Mechanics Transport Communications
Academic journal

ISSN 1312-3823
volume 12, issue 3/1, 2014
article № 0968
http://www.mtc-aj.com

INVESTIGATION OF THE SOFIA'S SUBWAY STATIONS USING CLUSTER ANALYSIS

Svetla Stoilova, Veselin Stoev
stoilova@tu-sofia.bg, stoev@tu-sofia.bg

Technical University of Sofia
Sofia, PostCode 1000, boulevard Kliment Ohridski 8
BULGARIA

Key words: *metro system, metrostation, subway, cluster analysis, passenger, classification*
Abstract: *The metrostations are part of the metro system of transport service of the population. They are different in infrastructure, operational and other indicators. This paper presents an investigation for classification the Sofia's metrostations. In research are determined the indicators for classifying. To make the classification of metrostations in groups, using various types and dimension factors is applied a multidimensional statistical method named cluster analysis. To perform clustering has been used statistical software package SPSS. In the research have been applied 30 factors for cluster analysis. The Sofia's metrostations has been classified into two groups. The first is the group of main metro station. It includes 8 metrostations that are situated only in first metro line. The second includes 19 metrostations that are situated in both metro lines. The proposal classification of the metrostations in groups (clusters) would allow implementation of adequate technical and technological solutions related to transportation.*

1. INTRODUCTION

Metrostations are the main points to form a passenger flows in the metropolitan. They are classified usually by one indication such as operational characteristics (starting, intermediate and interchange metrostations); infrastructural indicators (with one platform, with two or more platforms; with and without track development and non-track developmen etc.). Different stations generate a different passenger flow volume. This is determined by the location of the metrostations in the urban and housing areas, contact with other types of public transport, integration in the surrounding metrostation infrastructure. Subway stations that serve important locations, interchange with other lines or serve a large number of passengers daily have more relative importance than other subway stations. On the one hand the commercial and business buildings that are developed near the metro station are integrated regional connectivity. They easily and quickly should bring traffic of big flow of people. On the other hand schools and universities located near to metrostations also generate a large passenger flow.

In systematization of subway stations by one indicator the classification of metrostations is different. Metrostations, which are important for one indicator might not be important in another.

The object of the research is 27 real metrostations in Sofia's metro network. The aim of the study is:

- ♦ To determine the factors for investigate and classification the metrostations.
- ♦ To classify the studied subway stations by using their characteristics.
- ♦ To apply the Cluster Analysis for studying metrostations.
- ♦ To propose an classification of metro station.

2. LITERATURE REVUE

Some authors have investigated the factors characterizing the subway station. In [1] are investigated 268 metrostations in nine US cities. Certain factors characterizing travels from the stations of the "light metro" have been determined such as: employment, population, availability of an airport connection to other lines, intermodal connection with bus, the availability of parking spaces at the station, availability of school with pedestrian accessibility and other. The study [2] is investigated the factors affecting Metro demand at a station level. These factors were examined in the Seoul metropolitan area, which is one of the most densely populated regions in the world. Twenty-four independent variables were chosen based on insights and findings from previous studies. In [3] is introduced several new variables significantly associated with station boarding, such as: employment, commercial floor area, office floor area, net population density, number of transfers, number of feeder bus lines, and a dummy variable indicating transfer stations. A regression analysis was conducted with weekly average of station boarding as the dependent variable.

In [2] factors that were expected to affect the station ridership were chosen based on results from previous studies such as [1,3,4]. The factors were categorized into four areas: built-environment variables, travel impedance variables, intermodal connection variables, and other variables.

In [5] were examined number of passengers boarding at each station in the Madrid Metro network and were developed a ridership forecast model, based on the combined use of Geographic Information Systems, distance-decay functions and multiple regression models. The factors that were studied are: of the characteristics of the stations (type, number of lines, accessibility within the network, etc.) and of the areas they serve (population and employment characteristics, land-use mix, street density, presence of feeder modes, etc.).

All these studies investigate metrostations with purpose to determine the regression model for station-to-station boarding according to different factors.

Some authors have made research to classify the subway stations for rehabilitation. For example [6] is concerned with developing global rating system for subway stations' networks in order to set the rehabilitation strategies. In the research is applied Simos' ranking method. For determine the importance of subway stations were examined six attributes influence: the number of passengers per day, the age of the subway station, the number of years from last full rehabilitation, type of station whether it is above ground or underground and whether the subway station is an interchanging station or not. In [7] the different stations in a subway network are rated according to their relative importance with deferent attributes such as station size, location and passenger capacity, criticality of a station that depending on a set of identified factors as structural, architectural, electrical, mechanical, and communication systems. Both methods [6,7] are applied to Cairo's Metro.

All these studies indicate the importance of metrostations. There is no proposed a comprehensive study that compares and classifies metrostations for technological purposes by different infrastructural, operational and other indicators.

3. METHODOLOGICAL APPROACH

3.1. Factors for study the subway stations

The factors for the research were divided into four groups: the factors characterizing the trips; factors which characterize the intermodality of metropolitan to other types of public transport; infrastructure factors; factors that determine the type of the built environment in the area of metrostation.

1. Factors for the volume of travel from the metrostation.

♦ P_h – The mobility of traveling by metrostation. This factor indicates the effects of the number of daily trips from a metrostation on 1000 inhabitants of the population of Sofia.

(1) $$P_h = \frac{1000.P_S}{H}$$, pass. /1000 inhabitants

where: P_S is the average daily passenger flow from metro station, pass. / day; H is the e number of city inhabitants; H_q is the number of registered inhabitants in the area of the metrostation..

♦ P_q – The average number of trips per day for person that lives in the region where is situated the metrostation.

(2) $$P_q = \frac{P_S}{H_q}$$, pass./ 1000 inhabitants in region

where: H_q is the number of inhabitants in region to the metrostation.

♦ P_S – The average daily passenger flow, pass. / day.

2. Factors that indicate intermodality with other modes of public transport. The research analyzes the bus, trolley and tram transport which have contact with subway in the area on metrostations.

♦ n_L^A – The number of bus lines in the area to the metrostation.

♦ N_L^A – The intensity of bus routes per day for a metrostation.

♦ I_C^A – The intermodal coefficient of bus routes. It shows the relationship between bus and underground for a metrostation. It is defined as the ratio of the intensity of the bus to the intensity of a metro trains for a metrostation.

♦ n_L^{TB} – The number of trolley bus lines in the area to the metrostation.

♦ N_L^{TB} – The intensity of trolley bus routes per day for a metrostation.

♦ I_C^{TB} – The intermodal coefficient of trolley bus routes. It shows the relationship between trolley bus and underground for a metrostation. It is defined as the ratio of the intensity of the trolley bus to the intensity of a metro trains for a metrostation.

♦ n_L^{TM} – The number of tram lines in the area to the metrostation.

♦ N_L^{TM} – The intensity of tram routes per day for a metrostation.

♦ I_C^{TM} – The intermodal coefficient of tram routes. It shows the relationship between tram and underground for a metrostation. It is defined as the ratio of the intensity of the tram to the intensity of a metro trains for a metrostation.

The intermodal coefficients by type of urban transport is determined by the formulas:

(3) $$I_C^A = \frac{n_L^A.N_L^A}{N_M} \; ; \; I_C^{TB} = \frac{n_L^{TB}.N_L^{TB}}{N_M} \; ; \; I_C^{TM} = \frac{n_L^{TM}.N_L^{TM}}{N_M}$$

♦ M_T - The number of possible transfers of passengers from one metro line to another. M_T=1 when changing the route; M_T=0 otherwise.

♦ I_T – The intermodal coefficient for routes. It shows the total intensity of the routes of the public transport in the contact points of the metrostation. It shows the degree of connection of underground with other types urban transport. The intermodal coefficient determines the degree of satisfaction of urban transport for a metrostation. It is indicates the degree of satisfaction for a metrostation compared to other metrostations in metro network. It is calculated separately for each metrostation by formula:

$$(4) \qquad I_T = \frac{n_L^A + n_L^{TB} + n_L^{TM}}{n_S - 1}$$

where: n_S is the total number of metrostations in metro network.

3. Infrastructural factors relating to the entry of passengers. The research analyze the following factors: A_P – the area of the platform, m^2; n^P – the number of platforms; n^V – the number of entrance turnicators with validating equipment in the metrostation.; n^{ET} – the number of exit turnicators; n^E – the number of escalators in metrostation; n^C – the number of ticket office and automatic ticketing in metrostation; n^{EN} – the number of inputs for entry / exit to the metrostation .

4. Factors that determine the type of the built environment in the area of metrostation. The extent of this zone was adopted with a radius of 400 meters from the metrostation, given the Regulation 2 / 29.06.2004 " Planning and design of communication and transport systems of urban areas in Bulgaria", which defines the pedestrian movements of the buildings and workplaces to public transport stops. In the research are considered the following factors that determine the existence of a type of built environment in the area to the metrostation; B^A – administrative buildings; B^R – housing buildings; B^S – schools and universities; B^B – business centers and commercial complexes; B^C – cultural and sports centers; B^P – public buildings.

If there is a relevant type of infrastructure, the factor has a value of 1, otherwise it is 0.

3.2. Cluster analysis. An overview

The Cluster analysis is a suitable method for a classification of the examined metro networks into groups by using different factors. It is multi-measurable statistical analysis for a classification of units into groups, preliminarily unknown, based on numerous characteristics in relation to these units [8]. The number of examined factors is greater than 2. The statistics theory suggests different methods of clusterisation.

The dispersion analysis could be used for an approximate evaluation of the clusterisation's results as well as for determining the roles of each variables used for clusters' establishment. The determination of the statistical importance of different factors is done by using the F criterion (Fisher's criterion).

$$(5) \qquad F \geq F_T$$

where: F (α , k_1, k_2) is the empirical value of the criterion resulted from the dispersion analysis, F_T is the theoretical value when the level of risk $\alpha = 0,05$ and the number of degrees of freedom, $k_1 = n- 1, k_2 = m - n$; m is the number of observations (in research there are the observations of 27 factors for 27 metro station), n is the number of examined units (in research it is the number of metrostations).

On one hand, the Fisher' criterion's evaluation determines which factors are significant for the study, on the other it do not dismiss those other factors which are used for clusterisation but does not satisfy the condition (5). The F test in Cluster analysis should be used only for descriptive purposes because the clusters have been chosen to maximize the differences among cases in different clusters.

4. GROUPING OF METRO NETWORKS IN AN INDICATOR

Systematization of the investigated stations can be performed by each of the above factors. For this purpose, subway stations are classified into groups on conditional intervals of alteration of the factors.

In the figures 1, 2 and 3 is shown examples for grouping the metrostation for some of the investigated factors. Figure 1 shown grouping of metrostations by trips per day per 1000 inhabitants of the region which is located metrostation. Conditions, they can be divided into six groups, with the interval of grouping 100 pass./1000 inhabitants of the region. In figure 2 is done grouping by indicator trips of 1000 people of the population of the capital. Conditional metrostations can be divided into 4 groups at an interval of 5 pass. / 1000 per capita. Figure 3 illustrates grouping of metrostations depending on the coefficient of intermodal routes. Conditional on this indicator the metrostation can systematized in 6 groups. Figure 4 shows the change of intermodal coefficients for modes of public transport. The distribution in groups of metrostations is based on the number of modes of transport contacting in the metro station. Classifying the metrostations only by one indicator does not give a complete characterization for metropolitan. In each of the studied indicators can be compiled a different grouping of metrostation.

4. APPLICATION OF CLUSTER ANALYSIS FOR A METRO NETWORK

A method for hierarchical clustering has been used in the study. The main advantage of this method is that the determination of a unit into a specific cluster is definitive. Hierarchical clustering is performed by the agglomerative method of average linkage between groups. For the distance-type measures it is chosen the Squared Euclidean distance, [8].

Table 3 shows the value of F criterion for the examined indicators. The theoretical value of F criterion is F_T=1, 46.

The indices of the travel have F criterion bigger than theoretical. They are important for clusterisation. The criterion for possibility of chance the metro line is important for grouping. All indices on intermodality have value of F criterion smaller than theoretical. Four criterions on infrastructure as the area of platform, number of platforms, number of escalators, and number of cases and apparatus for automatic ticketing have F criterion bigger than theoretical. Four criterions for type of built environment have F criterion bigger than theoretical. Those are availability of schools and universities, commercial and business centres, cultural and sports centers, public buildings. The dispersion analysis only gives an idea of the importance of the factors in the cluster analysis and does not exclude any factor.

SPSS (Statistical Package for Social Science) software has been used for carrying out the study with a cluster analysis. A dendrogram of the formed clusters and their respective elements are shown in figure 5.

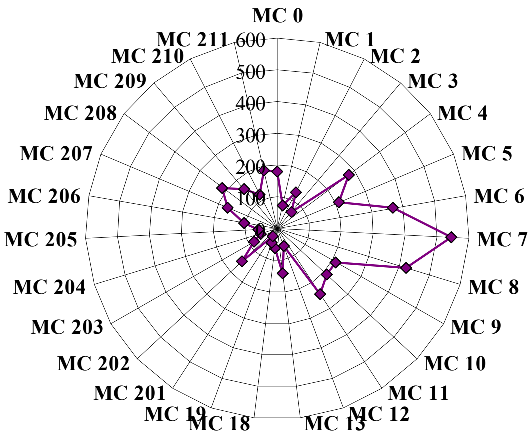

Fig.1. Travel by subway stations per day per 1000 inhabitants of the residential area

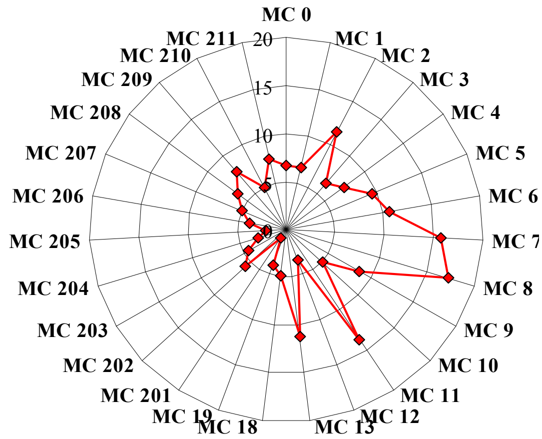

Fig.2. Travel by subway stations per day per 1000 inhabitants of the population of the city

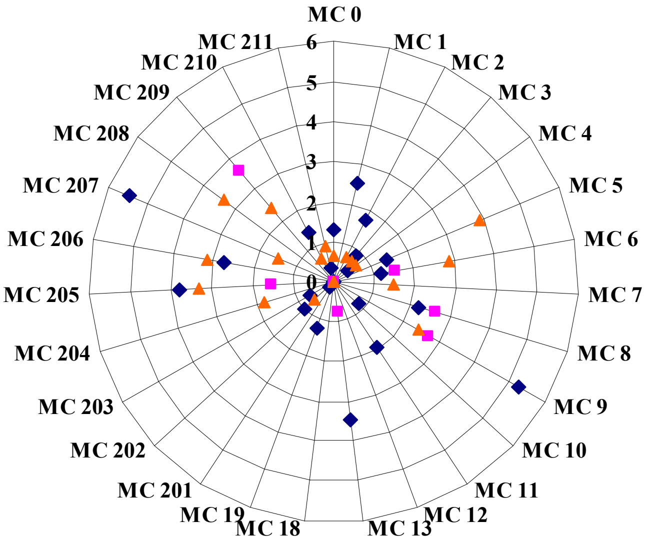

Fig..3. Intermodal coefficient by modes of transport

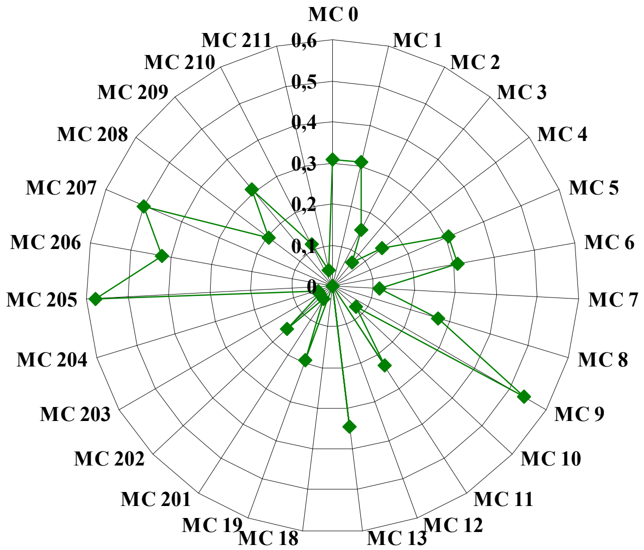

Fig.4. Intermodal coefficient of routes

Indications:

MC 0	Obelya
MC 1	Slivnitsa
MC 2	Lyulin
MC 3	Zapaden park
MC 4	Vardar
MC 5	Konstantin Velichkov
MC 6	Opalchenska
MC 7	Serdika
MC 8	SU "Kliment Ohridski"
MC 9	"Vasil Levski" stadium
MC 10	Joliot Curie
MC 11	G.M. Dimitrov
MC 12	Musagenitsa
MC 13	Mladost 1
MC 18	Mladost 3
MC 19	Tsarigradsko shose
MC 201	Lomsko shose
MC 202	Beli Dunav
MC 203	Nadezhda
MC 204	Khan Kubrat
MC 205	Maria Luiza
MC 206	Cetral Rail Stantion
MC 207	Lavov most
MC 208	Serdika 2
MC 209	NDK
MC 210	Europren union
MC 211	James Bouchier

◆ **autobus** ■ **trolley bus** ▲ **tram**

Table 1.Values of F criterion for the examined indicators

Factors of the travels	P_h	P_q	P_j								
F criterion	17,76	62,37	62,37								
Factors on intermodality	n_L^A	N_L^A	I_C^A	n_L^{TB}	N_L^{TB}	I_C^{TB}	n_L^{TM}	N_L^{TM}	I_C^{TM}	M_T	I_T
F criterion	0,01	0,02	0,02	1,41	1,32	1,32	0,84	1,20	1,20	1,71	0,58
Factors on infrastructure	A_P	n^P	n^V	n^{ET}	n^E	n^C	n^{EN}				
F criterion	4,79	7,06	0,25	0,65	1,57	2,08	0,69				
Factors for type of built environment	B^A	B^R	B^S	B^B	B^C	B^P					
F criterion	0,12	0,61	2,32	2,28	2,02	3,33					

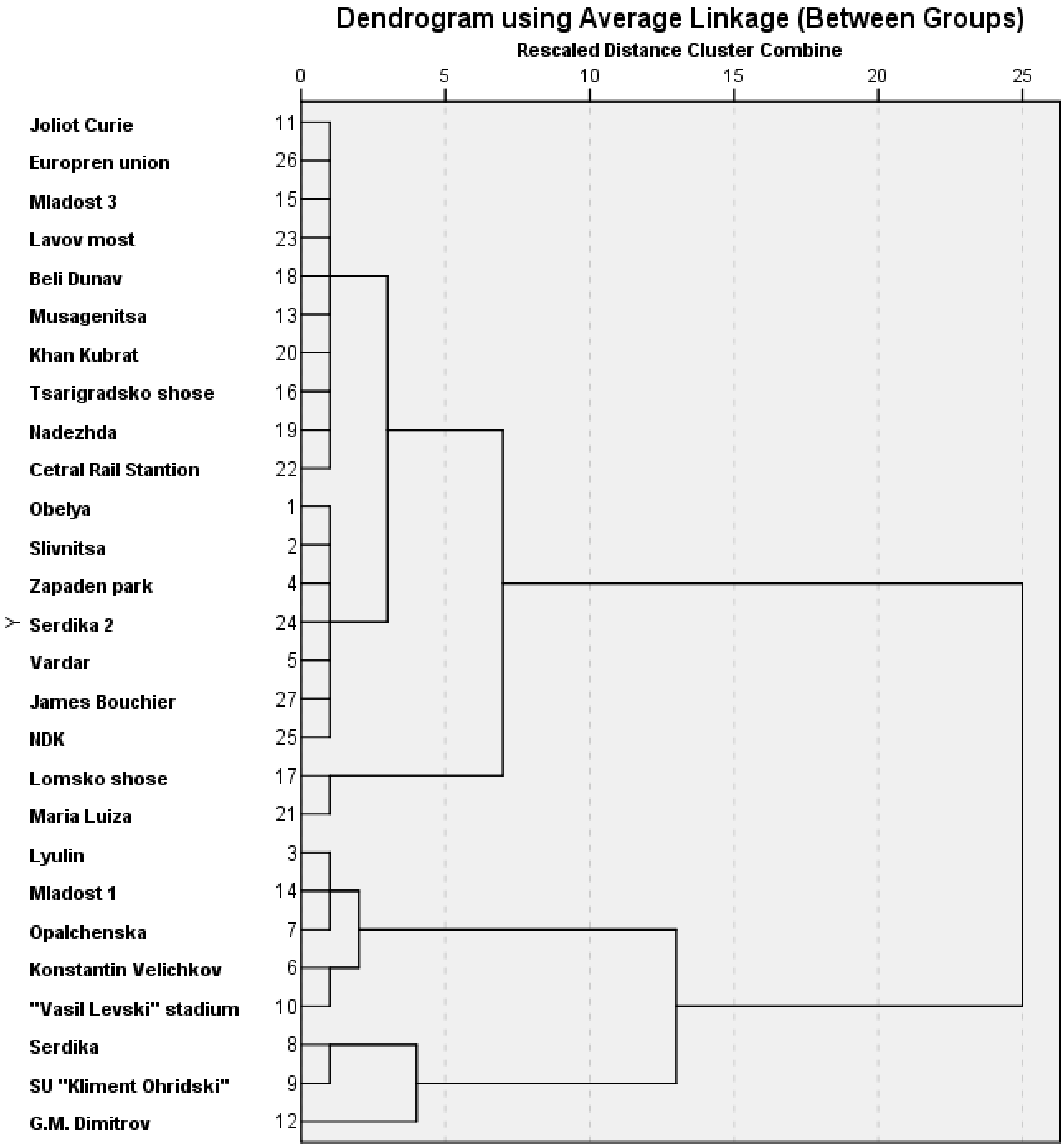

Fig. 5. Dendrogram of clusters for the subway stations

The results indicate that metro station of the considered Sofia's metro network can be classified into two groups:

♦ A cluster of basic metrostations. This cluster contents 8 subways stations. In this group are: Lulin, Konstantin Velichkov, Opylchenska, Serdica, Sofia University "St. Kliment Ohridski", Vasil Levski Stadium, G.M. Dimitrov, Mladost 1.

♦ A cluster of secondary metrostations. This cluster contents 19 subways stations. Obelya, Lomsko shose, Beli Danube, Nadejda, Han Kubrat, Maria Louisa, Central Railway Station, Lavov most, Serdica 2, NDK, European Union, James Boucher, Slivnitza Zapaden Park, Vardar, Joliot Curie, Mladost 3, Musagenitsa, Tzarigradsko shose.

In the cluster of principal metrostations are formed two sub-clusters, which indicate the degree of proximity of the studied metrostations. In the first sub-cluster are three most - loaded metrostations: Serdica, Sofia University "St. Kliment Ohridski" and G.M.Dimitrov. The first two stations are with a central platform and are located in the city centre, while the third has two platforms and is located outside the central urban area, but serves large passenger flow from schools and universities. In the other sub-cluster are metrostations: Lyulin, Mladost 1, Opalchenska, Konstantin Velichkov and "Vasil Levski" Stadium. All metrostations of the basic cluster are situated at the first metro line, which has a large passenger flow and passes through important urban areas. In the cluster of secondary metrostations have been formed three subgroups.

CONCLUSIONS

The study has shown the following results:

♦ The factors for classifications of metrostations have been defined-these factors allow us to evaluate the stage of development of the examined metrostations.

♦ A cluster analysis has been used for classification of the metrostations. The classification has been conducted by using 27 different factors.

♦ Sofia' metrostations are divided into two groups.

♦ The application of the Cluster analysis allows us to evaluate the stage of development of metrostations.

♦ The classification obtained by cluster analysis could be used for assessment of rehabilitation; for elaboration timetable of metro trains with passing up metrostations on certain days and / or times periods without violation the technological windows in the the traffic schedule; for the introduction of categorization in pay for the employees according to the metro station at which they work; for further development of technological processes in metrostations.

♦ In the research has been determined the intermodal coefficient of the routes of given types of public transportat. This coefficient indicates the satisfaction of passenger transport according to the metrostation, which the passengers want to do intermodal connection. This coefficient can be used to assess the load of the incoming routes of urban transport depending on the places where he made intermodal connection with subway.

♦ The proposed factors and methodology for classification the metrostations can be applied also for planned to build and operate metrostations by using forecasts and expert data.

ACKNOWLEDGEMENT

This research is conducted in relation to the execution of a contract № 142ПД0019-04" A simulation modelling of technological processes in main metrostations of the Sofia's subway". The research has been funded by Technical University of Sofia, Bulgaria.

REFERENCES
[1]Kuby, M., Barranda, A., Upchurch, C. Factors influencing light-rail station boarding in the United States. Transport Research Part A 38(3), 2004, pp.223-247
[2]Choi J, Y. Lee, T. Kim, K. Sohn. An analysis of Metro ridership at the station-to-station level in Seoul. Transportation, 2012. vol. 39. pp.705-722.
[3]Sohn K, Shim, H. Factors generating boarding at metrostations in the Seoul metropolitan area. Cities 27 , 2010, pp. 358–368.
[4]Estupinan, N., Rodrıguez, D., The relationship between urban form and station boarding for Bogota's BRT. Transportation Research Part A. 42, 2008, pp. 296–306.
[5]Gutiérrez, J., Cardozo, O., Palomares, J. Transit ridership forecasting at station level: an approach based on distance-decay weighted regression. Journal of Transport Geography 19, 2011 pp. 1081–1092.
[6]Marzouk, M., Abdelaty, A. BIM-based framework for managing performance of subway stations. Automation in Construction, Volume 41, May 2014, pp. 70–77.
[7]Abouhamad, M., T. Zayed. Criticality-based model for rehabilitating subway stations. Concordia University, Proceedings of the 30th ISARC, Montreal, Canada, 2013, pp.1296-1304.
[8]Tan, P., Steinbach, M., Kumar, U. Introduction to Data mining. Minnesota: Iniversity of Minnesota. 2006. p. 769.

ИЗСЛЕДВАНЕ НА МЕТРОСТАНЦИИТЕ НА СОФИЙСКИЯ МЕТРОПОЛИТЕН С ИЗПОЛЗВАНЕ НА КЛЪСТЪРЕН АНАЛИЗ

Светла Стоилова, Веселин Стоев
stoilova@tu-sofia.bg, stoev@tu-sofia.bg

Технически Университет – София,
София 1000, бул."Кл. Охридски" 8
БЪЛГАРИЯ

Ключови думи: *метросистема, метростанция, клъстърен анализ, пътници, класификация*

Резюме: *Метростанциите са част от метросистемата за транспортно обслужване на населението. Те се различават по инфраструктурни, експлоатационни и други показатели. Тази статия представя изследване за класифициране на метростанциите на Софийския метрополитен. В изследването са определени показатели за класифициране.За да се класифицират метростанциите в групи, чрез използване на различни по тип и дименсия фактори е приложен многомерният статистически метод Клъстърен анализ. За определянето на клъстърите е използван статистическия софтуер SPSS. В изследването за клъстърен анализ са приложени 27 фактора. Метростанциите на Софийския метрополитен са класифицирани в две групи. Първата е групата на основните метростанции. Тя включва 8 метростанции, разположени на първия метродиаметър.Втората група включва 19 метростанции, които са разположени и на двете метролинии. Предложената класификация на метростанциите в групи би позволило прилагане на адекватни технически и технологични решения при организацията на транспорта.*