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

Aleksandr I. Fadeev1, Sami Alhusseini1∗

1 Transport Department, Siberian Federal University, 79 Svobodny pr., 660041 Krasnoyarsk, Russia
∗ Corresponding author, e-mail: salkhussayni-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 to determine 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.
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.

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.
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.

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):

\[
M = \{M^k : k = a, b\}; \; M \subseteq M; \; |M| = m; \quad (1)
\]

\[
M^k = \{w_1, \ldots, w_i, \ldots, w_j\}; \; M^k \subseteq M; \; w_i \in W. \quad (2)
\]

The AFC system records the execution time of the \( k \)-th tapping on operation of the \( e_i \) 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 \( E = \{e_1, e_2, \ldots, e_k\} \), each element of which reflects the passenger trip \( p_i \) 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 \( W_c \) which are located in the \( k \)-th journey before the operation of tapping on the

<table>
<thead>
<tr>
<th>Table 1 Sample's volume and the share of interpreted e-ticket validations from different works</th>
</tr>
</thead>
<tbody>
<tr>
<td>Author</td>
</tr>
<tr>
<td>-------------------</td>
</tr>
<tr>
<td>Alsger et al. (2016)</td>
</tr>
<tr>
<td>Alsger et al. (2015)</td>
</tr>
<tr>
<td>Nassir et al. (2011)</td>
</tr>
<tr>
<td>Zhao et al. (2007)</td>
</tr>
<tr>
<td>Cui (2006)</td>
</tr>
<tr>
<td>Fadeev and Alhusseini (2021b)</td>
</tr>
</tbody>
</table>

Fig. 2 Calculation scheme of passenger trip, where \( w_i, w_{i+1}, \ldots \) are the PT stops of the route; \( g_i, g_{i+1}, \ldots \) are the navigation marks of the satellite positioning system; \( e_i \) is the e-ticket validation.
electronic ticket, and $W_i^+$ - after the validation, where $W_i^- \subseteq W; W_i^+ \subseteq W$. Thus, the boarding stop of passenger trip is an element of the set $W_i^-$, and the alighting stop is an element of the set $W_i^+$.

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:

$$R = G \times M = \{(g, w) : g \in G, w \in M, l(g, w) \leq l_u\},$$  \hspace{1cm} (3)

where $l(g,w)$ is the distance between the navigation mark $(g)$ and the PT stop $(w)$; $l_u$ 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 $j$-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 $X_{k+1}^i$ between completing $k$-th trip and starting $(k+1)$-th trip:

$$X_{k+1}^i = W_i^+ \times W_i^+; \quad l(w_i^+, w_{k+1}^+) \leq L_p; \quad w_{k+1}^+ \subseteq W_i^+; \quad w_{k+1}^+ \subseteq W_i^+,$$

where $l(w_i^+, w_{k+1}^+)$ is the Euclidean distance between PT stops $w_i^+$ and $w_{k+1}^+$; $L_p$ is the typical walking distance.

Possible variants (Fig. 5) of the $k$-th passenger trip are the Cartesian product of the sets $X_{k+1}^i$ and $X_{k+1}^{i+1}$.

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 $y_j = f_j(x), \ldots, y_j = f_j(x)$, which make up the vector of estimates. $Z_j$ is the scale (set of values) of criterion $f_j$.

We will assume that the criteria are positively oriented, i.e., as the values of each criterion increase, the preferences increase.
In the problem under consideration, the following indicators are used to evaluate the elements of the set $X_k$:

- 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):

\[
F(f_{vi}) = \sum_{j=1}^{\gamma} v_{ij} f_{i} \\
F(f_{vi}) = \prod_{j=1}^{\gamma} v_{ij} f_{i} \\
F(f_{vi}) = \prod_{j=1}^{\gamma} v_{ij} f_{i}^*.
\]

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 (Fetinina et al., 2003; Podinovsky and Potapov, 2013), for example:

\[
\phi(f_i) = \frac{f_i - f_{i\text{min}}}{f_{i\text{max}} - f_{i\text{min}}} \\
\phi(f_i) = \left(\frac{f_i - f_{i\text{min}}}{f_{i\text{max}} - f_{i\text{min}}}\right)^\alpha,
\]

where $f_{i\text{max}}$, $f_{i\text{min}}$ 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:

\[
\phi(f_i) = 1 - \frac{f_i - f_{i\text{min}}}{f_{i\text{max}} - f_{i\text{min}}}.
\]

The following functions were used for normalizing the indicators:

1. the walking distance between the alighting and boarding stops

\[
\phi(l_i) = 1 - \frac{l_i}{2L_p}, \quad l_i \leq 2L_p;
\]

2. stop number from the tapping on operation

\[
\phi(n_i) = 1 - \frac{n_i}{n^*}, \quad n_i \leq n^*;
\]

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

\[
\phi(w_i) = \frac{n_i}{n^*},
\]

where $n^*$ is the the largest stop point number from the tapping on operation; for PT stops, the number of which is
greater than \( n^{+} \), criterion Eq. (11) takes on a zero value; \( n^{-} \), \( n^{\pm} \) 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 \( X_{i+1}^k \) 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):

\[
K^i = v_i \varphi(l_i) + v_w \varphi(w_i) = \max, \tag{13}
\]

where \( v_i, v_w \) are the criteria weights in Eqs. (10)–(12).

Consider the problem of determining the values of the criteria weight coefficients \( v_i, v_w \) 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 \( y_i, i = 1, \ldots, N \), which can be determined (measured) with sufficient accuracy. The \( f_i(v_i, v_w) \), \( i = 1, \ldots, N \) 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):

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

It is necessary to determine the minimum of the function (Eq. (14)) by the variable weights coefficients \( v_i, v_w \).

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.

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:

\[
R(A, T, W, M', M'', M'), \tag{15}
\]

where \( T \) is the time of the navigation mark; \( A \) is the vehicle; \( M', M'' \) are the route number and direction of movement, respectively; \( W \) is the PT stop; \( M' \) 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:

\[
E\left(I^*, I^+, T, M', M'', A\right). \tag{16}
\]

where \( I^* \) is the identifier of the electronic ticket-tapping operation \( I^+ \); \( T \) is the operation time; \( M', M'' \) are the route number and direction of movement, respectively; \( A \) is the vehicle identifier.

An element of this set is denoted as \( e(i^*, i^+, t, m', m'', a) \) or \( e_k \) \((k)\)-tapping on operation).

The route number \( m_k' \) is known from the e-ticket operation. The travel direction of the passenger \( m_k' \) 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 \( X_{i+1}^k \) 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 \( w_{i+1}^j \left( w_{i+1}^j \subseteq W_{k-1}^j \right) \)
   located before the point \( e_k \), where the e-ticket

![Fig. 6 Scheme estimated passenger-walking distance between alighting PT stop and next boarding one](image-url)
tapping on operation took place and the first PT stop after this point \( w_{i1} \), \( w_{i2} < W_{i1} \). If the tapping on operation is performed at the PT stop, \( w_{i2} = w_{i1} \). The \( w_{i1} \) and \( w_{i2} \) are determined from the actual trajectory of the vehicle by the time of the operation:

Nearest PT stop before tapping on operation
\[
\text{min}\left(r^e\right); \ r^e \leq t^i; \ r_e \subset R_o
\]  

First stop after tapping on operation
\[
\text{max}\left(r^e\right); \ r^e \geq t^i; \ r_e \subset R_o
\]

where \( r_o \) is the actual trajectory of the \( a \)-th vehicle \( R_o \); \( t^i \) is the time of the \( k \)-th tapping on operation.

2. By Eq. (5), a set of admissible connectivity options \( X_{k+1}^i \) 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:

\[
X\left(I^i, I_{k+1}^i, W_{k+1}^i, M_{k+1}^i, W_{k+1}^i, M_{k+1}^i, W_{k+1}^i, M_{k+1}^i, L^e\right),
\]

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

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

\[
\Phi\left(I^i, I_{k+1}^i, W_{k+1}^i, M_{k+1}^i, W_{k+1}^i, M_{k+1}^i, W_{k+1}^i, M_{k+1}^i, L^e\right),
\]

where \( \Phi', \Phi^*, \Phi^+ \) 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 \( X_{k+1}^i \), 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 \( v_i, v_{k+1}, v_w \) are assumed to be known.

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

\[
P\left(I^i, M^i, M_{k+1}^i, W_{k+1}^i, T^i, W^w, T^w, T^d, L\right),
\]

where \( P \) is the identifier of the tapping on operation; \( M^i, M_{k+1}^i \) are the route number and direction of movement, respectively; \( W^w, T^w \) is the identifier of the PT stop for the beginning and end of passenger trip, respectively; \( T^i \) is 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 \( n^w \) and alighting \( n^a \) 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., \( n^w = n^a \).

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

\[
\phi_i = \phi_i^* + \frac{1}{n^a}, \text{ if } n^w < n^a, \quad I(e_i, p_i^w) \leq L_p,
\]

\[
\phi_i = \phi_i^* + \frac{1}{n^w}, \text{ if } n^w < n^a, \quad I(e_i, p_i^w) \leq L_p,
\]

\[
\begin{cases}
\phi_i = \phi_i^* + \frac{1}{2n^a}, \text{ if } n^w = n^a, \quad I(e_i, p_i^w) \leq L_p, \\
\phi_i = \phi_i^* + \frac{1}{2n^w}, \text{ if } n^w = n^a, \quad I(e_i, p_i^w) \leq L_p,
\end{cases}
\]
if \( n' ≠ 0 \) and \( n'' ≠ 0 \), where \( \phi' \) is the current value of the balancing factor of the \( k \)-th trip for e-ticket (at the beginning of the calculation \( \phi' = 1 \)); \( n'' \) is the number of e-ticket trips with the boarding point at an allowed walking distance from \( e' \); \( n''' \) is the number of e-ticket trips with the alighting point within allowed walking distance from \( e'' \); \( l(e', p_k') \) is the distance between PT stops \( e' \) and \( p_k'' \) (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 \( n'' \) than boarding \( n''' \);
- if there are fewer alighting than boarding, the number of alighting increases by 1;
- if \( n'' = n''' \), the number of boarding and alighting is increased by 0.5.

For the cases where \( \text{or} \), 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:

\[
Q = \sum_{i,j} T_{ij}; \quad i,j = 1,...,n, \tag{25}
\]

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:

\[
T_{ij} = \sum_{m} \sum_{k \neq m} \frac{\phi_{k}}{\alpha_{m}^{*}} \alpha_{m}' \tag{26}
\]

where \( \phi_{k} \) is the balancing factor of the \( k \)-th passenger trip, \( \alpha_{m}' \) is the share of operations by e-tickets on the \( m \)-th route (Fadeev and Alhusseini, 2021a) \( \alpha_{m}' ≤ 1 \); \( \alpha_{m}^{*} = (Q_{m}' - q_{m})/Q_{m} \),

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' | 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 | 92.782723 | 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.
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 $X_{k+1}^i$ 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.

For each connectivity option, evaluation indicators are calculated (Table 9). With known weight coefficients $v_i$, $v_e$, $v_w$, it is possible to calculate the integral criterion $\Phi^\Sigma$, according to the largest value of which the connectivity option $X_{k+1}^i$ is selected and on this basis, passenger trip is formed (see Fig. 5).
To calculate passenger trip, it is required to determine the weight of coefficients $v_i$, $v_a$, $v_w$ 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 number of the PT stop from the tapping on operation $v_l$ is the second factor influencing the criterion of compliance. With the exclusion of the factor ($v_l = 0$), the criterion of compliance dropped by 2 times compared to the minimum point.

The number of the PT stop from the tapping on operation $v_l$ is the second factor influencing the criterion of compliance. With $v_l = 0$, the compliance criterion dropped by 75% (from 0.355 to 0.203).

According to the results of the study, it can be concluded:
1. The minimum point of the function is when $v_l = 1$; $v_a = 1$; $v_w = 0$. 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 ($v_w$). With the exclusion of the factor ($v_w = 0$), the criterion of compliance dropped by 75% (from 0.355 to 0.203).
3. 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 $v_i$, $v_a$, $v_w$, the calculation of passenger trips was performed, the results of which are comparable with the previous version.

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 $v_i$, $v_a$, $v_w$.

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 $v_i$, $v_a$, $v_w$ 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 $v_i = 1$; $v_a = 1$; $v_w = 0$. 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 ($v_w$). With the exclusion of the factor ($v_w = 0$), the criterion of compliance dropped by 75% (from 0.355 to 0.203).
3. 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 $v_i$, $v_a$, $v_w$, the calculation of passenger trips was performed, the results of which are comparable with the previous version.

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

<table>
<thead>
<tr>
<th>$F^2$</th>
<th>$\Phi^2$</th>
<th>$\Phi^0$</th>
<th>$\Phi^v$</th>
<th>$\Phi^w$</th>
</tr>
</thead>
<tbody>
<tr>
<td>29030</td>
<td>0.96</td>
<td>1.00</td>
<td>0.03</td>
<td>1.98</td>
</tr>
<tr>
<td>30029</td>
<td>0.95</td>
<td>0.80</td>
<td>0.00</td>
<td>1.75</td>
</tr>
<tr>
<td>28030</td>
<td>0.55</td>
<td>1.00</td>
<td>0.03</td>
<td>1.57</td>
</tr>
<tr>
<td>31028</td>
<td>0.96</td>
<td>0.60</td>
<td>0.00</td>
<td>1.56</td>
</tr>
<tr>
<td>30030</td>
<td>0.41</td>
<td>1.00</td>
<td>0.03</td>
<td>1.43</td>
</tr>
<tr>
<td>32028</td>
<td>0.76</td>
<td>0.60</td>
<td>0.00</td>
<td>1.36</td>
</tr>
<tr>
<td>32027</td>
<td>0.93</td>
<td>0.40</td>
<td>0.00</td>
<td>1.33</td>
</tr>
<tr>
<td>31029</td>
<td>0.47</td>
<td>0.80</td>
<td>0.00</td>
<td>1.27</td>
</tr>
<tr>
<td>29029</td>
<td>0.45</td>
<td>0.80</td>
<td>0.00</td>
<td>1.25</td>
</tr>
<tr>
<td>33026</td>
<td>0.97</td>
<td>0.20</td>
<td>0.00</td>
<td>1.17</td>
</tr>
<tr>
<td>30028</td>
<td>0.51</td>
<td>0.60</td>
<td>0.00</td>
<td>1.11</td>
</tr>
<tr>
<td>31027</td>
<td>0.71</td>
<td>0.40</td>
<td>0.00</td>
<td>1.11</td>
</tr>
<tr>
<td>32029</td>
<td>0.29</td>
<td>0.80</td>
<td>0.00</td>
<td>1.09</td>
</tr>
<tr>
<td>34024</td>
<td>1.00</td>
<td>0.00</td>
<td>0.00</td>
<td>1.00</td>
</tr>
</tbody>
</table>

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

![Fig. 7 Dependence the function Eq. (28) on the variables $v_i$, $v_a$](image-url)
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.

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 passed before 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

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

<table>
<thead>
<tr>
<th>Day of the month</th>
<th>Total</th>
<th>Interpreted (Method 1)</th>
<th>Interpreted (Method 2)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Amount</td>
<td>weight, %</td>
</tr>
<tr>
<td>1</td>
<td>238789</td>
<td>152391</td>
<td>63.8</td>
</tr>
<tr>
<td>2</td>
<td>253128</td>
<td>163643</td>
<td>64.6</td>
</tr>
<tr>
<td>3</td>
<td>250250</td>
<td>163477</td>
<td>65.3</td>
</tr>
<tr>
<td>4</td>
<td>253132</td>
<td>162679</td>
<td>64.3</td>
</tr>
<tr>
<td>5</td>
<td>252721</td>
<td>160034</td>
<td>63.3</td>
</tr>
<tr>
<td>6</td>
<td>164948</td>
<td>101464</td>
<td>61.5</td>
</tr>
<tr>
<td>7</td>
<td>124539</td>
<td>77113</td>
<td>61.9</td>
</tr>
<tr>
<td>8</td>
<td>238016</td>
<td>155675</td>
<td>65.4</td>
</tr>
<tr>
<td>9</td>
<td>248460</td>
<td>163578</td>
<td>65.8</td>
</tr>
<tr>
<td>10</td>
<td>247446</td>
<td>161197</td>
<td>65.1</td>
</tr>
<tr>
<td>11</td>
<td>247844</td>
<td>163009</td>
<td>65.8</td>
</tr>
<tr>
<td>12</td>
<td>242969</td>
<td>158770</td>
<td>65.3</td>
</tr>
<tr>
<td>13</td>
<td>155113</td>
<td>101079</td>
<td>65.2</td>
</tr>
<tr>
<td>14</td>
<td>116213</td>
<td>76060</td>
<td>65.4</td>
</tr>
<tr>
<td>15</td>
<td>238363</td>
<td>155945</td>
<td>65.4</td>
</tr>
<tr>
<td>16</td>
<td>255730</td>
<td>167073</td>
<td>65.3</td>
</tr>
<tr>
<td>17</td>
<td>257936</td>
<td>168971</td>
<td>65.5</td>
</tr>
<tr>
<td>18</td>
<td>239022</td>
<td>156863</td>
<td>65.6</td>
</tr>
<tr>
<td>19</td>
<td>232266</td>
<td>149428</td>
<td>64.3</td>
</tr>
<tr>
<td>20</td>
<td>160447</td>
<td>102859</td>
<td>64.1</td>
</tr>
<tr>
<td>21</td>
<td>129981</td>
<td>85581</td>
<td>65.8</td>
</tr>
<tr>
<td>22</td>
<td>241130</td>
<td>156714</td>
<td>65.0</td>
</tr>
<tr>
<td>23</td>
<td>259047</td>
<td>164047</td>
<td>63.3</td>
</tr>
<tr>
<td>24</td>
<td>235845</td>
<td>151504</td>
<td>64.2</td>
</tr>
<tr>
<td>25</td>
<td>243887</td>
<td>156387</td>
<td>64.1</td>
</tr>
<tr>
<td>26</td>
<td>235609</td>
<td>147355</td>
<td>62.5</td>
</tr>
<tr>
<td>27</td>
<td>139348</td>
<td>85765</td>
<td>61.5</td>
</tr>
<tr>
<td>28</td>
<td>116055</td>
<td>69071</td>
<td>59.5</td>
</tr>
<tr>
<td>29</td>
<td>216090</td>
<td>125869</td>
<td>58.2</td>
</tr>
<tr>
<td>30</td>
<td>231156</td>
<td>88322</td>
<td>38.2</td>
</tr>
<tr>
<td>Σ</td>
<td>6465480</td>
<td>4091923</td>
<td>63.3</td>
</tr>
</tbody>
</table>
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.

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

<table>
<thead>
<tr>
<th>Methods</th>
<th>ID</th>
<th>Rout</th>
<th>Start</th>
<th>End</th>
<th>Distance, km</th>
<th>Ride</th>
<th>Walk</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>13119511</td>
<td>58 B</td>
<td>327</td>
<td>267</td>
<td>2.8</td>
<td>0.48</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>13119511</td>
<td>58 B</td>
<td>327</td>
<td>267</td>
<td>2.8</td>
<td>0.48</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>13119512</td>
<td>95 A</td>
<td>190</td>
<td>379</td>
<td>10.8</td>
<td>0.09</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>13119512</td>
<td>95 A</td>
<td>184</td>
<td>379</td>
<td>12.0</td>
<td>0.09</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>13119513</td>
<td>85 A</td>
<td>428</td>
<td>369</td>
<td>2.9</td>
<td>0.12</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>13119513</td>
<td>85 A</td>
<td>428</td>
<td>369</td>
<td>2.9</td>
<td>0.12</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>13119514</td>
<td>61 B</td>
<td>368</td>
<td>683</td>
<td>9.7</td>
<td>0.05</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>13119514</td>
<td>61 B</td>
<td>368</td>
<td>683</td>
<td>9.7</td>
<td>0.05</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>13119515</td>
<td>61 B</td>
<td>386</td>
<td>685</td>
<td>11.2</td>
<td>0.02</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>13119515</td>
<td>61 B</td>
<td>386</td>
<td>685</td>
<td>11.2</td>
<td>0.02</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>13119516</td>
<td>95 A</td>
<td>413</td>
<td>379</td>
<td>2.3</td>
<td>0.09</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>13119516</td>
<td>95 A</td>
<td>413</td>
<td>379</td>
<td>2.3</td>
<td>0.09</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>13119517</td>
<td>85 B</td>
<td>382</td>
<td>416</td>
<td>7.0</td>
<td>0.02</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>13119517</td>
<td>85 B</td>
<td>382</td>
<td>416</td>
<td>7.0</td>
<td>0.02</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>13119518</td>
<td>78 A</td>
<td>417</td>
<td>353</td>
<td>2.0</td>
<td>0.33</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>13119518</td>
<td>78 A</td>
<td>417</td>
<td>353</td>
<td>2.0</td>
<td>0.33</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>13119519</td>
<td>61 A</td>
<td>327</td>
<td>375</td>
<td>6.3</td>
<td>0.44</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>13119519</td>
<td>61 A</td>
<td>327</td>
<td>375</td>
<td>6.3</td>
<td>0.44</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>13119520</td>
<td>95 B</td>
<td>380</td>
<td>416</td>
<td>3.6</td>
<td>0.02</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>13119520</td>
<td>95 B</td>
<td>380</td>
<td>416</td>
<td>3.6</td>
<td>0.02</td>
<td></td>
</tr>
</tbody>
</table>

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
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):

\[ t_s = \frac{\bar{x} - \bar{y}}{\sigma_{x-y}}, \]  

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

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

(30)

We compare \( t_s \) 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 \( t_k < t_s \).

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 ticket 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>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>26</td>
<td>A</td>
<td>0.334</td>
<td>2.450</td>
<td>7</td>
<td>A</td>
<td>0.465</td>
<td>0.553</td>
</tr>
<tr>
<td>26</td>
<td>B</td>
<td>0.603</td>
<td>1.219</td>
<td>7r</td>
<td>B</td>
<td>0.536</td>
<td>0.744</td>
</tr>
<tr>
<td>31</td>
<td>A</td>
<td>1.692</td>
<td>1.368</td>
<td>3</td>
<td>A</td>
<td>1.368</td>
<td>1.692</td>
</tr>
<tr>
<td>31</td>
<td>B</td>
<td>1.067</td>
<td>2.378</td>
<td>3</td>
<td>B</td>
<td>1.655</td>
<td>0.379</td>
</tr>
<tr>
<td>37</td>
<td>A</td>
<td>0.379</td>
<td>1.655</td>
<td>37</td>
<td>A</td>
<td>0.522</td>
<td>0.980</td>
</tr>
<tr>
<td>37</td>
<td>B</td>
<td>0.771</td>
<td>0.037</td>
<td>52</td>
<td>A</td>
<td>1.205</td>
<td>0.079</td>
</tr>
<tr>
<td>52</td>
<td>B</td>
<td>1.205</td>
<td>0.079</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
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 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


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)


