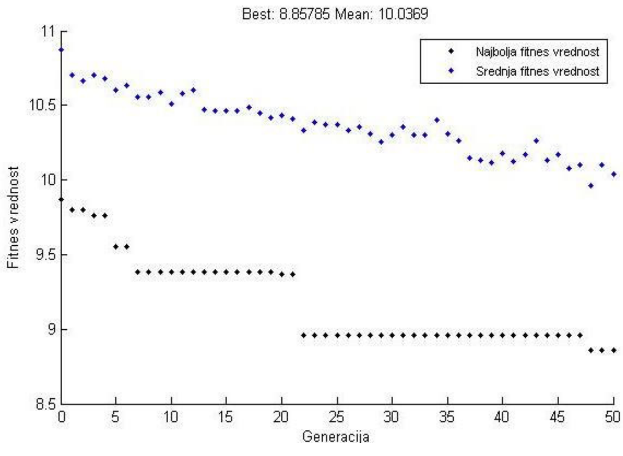

Figure 3. Values of accumulation parameter during GA tool optimization

This value is much lower compared to the initial (when the accumulation process started from the first group of wagons for all directions) which had a value of 9.97 hours. The final value of accumulation parameter was obtained after completion of all stages of the genetic algorithm (generating of initial population, completion of stages of selection, crossover, mutation and fulfilling of stopping criteria). As a stopping criterion the maximum number of generations was used (in this example 50 generations). Appearance of the best solution that is generated in all populations, or at the moment of reaching the final value of the fitness function, is shown in Table II. Namely, the solution obtained by genetic algorithm optimization shows the serial number of incoming group, for each direction, that determines the initial group for train accumulation. Namely, accumulation for the first route of referral (Požega) should start from the 6th group of wagons, for the second route (Prijepolje teretna) from the 3rd group of wagons, for third (Niš and Dimitrovgrad) 5th, for fourth (Preševo) 5th, for fifth (Preševo) 8th, etc.

TABLE II. THE SEQUENCE OF INITIAL GROUPS OF WAGONS FOR ACCUMULATION ON DIRECT ROUTES OF REFERRAL

Route of Refferal	Požega	Prijepolje	Niš	Preševo	Preševo	Dimitrovgrad	Dimitrovgrad	Dimitrovgrad	Subotica	Šid	Šid	Šid
Initial group	6	3	5	5	8	8	11	6	17	5	17	5

In order to check the obtained value of the accumulation parameter, the primary appearance of chromosome was tested for different initial values [15]. This change has produced results of similar error threshold. The best value of accumulation parameter obtained by experimenting with initial value of the chromosome was 8.8578 hours (Table III.). New value of accumulation parameter is reached for the population of 100 individuals and 50 generations. To further verify the final solution graphical representation of the accumulation process was produced (Fig. 4). The accumulation is carried out for each route of referral from the group of wagons that is defined in the final solution. The value of accumulation parameter for Belgrade marshalling yard, which is obtained from graphical method, was 8.8578 hours. This value is the same as in the GA model and thus verifies the solution obtained by GA.

TABLE III. RESULTS FOR THE BELGRADE MARSHALLING YARD

No.	Routes of Refferal (Directions)	Wagons flow N_{di}	Number of Trains N_{vi}	Accumu-lation Wagonhours (B_{nd})	Accumu-lation parameter C_{ni}	$C_{ni}(N_{di}/\Sigma N_{di})$
1	Požega	51	3	131	7.71	0.39
2	Prijepolje teretna	87	5	136	7.53	0.65
3	Niš ranžirna i Dimitrovgrad	74	4	158	8.33	0.61
4	Preševo	66	4	122	7.16	0.47
5	Preševo	70	4	163	9.03	0.62
6	Dimitrovgrad	102	6	154	9.03	0.91
7	Dimitrovgrad	114	6	181	9.51	1.07
8	Dimitrovgrad	76	4	179	9.42	0.71
9	Subotica	76	4	176	9.24	0.69
10	Šid	96	4	241	10.04	0.95
11	Šid	91	4	197	8.56	0.77
12	Šid	110	5	208	9.43	1.02
		1013	53			**8.86**

V. CONCLUSION

The current status of the Belgrade marshalling yard is such that its capacity is not fully exploited. However, in case that the number of freight trains increases, optimization of processes in marshalling yards will be required as the process of accumulation is the most important process regarding the unproductive time for wagons. In technical freight stations the most time is lost for holding up the wagons in the accumulation process. Even with the current number of trains and flow of wagons, with the use of high quality reports on incoming freight trains, optimization of accumulation process could reduce the overall costs of wagons delays at the technical freight stations.

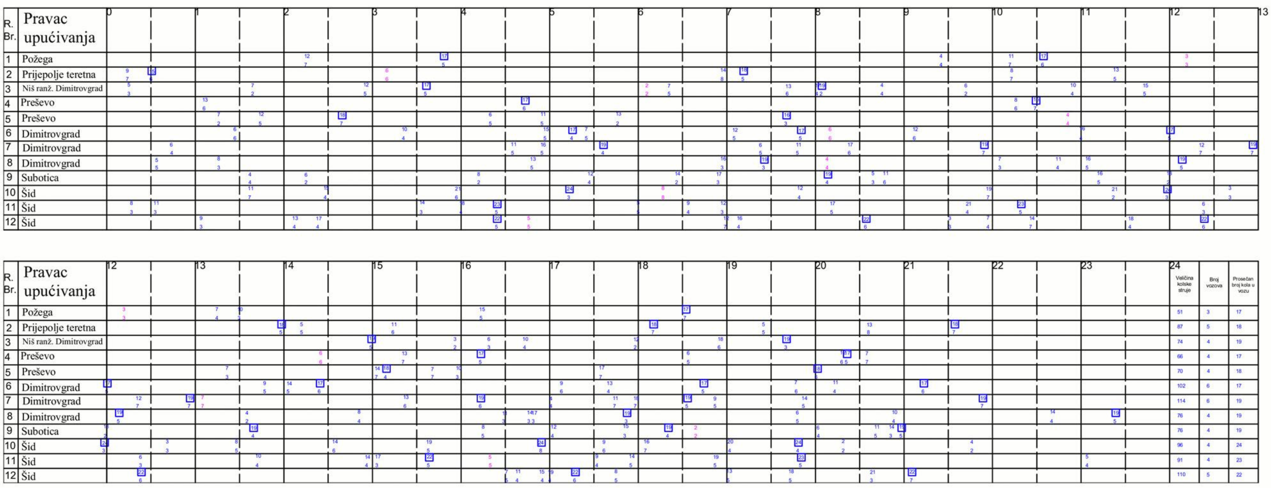

Figure 4. Results obtained by graphical method

Based on the obtained results it can be concluded that the application of genetic algorithms to optimize the accumulation of wagons gives satisfactory results. The obtained results show a significant decrease in the accumulation parameter when compared to initial value that is often in use for creating operational plans. The solution obtained by GA indicates groups of wagons from which the accumulation should begin. As a metaheuristic method, genetic algorithms can produce results that may not represent the optimal solution, but results are close to optimal with small (if any) errors, which is satisfactory for this problem size and structure.

One of the advantages of genetic algorithms is its ability to produce satisfactory results for accumulation parameter that does not require large financial investments. Also, the use of genetic algorithms is not complicated. In further research we will focus on increasing the speed of the model, (i.e. the code that should calculate accumulation parameter within the GA Toolbox) and on the applicability by improving the structure and quality of input data on the arriving trains and wagons. The accuracy of the created model depends on the quality of the input data, i.e. with more precise data on train arrivals and train composition at the station for a certain period of observation, model will be able to produce more precise results.

Acknowledgment

This paper is supported by The Ministry of Educations and Science of the Republic of Serbia, within the research projects No. 36012.

References

[1] M. Čičak, S. Vesković, "*Organizacija železničkog saobraćaja II*", Beograd, Srbija, Saobraćajni fakultet, 2005.

[2] N. Boysen, M. Fliedner, F. Jaehn, & E. Pesch, "Shunting yard operations: Theoretical aspects and applications". *European Journal of Operational research*, 220(1), 1-14., 2012.

[3] M. Čičak, "*Modeliranje u železničkom saobraćaju*", Beograd, Srbija, Saobraćajni fakultet, 2003.

[4] M. Mitchell, "*An introduction to genetic algorithms*". MIT press, 1998.

[5] D. Teodorović, M. Šelmić, "*Računarska inteligencija u saobraćaju*", Beograd, Srbija, Saobraćajni fakultet, 2012.

[6] N. Marković, N. Bešinović, and Paul Schonfeld. "Simulation-Based Optimization of Recovery for Multiterminal Freight Transportation System." *Transportation Research Board 91st Annual Meeting*. No. 12-2650. 2012.

[7] S. Dündar, and İ. Şahin., "Train Re-Scheduling with Genetic Algorithms and Artificial Neural Networks for Single-Track Railways". *Transportation Research Part C: Emerging Technologies*, Vol. 27, No. 0, 2013, pp. 1-15.

[8] K. Nachtigall, S. Voget, "A genetic algorithm approach to periodic railway synchronization". *Computers & Operations Research*, 23(5), 453-463., 1996.

[9] P. Tormos, A. Lova, F. Barber, L. Ingolotti, M. Abril, M. Salido, "A genetic algorithm for railway scheduling problems". In *Metaheuristics for Scheduling in Industrial and Manufacturing Applications* (pp. 255-276). Springer Berlin Heidelberg., 2008.

[10] A. J. Chipperfield, and P. J. Fleming. "The MATLAB genetic algorithm toolbox." *Applied Control Techniques Using MATLAB, IEE Colloquium on. IET*, 1995.

[11] P. Marton, N. Adamko, "Villon - a tool for simulation of operation of transportation terminals", *Communication* vol. 10(2)/2008; pp.10-14. 2008.

[12] R. Jacob, P. Márton, J. Maue, M. Nunkesser, "Multistage methods for freight train classification". *Networks*, 57(1), 87-105., 2011.

[13] M. Ivić, A. Marković, S. Milinković, I. Belošević, M. Marković, S. Vesković, N. Pavlović, M. Kosijer, "Simulation model for estimating effects of forming pick-up trains by simultaneous method". In *Proceedings of 7th EUROSIM Congress on Modelling and Simulation*, Prague. 2010.

[14] M. Ivić, I. Belošević, S. Milinković, M. Kosijer, N. Pavlović, "Track properties for formation of pick-up trains". *Građevinar*, 65(02.), 123-134., 2013.

[15] R. Karličić, "*Optimizacija nakupljanja kola primenom genetskog algoritma*". Master rad, Univerzitet u Beogradu – Saobraćajni fakultet, Beograd. 2014.