***E. V. Sinitsyn*** [a)], ***A. V. Tolmachev*** [a, b)], ***D. A. Brusyanin*** [c)]
[a)] Ural Federal University (Ekaterinburg, Russian Federation; e-mail: sinitsyn_ev@mail.ru)
[b)] Datatel-Ural LLC (Ekaterinburg, Russian Federation)
[c)] Ministry of Transport of the Sverdlovsk Region (Ekaterinburg, Russian Federation)

# MATHEMATICAL MODELLING OF REGIONAL CARGO AND PASSENGER FLOWS [1]

*The creation and implementation of the strategies for economic and social development in the Russian regions for the period up to 2035 implies an adequate development of transport services affecting all economic sectors and segments of the population. In this regard, we propose a model connecting the characteristics of passenger and cargo flows with the parameters of economic and social development, as well as with the region's demography. This model allows specifying the congestion of the transport system resulting from the implementation of plans for social and economic development and planned decisions in the sphere of economic activity. For developing the model, we selected parameters describing the economic situation, labour market, demography, living standards and social situation in the analysed subject. These parameters have the highest correlation coefficients with the analysed characteristics of the transport infrastructure. Further, we conducted a step-by-step regression analysis, adding to the already existing variables new ones that gave the greatest increase in the determinacy coefficient $R^2$. The model shows that the main factor determining the amount of passengers transported by public buses is the annual average number of employed persons. The passenger turnover is mostly affected by the population size. The volume of goods transported by trucks is determined by parameters characterising the level of the production development (investments in fixed assets, fixed capital in the economy, and the volume of shipped goods of domestic production). The use of nonlinear models and networks did not significantly reduce the model's errors. Additionally, we clustered the Russian regions by indicators of socio-economic development and the characteristics of transport infrastructure affecting traffic flows. Then we assessed the efficiency of transport infrastructure's exploitation in various clusters. This allows the targeted benchmarking, namely the selection of regions mostly appropriate for comparison with the analysed one.*

**Keywords:** passenger and cargo flows, passenger turnover, cargo turnover, socio-economic development, correlation coefficients, multidimensional regression, determinacy coefficients, data mining, clustering, Kohonen self-organizing map, *k*-means method, hierarchical structure of clusters

## 1. Introduction

Currently, the processes of defining strategic development directions for various industries for the period up 2035 are being complete. That fact, in particular, created a need for connecting strategic targets in different areas and a task of forecasting passenger and cargo flows. They aim to ensure the connection of transport infrastructure, public transportation routes, and rolling stock with the defined and targeted values of socio-economic development. Moreover, it is necessary to consider social impact of transport services that affect almost all industries, and a whole population [1], due to the fact that problems related to passenger and cargo haulage can become a source of social strain and macroeconomic problems.

It is necessary to move from simple tracking, used to plan logical relations among the target values of a region's social and economic development with transport operation parameters, towards applying more robust and statistically defined dependencies. Without claiming to resolve all the problems that occur along this way, we present a relatively simple model that allows connecting economy, demographics and other regional factors. It allows forecasting the results of transport operations on the base of socio-economic forecasts and transport infrastructure's development plans. We used similar models in course of defining the transport development strategy for Sverdlovsk Oblast.

Mathematical modelling of transport systems traditionally attracts interest of the researchers. For example, [2] used generalizing mathematical models to forecast traffic jams and to select optimal topology of transport networks and approaches to modelling transport systems based on the graph theory. Approaches to solving problems of forecasting short-term traffic flows for existing road networks based on regression methods were described in [3–5]. Box-Jenkins method was reviewed in



Translation

[6], and its analogues for time series were described in [7–9]. Also, for forecasting short-term flows in transport systems a number of works used the methods of intellectual data analysis, namely, support vectors method [10], and neural networks [11–14].

The aforementioned works were mostly concerned with solving problems of forecasting short-term traffic flow in the context of a given transport infrastructure, for example, a specific road and street network. That is unquestionably interesting from the point of mathematical modelling; however, the above-mentioned approaches are inconvenient for solving the problems of long-term development planning of transport systems. Obtainment of the necessary results for passenger and cargo flows at regional level is quite complicated both in terms of collecting initial data and running computational procedures.

In this regard, similarly to our previous paper [15], we suggest using an explicit and relatively simple model based on multivariate regression for connecting the characteristics of passenger and cargo flows with a region's economic and other characteristics (reflected in the programmes of socio-economic development) [16–19]. This model is described in 4.1. Moreover, we provide the model's extension for considering possible non-linear dependencies of the included parameters.

In the most general problem statement, it is necessary to examine various types of transport, including public, private, land, air, river transports, etc. However, in such case, the adequate model's design would require collection and processing of data that is not available in full. Further, organisation of these data's collection, ensuring its completeness and integrity, is an independent and complicated problem. So, the scope of this paper was limited by reviewing passenger and cargo flows related to motor transport.

Another task worth mentioning in relation to the long-term planning of transport development is benchmarking. That is a search for benchmarks or regions that are in some way similar to the one in question, so comparison of the target region with benchmarks can be used as a base for making management decisions. As an example, when analysing development parameters of transport services, Sverdlovsk oblast is usually compared with Tyumen and Chelyabinsk Oblast, Perm region, and the Republic of Tatarstan. Obviously, selection of benchmarks has a risk of subjectivity. Thus, it is necessary to review the possibility of such clustering and choosing the regions suitable for benchmarking by using intellectual methods of data analysis [19–25]. That procedure is done in 4.2.

Thereby, the paper aims to design mathematical models for connecting the characteristics of passenger and cargo flows related to motor vehicles with the parameters that characterise a region. As a source of data, we used the information from Federal State Statistics Service (Rosstat) on Russian regions[2] for the end of 2017.

## 2. Analysis methods

The considered models were based on using statistical analysis methods:
— Correlation analysis and multivariate regression [16–19],
— Intellectual data analysis [20–25], that includes:
— Clustering using k-means method [19, 23];
— Clustering using Kohonen self-organizing maps [25];
— Neural networks technology [20–22].

## 3. Results

### *3.1. General information and selecting parameters for analysis*

We considered the following parameters characterising passenger and cargo flows:

— Cargo haulage, mln tons;
— Cargo turnover, mln tons/km;
— Passengers carried by public buses, mln;
— Passenger turnover for public buses, mln passengers/km. (1)

[2] Regiony Rossii. Sotsialno-ekonomicheskie pokazateli [Regions of Russia. Social and Economic Parameters]. (2018). R32 Statistical Digest. Moscow, Rosstat, 1162. Retrieved from: http://www.gks.ru/wps/wcm/connect/rosstat_main/rosstat/ru/statistics/publications/catalog/doc_1138623506156 (Date of access: 01.08.2019).

At the first step, from the list of parameters presented on the site of Federal State Statistics Service, we removed the ones that have no logical connection with the above-mentioned parameters of passenger and cargo flows (1). Correlation coefficients were calculated for the remaining parameters describing regions and characteristics of passenger and cargo flows (1) for all federal entities of the Russian Federation. Then, we discarded the parameters having zero correlation coefficients with parameters from (1) for a significance level of 5 % [26]. As a result, a list of parameters for each entity, displayed in Table 1, was selected for further analysis.

Table 1

**Parameters of the constituent entities of the Russian Federation used to develop mathematical models**

| Item | Parameter | Measurement units |
|---|---|---|
| 1 | Railroads density for the end of the year | Kilometres of railroad tracks per 10,000 sq. km |
| 2 | Density of hard-top public motorways for the end of the year | Kilometres of roads per 10,000 sq. km |
| 3 | Number of public buses per 100,000 population, end of the year | |
| 4 | Number of motor vehicle accidents per 100,000 population | |
| 5 | Area | Thousand sq. km |
| 6 | Population, as of 01.01.2018 | Thousands |
| 7 | Total number of municipal entities in a region | |
| 8 | Gross regional product for a federal entity in current basic prices* | mln roubles |
| 9 | Capital investments | mln roubles |
| 10 | Capital assets in economy (gross book value at the end of the year) | mln roubles |
| 11 | Amount of delivered goods, locally produced, work performed, and services provided locally | mln roubles |
| 12 | Gross agricultural output | mln roubles |
| 13 | Commissioned residential buildings | Thousands sq. meters of total residential area |
| 14 | Retail turnover | mln roubles |
| 15 | Balanced financial result (profit minus losses) of entity operations | mln roubles |
| 16 | Average annual number of employed | thousands |
| 17 | Average monthly income per capita | roubles |
| 18 | Average monthly consumer expenses per capita | roubles |
| 19 | Number of private cars | Cars per 1000 population |
| 20 | Total number of students (bachelor, specialist and master programs) | thousands |
| 21 | Number of pensioners | Per 1000 population |
| 22 | Number of medical professionals (all specialties) | Per 10,000 population |

* For the end of 2016.

These parameters can be divided into five groups:

1. Transport infrastructure characteristics (1–4).
2. Overall characteristics of a federal entity (5–7).
3. Macroeconomic parameters (8–15).
4. Parameters describing population’s quality of life (16–18).
5. Other social parameters that can affect passenger flows (19–22).

Density of railroads was included in the analysis due to the fact that goods are often delivered to final user using motor vehicles. Therefore, potential development of railroad infrastructure can cause increase in motor transport’s cargo turnover. Correlation coefficient for these two parameters is positive, though quite small (0.15).

It can be noted that the Parameter 3 from Table 1 does not have significant correlation with passenger turnover parameters, which raises questions about efficiency of using existing bus stock.

### *3.2. Transport infrastructure use efficiency*

Maintenance and improvement of transport infrastructure's elements requires significant capital and operational expenses that are partially covered from the budget of different levels. Therefore, efficiency of using the infrastructure elements becomes important. For example, in order to access road network's efficient use for different federal entities, it is possible to use the following characteristic:

$$I = \frac{Q}{L}. \tag{2}$$

where *I* is the use intensity (load) of public motorways; *Q* is the average amount of transported cargo per 1 kilometre of roads; *L* is the length of roads. Parameter (2) and comparison of its values for different federal entities can be helpful while defining amounts of financing allocated to regions through federal programs for the purpose of road maintenance.

For analysing passenger carriage, it makes sense to start from assessing transport mobility of regional population for different federal entities. In order to do that, we used a parameter demonstrating a number of passenger-kilometres per capita:

$$M = \frac{PT}{N_P}, \tag{3}$$

where *PT* is a passenger turnover (passenger-kilometres), and $N_P$ is population of a federal entity. Figure 1 shows distribution of *M* value for the entities of the Russian Federation.

One can see that distribution presented in Figure 1 is relatively symmetrical, which allows speculation that transport mobility is substantially affected by random causes.

The efficiency of operating buses can be characterised by the average number of passengers (q) transported by a single bus in a year:

$$q = \frac{Q_p}{N_B}, \tag{4}$$

where $Q_p$ is a total number of passengers transported by buses within a federal entity, and $N_B$ is a number of buses. It is also possible to assess an average annual mileage per bus:

$$l_B = \frac{PT}{q}. \tag{5}$$

Like in (3), *PT* is a passenger turnover, and *q* is derived from (4).

Distribution of $l_B$ for the entities of the Russian Federation is shown in Figure 2.

Average number of passengers transported annually by a single bus is shown in Figure 3.

Of course, extensive use of buses (large amounts of transported passengers, and large annual mileage) causes their excessive wear. So, it can be recommended to use data presented in Figures 2 and 3 for planning allocation of funds from various budgets for stock renovation and modernization.

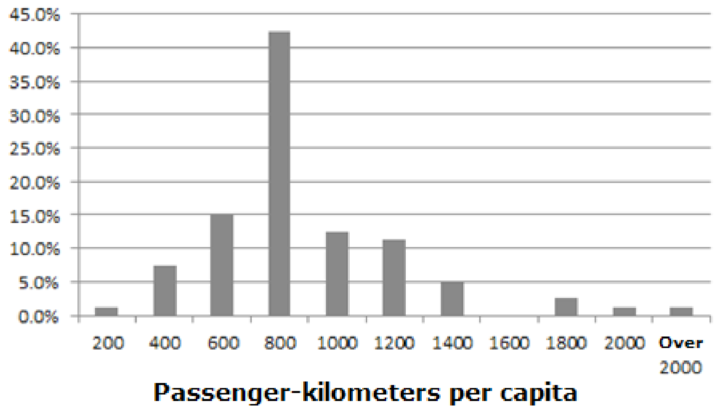



**Fig. 1.** *Distribution of the entities of the Russian Federation by mobility of population (3)*

Translation

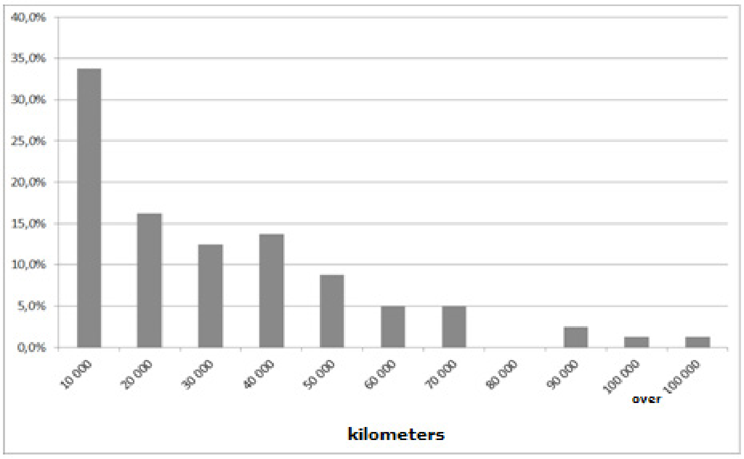



**Fig. 2.** *Distribution of average annual mileage for a bus (5) in the entities of the Russian Federation*

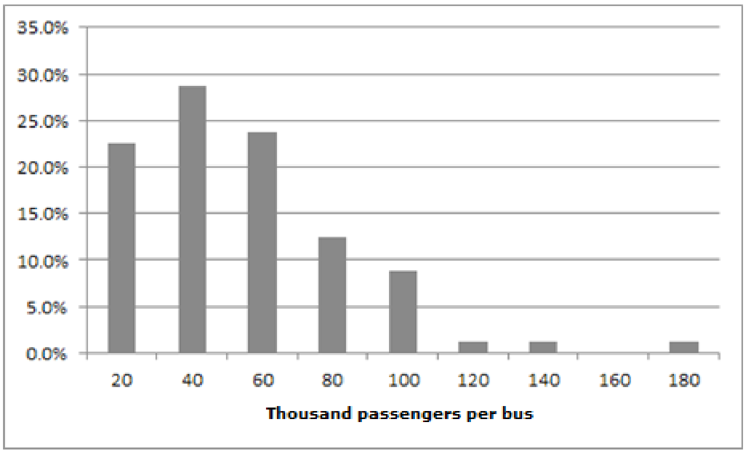



**Fig. 3.** *Average number of passengers transported annually by a single bus*

## 4. Mathematical modelling of passenger and cargo flows. Discussion of the results

The previous section demonstrates that the key parameters that characterise efficiency of using transport infrastructure can be calculated based on key characteristics of transport flows listed in (1). This section covers possible mathematical models to be used for calculations.

### *4.1. Multivariate regression*

For determining dependencies of characteristics listed in (1) we used stepwise multivariate linear regression. At the first step we used a parameter from Table 1, which has the largest correlation coefficient with a characteristic of interest, as an initial variable for calculating the said characteristic. At the second step we sequentially considered remaining parameters from Table 1 and selected a parameter providing the largest increase of determination factor and decrease of residual variance [16–19]. The procedure is then repeated for the remaining parameters from Table 1 while the increase of determination factor after adding new parameter remains significant. The obtained results are included in Table 2.

The data from Table 2 can be used to assess characteristics of transport flows that can be used, for example, in strategic planning processes where needs for passenger and cargo haulage must correlate with the target parameters of social and economic development, demographic forecasts, etc. Table 2 demonstrates that determination factors for all models are quite high. It is necessary to note that probability distributions for residual values (differences between the actual values and values determined using models from Table 2) with significance level equal to 5 % are subject to normal distribution. Figure 4 shows distribution of residual values for the number of passengers carried in different regions.

Table 2

**Results of developing multivariate regression models for the characteristics of passenger and cargo flows (1)**

| Parameters | Coefficients for independent variables in regression equations | | | | | | | | | Constant |
|---|---|---|---|---|---|---|---|---|---|---|
| | Railroads density | Density of hard-top public motorways | Commissioned residential buildings | Capital investments | Capital assets in economy | Amount of delivered goods, locally produced | Average annual number of employed | Number of motor vehicle accidents per 100,000 population | Population | |
| 1. Passengers carried by public buses | | | | | −1.08E−05 | 0.171 | 2.94E−05 | | | −8.98 |
| Determination coefficient $R^2$ | | | | | 0.933 | 0.930 | 0.925 | | | |
| $F$ -statistic | | | | | 511.5 | 355.0 | 964.7 | | | |
| Ratio of a critical $F$-statistic value for significance level equal to 1 % to the actual value of $F$-statistic | | | | | 0.011 | 0.010 | 0.007 | | | |
| 2. Passenger turnover for public buses | | | | | | 0.00020 | −2.14 | 5.23 | 1.88 | −796.7 |
| Determination coefficient $R^2$ | | | | | | 0.76 | 0.74 | 0.75 | 0.65 | |
| $F$-statistic | | | | | | 60.3 | 110.5 | 77,3 | 150.5 | |
| Ratio of a critical $F$-statistic value for significance level equal to 1 % to the actual value of $F$-statistic | | | | | | 0.059 | 0.044 | 0.052 | 0.046 | |
| 3. Cargo haulage | −0.026 | | | 0.00011 | −1.04 E−05 | 3.26E−05 | | | | |
| Determination coefficient $R^2$ | 0.764 | | | 0.531 | 0.651 | 0.705 | | | | |
| $F$-statistic | 60.6 | | | 88.3 | 71.8 | 60.6 | | | | |
| Ratio of a critical $F$-statistic value for significance level equal to 1 % to the actual value of $F$-statistic | 0.059 | | | 0.079 | 0.068 | 0.067 | | | | |
| 4. Cargo turnover | | −2.02 | 2.07 | | | 0.00044 | | | | |
| Determination coefficient $R^2$ | | 0.792 | 0.738 | | | 0.817 | | | | |
| $F$-statistic | | 146.2 | 219.6 | | | 112.9 | | | | |
| Ratio of a critical $F$-statistic value for significance level equal to 1 % to the actual value of $F$-statistic | | 0.033 | 0.031 | | | 0.036 | | | | |

Note: increase of $R^2$ coefficient value corresponds with the sequence of adding independent variables in stepwise multivariate regression.

## Translation

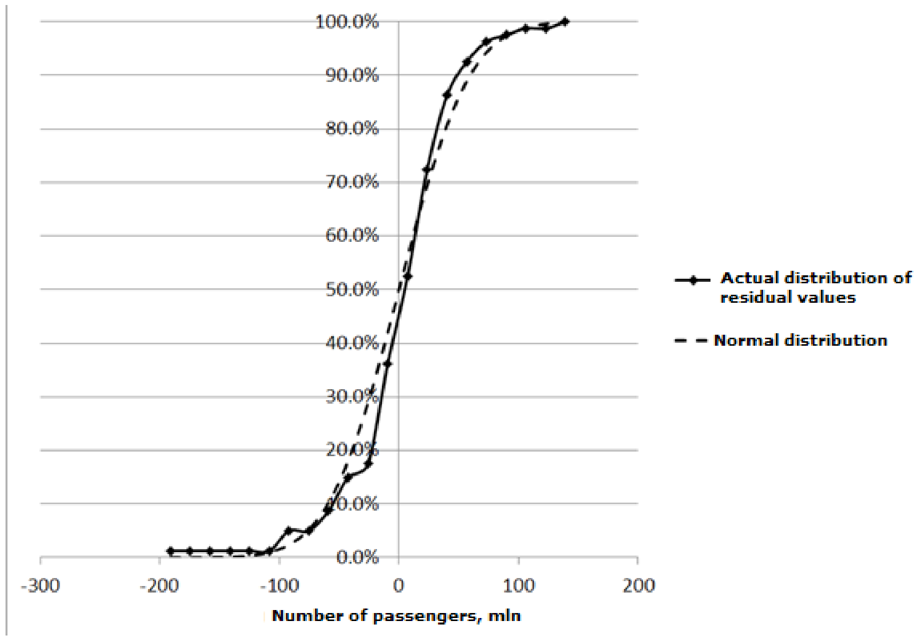



**Fig. 4.** *Integral distribution function for residual values in the model of passenger carriage from Table 2. Normal distribution function with mean value equal to zero and variation corresponding with actual value are shown for comparison purposes*

Therefore, the models presented in Table 2 can be used to assess characteristics of transport flows, including ones used in long-term strategic planning. In this case, the parameters specified in the regional socio-economic development strategy allow determining the expected values of cargo and passenger flows and planning the transport system's development correspondingly.

At the same time, a substantial value of random components in the parameters of passenger and cargo haulage (Figure 4) demonstrates presence of random processes affecting transport flows that are not considered by a model. Additional sampling correlation analysis of random residual values with the number of the Russian regions' parameters[3] (not included in Table 1) did not reveal significant correlations. We assumed that the reasons for the processes that cause appearance of random components in parameters of cargo and passenger haulages are outside of the economy and demographics sphere. Unfortunately, at the moment, there is not enough data to verify this statement.

It can be suggested that quality of the models can be improved by considering the non-linear dependencies of the traffic flows' parameters from the characteristics listed in Table 2. For assessing possible presence of heteroscedasticity [18] we created the logarithmic model (linear dependency of a logarithm of an examined value from the logarithms of independent variables). The determination coefficient for logarithmic model was usually smaller than the one of the corresponding models mentioned in Table 2; distribution of residual value remained almost the same.

Neural networks technologies can be also applied for non-linear modelling of parameters of passenger and cargo flows. According to A well-known Kolmogorov-Arnold theorem [27, 28], if there is a functional relationship among the calculated parameter and model variables, it is possible to build an approximation of a modelled value using the multilayer perceptron [20–22]. Perceptrons were built using Deductor Studio software by Base Group, academic version 5.3 0.88.

Architecture of the used multilayer perceptrons is not included here due to complexity of the patterns. We used the independent variables from Table 2 as input variables for the perceptrons. Determination coefficients for neural network models and distribution functions for random residual values did not improve significantly when compared with models from Table 2.

Thus, the results demonstrate possibility of assessing the characteristics of passenger and cargo flows based on limited number of parameters that describe the regional transport infrastructure, economy and demographics by means of applying simple linear models presented in Table 2.

[3] Regiony Rossii. Sotsialno-ekonomicheskie pokazateli [Regions of Russia. Social and Economic Parameters]. (2018). R32 Statistical Digest. Moscow, Rosstat, 1162. Retrieved from: http://www.gks.ru/wps/wcm/connect/rosstat_main/rosstat/ru/statistics/publications/catalog/doc_1138623506156 (Date of access: 01.08.2019).

### *4.2. Clustering and benchmarking*

As it was noted in the introduction, a common practice for studying region strategic goals is to examine experience of its analogues in order to perform comparison and make required management decisions. Basically, it is a classification problem, a search for the entities of the Russian Federation that are in some aspects close to the target ones. Due to the fact that tasks of transport development arise out of the socio-economic goals, and the corresponding parameters, as it was demonstrated above, are defined by economic parameters and demographics, it is possible to classify the entities of the Russian Federation using parameters from Table 2 as classification variables. Of course, such classification can reveal regions that are "similar" in terms of transport systems as it is not absolute, so it makes no sense to generalise it for the whole spectrum of socio-economic characteristics.

We used the k-means algorithm [19, 23, 24] for solving the classification problem.

Figure 5 presents structure of clusters, uniting federal entities in such way that elements of each group are closer to the elements of the same group than to the elements from other groups. In this case, proximity measure is the distance calculated in the space of all independent variables that describe an entity (Table 2). Each cluster can be described using the so-called centroid—"centre of masses" for a cluster defined by a multidimensional vector $\overline{X_C} = \{X_{1C}, X_{2C}, \ldots, X_{nC}\}$, where

$$X_{iC} = \sum_{j=1}^{M} X_i(j). \tag{6}$$

$X_i$ is the independent variable from Table 2, $j$ is a number of an entity ($M$ is a number of entities), $i$ is a number of a variable ($n$—total number of variables).

The closer the entity is located to the centre of a cluster, the higher is the confidence that the entity was correctly placed in a cluster, thus, the higher is classification reliability. In Figure 5 the entities of the Russian Federation are listed in order of increasing distance from the centroid[4].

In the k-means algorithm clusters are formed iteratively. During the first step centroids are selected randomly, then the process of consecutively adding point-vectors $\overline{X}(j)$ in the space of characteristics takes place. After adding each point, the location of centroids is recalculated. The process ends after stabilization of centroids' positions (when the relocation for a given step falls under a predefined threshold). The number of clusters in this case was defined automatically using the g-means algorithms (sequential examination of the fact that data within each of the selected cluster falls under a definite gaussian distribution). In case such examination produces negative results, each cluster is broken into two new clusters. The process is repeated at each step of a procedure until the number of clusters and centroid positions are stabilized.

Hierarchic structures presented in Figure 5 were built using minimum distance rule [19, 29]. At the first step, the clusters closest to each other are combined into groups; then more remote clusters are added until the whole hierarchy is formed. The distance among the clusters (groups of clusters) corresponds with the length of a vertical line to the closest hierarchy node. Similar results were obtained during clustering with the help of Kohonen self-organizing map [25]

It is necessary to point out that the regions traditionally compared with Sverdlovsk Oblast are located in either the same or close clusters, independently of which clustering algorithm is used. It is also necessary to note that Moscow and Moscow oblast constitute a separate cluster.

We consider differences in transport operations for the regions from different clusters. As Moscow and Moscow oblast form a single cluster that substantially differs from clusters 1 and 2, only the latter ones will be compared.

Distribution functions of parameters describing transport infrastructure for clusters 1 and 2 are quite close[5].

Nevertheless, there are substantial differences in the results of using transport infrastructure in clusters 1 and 2. For example, Figure 6 presents data on cargo haulage, cargo turnover and transport mobility of the population (for public buses). It is evident that distributions of these parameters in cluster 1 are skewed towards greater values when compared with distributions for cluster 2.

Thus, the differences between the groups of the entities of the Russian Federation (grouped using the *k*-means method) were caused by parameters of using transport infrastructures with similar

[4] Clustering using the algorithms was performed using Deductor Studio Academic version 5.3 0.88.

[5] For example, distributions for "railroads density" parameter for clusters 1 and 2 are the same with 5 % significance level, for "density of hard-top public motorways" — with 1 % significance level, and so forth.

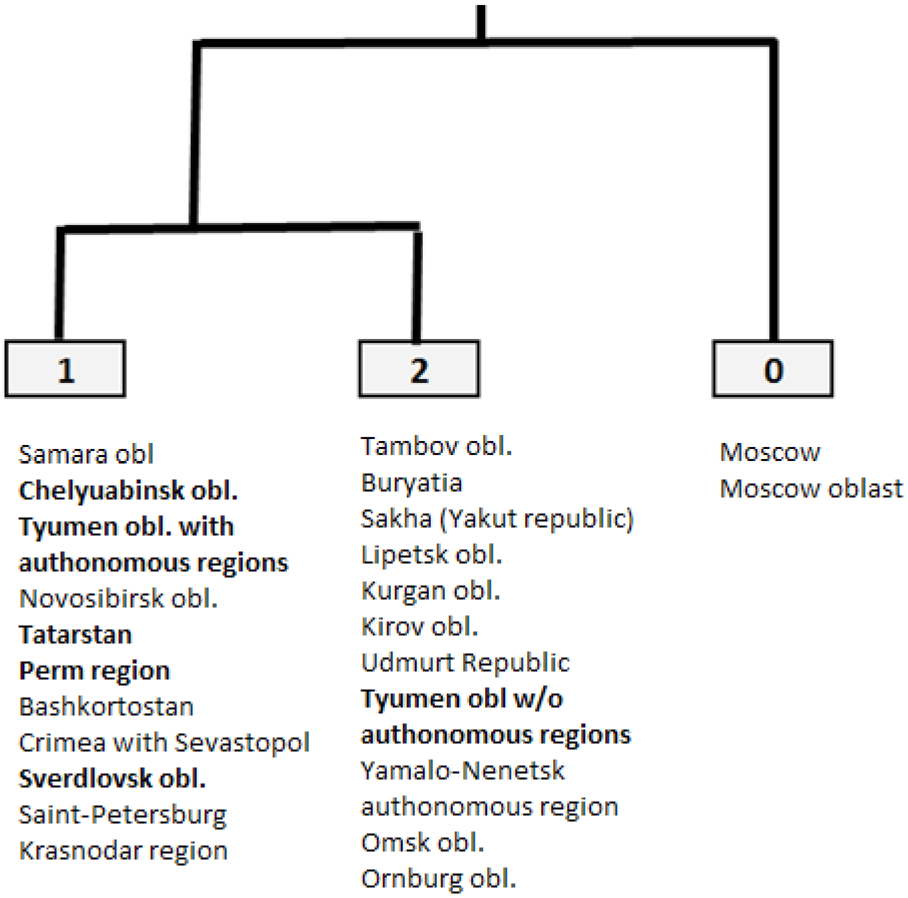
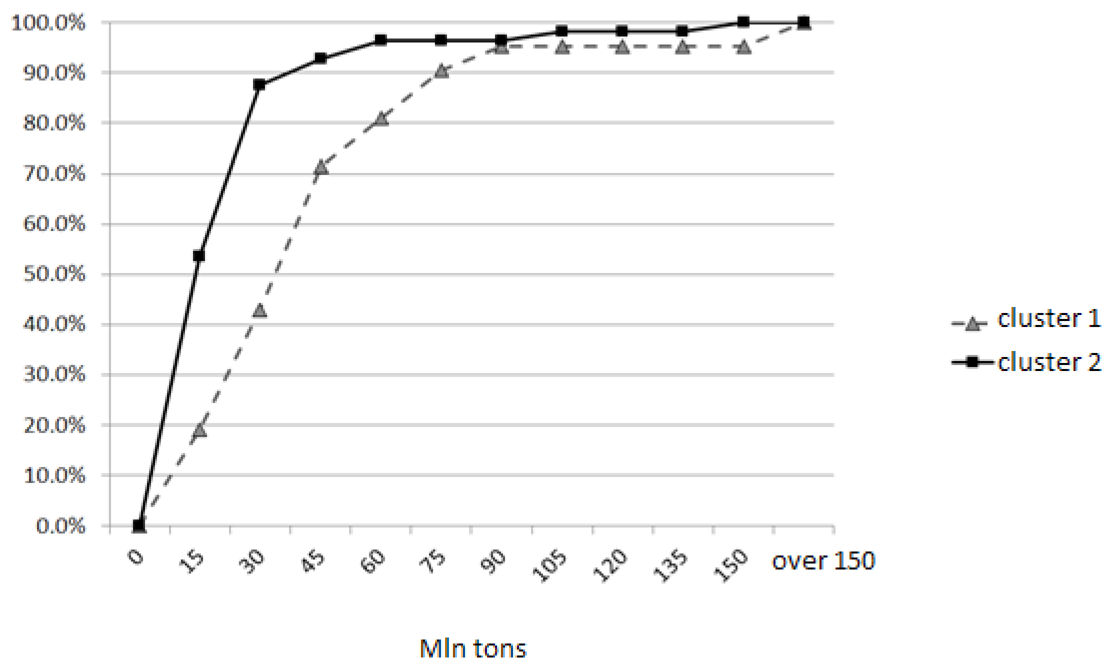
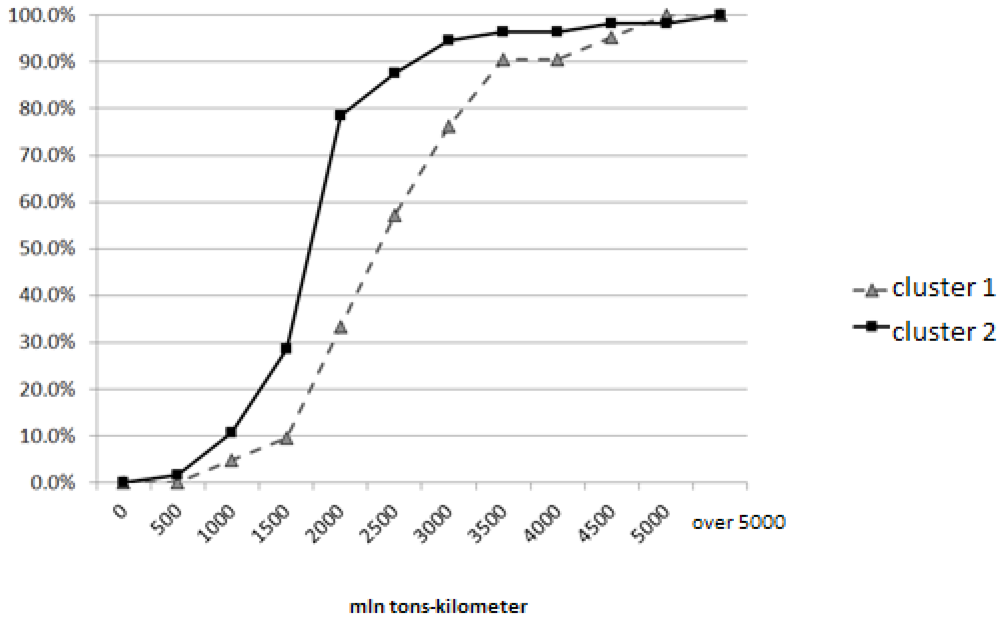



**Fig. 5.** *Results of clustering the entities of the Russian Federation in the space of independent variables (Table 2) using the k-means method. Regions traditionally compared with Sverdlovsk oblast are shown in bold*

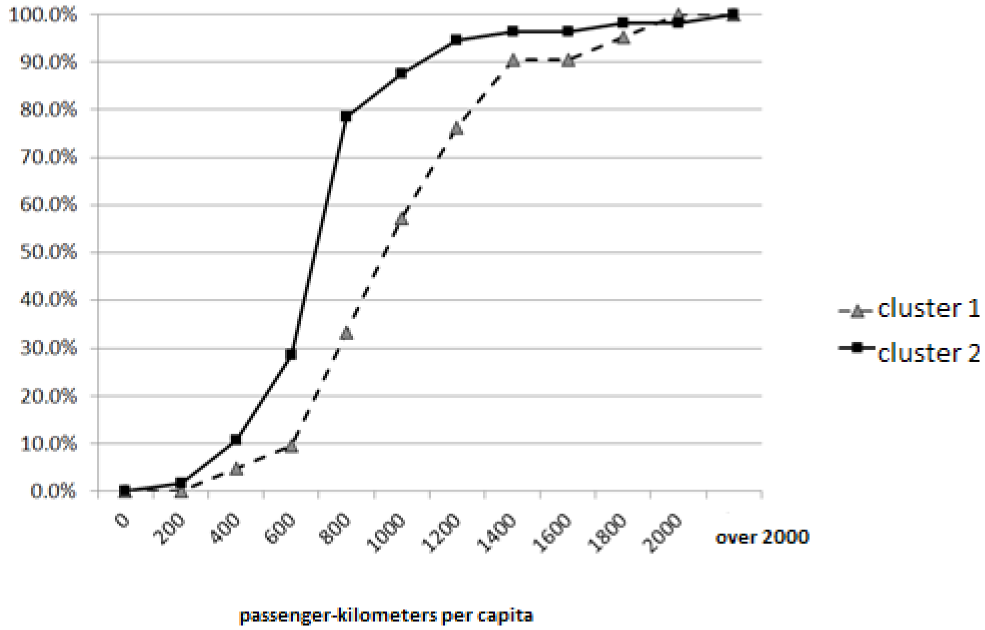



*a)*

Translation




*b)*




*c)*

**Fig. 6.** *Distributions of parameters for using transport infrastructure in clusters 1 and 2 (a) cargo haulage; б) cargo turnover; в) transport mobility of the population (public buses)*

parameters. That means that it is reasonable to use this clustering algorithm for benchmarking (selecting subjects for comparison with the one in question).

## 5. Conclusion

The paper presents a model for indirect assessment of motor transport passenger and cargo flows. The model allows estimating haulage, passenger, and cargo turnover without direct measurements of the numbers of transported passengers and amounts of cargo, as it is are complicated, expensive, and not always possible. Instead, our model uses indirect parameters that describe social and economic development of a territory in question. Similar models have a wide application area: from current strategic planning of territorial development for the period up to 2035, to operative and tactical resolution of transport management-related issues. For example, such models can be used to determine issues that transport infrastructure will encounter if the target economic and social development goals are met (new industries, equipment modernization, residential construction increase, etc.)

Translation

The analysis demonstrated that characteristics of passenger and cargo haulage over motorways are mainly affected by parameters describing economic development of a territory (Table 2), such as:

— Capital investments;

— Capital assets in economy;

— Amount of delivered goods;

— Commissioned residential buildings.

Transport infrastructure and demographic characteristics are the second-tier factors, while the effects of social factors are insubstantial.

At the same time, this macroanalysis does not provide answers for key questions of transport planning, including definition of optimal routes and stop points, bus schedules, and cargo delivery routes that minimize the strain on road network, etc. In order to solve these questions, it is necessary to solve a number of problems related to the collection and storage of complete and adequate data on the current status of transport flows, on both regional and interregional levels. Discussion of the related issues can be a subject of a separate paper.

In is also necessary to emphasise the perspectives of using the methods of intellectual data analysis for passenger and cargo flows in order to both make forecasts, and reveal the most typical scenarios of connections between social, economic, and transport systems, to reach the goals of their balanced development.

## References


1. Leontief, W. W. (1986). *Input-Output Economics.* 2nd ed. New York: Oxford University Press, 448.

2. Gasnikova, A. V. (2013). *Vvedenie v matematicheskoe modelirovanie transportnykh potokov [Introduction to mathematical modeling of traffic flows].* Moscow, 427. (In Russ.)

3. Sun, H., Liu, H., Xiao, H., He, R. & Ran, B. (2003). Use of Local Linear Regression Model for Short-Term Traffic Forecasting. *Journal of the Transportation Research Board, 1836,* 143–150. DOI: 10.3141/1836–18.

4. Oswald, R., Scherer, T. & Smith, B. L. (2001). *Traffic flow forecasting using approximate nearest neighbor nonparametric regression.* USA: Center for Transportation Studies, University of Virginia, F115.

5. Min, W. & Wynter, L. (2011). Real-time road traffic prediction with spatio-temporal correlations. *Transportation Research Part C: Emerging Technologies, 19(4),* 606–616. DOI: 10.1016/j.trc.2010.10.002.

6. Box, G. E., Jenkins, G. M. & Reinsel, G. C. (2008). *Time Series Analysis: Forecasting and Control.* 4th edition. Wiley, 784.

7. Mai, T., Ghosh, B., & Wilson, S. (2014). Short-term traffic-flow forecasting with auto-regressive moving average models. *Proceedings of the Institution of Civil Engineers — Transport, 167(4),* 232–239. DOI: 10.1680/tran.12.00012.

8. Stathopoulos, A. & Karlaftis, M. G. (2003). A multivariate state space approach for urban traffic flow modeling and prediction. *Transportation Research Part C: Emerging Technologies, 11(2),* 121–135. DOI: 10.1016/S0968–090X(03)00004–4.

9. Lin, S.-H., Huang, H.-Q., Zhu, D.-Q. & Wang, T.-Z. (2009). The application of space-time ARIMA model on traffic flow forecasting. *2009 International Conference on Machine Learning and Cybernetics, 6,* 3408–3412. DOI: 10.1109/ICMLC.2009.5212785.

10. Jin, X., Zhang, Y. & Yao, D. (2007). Simultaneously Prediction of Network Traffic Flow Based on PCA-SVR. *Lecture Notes in Computer Science,* 1022–1031. DOI: 10.1007/978–3–540–72393–6_121.

11. Vlahogianni, E. I., Karlaftis, M. G. & Golias, J. C. (2014). Short-term traffic forecasting: Where we are and where we're going. *Transportation Research Part C: Emerging Technologies, 43,* 3–19. DOI: 10.1016/j.trc.2014.01.005.

12. Guorong, G. & Yanping, L. (2010). Traffic Flow Forecasting based on PCA and Wavelet Neural Network. *Information Science and Management Engineering (ISME), 1,* 158–161. DOI: 10.1109/ISME.2010.10.

13. Zheng, W., Lee, D.-H. & Shi, Q. (2006). Short-Term Freeway Traffic Flow Prediction: Bayesian Combined Neural Network Approach. *Journal of Transportation Engineering, 132(2),* 114–121. DOI: 10.1061/(ASCE)0733–947X(2006)132:2(114).

14. Zhang, X. & He, G. (2007). Forecasting Approach for Short-term Traffic Flow based on Principal Component Analysis and Combined Neural Network. *Systems Engineering: Theory & Practice, 27(8),* 167–171. DOI: 10.1016/S1874–8651(08)60052–6.

15. Sinitsyn, E., Vikharev, S. & Brusyanin, D. (2019). Economic and Mathematical Model for Forecasting Passenger Traffic on a Long Term Basis Case of Study Russia. *Journal of Engineering and Applied Sciences, 14(3),* 773–779. DOI: 10.3923/jeasci.2019.773.779.

16. Draper, N. & Smith, G. (1986). *Prikladnoy regressionnyy analiz [Applied Regression Analysis].* Trans. from English. Moscow: Finance and statistics, 366. (In Russ.)

17. Dougherty, K. (1999). *Vvedenie v ekonometriku [Introduction to Econometrics].* Trans. from English. Moscow: Infra-M, 402. (In Russ.)

18. Hanke, J. E., Reitsch, A. G. & Wichern, D. W. (2003). *Biznes prognozirovanie [Business Forecasting].* Trans. from English. Moscow: Williams, 656. (In Russ.)

19. Barsegyan, A. A., Kupriyanov, M. S., Stepanenko, V. V. & Kholod, I. I. (2007). *Tekhnologii analiza dannykh. Data Mining, Visual Mining, Text Mining, OLAP [Data Analysis Technologies. Data Mining, Visual Mining, Text Mining, OLAP].* St. Petersburg: BHV, 384. (In Russ.)

20. Osovskiy, S. (2002). *Neyronnye seti dlya obrabotki informatsii [Neural networks for information processing].* Moscow, 344. (In Russ.)

21. Nikolenko, S. I., Kadurin, A. A. & Arkhangelskaya, E. O. (2018). *Glubokoe obuchenie. Pogruzhenie v mir neyronnykh setey [Deep learning. Immersion into the world of neural networks].* St. Petersburg: Piter, 480. (In Russ.)

22. Haykin, S. S. (2006). *Neyronnye seti. Polnyy kurs [Neural Networks: A Comprehensive Foundation].* Trans. from English. Moscow: Williams, 1104. (In Russ.)

23. Leskovec, J., Rajaraman, A. & Ullman, J. D. (2014). *Mining of massive datasets.* Stanford University: Milliway Labs, 516.


Translation

24. Zamyatin, A. V. (2016). *Intellektualnyy analiz dannykh [Intellectual data analysis].* Tomsk: Tomsk State University Publishing House, 120. (In Russ.)

25. Deboeck, G. & Kohonen, T. (2001). *Analiz finansovykh dannykh s pomoshchyu samoorganizuyushchikhsya kart [Visual explorations in finance with self-organizing maps].* Trans. from English. Moscow: Alpina, 317. (In Russ.)

26. Mitropolskiy, A. K. (1971). *Tekhnika statisticheskikh vychisleniy [Statistical Computation Technique].* Moscow: Science, 576. (In Russ.)

27. Kolmogorov, A. N. (1957). O predstavlenii nepreryvnykh funktsiy neskolkikh peremennykh superpozitsiyami nepreryvnykh funktsiy odnogo peremennogo i slozhenya [On the representation of continuous functions of several variables as a superposition of continuous functions of one variable and addition]. *Doklady AN SSSR [Reports of the Academy of Sciences of the USSR], 114,* 953–956. (In Russ.)

28. Arnold, V. I. (1957). O funktsii trekh peremennykh [About the function of three variables]. *Doklady AN SSSR [Reports of the Academy of Sciences of the USSR], 114,* 679–681. (In Russ.)

29. Faure, A. (1989). *Vospriyatie i raspoznavanie obrazov [Perception et reconnaissance des formes].* Trans. From French. Moscow: Mashinostroenie, 272. (In Russ.)

## Authors

**Evgeny Valentinovitch Sinitsyn** — Doctor of Physics and Mathematics, Professor, Academic Department of Systems Analysis and Decision Making, Graduate School of Economics and Management, Ural Federal University; Scopus Author ID: 7003263555 (19, Mira St., Ekaterinburg, 620002, Russian Federation; e-mail: sinitsyn_ev@mail.ru).

**Alexander Vladimirovich Tolmachev** — CEO, Datatel-Ural LLC; Senior Lecturer, Academic Department of Systems Analysis and Decision Making, Graduate School of Economics and Management, Ural Federal University; Scopus Author ID: 57204904436 (12B, Sibirskiy tract, Ekaterinburg, 620100; 19, Mira St., Ekaterinburg, 620002, Russian Federation; e-mail: at@idtu.ru).

**Brusyanin Dmitrii Alekseevitch** — Deputy Minister of Transport of the Sverdlovsk Region; Scopus Author ID: 56369168600 (1, Oktyabrskaya Sq., Ekaterinburg, 620031, Russian Federation; e-mail: dbrusyanin@mail.ru).