Periodica Polytechnica Transportation Engineering, 51(4), pp. 394–408, 2023

Determination of Urban Public Transport Demand by Processing Electronic Travel Ticket Data

Aleksandr I. Fadeev[1], Sami Alhusseini[1*]

[1] Transport Department, Siberian Federal University, 79 Svobodny pr., 660041 Krasnoyarsk, Russia
[*] Corresponding author, e-mail: salkhussyayni-a17@stud.sfu-kras.ru

Received: 02 November 2022, Accepted: 21 June 2023, Published online: 14 August 2023

Abstract

Determination of transport demand is one of the key factors to solve many transportation problems. Existing methods for obtaining information about passenger mobility have significant shortcomings; currently, methods based on the collection, integration and analysis of big data (Urban computing, big data) are being increasingly used.

Within the framework of this approach, a methodology has been developed for determining the passenger traffic by public transport from the operations of validating electronic travel tickets: smart cards, transport cards, magnetic cards, mobile phones or other electronic devices (Electronic Gadgets), their details are recorded in the Automated Fare Collection (AFC).

In this work we have taken into account that the passenger can travel one, two or more segments before paying for the trip. On some routes, payment is made at the end of the trip.

The article presents a methodology based on defining and evaluating the set of acceptable variants of the connectedness of the passenger trips' sequence, which takes into account many factors that influence the choice of travel routes by the passenger. For example, unlike existing solutions, the possibility of paying for travel at any point of the route is taken into account, not necessarily immediately after boarding the vehicle.

Approbation of the considered methodology was carried out on the data of the Krasnoyarsk public transport system for April 2019. It has been proved that passenger traffic from the validation of electronic travel tickets allows us to estimate the parameters of public transport demand within the limits of acceptable statistical errors.

Keywords

passenger flow, passenger trip, matrix of passenger journey, urban public transport

1 Background of the materials and methods

The existing approaches to solving the problem under study have been considered in many studies (Barry et al., 2009; Chu and Chapleau, 2008; Munizaga et al., 2010; Nassir et al., 2011 etc.). This allows by using AFC data to determine the following parameters of passenger trips: PT (Public Transport) stop, trip boarding and alighting time, mode of transport, trip distance.

In most AFC systems of urban public transport, a passenger has to tap on an e-ticket once per trip. The fare is paid at the station, stop point or in the vehicle. In the vehicle, the passenger taps on the ticket, usually immediately after boarding, although in some cases they may travel one or two stops before paying. In Russia, there are some systems where passengers place payments (tap on an e-ticket) for travel at the destination of their trip. This fact is not considered in the studies of foreign authors. When the passenger taps on their e-ticket, the following data is recorded (Barry et al., 2002; 2009; Li et al., 2011; Zhao et al., 2007): card number, validation time, validation device ID, which allows for determining the location of the operation (station, metro station, and metro station turnstile), mode of transport, etc.

Some systems also store additional data, for example, the Transantiago system (Santiago, Chile) records the card type (Munizaga and Plama, 2012): student, discounted, free, etc. In the work of Li et al. (2011), cards are divided into the following classes: a fixed monthly payment, unlimited number of trips, free, with 50% discount, standard (with 10% discount). These kinds of data can be used to classify passengers into social groups.

Cite this article as: Fadeev, A. I., Alhusseini, S. (2023) "Determination of Urban Public Transport Demand by Processing Electronic Travel Ticket Data", Periodica Polytechnica Transportation Engineering, 51(4), pp. 394–408. https://doi.org/10.3311/PPtr.21447

Because of technical reasons, the raw data of AFC often contain errors that need to be localized, corrected, if possible, or excluded from consideration (Li et al., 2011).

In some AFC systems, such as Queensland Australia (Alsger et al., 2016), tapping on an e-ticket is performed at the start and end of trips, i.e., in such transactions fixed card ID, route, direction, time, and stop of boarding and alighting were recorded.

Many studies have considered data from AFC system like (Zhao et al., 2007) subway transportation in New York (Barry et al., 2002; 2009), and others (Chu and Chapleau, 2008; Cui, 2006; Farzin, 2008; Gordon et al., 2013 etc.). In most of them, the location of the e-ticket validation (tap on) is not recorded, and the coordinates of the operation site are set according to the time of the operation.

In most cases, tapping on an e-ticket is carried out at the beginning of the trip. The start point of the trip is considered to be the PT stop, where the previous tapping on a ticket occurred. The following approaches are used to determine this PT stop:

- according to the bus schedule (Barry et al., 2002; 2009) where a separate problem has been solved related to the fact that in the Metrocard system the transaction time is ranged up to a 6-minute interval, which leads to inaccuracies in determining the passenger's boarding point);
- through additional processing of data from Automatic Vehicle Location systems (AVL), in which the location and time stamps of the vehicle's movement along the route are recorded (Agard et al., 2006; Munizaga and Plama, 2012); Zhao et al., 2007); the essence of this approach is to link the databases of AFC and AVL.

In some modern systems, AFC and AVL are integrated, which allows one to get the location of the tapping on (i.e., validation) operation, for example, GoCard (Alsger et al., 2016), Andante (Nunes et al., 2016), etc.

Human life activity is accompanied by related trips (Fig. 1), which makes it possible to identify passenger journeys by analyzing information from e-ticket tapping on operations. The trip chain method is used, which consists of building a sequence of passenger trips by connecting the data recorded in the e-ticket tapping on operations (Alsger et al., 2016; Barry et al., 2002; 2009; Gordon et al., 2013; Farzin, 2008; Ma et al., 2013; Munizaga et al., 2010; 2014; Nassir et al., 2011; Wang, 2010; Zhao, 2004; Zhao et al., 2007) as follows:

1. The PT stop is located on the route after tapping on the smart card.

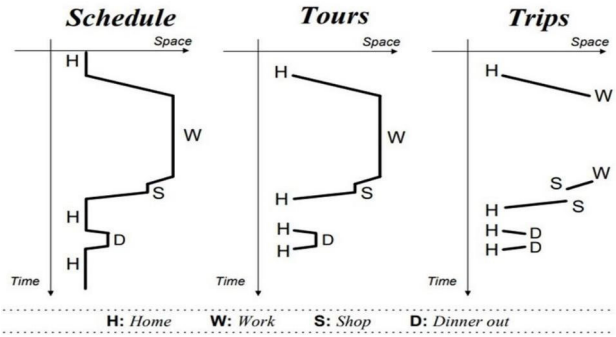


Fig. 1 Transport behavior of a passenger (Ben-Akiva, 2008)

2. It is within walking distance of the starting point of the next trip. Walking distance is calculated as the Euclidean length of a straight line between stops (the end of the current trip and the start of the next trip).
3. Passengers end their last trip of the day at the stop where they started their first ride of the day.

The work of Barry et al. (2009) was the first to formulate and test these assumptions. At the initial stage, their research was limited to metro stations; in 2009, they expanded the scope to bus routes.

Passengers made some trips using a mode of transport where an e-ticket is not used to pay for travel, such as taxis. Such gaps in travel chains need to be localized. To identify such gaps (unconnected trips) the following criteria are applied (Fadeev and Alhusseini, 2021a; 2021b):

- single trip per day;
- trips, in case of which it is not possible to determine the alighting time of the first trip because two consecutive tapping on validations occur at the same stop because two or more passengers pay with the same card;
- last trips of the day for which the alighting time cannot be determined because the first and last ticket validations of the day are performed at the same stop (Barry et al., 2009).

The general approach for determining the alighting time of a trip is to estimate the moment of arrival of the vehicle at the PT stop using the AVL system data or according to the route schedule.

The developed algorithms for generating passenger trips make it possible to interpret from 60% to 88% of e-ticket tapping on procedure (Table 1). The presence of unrecognized transactions is because as mentioned above, some trips are performed without using an electronic ticket. As a result, trips' chain falls out and a certain part of the operations cannot be interpreted.

Table 1 Sample's volume and the share of interpreted e-ticket validations from different works

Author	Sample size	Interpreted operation, %
Alsger et al. (2016)	161 446	76–84
Alsger et al. (2015)	473 525	88
Nassir et al. (2011)	84 413	61
Zhao et al. (2007)	2 500 000	72
Cui (2006)	2 736 454	79
Fadeev and Alhusseini (2021b)	6 340 518	65

When calculating the demand for public transport, it is required to take into account, firstly, misinterpreted tapping on operations and, secondly, the trips of passengers who pay for travel in a different way (not with electronic tickets). This problem was considered by Fadeev and Alhusseini (2021a). To calculate transport demand, it is necessary to establish the share of trips paid for with electronic travel tickets. In the paper of Fadeev and Alhusseini (2021a), it is shown that the share of trips paid for with electronic tickets should be determined for each route separately.

Misinterpreted tapping on operations is accounted by appropriate balancing factors.

After formatting passengers' trips, the travels of the passenger can be determined, i.e., movements from a starting point (for example, home) to a destination (for example, work), which can be performed by several trips with transfers.

Calculation of passenger's travels is performed based on successive trips which are ordered by time. For each pair of trips, the transfer conditions are analyzed and, if they are satisfied, the passenger's travel is formed (Alsger et al., 2015; Fadeev and Alhusseini, 2021a; Wang, 2010).

In works of Fadeev and Alhusseini (2021a; 2021b), a method for determining (restoring) passengers' travels by public transport have been introduced using an intellectual analysis of the electronic tickets tapping on operations, which in contrast to previous studies, provides:

- processing of e-ticket tapping on operations both at the beginning and at the end of the trip;
- calculation of passenger flow parameters, transport supply and O-D matrix taking into account unrecognized tapping on operations;
- representativeness assessment of the formed passenger's trips to the general demand on public transport.

In this method, it is considered that the PT stop, before the payment for the trip, is the beginning of the passenger's trip. However, in practice, some passengers can pay for the trip after passing one or two PT stops. As a result, there are errors in the determination of passenger trip.

This paper presents a method for calculating passengers' trips and travels by public transport from the operations of tapping on electronic travel tickets, which takes into account the mentioned disadvantage: it is considered that the passenger can pay for the fare anywhere on the route, not necessarily immediately after boarding the vehicle.

2 Methodology for calculating passengers' trips and travels by public transport from the operations of tapping on electronic travel tickets

It is necessary to determine the set of passengers' trips (P) and travels (H) made by using an electronic ticket.

On the transport network, there are stop points W, through which routes M pass, each of them consisting of two terminals (stations) and a sequence of intermediate stops. Movement along the routes is carried out in two directions: forward and reverse, i.e., there are two trips in the route, which are described by a pair of stop sequences and its trajectory depending on the direction of movement $k = \{a, b\}$ (where a, b are the forward and reverse directions, respectively):

$$M = \{M^k : k = a,b\};\ M \subset M;\ |M| = m; \quad (1)$$

$$M^k = \{w_1,\ldots,w_i,\ldots,w_t\};\ M^k \subset M;\ w_i \subset W. \quad (2)$$

The AFC system records the execution time of the k-th tapping on operation of the e_k e-ticket, the route and the vehicle by which the transportation was carried out. There is a sequence of tapping on operations for an electronic ticket $E = \{e_1, e_2, \ldots, e_k, \ldots\}$, each element of which reflects the passenger trip p_k on the k-th journey.

In a trip, the boarding point for the trip is located before the point at which the operation of ticket validation was performed, the end of the trip (alighting) - after the point of electronic ticket validation (see Fig. 2).

Forming a set of PT stops W_k^- which are located in the k-th journey before the operation of tapping on the

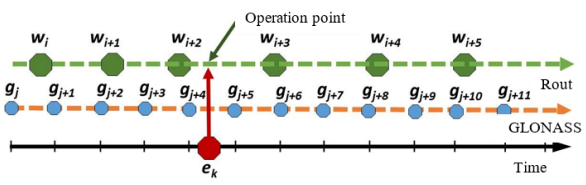


Fig. 2 Calculation scheme of passenger trip, where w_i, w_{i+1}, … are the PT stops of the route; g_i, g_{i+1}, … are the navigation marks of the satellite positioning system; e_k is the e-ticket validation

electronic ticket, and W_k^+ - after the validation, where $W_k^- \subset W$; $W_k^+ \subset W$. Thus, the boarding stop of passenger trip is an element of the set W_k^-, and the alighting stop is an element of the set W_k^+.

To determine tapping on points and alighting stops of the trips, the navigation marks from satellite positioning of the transport fleet are used, which contain the following data necessary to solve the problem under consideration: vehicle number, speed, time and coordinates (Latitude and Longitude) of the navigation mark.

The passage through the PT stop is fixed by navigation marks in its zone, which is set by the radius. For that, the trajectory of the vehicle's movement is formed:

$$R = G \times M = \{(g,w): g \in G, w \in M, l(g,w) \le l_w\}, \quad (3)$$

where $l(g,w)$ is the distance between the navigation mark (g) and the PT stop (w); l_w is the radius of the PT stop area.

The speed of the vehicle recorded in the navigation mark in the area of the PT stop may be greater than zero.

Usually, it is not always possible to establish the fact that the vehicle will stop at navigation marks. Fig. 3 shows that if the vehicle stop time does not exceed the interval between navigation marks, there may be cases in which a non-zero vehicle speed will be recorded in the navigation marks.

In some cases, the interval between navigation marks reaches 30 s considering that boarding and alighting at the stop was made regardless of the vehicle speed at the navigation mark.

The radius of the PT stop zone depends on several factors. It should be large enough to compensate for possible coordinate errors from the satellite positioning system, as well as the technological features of the transport station (e.g., bus stop), which may have several stops for vehicles. If the radius is insufficient, the navigation marks of the satellite system may not fall into the area of the PT stop. On the other hand, a large radius may cause false fixation of a PT stop that is not in the route of the vehicle.

Fig. 4 shows a diagram of a vehicle trajectory element. In the example under consideration, four navigation marks are located in the area of the PT stop. According to the

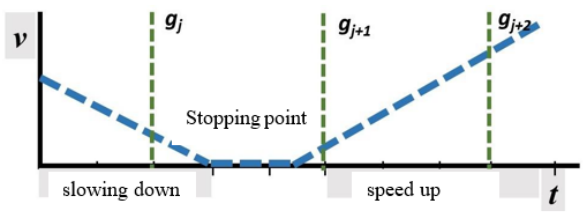


Fig. 3 Transport vehicle movement through the PT stop, where g_i, … are the navigation marks of the satellite positioning system

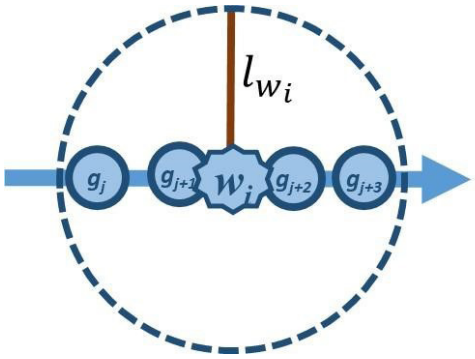


Fig. 4 Elements of the transport vehicle trajectory, where w_i is the PT stop; g_j, … are the navigation marks in the area of the PT stop; l_{w_i} is the Radius stop-point's zone

j-th navigation mark, the time of arrival at the PT stop is determined. Accordingly, the time of departure from the PT stop is set according to the (j + 3)-th mark.

In accordance with this scheme, the error in determining the time of arrival at the PT stop or departure from the PT stop does not exceed the interval between navigation marks.

As mentioned before, on urban transit, usually, payment for travel is carried out by a single tapping on or tapping off electronic tickets. In the previously developed algorithms (Fadeev and Alhusseini, 2021a; 2021b), one of the trip points determined by payment operation: trip's start point - if it is a tapping on operation (payment is made at the beginning of the trip); trip's end point - if it's a tapping off operation (payment is made before leaving the transport). These algorithms do not take into account cases when a passenger passes more than one stop before paying.

In this paper, for the calculation of passenger trip, an approach is applied based on the set of feasible options X_{k+1}^k between completing k-th trip and starting (k + 1)-th trip:

$$X_{k+1}^k = W_k^+ \times W_{k+1}^-;\ l\left(w_{kj}^+, w_{k+1i}^-\right) < L_p;\ w_{kj}^+ \subset W_k^+;\ w_{k+1i}^- \subset W_{k+1}^-, \quad (4)$$

where $l\left(w_{kj}^+, w_{k+1i}^-\right)$ is the Euclidean distance between PT stops w_{ki}^+ and w_{k+1i}^-; L_p is the typical walking distance.

Possible variants (Fig. 5) of the k-th passenger trip are the Cartesian product of the sets X_k^{k-1} and X_{k+1}^k. In the result set, it is required to determine the element that corresponds to the actual passenger trip. To do that, in the chain of passenger trips, the connectivity with the previous and subsequent trips is estimated by J criteria $y_1 = f_1(x), \ldots, y_J = f_J(x)$, which make up the vector of estimates. Z_{i^-} is the scale (set of values) of criterion f_i. We will assume that the criteria are positively oriented, i.e., as the values of each criterion increase, the preferences increase.

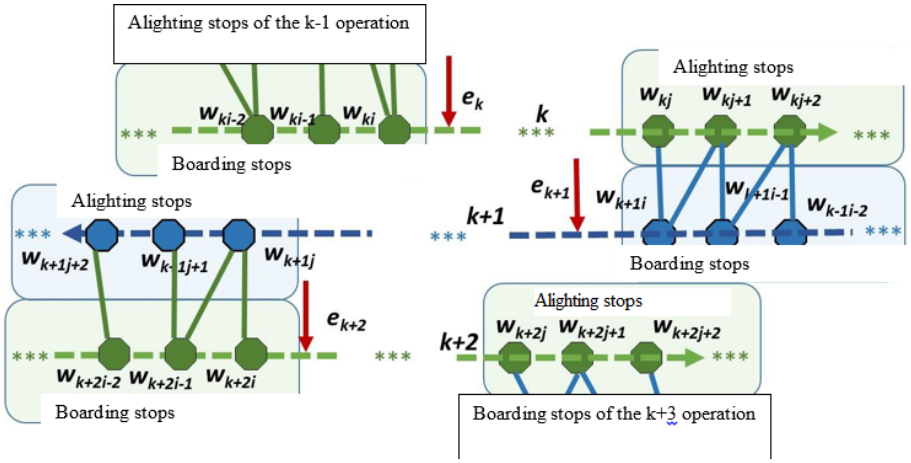


Fig. 5 Scheme of how variants of passenger correspondence been generated

In the problem under consideration, the following indicators are used to evaluate the elements of the set X_{k+1}^{k}:

- Euclidean distance between the alighting stops of the previous trip and the boarding stops of the next trips (walking distance);
- stop point number from the tapping on operation (it is considered that the closer the stop point is to the e-ticket tapping on operation, the higher the probability of starting the passenger trip from this point; in most cases, the passenger pays for the fare immediately after boarding the vehicle);
- frequency of use of the PT stop by the passenger (the passenger has clearly expressed places of attraction, for example, home, work, etc.).

To solve the selection problem, the vector criterion is "contracted" (Podinovsky and Potapov, 2013). An additive Eq. (5) or multiplicative Eqs. (6), (7) convolution of criteria is used (Fetinina et al., 2003):

$$F(f|v)=\sum_{j=1}^{J} v_j f_j \tag{5}$$

$$F(f|v)=\prod_{j=1}^{J} v_j f_j \tag{6}$$

$$F(f|v)=\prod_{j=1}^{J} v_j^{f_j}\,. \tag{7}$$

In the general case, the criteria can have different scales (for example, in our case, the walking distance and the number of the PT stops from the tapping on operation are measured differently). The criteria should be brought to a comparable form, i.e., normalized (Podinovsky and Potapov, 2013). Normalized criteria are dimensionless, their values, usually, are within the same limits, for example, from 0 to 1. Linear normalization functions are usually used (Fetnina et al., 2003; Podinovsky and Potapov, 2013), for example:

$$\varphi(f_i)=f_i/f_i^{*};\ \varphi(f_i)=(f_i-f_{i*})/(f_i^{*}-f_{i*}), \tag{8}$$

where f_i^{*}, f_{i*} are the largest and smallest values of the criterion $f_i(x)$, respectively.

For some indicators, the utility of the evaluated option increases as their values decrease. For example, in the problem under consideration, the most preferable option is the one with the smallest walking distance for a passenger. For such cases, a linear normalization function of the following form can be applied:

$$\varphi(f_i)=1-f_i/f_i^{*}\,. \tag{9}$$

The following functions were used for normalizing the indicators:

1. the walking distance between the alighting and boarding stops

$$\varphi(l_i^{p})=1-\frac{l_i^{p}}{2L_p},\ l_i^{p}\le 2L_p; \tag{10}$$

2. stop number from the tapping on operation

$$\varphi(n_i^{e})=1-\frac{n_i^{e}}{n^{e*}},\ n_i^{e}\le n^{e*}; \tag{11}$$

3. frequency of use of the PT stop (place of attraction for the passenger)

$$\varphi(w_i)=\frac{n_i^{w}}{n^{w*}}, \tag{12}$$

where n^{e*} is the the largest stop point number from the tapping on operation; for PT stops, the number of which is

greater than n^{e*}, criterion Eq. (11) takes on a zero value; n_i^w, n^{w*} is the number of operations with the *i*-th PT stop to the total number of passenger trips by e-ticket.

The use of double walking distance in Eq. (10) is explained in Fig. 6. It takes into account the movements between the passenger's point of attraction and the PT stops of boarding and alighting.

Thus, there are *I* variants X_{k+1}^k of connect the *k*-th and (*k* + 1)-th passenger trips. The preferred option is the one with the highest value of the aggregated criterion (additive convolution of criteria is used):

$$K^{\Sigma} = v_l \varphi\left(l_i^p\right) + v_n \varphi\left(n_i^e\right) + v_w \varphi\left(w_i\right) \Rightarrow \max, \tag{13}$$

where v_l, v_n, v_w are the criteria weights in Eqs. (10)–(12).

Consider the problem of determining the values of the criteria weight coefficients v_l, v_n, v_w which determine the choice of passenger trips generated from the electronic travel ticket operations. It is obvious that the resulting set must correspond to the general population of passenger trips by public transport. To estimate the general population, we use the parameters y_i, $i = \underline{1,z}$, which can be determined (measured) with sufficient accuracy. The $f_i\left(v_l, v_n, v_w\right)$, $i = \underline{1,z}$ is a function for calculating the *i*-th parameter from passenger trips paid for with an e-ticket. According to the least squares method (Johnson and Leone, 1980):

$$\sum_i e_i^2 = \sum_i \left[y_i - f_i\left(v_l, v_n, v_w\right)\right]^2 \Rightarrow \min. \tag{14}$$

It is necessary to determine the minimum of the function (Eq. (14)) by the variable weight coefficients v_l, v_n, v_w. To solve this problem of finding the unconditional extremum of a function of several variables, numerical methods are used (Prokopenko, 2018), for example, coordinate-wise descent, gradient, steepest descent, conjugate gradients, etc.

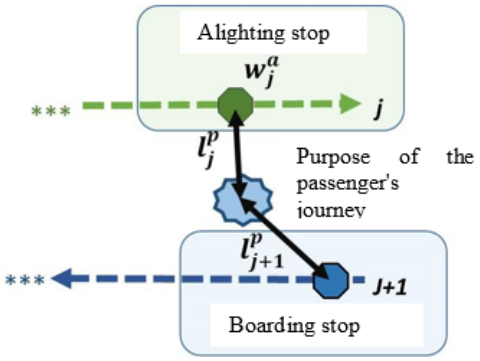


Fig. 6 Scheme estimated passenger-walking distance between alighting PT stop and next boarding one

3 Algorithm for calculating passenger trips and journeys by using electronic travel tickets

The formation of passenger trips from e-ticket tapping on operations involves the processing of big data: tens of millions of e-ticket tapping on operations and hundreds of millions of navigation data. In this work, for such an array of information, MS SQL Server is used as modern relational database management systems (DBMS).

The calculation of passenger trips consists of five stages.

Stage 1. The formation of the trajectory of transport vehicles movement is carried out from the navigation marks of the satellite navigation system in accordance with Eq. (3). For each vehicle, a query that implements Eq. (3) is executed using the SQL language. From the set of navigation marks, elements are selected that are located in the zone of PT stops of the served route (see Fig. 4). The trajectory of the movement of vehicles is described as:

$$R\left(A, T, W, M^r, M^k, M^i\right), \tag{15}$$

where *T* is the time of the navigation mark; *A* is the vehicle; M^r, M^k are the route number and direction of movement, respectively; *W* is the PT stop; M^i is the number of PT stops in the route.

Stage 2. Calculation of admissible connectivity options between end points of the current (*k*-th) trip with start points of the next (*k* + 1)-th trip

For an electronic travel ticket, there are many tapping on operations, which are described by the following:

$$E\left(I^o, I^t, T, M^r, M^k, A\right), \tag{16}$$

where I^o is the identifier of the electronic ticket-tapping operation I^t; *T* is the operation time; M^r, M^k are the route number and direction of movement, respectively; *A* is the vehicle identifier.

An element of this set is denoted as $e\left(i^o, i^t, t, m^r, m^k, a\right)$ or e_k (*k*-th tapping on operation).

The route number ($m_{e_k}^r$) is known from the e-ticket operation. The travel direction of the passenger $m_{e_k}^k$ is determined by the direction of movement along the route, which is established from the trajectory of the vehicle (formed in step 1).

According to Eq. (4), for the current (*k*-th) passenger trip, a set of admissible connectivity options X_{k+1}^k of the endpoints of the current (*k*-th) trip with the start points of the next (*k* + 1)-th trip is formed.

Stage 2 consists of the following operations:

1. Determination of the nearest PT stop $w_{k1}^- \left(w_{k1}^- \subset W_k^-\right)$ located before the point e_k, where the e-ticket

tapping on operation took place and the first PT stop after this point $w_{k1}^{+}\left(w_{k1}^{+} \subset W_k^{+}\right)$. If the tapping on operation is performed at the PT stop, $w_{k1}^{-} = w_{k1}^{+}$. The w_{k1}^{-} and w_{k1}^{+} are determined from the actual trajectory of the vehicle by the time of the operation:

Nearest PT stop before tapping on operation

$$\max\left(t^{r_a}\right);\ t^{r_a} \le t^{e_k};\ r_a \subset R_a \tag{17}$$

First stop after tapping on operation

$$\min\left(t^{r_a}\right);\ t^{r_a} \ge t^{e_k};\ r_a \subset R_a \tag{18}$$

where r_a is the actual trajectory of the a-th vehicle R_a; t^{e_k} is the time of the k-th tapping on operation.

2. By Eq. (5), a set of admissible connectivity options X_{k+1}^{k} of endpoints of the current (k-th) trip with start points of the next (k + 1)-th trip is formed, which are described as follows:

$$X\left(I^x, I_k^v, W_k, M_k^i, M_k^v, I_{k+1}^v, W_{k+1}, M_{k+1}^i, M_{k+1}^v, L^p\right), \tag{19}$$

where I^x is the variant identifier; I_k^v, I_{k+1}^v is the identifier of the k-th and (k + 1)-th tapping on operation, respectively; W_k, W_{k+1} is the identifier of the PT stop of the k-th and (k + 1)-th tapping on operation, respectively; M_k^i, M_{k+1}^i is the number of the PT stop along the route of the k-th and (k + 1)-th tapping on operations, respectively; M_k^v is the number of the nearest PT stop located in the route after the k-th operation; M_{k+1}^v is the number of the nearest PT stop located in the route before the (k + 1)-th operation; L^p is the Euclidean distance between points W_k, W_{k+1}.

3. For each element of the set X_{k+1}^{k}, the values of the evaluation criteria are calculated by Eqs. (10)–(12), which are represented by the following:

$$\Phi\left(I, \Phi^l, \Phi^n, \Phi^w\right), \tag{20}$$

where Φ^l, Φ^n, Φ^w is the value of the corresponding evaluation criterion: walking distance, number of the PT stop from the tapping on the operation and the proportion of passenger use of the PT stop, respectively.

Stage 3. Calculation of passenger trips. In accordance with Eq. (13), for each element of the set X_{k+1}^{k}, an aggregated criterion is calculated and based on its largest value, the endpoint of the k-th and the start point of the (k + 1)-th passenger trips are selected. The weights of the criteria v_l, v_n, v_w are assumed to be known.

As a result, passenger trips paid for with electronic ticket are formed, which are described as:

$$P\left(I^v, M^r, M^k, W^a, T^a, W^d, T^d, L\right), \tag{21}$$

where I^v is the identifier of the tapping on operation; M^r, M^k are the route number and direction of movement, respectively; W^a, W^d is the identifier of the PT stop for the beginning and end of passenger trip, respectively; T^a, T^d are the start and end times of passenger trip, respectively; L is the distance of the trip.

The distance of a passenger trip is calculated as the sum of the lengths of route parts between the points of beginning and end of trip in accordance with the journey of the vehicle. The start and end time of the trip is determined from the actual trajectory of the vehicle.

As mentioned above, a passenger can take trips that are not recorded in e-ticket operations, such as by taxis. In this case, the chain of trips based on an electronic ticket data is interrupted, so some of the tapping on operations remain misinterpreted.

Stage 4. Calculation of balancing coefficients to account for misinterpreted passenger trips.

For each misinterpreted passenger trip, the location of the e-ticket tapping on operation is known. The balancing coefficients are calculated based on the number of passengers boarding within the transport analysis zones (TAZ) of the tapping on operation. The balancing coefficients are calculated for each misinterpreted validation operation as follows:

1. Calculate the number of boarding n^a and alighting n^d of passenger journeys at the transport area of the e-ticket tapping on operation, which is determined by radius equal to the L_p allowed walking distance.
2. When calculating the balancing coefficient, the condition taken into account is that the number of passenger boarding from the transport zone must correspond to the number of passenger alighting at the transport zone, i.e., $n^a = n^d$.

The balancing coefficient is determined for the interpreted e-ticket operations as follows:

$$\phi_k = \phi_k' + \frac{1}{n^a},\ \text{if } n^a < n^d,\ l\left(e_k, p_k^{i^a}\right) \le L_p, \tag{22}$$

$$\phi_k = \phi_k' + \frac{1}{n^d},\ \text{if } n^a < n^d,\ l\left(e_k, p_k^{i^d}\right) \le L_p, \tag{23}$$

$$\begin{cases} \phi_k = \phi_k' + \dfrac{1}{2n^a},\ \text{if } n^a = n^d,\ l\left(e_k, p_k^{i^a}\right) \le L_p \\ \phi_k = \phi_k' + \dfrac{1}{2n^d},\ \text{if } n^a = n^d,\ l\left(e_k, p_k^{i^d}\right) \le L_p \end{cases}, \tag{24}$$

if $n^a \neq 0$ and $n^d \neq 0$, where ϕ'_k is the current value of the balancing factor of the k-th trip for e-ticket (at the beginning of the calculation $\phi'_k = 1$); n^a is the number of e-ticket trips with the boarding point at an allowed walking distance from e_k; n^d is the number of e-ticket trips with the alighting point within allowed walking distance from e^k; $l\left(e_k, p_k^{i^a}\right)$ is the distance between PT stops e^k and $p_k^{i^a}$ (Euclidean length of a straight line between PT stops).

In accordance with Eqs. (22)–(24), each misinterpreted operation is balanced through the interpreted operations of the transport region as follows:

- the number of boarding is increased by 1 if there are more alighting n^a than boarding n^d;
- if there are fewer alighting than boarding, the number of alighting increases by 1;
- if $n^a = n^d$, the number of boarding and alighting is increased by 0.5.

For the cases where or , the balancing factor is calculated within the itinerary by appropriately adjusting the share of e-ticket trips.

Stage 5. Based on passenger trips (Eq. (21)), passenger journeys are calculated. The calculation algorithm is given in the works of Fadeev and Alhusseini (2021a; 2021b).

Determining the demand for public transport. Transport demand is presented in the form of matrices of origin - destination (O-D matrix), which are formed as a kind of averaged model of the population's need for movement through the transport network:

$$Q = \left\| T_{ij} \right\|; \; i, j = 1, \ldots, n, \tag{25}$$

where T_{ij} is the number of passenger movements made during the period of interest between points i, j.

Consider the procedure for determining the O-D matrix from processing electronic travel tickets data. When calculating the O-D matrix, it is necessary to take into account the share of trips paid for with electronic tickets and unrecognized operations. The calculation of the number of trips between points i, j and is carried out as follows:

$$T_{ij} = \sum_{m} \sum_{\substack{p_k^{i^a} = i \\ p_k^{i^d} = j \\ p_k^{m^r} = m}} \frac{\phi_k}{\alpha_m^r} \tag{26}$$

where ϕ_k is the balancing factor of the k-th passenger trip, using of which misinterpreted operations are taken into account ($\phi_k \geq 1$); p_k is the k-th passenger trips; α_m^r is the share of operations by e-tickets on the m-th route (Fadeev and Alhusseini, 2021a) ($\alpha_m^r \leq 1$);

$$\alpha_m^r = \left(Q'_m - q_m\right) / Q_m, \tag{27}$$

where Q'_m, Q_m is the number of passengers holding electronic tickets and the total number of passengers on the m-th route, respectively; q_m is the number of unbalanced misinterpreted trips along the m-th route.

4 Practical implementation

The methodology was carried out on the data of the Krasnoyarsk public transport system for April 2019, provided by the municipal state institution of the city of Krasnoyarsk "Krasnoyarskgortrans". This information was used in the works of Fadeev and Alhusseini (2021a; 2021b) with a description of the previous version of the methodology for monitoring public transport demand by processing electronic travel tickets data, which makes it possible to objectively evaluate the effectiveness of the new version of the methodology.

"Krasnoyarskgortrans" provided the following data in MS SQL Server DBMS format:

- navigation data of satellite positioning of vehicles (Table 2);
- description of the network route (Tables 3 and 4);
- e-ticket tapping on operations (Table 5).

Data volume: 144 million satellite positioning navigation marks and more than 6 million e-ticket tapping on operations.

A fragment of the actual trajectory of the movement of the vehicle through the PT stops of the served route (stage 1 of calculations) is shown in Table 6. In the calculations, the radius of the PT stop zone was taken 175 m, if the interval between navigation marks is more than 30 s, and 100 m for smaller values of the interval.

Table 2 Navigation marks of the satellite positioning system (Fragment), where *Lat*, *Long* are the coordinates (latitude and longitude); *V* is the vehicle speed*

A	M^r	*T*	*Lat*	*Long*	*V*
EB 687	64	20.04.2019 5:15:19	56.04376	92.783055	7
EB 687	64	20.04.2019 5:15:29	56.04374	92.782662	10
EB 687	64	20.04.2019 5:15:39	56.04387	92.78243	9
EB 687	64	20.04.2019 5:15:49	56.04395	92.782273	6
EB 687	64	20.04.2019 5:15:59	56.04405	92.78205	2
EB 687	64	20.04.2019 5:16:09	56.04419	92.781848	7
EB 687	64	20.04.2019 5:16:19	56.04432	92.781655	7
EB 687	64	20.04.2019 5:16:29	56.04441	92.78137	7
EB 687	64	20.04.2019 5:16:39	56.04433	92.781105	7
EB 687	64	20.04.2019 5:16:49	56.04418	92.780775	8

* The table has a sequel, and this is only a fragment.

Table 3 PT stops list (Fragment), where *id* is the identifier; *Nm* is the name; *Lat*, *Long* are the coordinates (latitude and longitude)*

id	*Nm*	*Lat*	*Long*
25	School (Sudostroitelnaya St.)	55.982063	92.833626
27	LDK	55.983582	92.883812
29	Student (Sverdlovskvaya st.)	55.983150	92.879295
32	Art school	55.981358	92.862823
33	Khlebozavod (Sverdlovskaya st.)	55.980495	92.856873
35	Yubileinaya (Sverdlovskaya street)	55.979836	92.850945
37	Oktyabrskaya (Sverdlovskaya street)	55.978786	92.841217
39	Station "Yenisei"	55.978127	92.835556
41	OJSC "Kraspharma" (Sverdlovskaya street)	55.976963	92.820518

* The table has a sequel, and this is only a fragment.

Table 4 Route description (Fragment), where *L* is the length of the section L^{Σ} is the accumulated length from the beginning of the route*

M^r	M^k	M^i	*W*	*L*	L^{Σ}
10	A	8	66	0.2644	3.0087
10	A	9	95	0.2760	3.2848
10	A	10	116	0.4940	3.7788
10	A	11	118	0.8018	4.5806
10	A	12	120	0.5833	5.1640
10	A	13	122	0.4393	5.6033

* The table has a sequel, and this is only a fragment.

Table 5 Fragment of e-ticket tapping on operations*

I^o	I^t	*T*	M^r	M^k**	*A*
13119511	100010324	10.04.19 15:34:46	58		M 931 KK
13119512	100010324	10.04.19 17:23:06	95		EB 512
13119513	100010324	17.04.19 15:09:36	85		EE 183
13119514	100010324	18.04.19 9:23:22	61		K 119 OE
13119515	100010324	23.04.19 8:51:17	61		K 721 HP
13119516	100010324	02.04.19 17:58:22	95		EB 986
13119526	100010324	12.04.19 14:41:06	95		EB 988
13119527	100010324	14.04.19 14:49:29	95		EB 976
13119528	100010324	18.04.19 10:18:40	94		X 724 HX

* The table has a sequel, and this is only a fragment.
** The direction of movement along the route M^k is determined in the process of calculation from the trajectory of the vehicle (Table 6).

The developed computer program for calculating passenger trips by processing electronic travel tickets data consists of the main sections presented in Table 7.

Table 8 shows a fragment of the possible connectivity (linked) options X_{k+1}^k between the endpoints of the current trip (*k*) and the start points of the next trip (*k* + 1) obtained as a result of the 2nd stage of the calculation.

Table 6 Fragment of the vehicles movement*

A	*T*	*W*	M^r	M^k	M^i
3	21.04.19 12:16:00	335	6тр	A	19
3	21.04.19 12:16:00	336	6тр	B	11
3	21.04.19 12:16:10	335	6тр	A	19
3	21.04.19 12:16:10	336	6тр	B	11
3	21.04.19 12:16:20	335	6тр	A	19
3	21.04.19 12:16:20	336	6тр	B	11
3	21.04.19 12:16:30	335	6тр	A	19
3	21.04.19 12:16:30	336	6тр	B	11

* The table has a sequel, and this is only a fragment.

Table 7 Sections of a computer program for calculating passenger trips by processing electronic travel tickets data for urban public transport

	Description	Calculation time, hour*
Stage 1	Formation of the actual trajectory of the vehicles' movement through the PT stops of the served route.	5
Stage 2	Determination of valid connectivity options X_{k+1}^k between the endpoints of the current trip (*k*) and the start points of the next trip (*k* + 1). Calculation of estimated indicators.	50
Stage 3	Calculation of passenger trips.	4
Stage 4	Calculation of balancing coefficients to account the misinterpreted e-ticket transactions.	4
Stage 5	Calculation of passenger journeys.	5

* Time of the considered test case on a computer Intel Core i7 2.80 GHz, 16.0 GB RAM, Windows 10 Pro

Table 8 Possible options for connecting X_{k+1}^k between the endpoints of the current trip and the start points of the next trip (fragment)*

I^x	I_k^v	W_k	M_k^i	M_k^v	I_{k+1}^v	W_{k+1}	M_{k+1}^i	M_{k+1}^v	L^p
29030	13119546	741	29	9	13119524	742	30	30	0.04
30029	13119546	743	30	9	13119524	744	29	30	0.05
28030	13119546	739	28	9	13119524	742	30	30	0.45
31028	13119546	745	31	9	13119524	746	28	30	0.04
30030	13119546	743	30	9	13119524	742	30	30	0.59
32028	13119546	751	32	9	13119524	746	28	30	0.24
32027	13119546	751	32	9	13119524	752	27	30	0.07
31029	13119546	745	31	9	13119524	744	29	30	0.53
29029	13119546	741	29	9	13119524	744	29	30	0.55

* The table has a sequel, and this is only a fragment.

For each connectivity option, evaluation indicators are calculated (Table 9). With known weight coefficients v_l, v_n, v_w, it is possible to calculate the integral criterion Φ^{Σ}, according to the largest value of which the connectivity option X_{k+1}^k is selected and on this basis, passenger trip is formed (see Fig. 5).

Table 9 Results of calculating the evaluation criteria values (fragment)*

I^x	Φ^l	Φ^n	Φ^w	Φ^Σ
29030	0.96	1.00	0.03	1.98
30029	0.95	0.80	0.00	1.75
28030	0.55	1.00	0.03	1.57
31028	0.96	0.60	0.00	1.56
30030	0.41	1.00	0.03	1.43
32028	0.76	0.60	0.00	1.36
32027	0.93	0.40	0.00	1.33
31029	0.47	0.80	0.00	1.27
29029	0.45	0.80	0.00	1.25
33026	0.97	0.20	0.00	1.17
30028	0.51	0.60	0.00	1.11
31027	0.71	0.40	0.00	1.11
32029	0.29	0.80	0.00	1.09
34024	1.00	0.00	0.00	1.00

* The table has a sequel, and this is only a fragment.

To calculate passenger trip, it is required to determine the weight of coefficients v_l, v_n, v_w in such a way that the resulting set of passenger trips obtained from processing electronic travel tickets data corresponds to the maximum extent to the trips of the general population by public transport. We will evaluate the parameters of the general population based on the results of a field survey of passenger flows.

Field survey of passenger flows data, which was done by passengers' automated accounting, was received from the "Krasnoyarskgortrans". A sample survey of passenger flows on 5 routes was carried out in April 2019 (during the period in which e-ticket was processed). Passenger registration was carried out using special equipment mounted in vehicles. The volume of accounting is more than 281 thousand passengers, 6938 journeys were examined.

Forming two samples of passenger trips: From processing electronic travel tickets data (sample 1) and the field survey of passenger flows (sample 2). The square of the difference in the proportion of incoming and outgoing passengers at the PT stops of the routes is considered as a criterion for the compliance of these samples:

$$F\left(v_l, v_n, v_w\right) = \sum_i \sum_k \sum_j \left(q_{ikj}^v - q_{ikj}^*\right)^2 \Rightarrow \min , \qquad (28)$$

where q_{ikj}^v, q_{ikj}^* are the share of boarding or alighting passengers at the j-th stop of the k-th direction of the i-th route, determined from processing electronic travel tickets data and the field survey of passenger flows, respectively.

Thus, it is necessary to establish the minimum of function Eq. (28) with respect to the variable weight coefficients v_l, v_n, v_w. Let's study the function Eq. (28), for this we construct a graph of the dependence of the values of the function on the variables v_l, v_n, v_w.

As mentioned above, the calculation of passenger trips is carried out on the basis of admissible variants of connectivity between the endpoints of the current trip and the start points of the next trip. To study the function Eq. (28), stage 3 of the calculation of passenger trips is performed for various values of the variables v_l, v_n, v_w which vary from 0 to 2.5 (see Table 10, Fig. 7).

According to the results of the study, it can be concluded:

1. The minimum point of the function is when v_l = 1; v_n = 1; v_w = 0. The value of the integral criterion at the minimum point is 0.203.
2. It seems that the greatest influence is due to the factor of distance between endpoint of the previous and start point of the next trips (v_l). With the exclusion of the factor (v_l = 0), the criterion of compliance dropped by 2 times compared to the minimum point.
3. The number of the PT stop from the tapping on operation v_n is the second factor influencing the criterion of compliance. With v_n = 0, the compliance criterion dropped by 75% (from 0.355 to 0.203).
4. Function Eq. (28) does not depend on the estimate of the frequency of using the PT stop (see Table 10).

Using the obtained values of the weight coefficients v_l, v_n, v_w, the calculation of passenger trips was performed, the results of which are comparable with the previous version

Table 10 Function Eq. (28) results

v_i	$f\left(v_l \middle\| v_n = 1;\ v_w = 1\right)$	$f\left(v_n \middle\| v_l = 1;\ v_w = 1\right)$	$f\left(v_w \middle\| v_l = 1;\ v_n = 1\right)$
0	0.423	0.355	0.203
0.5	0.216	0.226	
1	0.204	0.204	0.204
1.5	0.219	0.205	
2	0.227	0.207	
2.5		0.207	
Extremum: v_l = 1; v_n = 1; v_w = 0			0.203

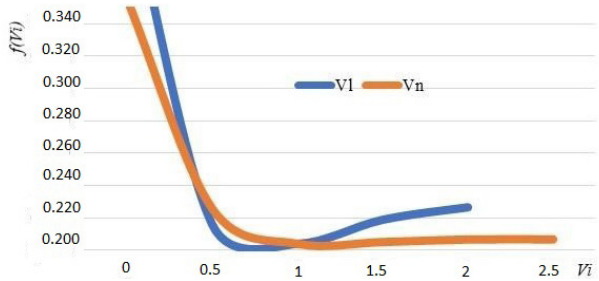


Fig. 7 Dependence the function Eq. (28) on the variables v_l, v_n

of the methodology (Fadeev and Alhusseini, 2021a; 2021b), which we will call Method 1, in contrast to the methodology considered in this article (Method 2). Table 11 shows the distribution of e-ticket tapping on operations by the days of April 2019, the number of interpreted operations by both methods. Table 11 shows that both methods provide the determination of almost the same number of passenger trips: the share of interpreted transactions is about 65%, if we exclude the last three days of the period. By the end of the period (month), there is a tendency to reduce the share of recognized transactions, which is explained by the absence in the database of tapping on transactions for the next month associated with trips of the current period.

Table 11 The results of processing tapping on operations of e-tickets in public transport of Krasnoyarsk city (April, 2019)

Day of the month	Total	Processed transactions			
		Interpreted (Method 1)		Interpreted (Method 2)	
		Amount	weight, %	Amount	weight, %
1	238789	152391	63.8	155108	65.0
2	253128	163643	64.6	165786	65.5
3	250250	163477	65.3	166549	66.6
4	253132	162679	64.3	165500	65.4
5	252721	160034	63.3	162598	64.3
6	164948	101464	61.5	102695	62.3
7	124539	77113	61.9	76973	61.8
8	238016	155675	65.4	156684	65.8
9	248460	163578	65.8	164200	66.1
10	247446	161197	65.1	162488	65.7
11	247844	163009	65.8	163103	65.8
12	242969	158770	65.3	158524	65.2
13	155113	101079	65.2	101240	65.3
14	116213	76060	65.4	75782	65.2
15	238363	155945	65.4	155781	65.4
16	255730	167073	65.3	166747	65.2
17	257936	168971	65.5	168661	65.4
18	239022	156863	65.6	156908	65.6
19	232266	149428	64.3	149298	64.3
20	160447	102859	64.1	102516	63.9
21	129981	85581	65.8	85938	66.1
22	241130	156714	65.0	157424	65.3
23	259047	164047	63.3	166277	64.2
24	235845	151504	64.2	153254	65.0
25	243887	156387	64.1	158725	65.1
26	235609	147355	62.5	149348	63.4
27	139348	85765	61.5	86398	62.0
28	116055	69071	59.5	69318	59.7
29	216090	125869	58.2	126819	58.7
30	231156	88322	38.2	85910	37.2
∑	6465480	4091923	63.3	4116552	63.7

As mentioned above, in Method 1, the start of passenger trips is considered to be the PT stop preceding the tapping on operation. This does not always correspond; in practice, passengers can travel one or two stops before paying for the trip.

On some routes, it is customary to collect the fare before the passenger exits the vehicle. In this case, the PT stop before the tapping on operation is the end of the passenger trip. Method 1 uses evaluation criteria to determine end-of-trip tapping on cases, which greatly complicates the algorithm.

In Method 2, there is no rigid binding to the tapping on operation of the trip start-end points. Fig. 8 shows the distribution of passenger trips by the number of the PT stops that have been passed until the validation of an electronic ticket. Fig. 8 shows that about 90% of passengers pay for the fare directly after boarding. 5.4% of passengers pass one stop before payment, 2.1% pass 2 stops.

Nowadays, in the city of Krasnoyarsk payment for travel at the end of the journey is used only on one or two routes on which small buses operate, and it's no more than 5% of e-ticket transactions. That is why it is not covered in this article.

Table 12 shows a fragment of passenger trips generated by the two methods under consideration. It can be seen from Table 12 that in some trips formed according to Method 2, the endpoints differ from Method 1. These are, for example, trips 13119511 and 13119515, the PT stops of which, determined according to Method 2, are located at a shorter walking distance to the beginning stop of the next trips compared to Method 1. In addition, some trips (e.g., 13119512) have different starting stops because in Method 2 the start of passenger trip is not necessarily the closest point to the e-ticket tapping on operation. Most of the passenger trips generated by Method 1 and Method 2 are the same, since, as

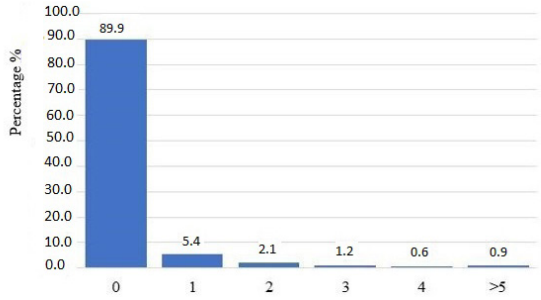


Fig. 8 Distribution of passenger trips by the number of the PT stops passed before the validation of the electronic ticket took place

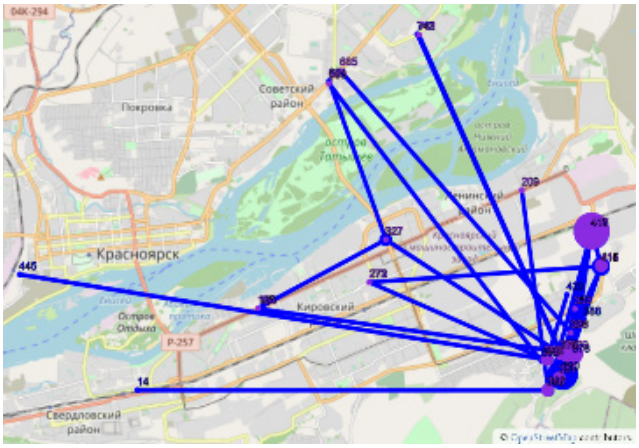
Table 12 List of generated trips using Methods 1 and 2 (fragment)

Methods	ID	Rout		Start		End		Distance, km	
				IdSt	Time	IdSt	Time	Ride	Walk
1	13119511	58	B	327	15:34	267	15:43	2.8	0.48
2	13119511	58	B	327	15:34	185	15:48	4.4	0.04
1	13119512	95	A	190	17:23	379	18:03	10.8	0.09
2	13119512	95	A	184	17:19	379	18:03	12.0	0.09
1	13119513	85	A	428	15:09	369	15:18	2.9	0.12
2	13119513	85	A	428	15:09	369	15:18	2.9	0.12
1	13119514	61	B	368	9:23	683	9:47	9.7	0.05
2	13119514	61	B	368	9:23	683	9:47	9.7	0.05
1	13119515	61	B	368	8:51	683	9:16	9.7	0.05
2	13119515	61	B	386	8:46	685	9:17	11.2	0.02
1	13119516	95	A	413	17:58				
2	13119516	95	A	415	17:56	379	18:04	2.3	0.09
1	13119517	85	B	382	8:22	416	8:45	7.0	0.02
2	13119517	85	B	382	8:20	416	8:41	7.0	0.02
1	13119518	78	A	417	14:41	353	14:45	2.0	0.33
2	13119518	78	A	417	14:40	353	14:45	2. 0	0.33
1	13119519	61	A	327	10:43	375	11:00	6.3	0.44
2	13119519	61	A	327	10:42	375	11:00	6.3	0.44
1	13119520	95	B	380	11:55	416	12:04	3.6	0.02
2	13119520	95	B	380	11:54	416	12:04	3.6	0.02

mentioned above, about 90% of passengers pay for the fare immediately after boarding the vehicle, which is reflected in both methods. The distance of passenger trip generated by Methods 1 and 2 differs insignificantly (by 1–2 km).

Fig. 9 shows the connection diagram of the boarding and alighting PT stops of passenger journeys of two electronic travel tickets. In the diagram, the radius of the circle is proportional to the number of arrivals or departures of the passengers, points of departure are indicated in pink, arrivals in blue. In the diagram, the PT stop with the largest number of departures (with the largest circle radius) is located near the passengers' place of residence. The PT stop with the largest number of arrivals is the main place of attraction for the passengers (work, study, etc.).

Fig. 10 shows the distribution of the number of passengers carried by weekday hours, determined from the validation of electronic tickets according to Methods 1 and 2 (2019) and a complete survey of passenger flows (2011) of public transport in the city of Krasnoyarsk. Fig. 10 shows that passenger trips paid for with electronic tickets correspond to the dynamics of passenger flows by hours of the day from the field survey. Some discrepancies are explained by changes in the structure of transport demand that have occurred since 2011.



(a)

(b)

Fig. 9 Scheme of connections between the boarding and alighting points of passenger trips from an electronic ticket transaction, (a) e-ticket 100010324, (b) e-ticket 100010341

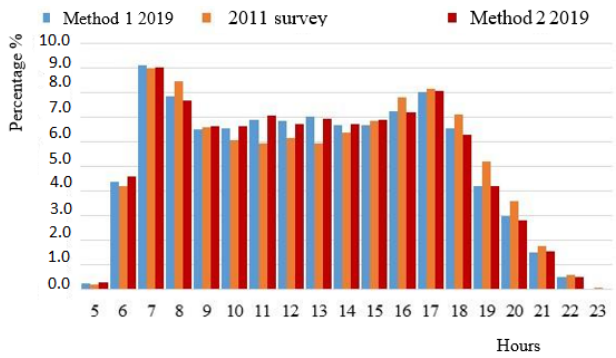


Fig. 10 Hourly distribution of passengers carried by urban public transport of the Krasnoyarsk city

Table 13 shows the distribution of passenger journeys by journey length. From Table 13, we can conclude that the share of trips obtained by Methods 1 and 2, as well as from field surveys of passenger flows, has the same dependence. The average travel distance for passengers is 6.29 km from processing e-ticket data and 6.66 km from

Table 13 Distribution of passenger traffic by trip length

Range, km	E-ticket operation 2019		Survey 2006	Survey 2011
	Method 1	Method 2		
0–5	49.9	49.8	49.5	46.3
5–10	32.2	32.2	29.2	32.0
10–15	12.1	12.1	13.2	14.4
15–20	4.2	4.2	5.6	5.6
20–25	1.3	1.3	1.9	1.4
25–30	0.3	0.3	0.5	0.2
30–35	0.0	0.0	0.1	0.0

the 2006 survey. The 5.6% difference between the average travel distance of a passenger in 2006 and 2019 is due to changes in the route system and public transport demand.

5 Representativeness of the results

It is necessary to determine the correspondence of passenger journeys, determined by processing electronic travel tickets data, to all journeys of passengers by public transport. The developed methodology, which allows to determine the compliance of the main parameters of the sample obtained from processing electronic travel tickets data, with the trips of the general population, is presented in the paper of Fadeev and Alhusseini (2021a).

The assessment of the representativeness of the trips sample obtained from processing electronic travel tickets data is carried out by comparing it with the results of an automated accounting of passengers, which was received from the municipal state institution of the city of Krasnoyarsk "Krasnoyarskgortrans".

Considering two samples of passenger trips: one obtained from processing electronic travel tickets data (sample 1) and the other from field survey of passenger flows (sample 2) we compare the distribution of boarding and alighting passengers along the length of the route.

The route is divided into k non-overlapping intervals. For each interval, the number of boarding and alighting passengers of samples 1 and 2 is calculated. Thus, two independent disconnected samples are formed. To compare them, we use the Student's t-test, which allows us to find the probability that both means in the samples belong to the same population.

Fig. 11 shows a histogram of the distribution of passengers on one route, which allows us to hypothesize that both samples belong to the same population.

Criteria statistics (Johnson and Leone, 1980):

$$t_e = \frac{\bar{x} - \bar{y}}{\sigma_{x-y}}, \tag{29}$$

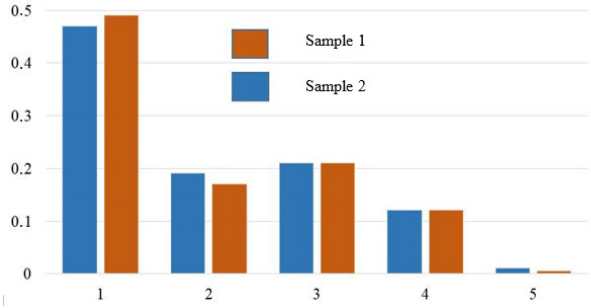


Fig. 11 Histogram of the passengers boarding number distribution along the length of route No. 26 (direct direction)

where $\bar{x}$, $\bar{y}$ are the arithmetic mean in the experimental and control groups; σ_{x-y} is the standard error of the difference between the arithmetic means:

$$\sigma_{x-y} = \sqrt{\frac{\sum(x_i - \bar{x})^2 + \sum(y_i - \bar{y})^2}{n_1 + n_2 - 1}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}, \tag{30}$$

We compare t_e with the theoretical value of Student's t-distribution t_k with the number of freedom degrees $n_1 + n_2 - 2$, the samples are equal at $t_e < t_k$.

The highest value of statistics (see Table 14) is 2.450 (route No. 26, direct direction, passengers getting off). Through analysis of the Excel data, we obtain a P-value of 0.02. Most of the criterion values do not exceed the critical value of 1.860.

Based on the foregoing, it can be concluded that passenger trips obtained from processing electronic travel tickets data are capable to estimate the parameters of public transport demand within the limits of permissible errors.

6 Conclusion

To ensure the mobility of the population, an urgent problem is the creation of a system for monitoring the demand on public transport, in order to constantly monitor the

Table 14 The results of calculating the value of Student's test for the compared samples

Rout	Direction	Boarding pass.	Alighting pass.	Rout	Direction	Boarding pass.	Alighting pass.
Bus				Trolleybus			
26	A	0.334	2.450	7т	A	0.465	0.553
26	B	0.603	1.219	7т	B	0.536	0.744
31	A	1.692	1.368				
31	B	1.067	2.378				
37	A	0.379	1.655				
37	B	0.522	0.980				
52	A	0.771	0.037				
52	B	1.205	0.079				

passenger flows, the parameters of which serve to justify management decisions on the formation of an optimal transport supply.

Existing methods for determining passenger flows, due to their complexity and limited efficiency, do not allow monitoring transport demand at the proper level. Today, technologies (Big data, Urban computing) based on the collection, integration and analysis of big data are widely used.

Within the framework of this approach, this article presents a solution to the problem of determining the trips of passengers by public transport by analyzing the information of electronic travel tickets (smart cards, transport cards, mobile phones or other electronic devices) tapping on operations which are recorded in the Automated Fare Collection system.

The developed method for calculating passenger trips from the operations of electronic travel ticket validations, integrated with the data of the global navigation satellite system, in contrast to previous studies, takes into account the practice of paying for travel at any point on the route, not necessarily immediately after boarding the vehicle.

A theoretically justified method for determining the demand for urban public transport, based on determining and evaluating the set of acceptable options for connecting a sequence of passenger trips, using a criterion formed from a vector of estimated indicators, allows for calculating the parameters of passenger trips, taking into account many factors that affect the choice of passenger's routes.

By using the developed methodology, about 65% of trips paid for with electronic travel tickets are interpreted.

The developed method for determining the weighting coefficients values of the evaluation indicators that determine the choice of passenger trips obtained from processing electronic travel tickets data, allows for calculating the set of passenger trips according to the criterion of compliance with the general population demand on public transport.

It has been proven that passenger trips obtained from processing electronic travel tickets data allows us to estimate the parameters of public transport demand within the limits of permissible errors.

The application of the developed methodology for calculating the demand for urban public transport from processing electronic travel tickets data ensures continuous monitoring of passenger flows, technical and operational indicators of the functioning of public transport and thus allows for implementing the concept of sustainable development of public transport by designing a transport supply that meets demand.

The direction of further research is:

- exploring the factors influencing on the choice of the actual passenger trip form the acceptable options, such as the equipment of PT stops, their location, for example, near shops, medical institutions, etc.;
- formation of a set of reporting forms (feedback forms) for use in the work of public transport management bodies and organizations;
- integration of existing software into the fare accounting system and traffic dispatch control to provide industrial technology for continuous monitoring of passenger flows in public urban transport.

Acknowledgments

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

References

Agard, B., Morency, C., Trépanier, M. (2006) "Mining public transport user behaviour from smart card data", IFAC Proceedings Volumes, 39(3), pp. 399–404. https://doi.org/10.3182/20060517-3-FR-2903.00211

Alsger, A. A., Mesbah, M., Ferreira, L., Safi, H. (2015) "Use of smart card fare data to estimate public transport origin–destination matrix", Transportation Research Record, 2535(1), pp. 88–96. https://doi.org/10.3141/2535-10

Alsger, A., Assemi, B., Mesbah, M., Ferreira, L. (2016) "Validating and improving public transport origin–destination estimation algorithm using smart card fare data", Transportation Research Part C: Emerging Technologies, 68, pp. 490–506. https://doi.org/10.1016/j.trc.2016.05.004

Barry, J. J., Newhouser, R., Rahbee, A., Sayeda, S. (2002) "Origin and destination estimation in New York City with automated fare system data", Transportation Research Record, 1817(1), pp. 183–187. https://doi.org/10.3141/1817-24

Barry, J. J., Freimer, R., Slavin, H. (2009) "Use of entry-only automatic fare collection data to estimate linked transit trips in New York City", Transportation Research Record, 2112(1), pp. 53–61. https://doi.org/10.3141/2112-07

Ben-Akiva, M. (2008) "Travel Demand Modeling", [pdf] Massachusetts Institute of Technology, Cambridge, MA, USA. Available at: https://ocw.mit.edu/courses/1-201j-transportation-systems-analysis-demand-and-economics-fall-2008/resources/mit1_201jf08_lec05/ [Accessed: 07 September 2020]

Chu, K. K. A., Chapleau, R. (2008) "Enriching archived smart card transaction data for transit demand modeling", Transportation Research Record, 2063(1), pp. 63–72. https://doi.org/10.3141/2063-08

Cui, A. (2006) "Bus passenger origin-destination matrix estimation using automated data collection systems", Master of Science in Transportation Thesis, Massachusetts Institute of Technology.

Fadeev, A. I., Alhusseini, S. (2021a) "Обследование пассажирских потоков путем анализа валидаций электронных проездных билетов" (Transit ridership survey by analysis validation of electronic pass tickets), The Russian Automobile and Highway Industry Journal, 18(1), pp. 52–71. (in Russian) https://doi.org/10.26518/2071-7296-2021-18-1-52-71

Fadeev, A. I., Alhusseini, S. (2021b) "Passenger trips analysis determined by processing validation data of the electronic tickets in public transport", IOP Conference Series: Materials Science and Engineering, 1061(1), 012001. https://doi.org/10.1088/1757-899X/1061/1/012001

Farzin, J. M. (2008) "Constructing an automated bus origin–destination matrix using farecard and global positioning system data in São Paulo, Brazil", Transportation Research Record, 2072(1), pp. 30–37. https://doi.org/10.3141/2072-04

Fetinina, E. P., Korablina, T. V., Solovyova, Y. A. (2003) "Типологические аспекты многокритериального выбора вариантов" (Typological aspects of multi-criteria choice of options), [pdf] Federal State Budgetary Educational Institution of Higher Education "Siberian State Industrial University", Novokuznetsk, Russia. Available at: https://www.freelancejob.ru/upload/546/69040338601917.pdf [Accessed: 24 June 2021] (in Russian)

Gordon, J. B., Koutsopoulos, H. N., Wilson, N. H. M., Attanucci, J. P. (2013) "Automated inference of linked transit journeys in London using fare-transaction and vehicle location data", Transportation Research Record, 2343(1), pp. 17–24. https://doi.org/10.3141/2343-03

Johnson, N. F., Leone, F. C. (1980) "Статистика и планирование эксперимента в технике и науке: методы обработки данных" (Statistics and Experimental Planning in Engineering and Science: Data Processing Methods), Mir, Moscow, Russia. (in Russian)

Li, D., Lin, Y., Zhao, X., Song, H., Zou, N. (2011) "Estimating a transit passenger trip origin-destination matrix using automatic fare collection system", In: Xu, J., Yu, G., Zhou, S., Unland, R. (eds.) Database Systems for Advanced Applications, Springer, pp. 502–513. ISBN 978-3-642-20243-8 https://doi.org/10.1007/978-3-642-20244-5_48

Ma, X., Wu, Y.-J., Wang, Y., Chen, F., Liu, J. (2013) "Mining smart card data for transit riders' travel patterns", Transportation Research Part C: Emerging Technologies, 36, pp. 1–12. https://doi.org/10.1016/j.trc.2013.07.010

Munizaga, M., Palma, C., Mora, P. (2010) "Public transport OD matrix estimation from smart card payment system data", In: Viegas, J. M., Macario, R. (eds.) General Proceedings of the 12th World Conference on Transport Research, Lisbon, Portugal, pp. 1–16. ISBN 9789899698604

Munizaga, M. A., Palma, C. (2012) "Estimation of a disaggregate multimodal public transport Origin–Destination matrix from passive smartcard data from Santiago, Chile", Transportation Research Part C: Emerging Technologies, 24, pp. 9–18. https://doi.org/10.1016/j.trc.2012.01.007

Munizaga, M., Devillaine, F., Navarrete, C., Silva, D. (2014) "Validating travel behavior estimated from smartcard data", Transportation Research Part C: Emerging Technologies, 44, pp. 70–79. https://doi.org/10.1016/j.trc.2014.03.008

Nassir, N., Khani, A., Lee, S. G., Noh, H., Hickman, M. (2011) "Transit stop-level origin–destination estimation through use of transit schedule and automated data collection system", Transportation Research Record, 2263(1), pp. 140–150. https://doi.org/10.3141/2263-16

Nunes, A. A., Galvão Dias, T., Falcão e Cunha, J. (2016) "Passenger journey destination estimation from automated fare collection system data using spatial validation", IEEE Transactions on Intelligent Transportation Systems, 17(1), pp. 133–142. https://doi.org/10.1109/TITS.2015.2464335

Podinovsky, V. V., Potapov, M. A. (2013) "Метод взвешенной суммы критериев в анализе многокритериальных решений: pro et contra" (The method of weighted sum of criteria in the analysis of multi-criteria decisions: pro et contra), Бизнес-информатика, 3(25), pp. 41–48. (in Russian)

Prokopenko, N. Y. (2018) "Методы оптимизации: Учебное пособие" (Optimization methods: textbook), Nizhny Novgorod State University of Architecture and Civil Engineering (NNGASU). ISBN 978-5-528-00287-3 (in Russian)

Wang, W. (2010) "Bus passenger origin-destination estimation and travel behavior using automated data collection systems in London, UK", Master of Science in Transportation Thesis, Massachusetts Institute of Technology.

Zhao, J. (2004) "The planning and analysis implications of automated data collection systems: rail transit OD matrix inference and path choice modeling examples", Master in City Planning and Master of Science in Transportation Thesis, Massachusetts Institute of Technology.

Zhao, J., Rahbee, A., Wilson, N. H. M. (2007) "Estimating a rail passenger trip origin-destination matrix using automatic data collection systems", Computer-Aided Civil and Infrastructure Engineering, 22(5), pp. 376–387. https://doi.org/10.1111/j.1467-8667.2007.00494.x