

METHODOLOGY FOR CLASSIFICATION SECTIONS OF RAILWAY NETWORK USING CLUSTER ANALYSIS

Svetla Stoilova

Technical University-Sofia, Bulgaria
stoilova@tu-sofia.bg

Abstract. Categorization of sections of the rail network is an important problem related to the development of railway infrastructure, the organization and management of train movements. In this study a methodology is elaborated to classify the sections of railway network. The methodology includes three steps. In the first step the sections and modules for research are determined. The modules are formed based on the type of railway transportation – passenger trains, fast and express trains, freight trains. The common module with all categories of trains also has been studied. The criteria for categorization, which include infrastructural and operational factors, have been defined in the second step. Nineteen criteria have been proposed in this study. The third step applies the hierarchical cluster analysis method to classify the sections for each of modules. The methodology was applied to Bulgarian railway network and 170 sections have been investigated. The cluster analysis has been used for each of the predefined modules. For each module clusters of sections have been determined. The impact of each criterion has been studied. It was found that the passenger network usage intensity (passenger·km·km⁻¹) is most important in the module related to the passenger transport; the freight network usage (ton·km·km⁻¹) is the main criterion for the freight transportation and the train network usage intensity (gross ton·km/km) is primary for the common module. The methodology can be used to analyse the state of the railway network; traffic assessment; evaluation of the significance of the section in relation to the load; defining infrastructure charges; determination of the capacity of the railway lines; planning of reconstructive events, etc.

Keywords: cluster analysis, railway transport, section, network, infrastructure.

Introduction

Organization of railway transport depends of the state of the railway infrastructure. Classification of the railway sections into groups would help make operational and strategical decisions about railway development, rehabilitation of sections or lines of the railway infrastructure, guidance of investments and management.

The subject of the study is railway sections in the railway network. Categorization of railway sections can be made on the basis of one or more indicators, such as the railway line profile, freight carried, passengers carried, load, etc. Typically, such categorizations use expert methods, where the sections are divided into groups on the basis of an increase or decrease in the indicators.

The purpose of the study is to analyse the sections of the railway network, to define the operational and infrastructure criteria for its classification and to determine different groups using multidimensional statistical analysis. Classification of the railway sections in groups would make it possible to define the main directions in the development of the railway infrastructure.

The classification of railway sections in order to rehabilitation was made in [1]. The criteria that were used are: rails, sleepers, fastenings and the ballast layer. In [2] the line accessibility classification model is elaborated for the classification of railway lines in a complex railway network. The impact of a line is assessed by pre-defined indicators using the graph theory. The model is applied to the Northeast China railway network. The railway passenger stations have been classified in [3], using the method of principal components and hierarchical cluster analysis. Four groups of factors: potential of the town, importance of the town, infrastructural factors, and characteristics of passengers have been studied to classify 98 passenger stations of the railway network in Bulgaria. The influence of criteria connected to disasters on transport infrastructure is investigated in [4;5]. Modelling of a railway station by using the GERT method is studied in [6].

The cluster analysis (K-means) in combination with Variable Neighbourhood Search (VNS) metaheuristic has been used to classify the railway tracks in [7]. Ten clusters have been found of total 75 elements by applying 14 criteria, as: length of track, train speed, number of trains, traffic load, traffic load of dangerous goods, type of traffic, axle load, integrated functional safety, age and quality, alternative route, time constrains, type of track, type of sleepers, and type of switches.

The principal component analysis and K-Means cluster analysis have been applied to classify the urban area in different French cities in [8]. Sixteen territorial clusters have been provided based on the

following criteria-population density; urban density; network density; type of housing; population incomes; land use; age of the population.

The cluster analysis with the Ward's hierarchical technique and squared Euclidean distances has been used to study passenger satisfactions in maritime transportation, [9]. The time-series data in Singapore public train transit are studied in [10] using hierarchical clustering. K-means clustering has been applied to study train delay in Danish railway line, [11]. The principal components analysis and the cluster analysis with the Ward's hierarchical method have been explored in [12] to evaluate the quality of tourist destinations in the context of customer satisfaction and loyalty. The road transport in statement centres in the Czech Republic has been studied by applying the cluster analysis, [13].

The travel behaviours in Nanjing City of China have been estimated by using the cluster analysis, [8]. The authors investigated two types of travel: the transit (bus and metro) and car. The criteria of estimate include personal information as: gender; occupation; income; car ownership, and travel information as: mode choice, walking time, waiting time, in vehicle time, fare, comfort, etc. As a result, three clusters are formed.

Most of these studies indicate that the cluster analysis may be successfully used for examining the different problems in transport.

The purpose of the study is to analyse the sections of the railway network and to perform a classification of operational and infrastructure indicators using the cluster analysis. The object of the research is 170 railway sections in the Bulgarian railway network. The methodology of the study includes the following steps: determination of the sections and modules for the research; defining the criteria of classification; application of the cluster analysis for studying railway sections and elaboration of the classification of the studied railway sections by using their pre-defined characteristics.

Materials and methods

The methodology for classification of the railway sections includes the following steps:

Step 1: Determination the railway section and modules for the research

The sections included in the study are formed between junction stations, determined by the intersection between the main railway lines, or between the main and secondary railway lines. The main railway lines are these that are included in international railway corridors.

The modules present the type of investigated transportation. The following modules are proposed:

- Passenger trains;
- Fast and express trains;
- Freight trains;
- Common module with all categories of trains.

Step 2: Choosing the criteria for assessment of the railway sections

The quantitative and qualitative criteria expressing infrastructural and operational factors of the sections are included in the study.

The infrastructural criteria present the transport potential of the sections. The study includes the following infrastructural criteria:

C_1 is the length of the section, km.

C_2 is the length of restrictive inter station, km. The restrictive inter station is characterized with maximal duration of travel of the trains. The section consists of one or more inter stations. This depends on the number of intermediate stations in the section.

C_3 is the number of stations in the section.

C_4 is the number of the main tracks in the railway line. $C_3 = 1$ for a single-track railway, $C_3 = 2$ for a double track railway.

C_5 is the coefficient of non-identity between stations,

$$C_5 = \frac{\sum_{i=1}^n t_i}{n} / t_r \quad (1)$$

where t_i – time travel for i -th inter-station in the section, min;
 t_r – time travel in restrictive inter-section, min.

C_6 is the type of railroad communications system. $C_5 = 1$ for automatic block signalling (ABS), which controls the movement of trains between the blocks using automatic signals, $C_5 = 2$ for semi-automatic block systems; $C_5 = 3$ for other systems.

C_7 is the type of traction. $C_5 = 1$ for electric traction, $C_5 = 2$ for diesel traction.

C_8 is the type of railway line. $C_7 = 1$ for the main railway line, $C_7 = 2$ for the secondary railway line, $C_7 = 3$ for the railway highway with the speed more $130 \text{ km}\cdot\text{h}^{-1}$.

The operational criteria present the transport resistance and the performance of a section for a given period. The study includes criteria related to the capacity of the sections and criteria characterizing the passenger and freight flow. The operational criteria are:

Capacity criteria

C_9 is the number of trains per day, trains/day.

C_{10} is the maximum capacity, number of trains/day.

Criteria characterizing the passenger and freight flow

C_{11} is the number of passengers carried per day, $\text{pass}\cdot\text{day}^{-1}$.

C_{12} is the number of realized passenger kilometers per day, $\text{pass}\cdot\text{km}\cdot\text{day}^{-1}$.

C_{13} is the passenger network usage intensity, $\text{pass}\cdot\text{km}\cdot\text{km}^{-1}$.

C_{14} is the number of tons carried per day, $\text{ton}\cdot\text{day}^{-1}$.

C_{15} is the number of realized ton- kilometers per day, $\text{ton}\cdot\text{km}\cdot\text{day}^{-1}$.

C_{16} is the freight network usage, $\text{ton}\cdot\text{km}\cdot\text{km}^{-1}$.

C_{17} is the number of realized train- kilometers per day, $\text{train}\cdot\text{km}\cdot\text{day}^{-1}$.

C_{18} is the number of realized gross ton- kilometers per day, $\text{gross ton}\cdot\text{km}\cdot\text{day}^{-1}$.

C_{19} is the intensity of usage of railway network by trains per day, $\text{grosston}\cdot\text{km}\cdot\text{km}^{-1}$.

Step 3: Application of the cluster analysis

The cluster analysis is a multi-measurable statistical analysis for classification of units into groups, preliminarily unknown, based on numerous characteristics in relation to these units [14]. The number of examined factors is greater than 2. The cluster analysis is a suitable method for classification of the examined sections into groups by using different criteria.

The impact of each of the applied criteria is determined by using dispersion analysis. It is used also for an approximate evaluation of the clusterization results. The determination of the statistical importance of different criteria is done by using the F criterion (Fisher's criterion).

$$F \geq F_T, \quad (2)$$

where F – empirical value of the criterion resulted from the dispersion analysis;

F_T – 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 – the number of observations (number of investigated sections);

n is the number of examined criteria (number 19).

The theoretical criterion F_T is determined by using the tables for F distribution. F is defined as a relationship between two independent values of dispersion.

$$F = \frac{S_1^2}{S_2^2}; S_1^2 = \frac{\sum_{i=1}^n (\bar{x}_i - \bar{x})^2 \cdot m_i}{n - 1}; S_2^2 = \frac{\sum_{i=1}^n \sum_{j=1}^m (x_{ij} - \bar{x}_i)^2}{m - n}, \quad (3)$$

where S_1^2 – between-group value of total dispersion;
 S_2^2 – inter-group value of total dispersion;
 x_{ij} – value of criterion j and for section i .

The Fisher criterion evaluation determines which criteria are significant for the study. It does not dismiss those other criteria, which are used for clusterization, but do not satisfy the condition (2). The F tests should be used only for descriptive purposes, because the clusters have been chosen to maximize the differences between cases in different clusters.

There are different methods of cluster analysis. 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 the Squared Euclidean distance is chosen, [14].

Results and discussion

The methodology was applied to Bulgarian railway network and 170 sections have been investigated. The following modules are investigated, taking into account the first and second steps of methodology.

- Passenger trains. This module includes 115 sections. The criteria connected to freight transport are not included in the cluster analysis.
- Fast and express trains. This module includes 54 sections. The criteria connected to freight transport are not included in the cluster analysis
- Freight trains. This module includes freight, yard and node stations. The number of surveyed sections is 170. Here, from the factors of the transport resistance are examined those, which concern the freight transport.
- All categories of trains. This module includes 87 sections between nodes. All criteria have been used for clusterization.

The data needed to conduct the study were structured in tables. The number of tables corresponds to the number of investigated modules. The rows of the tables are equal to the number of sections for the relevant module. The sections in each module were numbered. The columns correspond to the defined criteria.

In the first two modules, the sections are formed between stations that are start or final of the trains, as well as junction stations, determined by the intersection between the main or between the main and secondary railway lines. The main railway lines are the part of European transport corridors; the secondary railway lines are branches of the main railway lines.

The paper presents only a small part of the data due to their large volume in order to present their structure. Table 1 and Table 2 present an example about the type of data according to the methodology for the module of fast and express trains. The tables show the first ten data of a total of 54 elements.

Table 1

Part of data for Module: Fast and express trains (C1-C9)

| Section | C1 | C2 | C3 | C4 | C5 | C6 | C7 | C8 | C9 |
|---------|-----|-----|-----|-----|-------|-------|-------|-------|--------------------------|
| | km | km | no. | no. | coef. | coef. | coef. | coef. | trains·day ⁻¹ |
| 1 | 54 | 7.4 | 8 | 1 | 0.7 | 2 | 2 | 1 | 2 |
| 2 | 103 | 9.3 | 21 | 2 | 0.8 | 1 | 1 | 1 | 8 |
| 3 | 53 | 16 | 11 | 2 | 0.6 | 1 | 1 | 1 | 8 |
| 4 | 78 | 4.6 | 14 | 1 | 0.9 | 1 | 1 | 1 | 3 |
| 5 | 69 | 4.6 | 10 | 1 | 0.9 | 1 | 1 | 1 | 2 |
| 6 | 83 | 9 | 28 | 2 | 0.7 | 2 | 1 | 1 | 12 |
| 7 | 111 | 18 | 18 | 2 | 0.5 | 2 | 1 | 1 | 12 |
| 8 | 45 | 13 | 4 | 2 | 0.7 | 2 | 1 | 1 | 14 |
| 9 | 55 | 12 | 8 | 2 | 0.8 | 2 | 1 | 1 | 14 |
| 10 | 99 | 20 | 18 | 2 | 1 | 2 | 1 | 1 | 12 |

Table 2

Part of data for Module: Fast and express trains (C10-C19)

| Section | C10 | C11 | C12 | C13 | C17 | C18 | C19 |
|---------|--------------------------|------------------------|-----------|--------------------------|----------------------------|--------------------------------|-------------------------------|
| | trains·day ⁻¹ | pass·day ⁻¹ | trains·km | pass·km·km ⁻¹ | train·km·day ⁻¹ | gross ton·km·day ⁻¹ | gross ton·km·km ⁻¹ |
| 1 | 65 | 134 | 7236 | 2 | 108 | 33124 | 8.13 |
| 2 | 149 | 2335 | 240505 | 59 | 824 | 755351 | 185.50 |
| 3 | 149 | 1904 | 100912 | 25 | 424 | 386391 | 94.89 |
| 4 | 77 | 554 | 43212 | 11 | 234 | 121321 | 29.79 |
| 5 | 77 | 640 | 44160 | 1 | 138 | 76342 | 18.75 |
| 6 | 155 | 3645 | 302535 | 74 | 996 | 648604 | 159.28 |
| 7 | 99 | 3236 | 359196 | 88 | 1332 | 757420 | 186.01 |
| 8 | 113 | 2992 | 134640 | 33 | 630 | 332964 | 81.77 |
| 9 | 122 | 3019 | 166045 | 41 | 770 | 407105 | 99.98 |
| 10 | 96 | 2669 | 264231 | 65 | 1188 | 551123 | 135.34 |

The cluster analysis has been used for each of the defined modules. SPSS software has been applied to make the research by applying the cluster analysis, [15]. The impact of each criterion was determined by using dispersion analysis and the Fisher's criterion for each of the modules. The defined criteria have been used according to the type of trains in the module.

Table 3 presents the values of the Fisher's criterion. The last row of Table 3 presents the theoretical value of the Fisher's criterion. The values of the Fisher's criterion, which are smaller than the theoretical, present that the relevant criteria are not significant, but according to the cluster analysis methodology, they are not excluded from the survey.

Table 3

Values of F criterion

| Criterion | Modules | | | | |
|--------------------|-------------------------------|-------------------------|----------------|--------------------------|-------|
| | Passenger trains | Fast and express trains | Freight trains | All categories of trains | |
| C ₁ | km | 5.08 | 15.5 | 39.88 | 15.63 |
| C ₂ | km | 4.45 | 10.11 | 31.15 | 11.88 |
| C ₃ | number | 1.28 | 1.98 | 2.05 | 18.55 |
| C ₄ | number | 1.01 | 1.53 | 2.14 | 1.15 |
| C ₅ | coef. | 1.29 | 1.11 | 1.96 | 1.56 |
| C ₆ | coef. | 1.19 | 1.38 | 1.88 | 1.87 |
| C ₇ | coef. | 1.22 | 1.29 | 1.56 | 1.65 |
| C ₈ | coef. | 1.31 | 1.63 | 1.98 | 1.34 |
| C ₉ | trains·day ⁻¹ | 3.47 | 13.57 | 14.61 | 20.86 |
| C ₁₀ | trains·day ⁻¹ | 20.18 | 34.53 | 41.15 | 58.47 |
| C ₁₁ | pass·day ⁻¹ | 38.3 | 19.82 | - | 25.66 |
| C ₁₂ | pass·km | 138.15 | 130.24 | - | 63.12 |
| C ₁₃ | pass·km·km ⁻¹ | 138.22 | 132.66 | - | 72.51 |
| C ₁₄ | t | - | - | 21.02 | 35.76 |
| C ₁₅ | ton·km | - | - | 412.25 | 78.14 |
| C ₁₆ | ton·km·km ⁻¹ | - | - | 412.36 | 9.34 |
| C ₁₇ | trains·km | 57.85 | 61.34 | 214.75 | 57.12 |
| C ₁₈ | gross ton·km | 121.43 | 81.12 | 361.15 | 94.76 |
| C ₁₉ | gross ton·km·km ⁻¹ | 123.42 | 85.16 | 382.45 | 116.7 |
| Number of sections | | 115 | 54 | 170 | 87 |
| F _T | | 1.93 | 2.07 | 1.93 | 1.97 |

Figures 1, 2, 3 and 4 show the change of the empirical value of the Fisher's criterion for each of the investigated modules. The figures use the data given in Table 3.

Figure 1 and Figure 2 present that that the passenger network usage intensity C_{13} ($\text{pass}\cdot\text{km}\cdot\text{km}^{-1}$) and the realized passenger kilometers per day C_{12} are most important according to the Fisher's criterion in the module related to the passenger transport. Figure 3 shows that the freight network usage C_{16} ($\text{ton}\cdot\text{km}\cdot\text{km}^{-1}$) and the number of realized ton·km per day C_{15} are the main criteria according to the Fisher's criterion for the freight transportation. Figure 4 presents that the train network usage intensity C_{19} ($\text{gross ton}\cdot\text{km}\cdot\text{km}^{-1}$) and the realized gross ton·km per day C_{18} are the most important according to the Fisher's criterion for the common module.

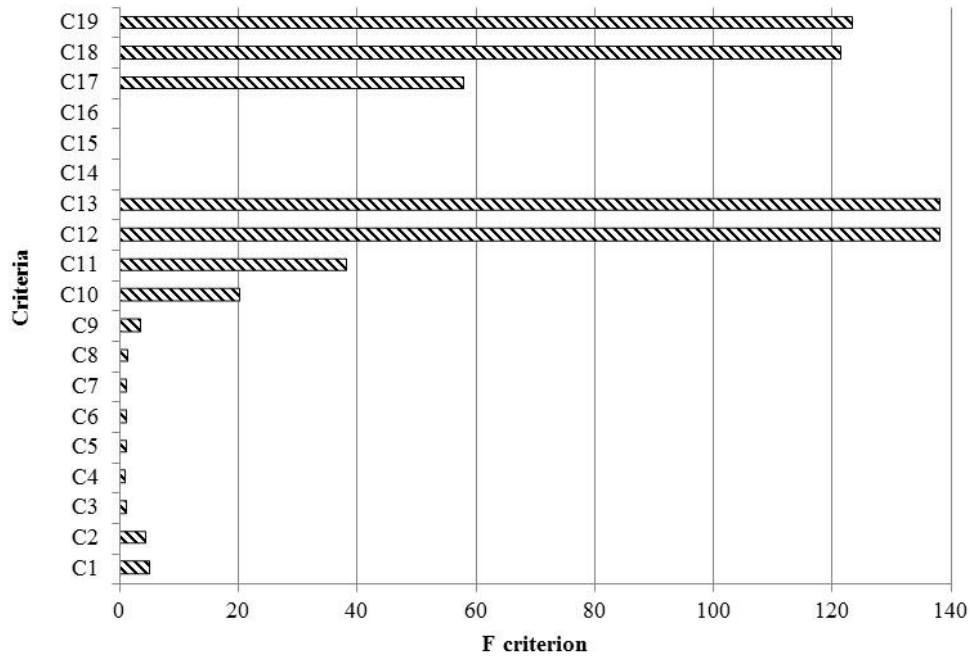


Fig.1. Module: Passenger trains

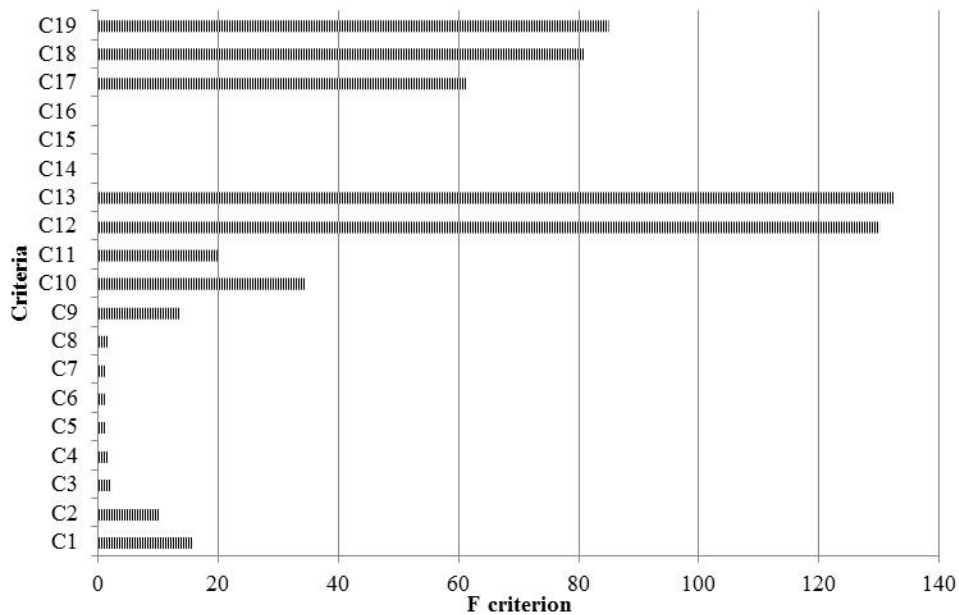


Fig.2. Module: Fast and express trains

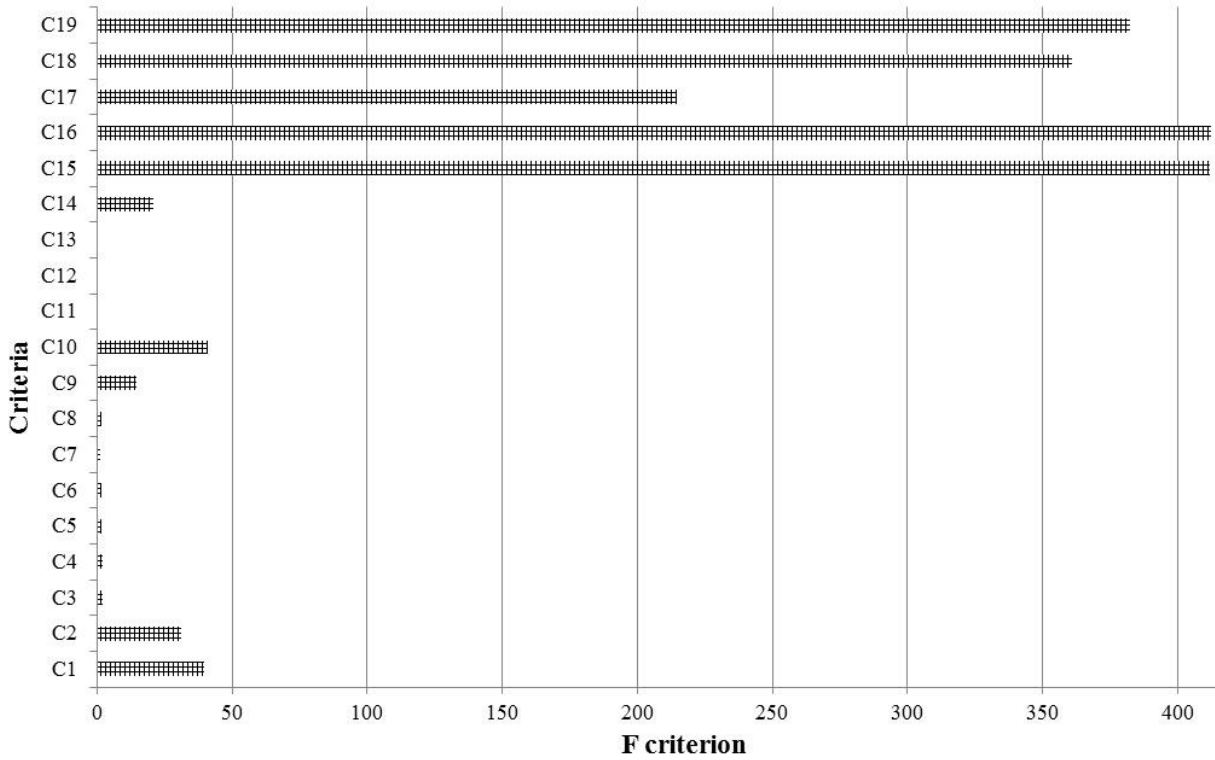


Fig.3. Module: Freight trains

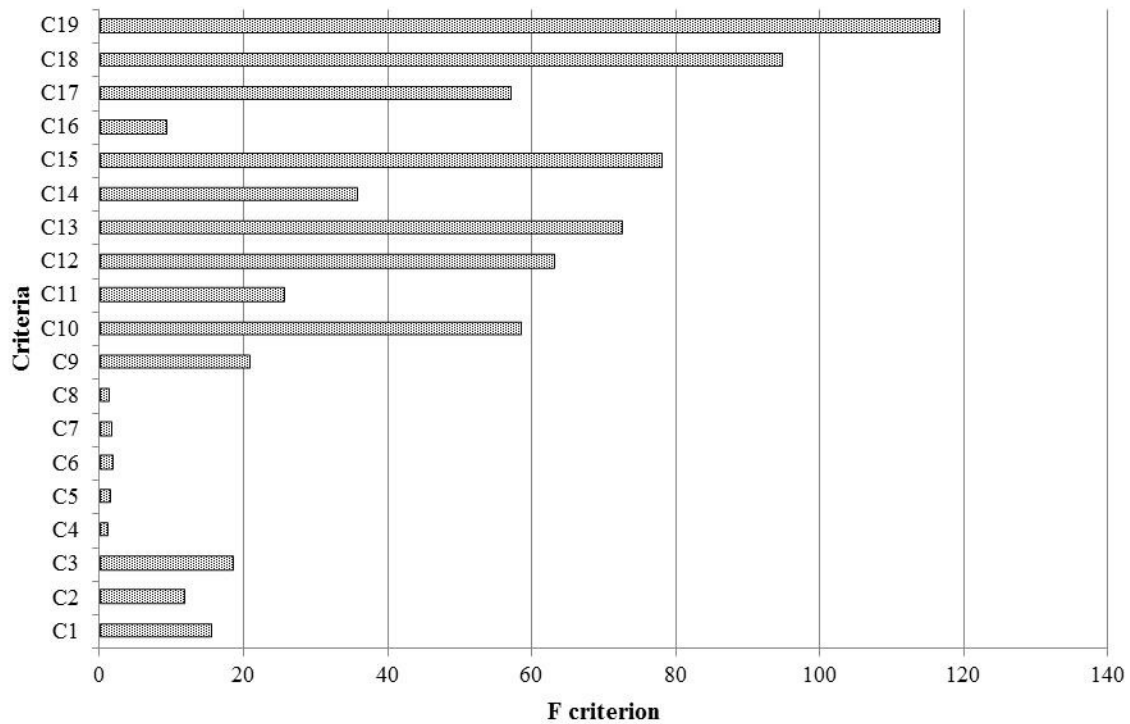


Fig.4. Module: All categories of trains

The groupings of sections according to the criteria are presented by a dendrogram. The dendrogram is a diagram that shows the hierarchical relationship between the sections. It is representing a tree. Vertically, the numbers of sections are indicated, horizontally the distances between the clusters are presented. For each module clusters of the sections have been determined.

Fig. 5 and Fig.6 present the dendrograms for module-passenger trains and module-fast and express trains. In the figures the clusters and they numbers are marked.

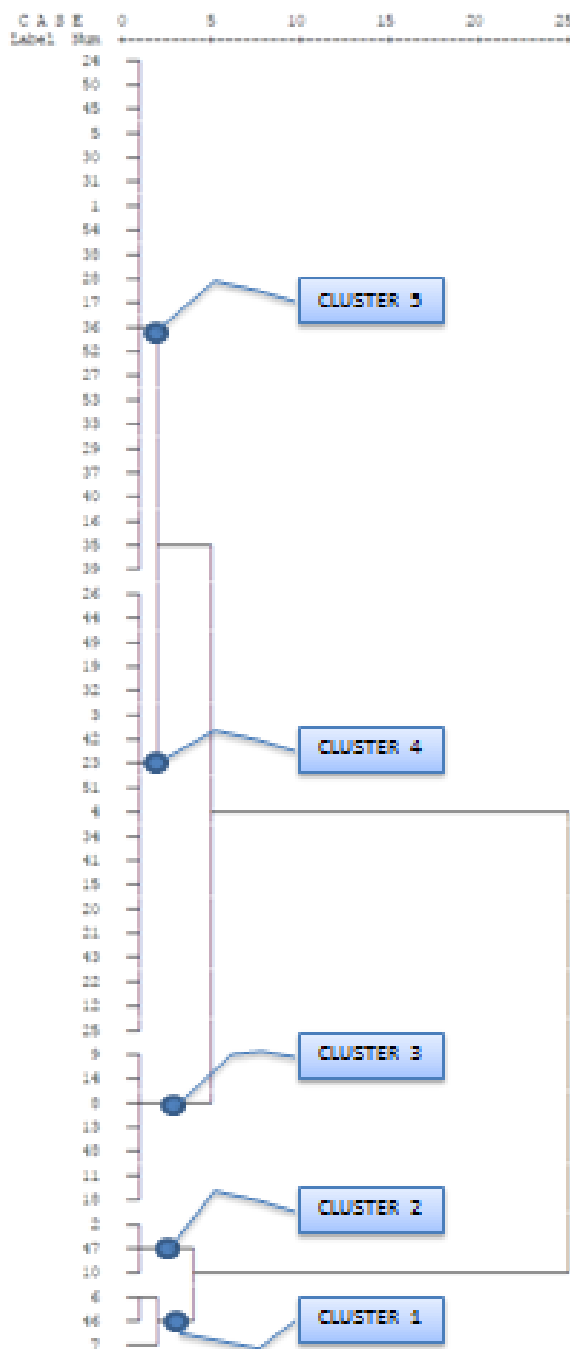


Fig. 5. Dendrogram given by SPSS Software for Module-Fast and express trains

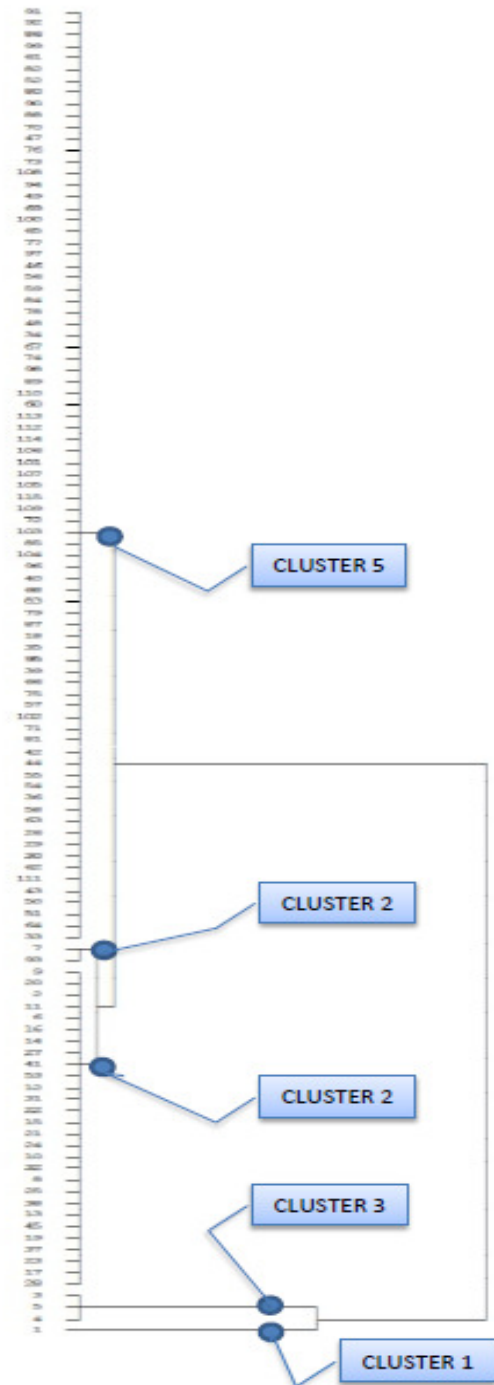


Fig. 6. Dendrogram given by SPSS Software for Module-Passenger trains

The information given in the dendrograms is summarized and presented in Tables 4-7. The first column of the tables shows the number of the cluster, the second column presents the number of elements for each cluster, the third column shows the initial and the final station of the included sections in the relevant cluster, the fourth presents the values of the main criterion for the module according to the Fisher's criterion.

Table 4 presents the number of elements and included sections for the formed clusters according to the dendrogram in Fig.5. The last column in the table shows the limits of the change of the criterion passenger network usage intensity (C13), which has highest value of Fisher's criterion (138.22).

Table 4

Module–Fast and express trains

| Cluster | Number of elements | Sections | Passenger network usage intensity, pass·km·km ⁻¹ |
|---------|--------------------|---|---|
| 1 | 3 | Sofia-Mezdra, Plovdiv-Stara Zagora, Mezdra-Pleven | upper45 |
| 2 | 3 | St.Zagora-Yambol, Sofia-Septemvri, G. Oryahovitsa-Targovishte | from 35 to 45 |
| 3 | 7 | Levski-Gorna Oryahovitsa, Yambol-Karnobat, Pleven-Levski, Sofia-Karlovo, Kaspichan-Sindel,and others | from 15 to 35 |
| 4 | 19 | September-Plovdiv, Sofia-Pernik, Karnobat-Bourgas, Rouse-G.Oryahovitsa, Plovdiv-Dimitrovgrad,and others | from 5 to 15 |
| 5 | 22 | Pernik-Radomir, Dabovo-Stara Zagora, St.Zagora-Dimitrovgrad, Dimitrovgrad-Svilengrad and others | below 5 |

Table 5

Module–Passenger trains

| Cluster | Number of elements | Sections | Passenger network usage intensity, pass·km·km ⁻¹ |
|---------|--------------------|---|---|
| 1 | 1 | Sofia-Kostenets | upper 30 |
| 2 | 3 | Sofia-Pernik, Karnobat-Bourgas, Sofia-Svoqe | from 20 to 30 |
| 3 | 2 | Plovdiv-Asenovgrad, Plovdiv-Dolnamahala | from 15 to 20 |
| 4 | 28 | Plovdiv-Chirpan, Septemvri-Stamboliyski, St. Zagora-Yambol, Pleven-Levski, Levski-G.Oryahovitsaand others | from 5 to 15 |
| 5 | 77 | Dimitrovgrad-Simeonovgrad, Sliven-Zimnitsa, D.Mahala-Hissar, Tulovo-S.Zagoraand others | below 5 |

Table 6

Module–Freight trains

| Cluster | Number of elements | Sections | Freight network usage, ton·km·km ⁻¹ |
|---------|--------------------|--|--|
| 1 | 2 | Dabovo - Sliven, Karnobat - Burgas | upper 400 |
| 2 | 6 | Vakarel-Septemvri; Simeonovgrad-Svilengrad; Zimnitsa-Karnobat, Karlovo-Pirdop | from 250 to 400 |
| 3 | 6 | Parvomai-Dimitrovgrad,Dimitrovgrad-Simeonovgrad,Dupnitsa-G.Todorov, Ilentsi-Mezdra, Krumovo-Parvomay | from 150 to 250 |
| 4 | 16 | Sliven-Zimnitsa, Kalotina-Dragoman, Ivanovo-Byala, Stamboliyski-Plovdivand others. | from 90 to 150 |
| 5 | 140 | Septemvri-Pazardjik, Byala-G.Oryahovitsa, Samuil-Kaspichan, St.Zagora-N.Zagoraand others | below 90 |

Table 5 presents the formed clusters for the module-passenger trains according to Fig. 6. The main criterion for this module is also the passenger network usage intensity (C13), which has the highest value of Fisher’s criterion (132.66). Table 6 shows the clusters for the module–Freight trains. The main criterion for this cluster is the freight network usage (C16), which has highest value of Fisher’s criterion (412.36). Table 7 presents the clusters for the module, which includes all categories of trains. The main criterion for this cluster is the intensity of usage of railway network by trains per day (C19), which has highest value of Fisher’s criterion (116.7).

Table 7

Module – All categories of trains

| Cluster | Number of elements | Sections | Intensity of usage of railway network by trains per day, gross ton-km/km |
|---------|--------------------|---|--|
| 1 | 1 | Dabovo - G. Oryahovitsa | upper800 |
| 2 | 2 | G. Oryahovitsa - Shumen, Stolnik - Karlovo | from 500to800 |
| 3 | 5 | Mezdra-Brusartsi, Dulovo-Zimnitsa, Mezdra-Yasen, Ilienci-Mezdra, Yassen-Levski | from 300to 500 |
| 4 | 16 | Karlovo-Tulovo, Rouse-Gorna Oryahovitsa, Zimnitsa-Karnobat, Krumovo-Dimitrovgrad, Karnobat-Komunariand others | from 150to300 |
| 5 | 63 | Shumen-Kaspichan, Tulovo-Dubovo, Kyustendil-Radomir, Kaspichan-Sindeland others | below 150 |

Figure 7 presents the number of clusters and the number of sections in each cluster given by the cluster analysis for the investigated modules.

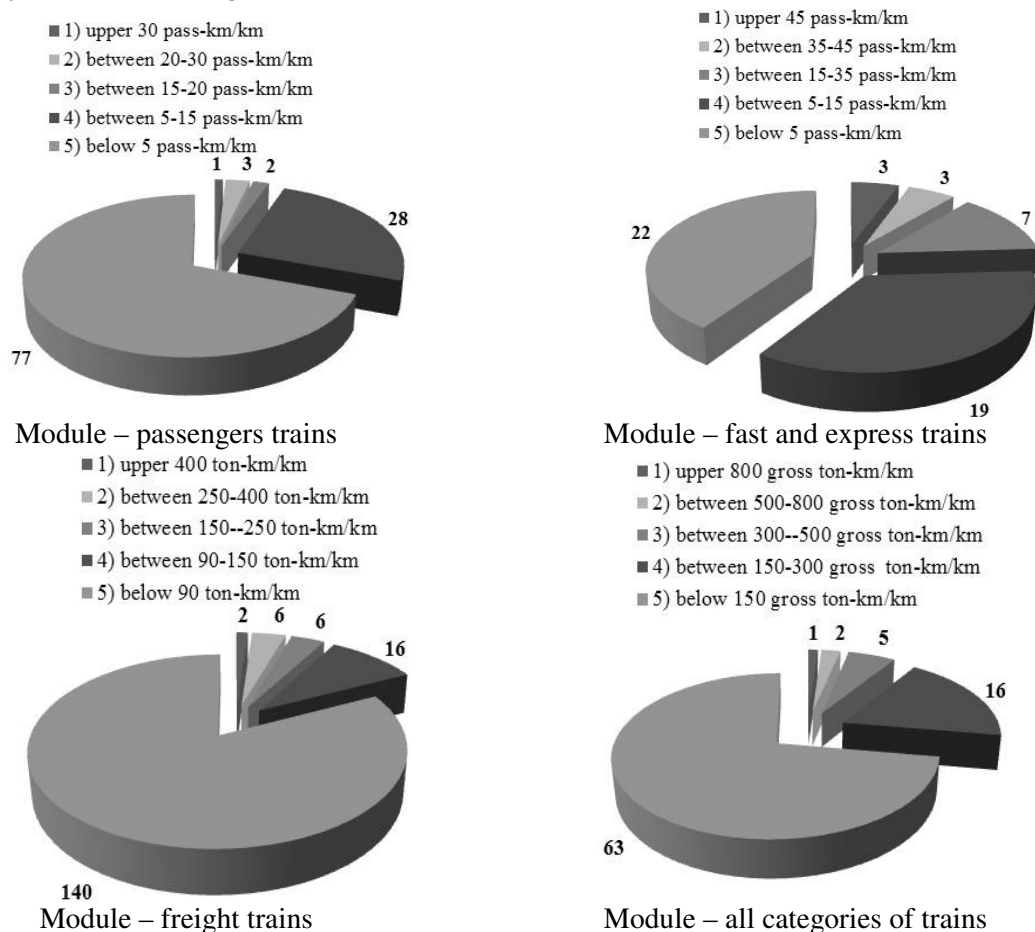


Fig.7. Number of clusters for modules

The information given in Tables 4-7 is used to present the diagrams given in Table 7. The number of clusters is determined according to the results in the dendrograms. The formed clusters correspond to the change of the values of the main criterion (these that have the main importance according to the Fisher's criterion).

Conclusions

1. In this research a methodology has been elaborated based on application of the cluster analysis to classify the sections in the railway network.
2. The criteria for categorization, which include factors of transport potential and transport resistance, have been defined. Nineteen criteria to assess the railway sections have been determined.
3. In total 170 sections of Bulgarian railway network have been studied. Four modules according to the type of carriage have been defined. The module "Passenger trains" includes 115 sections; the module "Fast and express trains" includes 54 sections. The module "Freight trains" includes 170 sections. The module "All categories of trains" includes 87 sections. The clusters have been determined for each of the studied modules
4. The operational criteria connected to transport resistance, which present the performance of the sections, are the most important for the classification. The most important factors are – the passenger network usage intensity, freight network usage and the intensity of usage of railway network by trains.
5. The most important factors for the categorization of the sections are determined by using dispersion analysis and the Fisher's criterion. The passenger network usage intensity with the value of the Fisher's criterion 138.22 is the most important for the module – passenger trains. The main criterion for the module fast and express trains is also the passenger network usage intensity, which has the highest value of Fisher's criterion 132.66. The freight network usage has highest value of Fisher's criterion 412.36 for the module– freight trains, and it is the most important criterion. The train network usage intensity is the most important according to the value of Fisher's criterion 116.7 for the module of all categories of trains.
6. The clusters for the investigated modules have been determined. Five clusters in each module have been determined using the hierarchical cluster analysis. The number of elements of the sections in the clusters for the investigated modules is different.
7. The classification of the sections could be used to make analysis of the state of the railway network; for forecasting of passenger and freight flows; for traffic assessment; for assessing the significance of the section in terms of load; to define the infrastructure charges; to determine the capacity of the railway lines; for planning of reconstructive events, and for other purposes.

Acknowledgements

This research has been supported by the contract No.KII-06-H27/12 of 11.12.2018 "Modelling and elaboration of complex system for selection of transport technology in transport network" funded by the National Science Fund of the Ministry of Education and Science of Bulgaria.

References

- [1] Shtompel A., Trykoz L., Borodin D., Ismagilov A., Chmuzh Y. Probabilistic evaluation of the railway track infrastructure components failure risk. MATEC Web of Conferences, vol. 230, [online] [01.03.2019]. Available at: <https://doi.org/10.1051/mateconf/201823001017>.
- [2] Yang M., Wang Y. J. Modelling the Accessibility Classification of Railway Lines: A Case Study of Northeast China Railway Network. *Promet – Traffic&Transportation*, vol. 25, no. 5, 2013, pp. 467-474
- [3] Stoilova S., Nikolova R. Classifying railway passenger stations for use transport planning - Application to Bulgarian railway network. *Transport Problems*, vol.11(2), 2016, pp.143-155, doi: 10.20858/tp.2016.11.2.14
- [4] Dimitrov D., Zlateva P. Information Model for the Registration of Disasters Affecting the Road Infrastructure, World Science, RS Global, № 1(29), vol.1, 2018, [online] [01.03.2019]. Available at: <http://archive.ws-conference.com/wp-content/uploads/2072.pdf>

- [5] Dimitrov D., Zlateva P., Velev D. A Methodology for Designing an Information System for Road Infrastructure Monitoring and Traffic Management in Disasters, 2018, IOP Conference Series: Earth and Environmental Science, vol.167, conference 1, [online] [01.03.2019]. Available at: <https://iopscience.iop.org/article/10.1088/1755-1315/167/1/012044>.
- [6] Todorova M. GERT model of a technical station for determine the passenger trainset delay. *Transport Problems*, vol.14, Issue 1, 2019, pp. 81-94, [online] [01.03.2019]. Available at: <http://transportproblems.polsl.pl/en/default.aspx>
- [7] Dorić D., El Cadi A A., Hanafi S., Mladenović N., Artiba A. Clustering approach in maintenance of capillary railway network. *Electronic Notes in Discrete Mathematics*, vol. 58, 2017, pp. 239-246, <http://dx.doi.org/10.1016/j.endm.2017.03.031>
- [8] Ducret R., Lemarié B., Roset A. Cluster analysis and spatial modeling for urban freight. Identifying homogeneous urban zones based on urban form and logistics characteristics. *Transportation Research Procedia*, vol. 12, 2016, pp. 301-313
- [9] Ekinçi Y., Uray N, Ülengin F., Duran C. A segmentation based analysis for measuring customer satisfaction in maritime transportation. *TRANSPORT*, vol.33(1),2018, pp. 104-118, doi:10.3846/16484142.2015.1079800
- [10] Roy Ka-Wei Lee, Tin Seong Kam. Time-Series Data Mining in Transportation: A Case Study on Singapore Public Train Commuter Travel Patterns. *IACSIT International Journal of Engineering and Technology*, vol. 6, no. 5, 2014
- [11] Cerreto F., Nielsen B F., Nielsen O A., Harrod S S. Application of Data Clustering to Railway Delay Pattern Recognition, *Journal of Advanced Transportation*, vol. 2018, Article ID 6164534, 2018, 18 p. [online] [01.03.2019]. Available at: <https://doi.org/10.1155/2018/6164534>
- [12] Vajčnerová I, Šácha J., Ryglová K., Žižan P. Using the Cluster analysis and the Principal component analysis in evaluating the quality of a destination. *Acta Universitatis agriculturae et silviculturae mendelianae brunensis*, vol.6479, no.2, 2016, [online] [01.03.2019]. Available at: <http://dx.doi.org/10.11118/actaun201664020677>
- [13] Kraft S. A transport classification of settlement centers in the Czech republic using Cluster analysis. *Moravian geographical reports*, vol. 203, 2012, 12 p.
- [14] Tan P., Steinbach M., Kumar U. *Introduction to Data mining*. Minnesota: University of Minnesota. 2006. 769 p.
- [15] SPSS Software. [online] [01.03.2019]. Available at: <https://spss.en.softonic.com>