

Using Automated Fare Collection System Data to Determine Transport Demand

A. Fadeev

Siberian Federal University (SibFU)
Krasnoyarsk, Russian Federation
fai@ak1967.ru

S. Alhousseini

Siberian Federal University (SibFU)
Krasnoyarsk, Russian Federation
Eng.sami20143@gmail.com

Abstract— Transport demand (a set of data about trips and passengers mobility patterns in the network) is a golden key for solving a wide range of issues of transportation and town planning problems. These are development of the road network, development and optimization of public transport routes, etc. To determine the transport demand, the complex four-stage model is traditionally used, in which traditional manual surveys are needed. Indirect methods of obtaining information represent a special perspective based on the collection, integration and analysis of large and mixed type of data which been generated by various sources of human life aspects (Urban computing [10], Big data [11, 12]). One of such methods is the analysis transactions of non-cash fare payment in public transport [9 - 24]. The results of the passenger trips are obtained by comparing the sequence of payment transactions. As a result of processing transactions of non-cash fare payment, a sample is drawn from the general mobility of citizens transported by public transport. The article considers the following tasks: 1. Estimation of the representativeness of the non-cash fare payment transactions sample to the general set of trips. 2. Calculation of the parameters of transport demand based on the trips set's sample, which obtained as a result of processing transactions of non-cash fare payment. We proposed calculation of the Origin-Destination (OD) matrices, which takes into account presence of non-decrypting transactions. The obtained results allow us to use automated fare collection system (AFC) data to study transport demand: to obtain information about passenger's trip between the stopping points of network route or between transport districts, as well as passenger correspondence, which consist of one or several trips (when traveling with transfers).

Keywords— *transport demand, passenger flows, passenger trip, Origin-Destination matrices.*

I. INTRODUCTION

Transport demand is a set of data about passenger trips and mobility patterns in the network which is influenced by a large number of factors; modeling and forecasting transport demand is a complicated task, which if completed properly directly affects the efficiency of transport planning [1 - 4].

For determining transport demand, four-stage transport model usually is used, which consisting of the following stage [4 - 8]:

- Trip generation;
- trip distribution;

- mode split, mode-share;
- traffic assignment, route assignment, route choice.

The most common approach to determining the number of trips (Trip generation) is a multiple factors analysis model, according to which the demand is distributed by trip purpose and the average number of trips per family which is calculated based on statistical studies [6].

Next, (Trip distribution) we determine the departure and arrival capacity of transport districts. The obtained capacities of the districts are used in the calculation of the OD matrices [5]:

$$T_{ij} = A_i B_j E_i D_j f(C_{ij}) \quad (1)$$

where:

T_{ij} - The number of correspondence between the transport districts i and j ;

E_i - capacity i district by departure;

D_j - capacity j district by arrival;

$A_i B_j$ - balancing factors for rows and columns;

$f(C_{ij})$ - attraction curve (C_{ij} - the cost of moving from district i to district j , for example, time costs).

At the third stage (Mode split), we carried out the probability prediction of choosing the method of movement from various available modes of transport. Models of this kind are based on disaggregated models of individual preferences of the population [5–8]. At this stage it may take into account some measures which aimed to redistribution the passenger flow between the means of transport modes (for example, ensuring the preference of public transport over private cars).

At the fourth stage (Traffic assignment), the correspondence is distributed over the network using certain procedures that take into account the travel time along the route and the intensity of transportation units movement.

Passenger flows field surveys are carried out in order to study the parameters of transport demand: the distribution of

passenger flows in directions, changes over time, and the evaluation of transport demand calculation methods [5].

Methods for surveying passenger flows are classified according to a number of signs:

- The duration of the period covered: regular, one-time.
- Width of coverage: massive (at the same time the entire transport network of the area is served), selective.
- Modes of movement: general (all modes of transport), specialized (one or several modes of transport).
- Survey method: questionnaire (Household Travel Survey); reporting and statistical; full-scale; automated.

The main flaw of passenger flows field surveys is the high cost in resources and high working intensity [9].

Nowadays, indirect methods of obtaining information represent a special perspective based on the collection, integration and analysis of large and mixed type of data which been generated by various sources of human life aspects (Urban computing [10], Big data [11, 12]). This approach is used to solve the main problems that facing cities: air pollution, energy consumption, traffic, etc.

Presently for urban public transport with the Urban computing system it is possible to [10]:

- provide information about the bus arrival in real time directly to passengers' mobile phones;
- study the passenger mobility based on the analysis of GPS tracks received from mobile phones;
- format the public transport routes by analyzing passenger travel data from a large set of GPS taxi trajectories.

II. APPLY DATA FROM NON-CASH FARE PAYMENT TRANSACTIONS (AFC TRANSACTIONS)

One of promising methods for studying transport demand is mining in transactions of non-cash fare payment, which were recorded by magnetic cards, mobile phones, etc. The task of determining transport demand using data from the systems of non-cash fare collection (AFC system) is discussed in the following papers [9 - 24]. In [13, 14, 17, 20, 22] the subway demand was investigated, [15, 18, 19] - the rail demand, [14, 22] - the road public transport (bus, tram).

The transaction of non-cash fare payment on urban public transport records the following information, which can be used to infer the passenger trips: time of the transaction; stopping point; route and direction. These data make it possible to calculate both the passenger's trip and the Origin & Destination of his journey, which consists of one or several trips with transfers [9, 13, 14, etc.].

The determination of the passenger's trip is carried out by analyzing the sequence of fare payment transactions (trip-chaining method). In the works mentioned above, it is assumed that payment is made at the boarding point, i.e. from the transaction, the parameters of trip starting point are

known: point and time of boarding. The alighting point is obtained based on assumption that it is within walking distance from the point of the beginning of the next passenger's trip.

For calculation of the OD matrix (passenger correspondences) in papers [13 - 24], similar algorithms were used, which differ in and variations in the structure of the AFC system in each study, as well as the walking distance (buffer zone) and the maximum allowable transfer time.

The methods described in [13-24] have the following limitations:

- a) the option that the fare payment occur at the end of the trip is not supported (before alighting from the vehicle);
- b) inferring trips by analyzed transactions were limited by data from one day. In this case, the last trip is inferred based on assumption that the passenger returns to the boarding point of the first trip in the day. Without considering, the possibility of inferring trips on the basis of an analysis of the fare collection transactions carried out on the following days.

An improved algorithm was proposed in [9, 25], free from the mentioned flaws: the fare may be paid both at the boarding or at the alighting point; trips inferred based on the analysis of a large array of transactions over a certain period. The algorithm allows interpreting in average about 72% of transactions. The calculation of passengers' trips, as well as passenger O-D journeys (passenger correspondences), are implemented in a computer program based on Delphi programming language with using a database management system (DBMS), and the SQL language which ensures the efficiency of calculations through the relational algebra engine. As a result, the software package allows the processing of large amounts of information. For example, the monthly volume of transport AFC transactions of the Krasnoyarsk city (more than 6 million transactions) in MS SQL Server DBMS is processed in a few hours of computer operation.

Thus, as a result of processing transport AFC transactions, a set of trips $P(D, K, G, V^b, T^b, V^e, T^e, L)$ has been obtained that has the following variables [9]:

- D – transaction ID (unique transaction number);
- K – the ID of the ticket;
- G – route number;
- V^b, V^e - the ID of the trip's boarding and alighting stops;
- T^b, T^e - timestamp at boarding and alighting;
- L – the trip distances.

A trip is a relationship tuple (an instance of an object), which is denoted as $p(d, k, g, v^b, t^b, v^e, t^e, l)$. In addition to travel, passenger O-D journey is calculated, which consist of one or more trips (journey stages) with transfers. O-D journey is interpreted by analyzing the sequence of trips (trip-chaining method). It is considered that the passenger made the transfer, if the boarding point of the next trip is within walking distance (buffer zone) and the time between the start of the next and the

end of the previous trips does not exceed the transfer time interval.

Thus, passenger journey is a set of sequentially ordered trips consisting of one, two or several elements $h(p_1, \dots)$.

In figure 1 shows the distribution of carried passengers by hour of the day, obtained based on AFC transactions and the results of a filed survey of passenger flows of the Krasnoyarsk city in 2006 (conducted by the SibFU) and 2011 (NIIAT). It can be seen from the figure that the distribution of trips, which obtained from data of non-cash fare payment transaction, repeat the dynamics of passenger flows by the hour of the day, which obtained from the results of a complete filed survey in 2006, 2011.

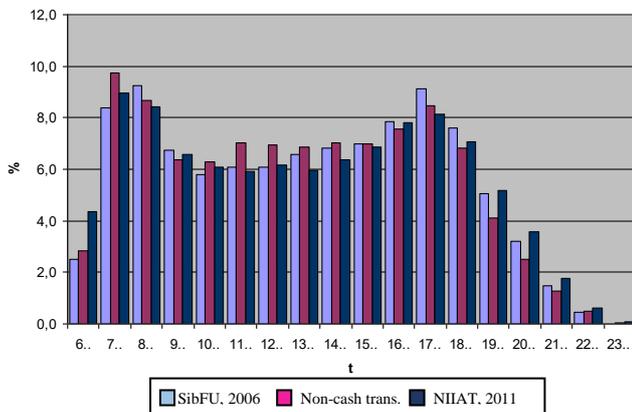


Fig. 1. Distribution of carried passengers by hour of the day

As a result of mining and processing AFC transactions, a trips sample was obtained from the general aggregate of public transportation trips. Which needed to check the following questions:

1. Evaluate the representativeness of the trips sample obtained from AFC transactions for the total amount of public transportation trips, i.e. the ability of the sample to reproduce certain characteristics of the general population mobility by public transportation within the limits of permissible errors.
2. Determine the method of calculating transport demand parameters based on the trips sample, which obtained during mining and processing AFC transactions.

It is necessary to determine the conformity of the trips sample obtained according to AFC transactions to the general set of trips (in cash and non-cash) which requires separate consideration. It is necessary to prove that the basic parameters of the trips sample (from AFC transactions) correspond to the general one with sufficient accuracy. It is also necessary to resolve the issue with the coefficient of non-cash payment: should it be calculated on average by the network or differentiation by type of transport (bus, tram trolleybus), routes, etc. or not.

As is well known [1, 2, 3, 27], the mobility of the population is determined by the social composition of the population and the journey purpose. When studying transport demand, the population is grouped according to certain criteria, such as citizens, the suburbs residents, visitors, workers, students, etc. The proportion of the non-cash fare payment today is less than 50 percent. There is a possibility that the structure of passengers who pay for travel by transport card differs from that of the general population, for example, it can be expected that the non-cash fare payment is less among visitors.

III. EVALUATION OF THE REPRESENTATIVENESS OF THE TRIPS SAMPLE OBTAINED FROM AFC TRANSACTIONS

To evaluate the representativeness of the trips sample obtained from AFC transactions, we compare it with the results of automated passenger counting which made on some routes in the same period under review. The data of automated passenger counting are provided by the (Krasnoyarskcitytrans) MSE.

A selective survey of passenger flow was carried out using special equipment installed in transport vehicles. For counting the passengers were used infrared detectors at the route stopping points. The volume of the sample automated passenger counting included more than 336 thousand passengers, 6094 transport vehicles trips surveyed in October 2016 on 6 routes of the Krasnoyarsk city (two trolleybus and 4 bus). The list of surveyed routes is given in Table 1.

Table I. List of surveyed routes

Route number	Direction of movement	Number of passengers	Run on route, km
<i>Bus</i>			
49	direct	16297	6241.5
49	return	16678	6426.4
52	direct	19101	7740.0
52	return	19615	6900.6
55	direct	39547	12152.4
55	return	71586	24553.2
61	direct	50535	19602.0
61	return	50407	19411.0
<i>Trolleybus</i>			
5T	direct	6787	3274.6
5T	return	5117	3096.0
15T	direct	20800	9562.5
15T	return	20219	9134.4

To estimate the representativeness of the trips sample obtained from AFC transactions, we compare the distribution of passengers boarding and alighting along the route.

Divide the route into k non-intersecting sections. For each section, we calculate the number of passengers boarding and alighting, which obtained from the results processing AFC transactions and automated passenger counting (see Table 2).

Table II. Samples of passenger flow obtained AFC transactions and automated passenger counting (2 routes out of 6 analyzed are shown)

Sections №	Boarding				Alighting			
	Automated survey		AFC transactions		Automated survey		AFC transactions	
	Passenger	Share	Passenger	Share	Passenger	Share	Passenger	Share
Bus, rout №49, direct direction								
1	7318	0.45	13852	0.36	581	0.04	2367	0.06
2	4359	0.27	10950	0.28	4211	0.26	8440	0.22
3	2731	0.17	8961	0.23	3641	0.22	8520	0.22
4	1623	0.10	3920	0.10	3703	0.23	9318	0.24
5	266	0.02	757	0.02	4161	0.26	9795	0.25
Sum	16297	1.00	38440	1.00	16297	1.00	38440	1.00
rout №49, return direction								
1	4002	0.24	10740	0.32	214	0.01	1574	0.05
2	4416	0.26	11165	0.33	2193	0.13	7160	0.21
3	4563	0.27	7303	0.22	3092	0.19	9216	0.28
4	2828	0.17	3552	0.11	2945	0.18	6120	0.18
5	869	0.05	707	0.02	8229	0.49	9397	0.28
Sum	16678	1.00	33467	1.00	16673	1.00	33467	1.00
Trolleybus, rout №15, direct direction								
1	5661	0.27	6432	0.23	1379	0.07	1380	0.05
2	4060	0.20	5489	0.20	2358	0.11	3919	0.14
3	4046	0.19	6464	0.23	3329	0.16	5009	0.18
4	4164	0.20	5739	0.20	3645	0.18	5003	0.18
5	2869	0.14	4000	0.14	10086	0.48	12813	0.46
Sum	20800	1.00	28124	1.00	20797	1.00	28124	1.00
rout №15, return direction								
1	6378	0.32	8741	0.28	1353	0.07	2380	0.08
2	4160	0.21	6040	0.19	2819	0.14	3918	0.13
3	4367	0.22	6950	0.22	3398	0.17	5352	0.17
4	3336	0.16	6189	0.20	5433	0.27	7753	0.25
5	1978	0.10	3295	0.11	7216	0.36	11812	0.38
Sum	20219	1.00	31215	1.00	20219	1.00	31215	1.00

Thus, two independent (not-connected) samples were formed, for comparison we use the Student's test (t-test), which allows us to find the probability that both average values in the sample belong to the same set.

Statistics criteria for the case of independent samples is equal to [28]:

$$t_e = \frac{\bar{x} - \bar{y}}{\sigma_{x-y}}, \quad (2)$$

where:

\bar{x}, \bar{y} - arithmetic means in the experimental and control groups;

σ_{x-y} - standard error of the difference in arithmetic means, which is calculated by the formula [28]:

$$\sigma_{x-y} = \sqrt{\frac{\sum(x_i - \bar{x})^2 + \sum(y_i - \bar{y})^2}{n_1 + n_2 - 1} \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}, \quad (3)$$

For $n_1 \ll n_2$, or :

$$\sigma_{x-y} = \sqrt{\frac{\sum(x_i - \bar{x})^2 + \sum(y_i - \bar{y})^2}{(n - 1)n}}, \quad (4)$$

For $n_1 = n_2$, where n - sample size.

Next, it is necessary to compare t_e with the theoretical value of the Student's t-distribution t_k with degrees of freedom equal $n_1 - n_2 - 2$. The hypothesis about the equality of the samples is accepted if $t_e < t_k$.

Table 3 shows the results of calculating the t-test of the compared samples. With a degree of freedom equal to 8, the

theoretical values of the t_k are [28]: 1,860 ($P \leq 0.05$); 2,896 ($P \leq 0.01$); 4,501 ($P \leq 0.001$). The highest value of the t_e (Table 3) is 2.923 for the direct direction of the route number 49 through the alighting passengers. Most of the calculation results do not exceed the critical value of 1.860.

Table III. The calculation results of the Student's t-test for compared samples

Rout	Direction	t_e	
		Boarding	Alighting
Bus			
49	direct	1.661	2.923
49	return	1.566	0.375
52	direct	1.822	1.637
52	return	1.693	1.637
55	direct	0.618	0.326
55	return	0.469	0.672
61	direct	0.188	0.164
61	return	0.188	0.217
Trolleybus			
5T	direct	1.249	0.817
5T	return	0.698	1.248
15T	direct	2.320	0.597
15T	return	1.936	1.184

In the light of the foregoing, it can be concluded that the samples obtained from AFC transactions and from the results of a complete filed survey of passengers' flows are the same statistically. AFC transactions are representative of the general population mobility, i.e. they can be used to assess the characteristics of the general population mobility within the limits of permissible errors.

To estimate the total transport demand, it is necessary to determine:

- traffic volume;
- passenger trips;
- passenger O-D journeys.

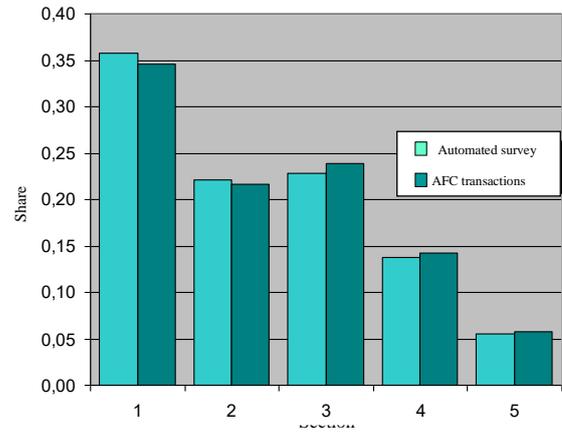
The number of passengers carried by public transport is calculated using the following formula:

$$Q = R / \alpha^r, \quad (5)$$

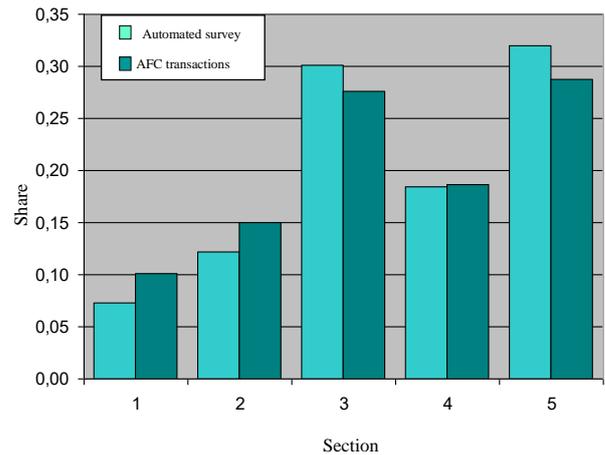
where:

α^r - coefficient of non-cash payment;

R - the number of passengers who paid fare by non-cash payment.



a) Boarding Passengers



b) Alighting Passengers

Fig. 2. Distribution of amount of (a) Boarding Passengers and (b) Alighting passengers along the route No. 61 (return direction)

For its application, it is necessary to determine the coefficient of non-cash payment. At the same time, it should be investigated whether this parameter is the same for all transport modes and for all routes. We will solve this problem as follows. We form a sample of the coefficients of non-cash payment for each routes, for calculating the elements of the sample we use:

$$\alpha_i^r = \frac{R_i z_i^s}{Q_i^s z_i^f}, \quad (6)$$

where:

R_i - the number of passengers who paid the fare by non-cash payment for the i element of the sample (per day for a certain route);

Q_i^s - number of passengers carried by transport trips which been surveyed;

z_i^s, z_i^f - the number of surveyed and actual transport trips on the route of the i element of the sample, respectively.

As a result, a sample of non-cash payment coefficients for October 2016 was formed, consisting of 122 elements for bus routes and 56 elements for trolleybus routes. Sample values for the bus range from 0.27 to 0.45, trolley buses - from 0.62 to 0.97. Taking into account, the significant difference between the values of coefficients for bus and trolleybus routes, it can be concluded that the coefficients of non-cash payment should be regulated and studied separately for each type of transport.

Sample factors for non-cash payment of buses' routes are given in Table. 4. As a result of statistical processing, we obtained an average value of 0.34, the confidence interval of coefficient of non-cash payment is from 0.33 to 0.35 for a significance level of 0.05.

Table IV. Sample factors for non-cash payment of buses' routes

№ Interval	Interval	qt.	Share
1	0.25	7	0.06
2	0.275	13	0.11
3	0.3	21	0.18
4	0.325	26	0.22
5	0.35	24	0.20
6	0.375	18	0.15
7	0.4	10	0.08
8	0.425	1	0.01
Sum	0.45	120	1.00
Average $\bar{\alpha}^r$			0.3429
Dispersion S^2			0.0017
Standard deviation σ			0.0418
Average confidence interval		0.05	0.3339...0.3519

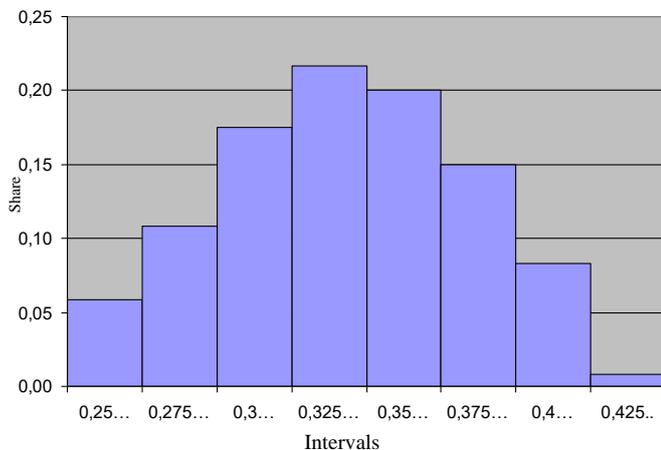


Fig. 3. Distribution of non-cash payment of buses' routes coefficients.

The viewed sample of non-cash payment coefficients is a composition of four samples (4 routes). We examine whether these samples differ among themselves. We use the criterion of Kruskal and Wallis [28] to test the hypothesis that the distribution of non-cash payment coefficients for the viewed routes do not differ:

$$\sum_{i=1}^k \{n_i [\bar{X}_i' - (N = 1) / 2]^2\} / [(N^2 - 1) / 12], \quad (7)$$

where: n_i - number of observations in the i sample;

\bar{X}_i' - the average value of the i sample.

We order and rank the N elements from the total sample, then divide the sample into four sets related to each of the viewed routes. As a result of the calculation, we obtain the statistics value which equal to 58.06. The upper 5% value of the criterion χ^2 with three degrees of freedom is 7.185. As a result, it is necessary to recognize the significant difference between the mathematical expectations of non-cash payment coefficient for different routes. A pairwise check of the variances of the samples under by using the Fisher criterion [28] allowed us to conclude that they are equal.

Thus, it was established that the coefficients of non-cash payment for different routes with equal dispersion have different mathematical expectation (see Figure 4), i.e. non-cash payment coefficient should be regulated and studied separately for each route. To determine the non-cash payment coefficient, we need to determine the number of passengers carried along the route per day. To do this, we can use the reporting and statistical method. The non-cash payment coefficient for the k route is calculated by:

$$\alpha_k^r = R_k / Q_k, \quad (8)$$

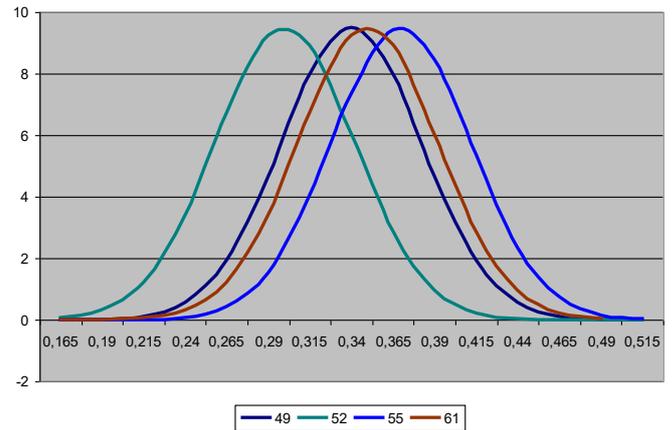


Fig. 4. Probability density of the non-cash payment coefficients distribution for different routes (59, ..., 61 - route numbers)

IV. CALCULATION OF PARAMETERS TRANSPORT DEMAND

Transport demand is usually represented in the form of OD matrix (passenger correspondences between districts) [2, 3,

etc.], which are formed as a kind of averaged model of the population's need for movement along the transport network:

$$A = \|q_{i,j}\|; i, j = 1, \dots, n, \quad (9)$$

where: $q_{i,j}$ - the number of passengers movements which made during the period of study between transport districts i, j .

The results of processing AFC transactions allow to format transport demand matrices of two types:

- O-D matrix for runs (passenger trips matrix);
- O-D matrix for the whole network (between origin points and destination ones, taking into account journey with transfers).

The O-D matrix for runs can be used to solve problems of improving the functioning of the transport system without changing the network routes, for example, changing movement intervals, structure of transport fleet, etc. The OD matrix for the whole network is necessary for tasks related to the improvement of the network routes.

The procedure for calculating the O-D matrix for runs is by defining the number of trips between stopping points (transport districts) which will be carried out as follows:

$$q_{ij} = \sum_{k=1}^m \frac{P_{kij}}{\alpha_k^r} \varphi_i \varphi_j, \quad (10)$$

where: q_{ij} - estimated number of trips between stopping points (transport districts) i, j ;

P_{kij} - the number of recognized trips between stopping points i, j along the k route ($k = 1, \dots, m$);

φ_i, φ_j - the balancing coefficients of departure for the i stop and of arrival for the j stop, respectively ($\varphi \geq 1$).

Balancing coefficients of departure can be calculated by:

$$\varphi_i = \frac{\sum_{j=0}^n P_{ij}}{\sum_{i=1}^n P_{ij}}, \quad (11)$$

where: $\sum_{j=0}^n P_{ij}$ - the total number of trips from the i stop (including unrecognized trips, the destination of which is considered to be equal to 0);

$\sum_{i=1}^n P_{ij}$ - the number of recognized trips from the i stop ($i = 1, \dots, n$).

In similar way, we can calculate the balancing coefficients of arrival.

By using balancing coefficients, unrecognized transactions are compensated, for example, if a passenger makes one or more trips by transport modes that does not use non-cash fare payment system (taxi, personal car, etc.). Thus, the chain of

non-cash payment transactions is broken and the previous transaction will not be able to be recognized.

Mining the AFC transaction is given the opportunity to form O-D matrix for runs, which has a wide range of required properties. For example, you can create an O-D matrix for runs for a working day, weighted average for a month. To do this, from the set of trips P we make a selection (subset) of trips $P' \subset P$ made on working day of the viewed month. The weighted average number of trips is defined as:

$$q_{ij} = \sum_{k=1}^m \frac{P_{kij}}{\alpha_k^r d} \varphi_i \varphi_j, \quad (12)$$

d - number of days sampled

The procedure for calculating the O-D matrix for the whole network. The number of correspondences between districts i, j is defined as:

$$q_{ij} = \sum_{k=1}^m \frac{h_{kij}}{\alpha_k^r \psi_k} \varphi_i \varphi_j, \quad (13)$$

where: h_{kij} - the number of correspondences between districts i, j , the first trip of which was made on the k route;

ψ_k - coefficient of transfer ability for the k route ($\psi_k \geq 1$).

Coefficient of transfer ability is ratio of the number of journey on the k route to the number of all carried passengers from i to j , i.e.:

$$\psi_k = \frac{\sum_{i=0}^n \sum_{j=0}^n P_{kij}}{\sum_{i=0}^n \sum_{j=0}^n h_{kij}}, \quad (14)$$

The balancing factors of departure and arrival are calculated for the correspondences as follows:

$$\varphi_i = \frac{\sum_{j=0}^n h_{ij}}{\sum_{j=1}^n h_{ij}}, \quad \varphi_j = \frac{\sum_{i=0}^n h_{ij}}{\sum_{i=1}^n h_{ij}}, \quad (15)$$

Table 5 shows a segment of the O-D matrix for runs, compiled for the weekday of October 2016 in accordance with the above methodology.

Table V. Segment of the O-D matrix for runs

Origin stop 647 Academgorodok					
Destination stop		Passengers	Destination stop		Passengers
<i>Id</i>	<i>Stops' name</i>		<i>Id</i>	<i>Stops' name</i>	
649	Institut (ul. A.Kirenskogo)	96.1	809	Mezhdugorodnyj avtovokzal	6.5
655	Gastronom	35.1	454	Muzej im. Surikova	6.3
657	Studencheskij gorodok	33.9	448	Krasnaja ploshhad'	6.1
661	stancija Junnatov	18.2	32	Hlebozavod (ul. Sverdlovskaja)	6.1
453	Dom tehniki	17.5	672	Medicinskij universitet	5.8
450	Organnyj zal	17.5	861	Koda «Medistal»	5.8
1004	Zapadnyj	15.4	665	ul. Lunacharskogo	5.8
1002	ul. Korneeva	14.3	651	Lesnaja	5.8
663	kinoteatr Udarnik	14.0	455	gostinica «Oktjabr'skaja»	5.6
663	Kopylovskij most	13.8	552	Torgovyj kvartal	5.6

V. CONCLUSIONS

1. Indirect methods of obtaining information represent a special perspective based on the collection, integration and analysis of large and mixed type of data which been generated by various sources of human life aspects (Urban computing [10], Big data [11, 12]). One of effective methods for studying transport demand is mining in transactions of non-cash fare payment in public transport.

2. The method of analysis the AFC transaction discussed in the article allows to recover information about passengers trips between stopping points of the route network (O-D matrix for runs), as well as O-D matrix for the whole network, which are formed from one or several trips (journey with transfers).

3. The article estimates the representativeness of the non-cash fare payment transactions sample to the general set of trips. As a result of comparing the samples obtained from AFC transactions and from the results of a complete filed survey of passengers' flows, it was established that passenger trips obtained from AFC transactions are representative of the total population trips, i.e. they can be used to assess the characteristics of the general population mobility within the limits of permissible errors.

4. To calculate the transport demand according to AFC transactions data, a non-cash payment coefficient has been proposed. It has been established that the coefficients of non-cash payment for different routes, with the same dispersion, have unequal mathematical expectation, i.e. non-cash payment coefficient should be regulated and studied separately for each route.

5. According to the results of processing AFC transactions, the article considers the procedure to format transport demand matrices of two types:

- O-D matrix for runs (passenger trips matrix);
- O-D matrix for the whole network (between origin points and destination ones, taking into account journey with transfers).

Acknowledgment

The authors are grateful to Municipal Public Institution of Krasnoyarsk City (MSE) "Krasnoyarskcitytrans" for providing the data for this research.

References

- [1] W. A. Martin, N. A. McGuskin, "Travel Estimation Techniques for Urban Planning" National Cooperative Research Program Report 365. Transportation Research Board. National Research Council, Washington, DC, 1998
- [2] A.A. Shtockaja, A.Ju. Mihajlov, Ocenka "the population mobility based on disaggregated models", "transportnoj podvizhnosti naselenija na osnove dezagre-girovannyh modelej" Vestnik Irkutskogo gosudarstvennogo tehniceskogo universiteta, 2017, Vol. 21, No. 5, pp. 199-207
- [3] The Online TDM Encyclopedia. Retrieved from: Available at: <http://www.vtpi.org/>
- [4] M. R. Jakimov, "Transport planning: creating transport models of cities", "Transportnoe planirovanie: sozdanie transportnyh modelej gorodov" monografija, M.: Logos, 2013
- [5] V.V. Semjonov, A.V. Ermakov, "Historical analysis of modeling transport processes and transport infrastructure", "Istoricheskij analiz modelirovanija transportnyh processov i transportnoj infrastruktury" Preprinty IPM im. M.V.Keldysha, 2015, No. 3, 36 p.
- [6] J. D. Ortuzar, L. G. Willumsen, "Modelling transport." John Willey & Sons, 2011
- [7] D. Lee, "Requiem for large-scale models" Journal of the American Institute of Planners, Vol 39, No 3, May, 1973
- [8] S. Atkins, "Transportation planning models: What the papers say" Traffic Engineering and Control, Vol. 27, No. 9, September, 1986
- [9] A.Fadeev, S.Alhuseini, E.N. Belova, "Monitoring Public Transport Demand Using Data From Automated Fare Collection System" Advances in Engineering Research, vol. 158, Proceedings of the International Conference "Aviamechanical engineering and transport" (AVENT 2018), 2018, pp. 5 - 12
- [10] Y. Zheng, L. Capra, O. Wolfson, H. Yang, "Urban computing: Concepts, methodologies, and applications" ACM Trans. Intell. Syst. Technol., vol. 5, no. 3, pp. 1-55, Sep. 2014
- [11] Min Chen, Shiwen Mao, Yin Zhang, Victor C.M. Leung, "Big Data. Related Technologies, Challenges, and Future Prospects" Spinger, 2014, 100 p.
- [12] I. M. Bljankinshtejn, A.I. Fadeev, A. V. Fjodorov, N. V. Shadrin, E. G. Mahova, "Expediency justification of studying population mobility by

- monitoring the mobile subscribers”, “Obosnovanie celesoobraznosti izuchenija transportnoj podvizhnosti naselenija na osnove monitoringa abonentov mobil'noj svjazi” Zhurnal Sibirskogo federal'nogo universiteta. Serija: Tehnika i tehnologii, 2015, Vol. 8, No. 2, pp. 254 - 263
- [13] J.J. Barry, R. Newhouser, A. Rahbee, S. Sayeda, “Origin and destination estimation in New York City with automated fare system data” *Transportation Research, Record* 1817, 2002, pp.183–187
- [14] J. Zhao, A. Rahbee, N. Wilson, “Estimating a rail passenger trip origin-destination matrix using automatic data collection systems” *Computer-Aided Civil and Infrastructure Engineering*, vol. 22, 2007, pp.376–387
- [15] M. Trépanier, N. Tranchant, R. Chapleau, “Individual trip destination estimation in a transit smart card automated fare collection system” *Journal of Intelligent Transportation Systems*, vol. 11, 2007, pp.1–14
- [16] F. Devillaine, M.A. Munizaga, M. Trépanier, “Detection activities of public transport users by analyzing smart card data” *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2276, 2012, pp.48–55
- [17] M. Munizaga, C. Palma, “Estimation of a disaggregate multimodal public transport OD matrix from passive smartcard data from Santiago, Chile” *Transportation Research Part C*, Vol. 24, 2012, pp.9-18
- [18] M. Bagchi, P. White, “What role for smart-card data from bus systems?” *Munic. Eng.*, vol. 157, 2004, pp.39–46
- [19] N. Nassir, A. Khani, S.G. Lee, H. Noh, M. Hickman, “Transit stop-level origin–destination estimation through use of transit schedule and automated data collection system” *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2263, 2011, pp.140–150
- [20] A. Cui, “Bus passenger origin–destination matrix estimation using automated data collection systems” master’s thesis - Massachusetts Institute of Technology, 2006
- [21] W. Wang, P. John, H.M. Nigel, “Bus passenger origin–destination estimation and travel behavior using automated data collection systems in London” *Journal of Public Transportation*, Vol. 14, No. 4, 2011
- [22] A. Alsgar, M. Mesbah, L. Ferreira, H. Safi, “Use of smart card fare data to estimate public transport origin–destination matrix” *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2535, 2015, pp.88–96
- [23] J.J. Barry, R. Freimer, H.L. Slavin, “Use of entry-only automatic fare collection data to estimate linked transit trips in New York City” *Transp. Res. Rec. J. Transp. Res. Board*, vol. 2112, 2009, pp.53–61
- [24] K. El. Mahrsi, E. Côme, L. Oukhellou, M. Verleysen, “Clustering smart card data for urban mobility analysis” *IEEE Transactions on intelligent transportation systems*, vol. 18, NO. 3, March 2017
- [25] A.I. Fadeev, S. Alhusseini, “Using data from the automated fare collection system to survey passenger flow”, “Ispol'zovanie dannyh sistemy avtomatizirovannoj sistemy oplaty proezda dlja obsledovanija passazhirskih potokov” “Aviamechanical engineering and transport”, sbornik statej X mezhdunarodnoj nauchno-tehnicheskoy konferencii. 2018, pp. 329 - 334
- [26] A.I. Fadeev, E.V. Belokurova, “Monitoring the population mobility through using the automated fare collection system data”, “Monitoring transportnoj podvizhnosti naselenija posredstvom ispol'zovanija dannyh avtomatizirovannoj sistemy oplaty proezda” *Vestnik transporta*. 2017. No.7, pp. 25 -32
- [27] I. S. Efremov, V. M. Kobozev, V. A. Judin, “Theory of urban public transportation”, “Teorija gorodskih passazhirskih perevozok”:Ucheb. posobie dlja vuzov Vyssh. shkola, 1980, 535 p.
- [28] N. Dzhonson, F. Lion, “Statistics and experiment planning in engineering and science”, “Statistika i planirovanie jeksperimenta v tehnike i nauke”: Metody pla-nirovanija jeksperimenta. Per. s angl. - M.: Mir, 1980, 510 p.