Transportation Research Interdisciplinary Perspectives 19 (2023) 100832

Contents lists available at ScienceDirect

Transportation Research Interdisciplinary Perspectives

journal homepage: www.elsevier.com/locate/trip

Public transport demand estimation by frequency adjustments

Victoria M. Orlando [a,b,c,*], Enrique G. Baquela [d], Neila Bhouri [c], Pablo A. Lotito [a,b]

[a] Universidad Nacional del Centro de la Provincia de Buenos Aires, Facultad de Ciencias Exactas, PLADEMA, Tandil, Argentina
[b] CONICET, Consejo Nacional de Investigaciones Científicas y Técnicas, Buenos Aires, Argentina
[c] Univ. Gustave Eiffel, COSYS/GRETTIA Lab, 77454 Marne-la-Vallee, France
[d] GISOI, Dpto. Ingeniería Industrial, Facultad Regional San Nicolás, UTN, San Nicolás, Argentina

ARTICLE INFO

Keywords:
Public transport demand
Bi-level optimization
Neural networks
Inverse problem

ABSTRACT

This article addresses the problem of estimating the demand for public transport from two approaches. First, we propose a bilevel optimization problem that allows estimating the demand using historical data and the observed bus frequencies. This model has been applied to small theoretical networks and the transit network of Tandil (a medium-sized city in Buenos Aires, Argentina), showing good results. However, from a practical point of view, the computation time of the algorithm used to solve the bilevel problem is long, reducing its applicability by traffic authorities. To solve this, we propose to use an artificial neural network module that allows to quickly detect if the change in demand is significant enough (for example, beyond a predefined threshold). If it is substantial, the operator can decide to run the algorithm to estimate the demand and take action to adapt the system to the new reality, for example, adapting vehicle frequencies or incorporating more vehicles into the system so that the current demand can be served. The machine learning approach allows it to be used as a fast change detection tool, avoiding running the expensive algorithm for false positives.

1. Introduction

The optimal operation of a transport network is fundamental in public transport planning. Providing a good service benefits current passengers and may attract new users, contributing to not increasing (or even decreasing) the individual traffic flow. This would be interesting as individual road transport is known to be a major contributor to air pollution. In this way, as expressed Lera-López et al. (2014) and Sánchez-García et al. (2021), it is essential that the authorities promote initiatives to increase environmental awareness and thus reduce individual pollution. A good action would be to provide reliable public transportation service, and one contributing factor to this is maintaining regular headways (and, therefore, regular frequencies). As highlighted in Tirachini et al. (2021), bus headways depend on many factors, including initial delays and the number of passengers using the service. Maintaining the same fixed frequency throughout the day could be insufficient when demand increases (generating discomfort for users due to overcrowding or inability to board) and unnecessary when demand decreases (deploying more vehicles than required, leading to inefficient use of resources and higher operating costs). For this reason, it becomes essential to detect changes in transport demand and then adapt the system (if necessary), maintaining the optimal operation of the network. Considering the importance of knowing how many passengers use the network, we will present two approaches to address the demand estimation problem.

As mentioned earlier, estimating public transport demand is a crucial aspect of urban planning, particularly when designing the network and determining bus frequencies. The demand estimation method can vary, among other things, depending on the available information in each system. As noted by Sun and Schmöcker (2021), it is essential to distinguish between cases where a new network is being designed (with no information on current demand) and cases where the goal is to estimate demand on an existing network (where current network usage information can be obtained). This paper focuses on the latter scenario. Knowing the public transport demand would allow for predicting the possible flow distribution on different bus lines and, eventually, adding vehicles to the service so that the frequency is greater and possible system saturation is avoided.

A wide variety of methods can be used to find out the demand between origin–destination pairs. Within the public transport area, one option is to deploy surveys at specific stops or onboard the buses and use this information to build the OD matrix using, for example, a gravity model (Asmael and Waheed, 2018, Sun and Schmöcker, 2021). However, it can be difficult or costly to develop surveys, in addition to not guaranteeing a reliable result since they could be biased or

* Correspondence to: Pinto 399, Tandil, Buenos Aires, Argentina.
E-mail addresses: vorlando@pladema.exa.unicen.edu.ar (V.M. Orlando), ebaquela@frsn.utn.edu.ar (E.G. Baquela), neila.bhouri@univ-eiffel.fr (N. Bhouri), plotito@exa.unicen.edu.ar (P.A. Lotito).

https://doi.org/10.1016/j.trip.2023.100832
Received 6 October 2022; Received in revised form 14 April 2023; Accepted 29 April 2023

inaccurate. Hence, it becomes necessary to apply other methods. At present, in those cases where the transport networks have information provided by the information and communication technologies, it would be possible to know the number of passengers that are using public transport in real-time (see Barcelo et al., 2012, Montero et al., 2012). However, this could be insufficient since it would be possible to know how many people are inside a bus but not its origin or destination. When the transportation system has smart cards with an ID associated with each passenger, the origin can be known, and the destination of each trip can be inferred using automated fare collection (AFC) data. As mentioned in Barry et al. (2002), the estimation can be made based on two assumptions. The first one is that the destination of a user's first trip can be established when he makes a second trip since the origin of this new trip usually coincides with the destination of the first one. The second specifies that the destination of the last trip of the day is typically the origin of the day's first trip. These models allow inferring the destination of passengers when only their origin information is available. In those networks that do not have smart card information, or in cases where the origin of the passengers is unknown, demand estimation must be based on previous information like historical demand and data that can be measured from the network. As has been made for the case of urban traffic planning (see Walpen et al., 2015), here we propose to estimate the demand through the solution of a bilevel optimization problem that uses information obtained from the network, such as the actual bus frequencies, and historical data such as previous demand measurements. This model is based on the one presented in previous works (see Bhouri et al., 2021, Orlando et al., 2021, Orlando et al., 2022), making some modifications that contribute to obtaining better results. In the first stage, we apply the analytical modeling to different examples based on the bus network of Tandil (a medium-sized city in Buenos Aires, Argentina), solving the bilevel problem using a derivative-free method. In the second stage, due to the calculation times being big for practical implementations, we use an artificial neural network (ANN) module to determine when the change in the demand is greater than a predefined threshold. If the threshold is well defined, it is avoided running the expensive optimization algorithm when the change in the demand is so small that no action is taken by the transportation authority.

Many factors can be considered when studying transport demand, as in Jaber et al. (2022), where the behavior of frequent and occasional passengers is distinguished. This work assumes that all network users behave similarly, aiming to minimize their total travel time. Other factors besides seeking the lowest total travel time could be considered, such as maintaining a certain level of user satisfaction considering their temporal and spatial variation (Echaniz et al., 2022). We approach the problem of public transport demand estimation from the perspective of network management rather than the viewpoint of network planning and design as previously studied (Ahern et al., 2022), where the main goal is to minimize passenger and operator costs.

The bilevel model we propose requires resolving a flow assignment problem within its constraints. For this, adopting a model that best represents reality is important. All factors involved in network operation (passenger behavior, travel times, bus frequencies, among others) must be as realistic as possible. Several authors have studied transit flow assignment over the years, assuming different network behavior, formulations, and resolutions of the assignment problem. Spiess and Florian (1989) consider that travel times, frequencies, and waiting times are independent of the flow, which may be unrealistic when demand is high and passengers cannot board the first vehicle that arrives at the stop due to lack of capacity. De Cea and Fernández (1993) contemplate network congestion by considering that waiting times at stops depend on the lines' occupation level. In some works (Cominetti and Correa, 2001, Cepeda et al., 2006), it is assumed that bus frequencies decrease as the flow on them approaches maximum capacity. In other cases (Oliker and Bekhor, 2020), capacity restrictions and online information accessible to travelers are considered. In Oliker and Bekhor (2020), the assignment problem is not formulated as an equilibrium problem, while in other cases (Cominetti and Correa, 2001, Cepeda et al., 2006), the model is based on the Wardrop Equilibrium. In this last case, the problem has been reformulated as a variational inequality problem (Codina, 2013, Codina and Rosell, 2017). We choose the model proposed in Cepeda et al. (2006), where bus frequencies depend on the number of traveling passengers and, consequently, on the bus capacity. This model assumes that when users must choose the routes to make their trip, they will consider the level of occupation of the vehicles, reflecting the influence of network congestion on travelers' decisions and making the model more realistic. The resolution of the flow assignment problem under these assumptions may require long calculation times if large networks are considered. This becomes an issue when solving the bilevel problem to estimate the demand. For this reason, this work proposes a modification in the algorithm implemented for the flow assignment, reducing the calculation times without altering the results. Other tools had previously been used to overcome the problem of calculation times, such as in Orlando et al. (2022), where a bilevel model similar to the one proposed here was used to estimate the demand on small networks coupled with agent-based simulations. In that work, we use the algorithm proposed by Cepeda et al. (2006) to perform the flow assignment, but also, bus trips and passenger assignments were modeled as a simulation. The main benefit of combining simulation and optimization is that complicated systems can be modeled more easily compared to a purely mathematical approach. These advantages are evident not only in public transport problems but also in other areas, as suggested in Sawik et al. (2022a,b) for the case of Automated Parcel Lockers.

The main contributions of this paper are the proposal of an improvement in the flow assignment algorithm and in the analytical model for demand estimation, the usage of an ANN as a tool for change detection, and the application of both models to a real-world public transport network using actual data. Demand estimation can be done in many fields, not just public transportation. Applications range from the demand for products from a store (Bajari et al., 2015), banking service demand (Dick, 2008), demand for general resources (Spinner et al., 2015), and traffic demand (Cipriani et al., 2021), to mention just a few. The models proposed in this work can be applied in various contexts beyond the previously mentioned. If there is historical information on the demand and it is defined what current information will be measured from the network (as in this case, it is the bus frequencies), the model should be able to be replicated without major inconveniences.

The paper is organized as follows. The methodology section describes the equilibrium problem that serves as a model of the transit distribution, the demand estimation problem formulated as the solution to a mathematical optimization problem with two levels, and the formulation using ANN. The results section provides a numerical example of a real network taken from Tandil (Argentina). With this real data, we obtain numerical results that allow the comparison with the machine learning approach. The discussion section exposes the main contributions of the work as well as the current limitations. Finally, the last section contains the conclusions and undergoing work.

2. Methodology

In this section, the transit model is first thoroughly described. The chosen is an equilibrium model that allows calculating the flow assigned to each network arc given the travel demand. Then, we propose a numerical procedure to solve the assignment problem that combines a standard successive average method with some heuristics for performance improvement. After that, we present the central issue in this work, which is the inverse of the assignment problem, i.e., given the frequencies obtain the demand. This problem is solved using a variation of the Nelder–Mead algorithm, also known as the simplex algorithm (it should be distinct from the simplex algorithm of Dantzig). To cope with some convergence issues, we also present a machine-learning approach that can be a good option when trading off precision to reduce computation times.

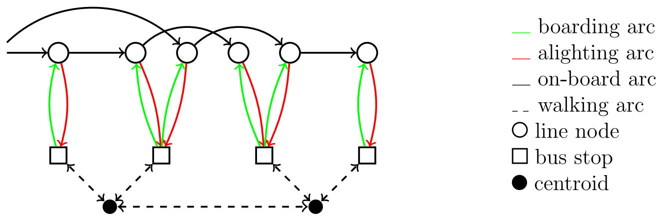


Fig. 1. Public transport network.

2.1. Transit equilibrium model

The transit network will be represented by a directed graph $G = (N, A)$ where N is the set of nodes and A is the set of arcs. The set N can be expressed as $N = N_s \cup N_l \cup N_c$, where N_s contains the bus-stop nodes, N_l represents the set of line nodes, and N_c is the set of centroid nodes. Bus-stop nodes represent where one or more lines stop, and line nodes are included and connected with bus-stop nodes to model the boarding and alighting of passengers. Centroid nodes are used to represent predefined urban areas, and they will be considered the source and the destination of travelers. Meanwhile, the arcs are divided into walking arcs (connecting bus-stop nodes with centroids and centroids between them), boarding and alighting arcs (connecting bus-stop nodes and line nodes), and on-board arcs (connecting line nodes). In this way, each traveler can walk from the centroid of their origin zone to the nearest bus stop that is served by those lines that allow them to make their trip, wait for a bus and board them to finally alight at the final bus stop and walk to their destination centroid. Including walking arcs allows for this travel mode, either as an alternative to using the congested transit network or as a combination of walking and using the bus, if necessary. An example of public transport network can be seen in Fig. 1, where two centroids and four bus stops are considered, along with two bus lines and the corresponding line nodes.

Continuing with the notation we will use throughout the work, we denote the set of arcs incoming to node $i \in N$ as A_i^- and the set of arcs emerging from node i as A_i^+. Some pairs $(i, d) \in N_c \times N_c$ have associated a non-negative demand g_i^d, leading to a set $D = \{d \in N_c : g_i^d > 0\}$ containing those nodes that are the destination of some demand.

We define the node-arc incidence matrix $M \in R^{|N|} \times R^{|A|}$ where $M_{ia} = 1$ iff $a \in A_i^+$, $M_{ia} = -1$ iff $a \in A_i^-$ and otherwise zero. Denoting v_a^d the flow through arc a with destination $d \in D$, and $v^d = (\ldots, v_a^d, \ldots)$, we define the set of feasible flows with destination d as

$$V^d = \left\{ v^d \in R_+^{|A|} \,:\, Av^d = g^d \right\}, \tag{1}$$

and the set of total feasible flows:

$$V = \left\{ v \in R_+^{|A|} \,:\, v = \sum_{d \in D} v^d, v^d \in V^d \right\}. \tag{2}$$

Also we call $V(g)$ the set of feasible flows for the demand g, composed by the flows $v_a^d \geq 0$ such that $v_a^d = 0$ if $a \in A_d^+$ and satisfying the flow conservation constraints:

$$g_i^d + \sum_{a \in A_i^-} v_a^d = \sum_{a \in A_i^+} v_a^d, \quad \forall i \neq d. \tag{3}$$

Once the transport network and the corresponding graph are defined, we must choose the equilibrium model that best fits the system and the users' behavior we want to represent. As we mentioned, numerous works are approaching the transit assignment problem, each assuming different passenger behavior and network design, as well as different available information about the transport system. Another aspect in which the works differ is how they consider the effects of transit network congestion when modeling the problem. We choose the equilibrium model proposed in Cepeda et al. (2006), considering that travel times t_a for each arc $a \in A$ are fixed and not depending on the flow through them, but assuming that the number of passengers on board influences bus frequencies. The impact of bus load on the frequency can be modeled by an effective frequency function $f_a : V \to [0, \infty]$ associated with each arc $a \in A$. This function is strictly decreasing, tends to zero when flow through arc a grows and can be defined as:

$$f_a(v) = \begin{cases} \mu \left[1 - \left(\frac{v_a}{\mu c - v_{a'} + v_a} \right)^{\beta} \right], & \text{if} \quad v_{a'} < \mu c, \\ 0, & \text{otherwise}, \end{cases} \tag{4}$$

where $v_a = \sum_{d \in D} v_a^d$ represents the total flow boarding at stop using arc a and $v_{a'}$ is the total flow on board after the stop, that is, the flow after the boarding and alighting of passengers. The constant μ is the nominal frequency of the lines (that is, the frequencies planned by the transport companies), and c is the physical capacity of the buses. We consider infinite frequencies $f_a(v) = \infty$ for all arcs, except for the boarding arcs, in which case $f_a(v) < \infty$ since that is where the wait occurs. The function (4) states that if the capacity is exceeded ($v_{a'} \geq \mu c$), the frequency will be equal to zero, leading to an infinite waiting time at the stop. To avoid it, in practice, we use a value small enough. This would give an excessive waiting time but allow the algorithm to not get stuck during initial iterations when the capacity is exceeded. If there is a feasible solution, it will be reached in later iterations (without considering arcs with excessive waiting time at the stops). Additionally, travel times are considered 0 for boarding and alighting arcs, and $t_a > 0$ for on-board and walking arcs.

The graph shown in Fig. 1, where all types of arcs are considered, is often modified in practice. As suggested in Spiess and Florian (1989), when a node i have only one incoming arc $a_1 \in A_i^-$ and one outgoing arc $a_2 \in A_i^+$, with travel times t_{a_1} and t_{a_2} and frequencies f_{a_1} and f_{a_2} respectively, the arcs a_1 and a_2 can be eliminated, and instead consider a single arc a such that $t_a = t_{a_1} + t_{a_2}$, and $f_a = f_{a_1}$. This allows working with a simplified graph, considering fewer nodes and arcs without altering the dynamic of the problem.

An important aspect of assignment models is the behavior of passengers. In this case, it is assumed that each passenger aims to minimize their total travel time and to achieve it, they must choose an arc to continue their trip at each node. A common line problem would be solved at each node $i \in N$: travelers select a nonempty subset of possible lines to continue their trip $s \subseteq A_i^+$ and board the first bus that arrives at the stop and belongs to this set. As a result, the chosen strategy (hyperpath) minimizes their total expected travel time. As mentioned earlier, walking arcs could be included in these strategies.

As stated in Cepeda et al. (2006), the equilibrium flow is the global solution of

$$\underset{v \in V(g)}{\text{Min}}\; G(v, g) = \sum_{d \in D} \left[\sum_{a \in A} t_a v_{a'}^d + \sum_{i \neq d} \max_{a \in A_i^+} \frac{v_a^d}{f_a(v)} - \sum_{i \neq d} g_i^d \tau_i^d(v) \right], \tag{5}$$

In this formulation $\tau_i^d(v)$ represents the expected travel time from i to d, defined by:

$$\tau_i^d(v) = \begin{cases} 0, & \text{if} \quad i = d, \\ \min\limits_{s \in \mathcal{S}_i} \frac{1 + \sum_{a \in s} [t_a + \tau_{j_a}^d] f_a(v)}{\sum_{a \in s} f_a(v)}, & \text{otherwise}, \end{cases} \tag{6}$$

where $\mathcal{S}_i$ is the set of nonempty subsets $s \subseteq A_i^+$. In practice, the method proposed in Nguyen and Pallottino (1988) can calculate τ_i^d.

The gap function $G(v, g)$ represents the difference between the total time experienced by travelers (this is, total travel time + maximum waiting time at stops) and the total expected travel time of the system. In Cepeda et al. (2006), it is proved that its optimal value is 0. Based on this, we will present in the following subsection an algorithm to solve the assignment problem (5).

2.2. Assignment problem

As discussed later, the efficient solution to the assignment problem will be essential for estimating the demand. As mentioned in Cepeda et al. (2006), it is not possible to solve the problem (5) by using descent methods. They proposed to solve it using a heuristic method based on the Method of Successive Averages (MSA) and taking advantage of the knowledge that the optimal value of $G(v, g)$ is 0.

We made flow assignments over different networks observing that when the size of the graph increases, the algorithm proposed in Cepeda et al. (2006) becomes very slow. This slow convergence rate may be due, among other factors, to MSA using a predetermined sequence of step sizes. Against this, MSA can be replaced by similar methods like the Method of Successive Weighted Averages (MSWA) or the self-regulated averaging method, as proposed in Liu et al. (2009) for the Stochastic User Equilibrium Problem. These methods differ in the choice of step α_k at each iteration. In the conventional MSA, the step is $\alpha_k = 1/k$, increasing k by one at each iteration regardless of how close the current solution is to the optimal solution. MSWA sets the steps so that later auxiliary flow has more weight. Finally, the self-regulated averaging method adjusts the step size based on the distance between the intermediate solution and the auxiliary point. As in MSA, at each iteration k the self-regulated method selects firstly an auxiliary point $\mathbf{y}^k = \mathbf{F}(\mathbf{x}^k)$ and then update the solution by:

$$\mathbf{x}^{k+1} = (1 - \alpha_k)\mathbf{x}^k + \alpha_k \mathbf{y}^k \tag{7}$$

The step size is set as $\alpha_k = 1/\beta_k$, where the increment of β_k is determined by the absolute error between current solution $\mathbf{x}^k$ and auxiliary point $\mathbf{y}^k$:

$$\beta_k = \begin{cases} \beta_{k-1} + \Gamma & if \quad \|\mathbf{x}^k - \mathbf{y}^k\| \geq \|\mathbf{x}^{k-1} - \mathbf{y}^{k-1}\| \\ \beta_{k-1} + \gamma & if \quad \|\mathbf{x}^k - \mathbf{y}^k\| < \|\mathbf{x}^{k-1} - \mathbf{y}^{k-1}\| \end{cases} \tag{8}$$

As expressed in Liu et al. (2009), the parameters Γ and γ can be freely chosen considering $\Gamma \in [1.5, 2]$ and $\gamma \in [0.01, 0.5]$, and their values influence the speed of convergence of the algorithm.

After comparing the results on different assignment examples using MSA, MSWA, and the self-regulated method, we observe a considerable reduction in computational times in the last case. In this way, the chosen scheme is similar to the one proposed in Cepeda et al. (2006) but considers a self-regulated averaging method instead of MSA. A detailed description is presented in Algorithm 1:

Algorithm 1: Assignment method based on self-regulated

Result: Flow at equilibrium
Let $k = 0$ and $\beta_0 = 1$;
Find $v^0 \in V(g)$;
while $G(v^k, g) > \varepsilon \cdot G(v^0, g)$ **do**
  Compute $f_a = f_a(v^k)$;
  Compute the shortest hyperpath for each $d \in D$;
  Compute the induced flows $\hat{v}_a^d$;
  Update β_k according to (8);
  Set $\alpha_k = 1/\beta_k$;
  Update $v^{k+1} = (1 - \alpha_k)v^k + \alpha_k \hat{v}$;
  Set $k = k + 1$;
end

The initial flow v^0 corresponds to an all-or-nothing assignment. Then at each iteration, the shortest hyperpath for each $d \in D$ and the induced flows must be found after calculating the effective frequencies. For this purpose, we use the assignment method proposed in Spiess and Florian (1989).

The self-regulated averaging method considerably reduces the time execution compared to the case where MSA is used, and the number of iterations performed can be further reduced by correctly choosing the parameters Γ and γ. To obtain the best values for these parameters, we perform several flow assignments choosing $\Gamma \in [1.5, 2]$ and $\gamma \in [0.01, 0.5]$ and compare the number of iterations performed until convergence. Generally, the best results (the cases where convergence is reached in fewer iterations) are obtained for the lowest values of Γ and γ. Fig. 2 shows the number of iterations performed during the flow assignment over two OD-pairs on the network exposed in the example presented in Section 3.1. This figure shows how the fewest number of iterations is performed when $\Gamma = 1.7$ and $\gamma = 0.01$. An important aspect to note is that whatever the value of Γ and γ, the assignments obtained are the same in all cases, with which a good choice of parameters manages to reduce calculation times without affecting the solution.

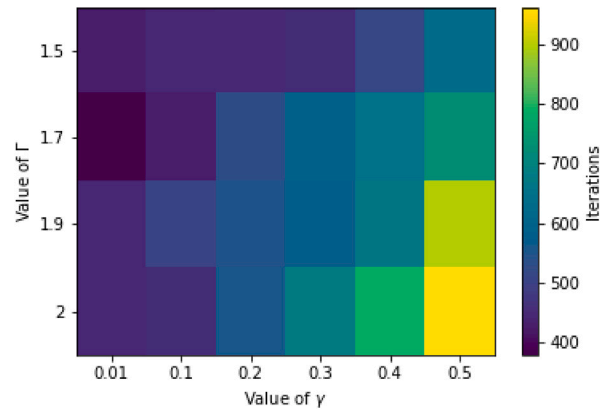


Fig. 2. Number of iterations performed during the flow assignment for different parameters.

2.3. Demand estimation

When the public transport demand is known (when you know how many people want to travel from each origin to each destination), it is possible to estimate their distribution between the different bus lines. For this, it is enough to choose any assignment model that represents the system's and travelers' dynamic, for example, the one proposed in Section 2.1. By having information on the flow distribution and passenger behavior, the network's performance can be monitored and its operation can be controlled to achieve optimal efficiency. This way, improvements can be planned when necessary, generating greater comfort for travelers and attracting new users to the system.

However, transport demand is only sometimes available, requiring its estimation most of the time. As previously mentioned, this can be done using various methods and data from the system. In our case, we propose to perform the estimation by using information about the effective frequencies of the busses and nominal demands (for example, historical data or a previous measurement). Supposing that A_{obs} contains those arcs for which we have information about their frequency and W represents the set of OD-pairs (i, d) for which we want to estimate the demand, the problem is similar to the one proposed in previous works (see Bhouri et al., 2021, Orlando et al., 2021, Orlando et al., 2022) and consists of finding the demand g that minimizes:

$$\min_{g,\,v} \quad \frac{\theta}{|A_{obs}|} \sum_{a \in A_{obs}} \left(\frac{\bar{f}_a - f_a}{\bar{f}_a} \right)^2 + \frac{1}{|W|} \sum_{(i,d) \in W} \left(\frac{\bar{g}_i^d - g_i^d}{\bar{g}_i^d} \right)^2 \tag{9a}$$

$$\text{s.t.} \quad v \in V(g), \tag{9b}$$

$$G(v, g) = 0 \tag{9c}$$

where $\bar{f}$ represents the observed (or measured) frequencies, $\bar{g}$ represents the nominal demand, f_a is the effective frequency for arc a associated with demand g_i^d and θ is a parameter that regulates the level of confidence on the data of observed frequencies compared to nominal demands. Unlike previous works, here we include an inverse proportionality factor at each term of the objective function to handle the difference between the amount of data for frequencies (scaled with

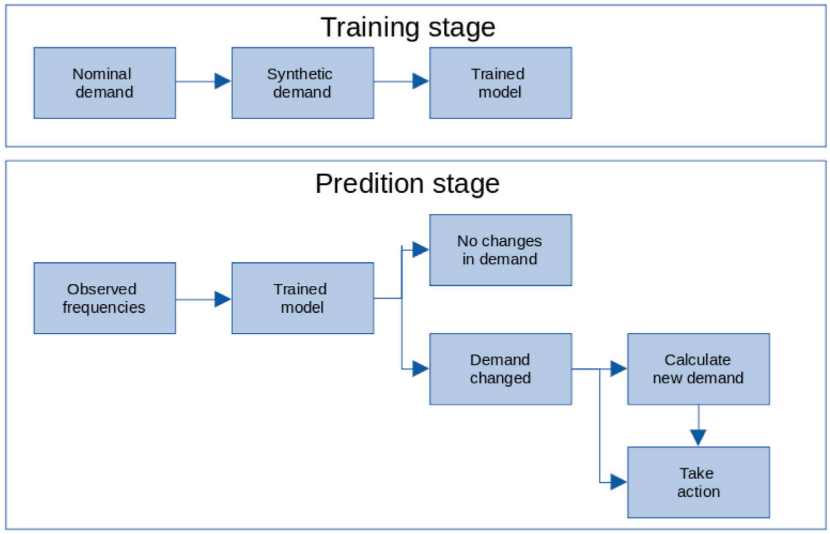


Fig. 3. Training and prediction stages.

$|A_{obs}|$, the number of observed arcs) and for demands (scaled with $|W|$ the number of OD pairs). Also, we apply the regularization parameter θ on the first term instead of using θ^{-1} in the second term to avoid working with minimal orders of magnitude. More details about the choice of θ will be treated in the example in Section 3.1.

To solve the problem (9a)–(9c), we used `optimize.fmin` from the library SciPy v.1.8.0 (Virtanen et al., 2020), available in Python v.3.8.5 (Van Rossum and Drake, 2009), a minimization method based on the Nelder–Mead simplex algorithm (see Lagarias et al., 1998). At each step of this algorithm, the assignment problem is updated and solved, and this is why, as stated in Section 2.2, it is crucial to solve it accurately and efficiently.

Since the network includes walking arcs between centroids and bus stops, people can travel by bus, walking, or combining both. Walking arcs have infinite capacity and frequency; thus, the amount of flow using them does not influence their frequencies. Consequently, different demands may generate the same arc frequencies leading to other possible solutions to the problem (9a)–(9c). Nominal demands play an important role at this point, given that the second term of the objective function (9a) ensures choosing the closest solution to the nominal demand. In mathematical terms, the smaller the parameter θ, the more (strictly) convex the optimization problem.

Section 3 presents a demand estimation example.

2.4. Demand change detection with machine learning

As mentioned above, in each step of solving the problem (9a)–(9c) must be solved an assignment problem by applying algorithm 1. Despite the time-saving improvements discussed in Section 2.2, the execution time for solving the problem (9a)–(9c) increases as the network size grows, making it challenging to use the demand estimation method for day-to-day planning. Therefore, it is essential to have a quick alternative when monitoring the public transport network for changes in demand. If a change is perceived (e.g. through variations in frequencies or smart card data), a fast method can be used instead of running the complete optimization algorithm to confirm or disprove the change. If the change is confirmed, only then should the problem (9a)–(9c) be solved. As mentioned in Bagloee et al. (2018), machine learning can be used as an auxiliary tool to an existing metaheuristic method, or it can also be combined with optimization techniques to design a new solution algorithm. Here, we propose to use an Artificial Neural Network (ANN) as a proxy for demand change detection. To train the ANN, we will need many scenarios composed of demands and the frequencies corresponding to their equilibrium assignment.

This trained model allows the traffic system manager to fast evaluate if it is necessary to take any action to regulate the system, and if it decides to take action, run the optimization algorithm to get an accurate demand estimation (Fig. 3).

2.4.1. Generating synthetic demand-scenarios

Given a network, assignments can be generated for different demand configurations in advance. Solving the assignment problem (5), the frequencies of the equilibrium state associated with an artificial vector of demand can be known.

To generate training data for the neural network, it is assumed that changes in demand patterns in the transit system are smooth (an implicit assumption taken by the inverse optimization problem developed in Section 2.3). Given the vector of all nominal demands $\bar{g}_i^d$ representing the historical measurements, new demand vectors can be generated using Gaussian noise, and the frequency vector can be calculated by solving the assignment problem for each of them. Repeating the process, we could create a dataset in which demands are random, but the frequencies comply with the equilibrium model.

2.4.2. Trained model

The trained model consists of predicting the Euclidean distance between the nominal demand vector (used as a base for dataset creation) and the demand vector calculated by the problem (9a)–(9c), given a vector of frequencies. With this model, the transit system manager could determine whether the changes in the system are significant enough to run the whole (expensive) demand estimation algorithm to get the current demand more accurately.

The trained model was a neural network with an input layer that collects all frequencies, five fully connected hidden layers with the same size as the input layer, a single output layer, and the root mean square error (RMSE) as the loss function. Relu was the activation function used in the hidden layers. The implementation was done using Flux (Innes, 2018) running under Julia 1.7 (Bezanson et al., 2017).

The model was trained over the whole network for demand values similar to the historical ones. So, it is powerful enough for change detection under the assumption of smooth and minor changes and no change in the transit network topology. If the frequency changes are abrupt, they will correspond to abrupt changes in the demand vector. In this case, the variation in historical demand would be evident, so using the ANN to detect demand changes would not be necessary.

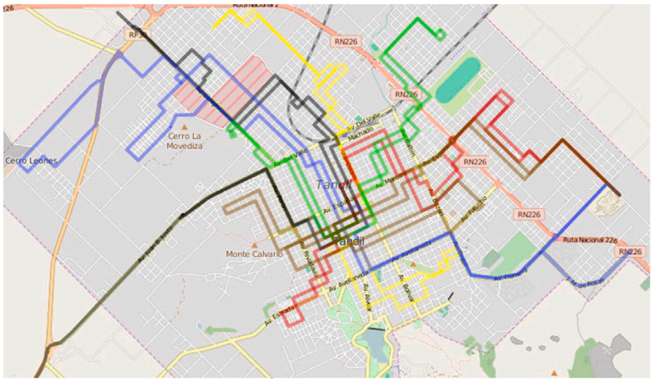


Fig. 4. Tandil bus network.

3. Results

Our objective is to perform the demand estimation on the transit network of Tandil (a medium-sized city in Buenos Aires, Argentina). This bus network comprises six lines serving an area of approximately 30 km^2 and connecting different city zones, enabling the combination of different lines during the trip. To distinguish the lines, each one has a name: $L1$, $L2$, $L3$, $L4$, $L5$, $L6$ and a color: yellow, red, black, blue, green, and brown, respectively, and their routes can be seen in Fig. 4.

Given the size of the entire network and intending to test our model and obtain results on a small and high scale, we selected sub-networks of increasing size and estimated the demand for each. Section 3.1 presents one of these examples. The selected sub-network includes the city center, where the largest line intersections occur. In addition, we used factual information about the nominal frequencies and capacities of the buses. By regulation, a bus leaves every 10 minutes, and each can carry up to 64 passengers, of which half can access a seat, and the rest must travel standing. Finally, travel times were estimated from the information provided by smart cards. For each line, we had temporal and spatial information about the place where each bus card registered a trip. Using the records between consecutive stops, we selected a day of the week and a time slot to estimate the average travel time between stops.

Demand estimation problem (9a)–(9c) requires knowing the nominal demand (that can correspond to previous measurements) and the observed frequencies. In the examples, we perform an assignment and use the resulting frequencies as the observed frequencies. This allows us to test our model since we will know if the estimated demands are correct because we know what demands generate the observed frequencies. Once the model has been tested and the appropriate parameters have been selected, it can be applied to the entire network using the data collected from the system.

3.1. Numerical example

The area under consideration consists of 90 blocks and is serviced by all the bus lines in the city's network. The graph representing this network contains 169 nodes (of which 26 are centroids and 143 correspond to bus stops and line nodes) and 847 arcs (including boarding, alight, on-board, and walking arcs). In Fig. 5(a) can be seen the distribution of the bus stops (green points) and centroids (blue points), while in Fig. 5(b) are represented the areas covered by the centroids, named traffic analyzing zones (TAZ). The graph illustrating the simplified network, considering only bus stops and on-board arcs, can be seen in Fig. 6, where the label accompanying each link indicates the bus line that serves that section.

We choose two pairs of centroids to estimate the demand between them: the pairs $p_1 = (1, 2)$ (red points in Fig. 5(a)) and $p_2 = (3, 4)$ (orange points in Fig. 5(a)). Centroids 1 and 2 are separated by a Manhattan distance of 1.41 km, and centroids 3 and 4 are separated by a Manhattan distance of 1.78 km. The assigned demands to obtain the information about measured frequencies were $g_1^2 = 120$ and $g_3^4 = 150$. Using nominal demands $\bar{g}_1^2 = 90$ and $\bar{g}_3^4 = 100$, we perform estimations for different values of θ, and we compare how close the estimated demand is to the objective. The results for these estimations can be seen in Table 1, where they are shown the estimated demand for each OD pair, the number of objective function evaluations, and the root mean square error (RMSE) between the objective demand [120,150] and the estimated demands. There it can be seen that the best results are obtained for θ between 10^3 and 10^4, more specifically for $\theta = 6 \cdot 10^3$. In this case, the RMSE is less than 1, which indicates great precision given that we are working with passengers (and, consequently, with integers). Table 1 also shows the calculation times for each example, which were carried out on a server using an Intel(R) Xeon(R) CPU E52683 v4 with 2.10 GHz and 4 Gb RAM. As observed, the calculation times can reach 80 hours for a relatively small network, highlighting the need for a complementary tool that enables quick preliminary analysis.

Table 1
Demand estimation for different θ for the example 1.

θ	Estimated g_1^2	Estimated g_3^4	Function evaluations	RMSE	Calculation time (hs)
10^1	92.30	106.55	38	36.44	6.32
10^2	101.81	119.38	53	25.19	8.36
10^3	118.47	142.74	76	5.25	15.51
$2\ 10^3$	119.63	145.55	67	3.16	13.4
$4\ 10^3$	120.42	147.92	72	1.51	14.06
$6\ 10^3$	119.79	148.96	68	0.75	13.6
$8\ 10^3$	123.54	150.34	65	2.52	12.28
10^4	123.54	150.34	65	2.52	12.56
10^5	134.59	150.00	400	10.31	80.06
10^6	130.75	149.76	403	7.60	79.09

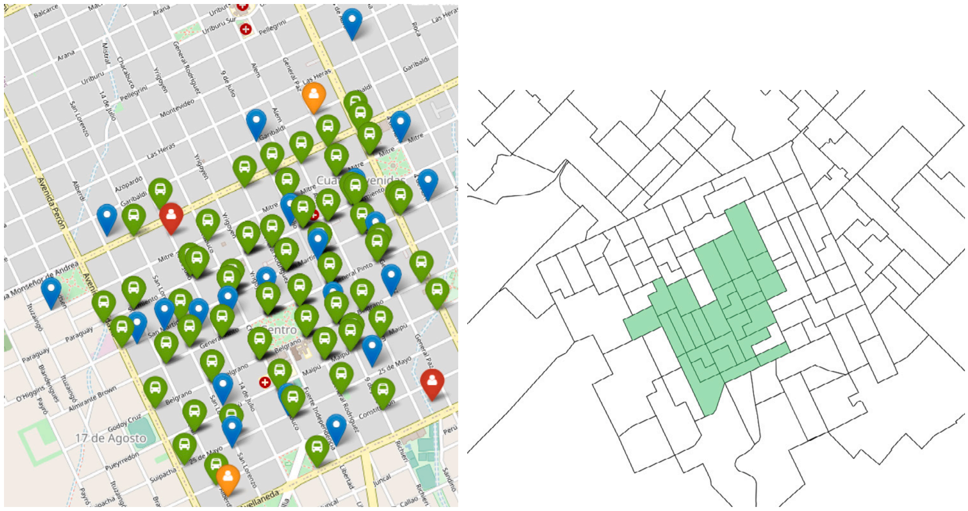
3.2. Results with the machine learning approach

To test the approach, we created a synthetic dataset of 10000 samples. Each sample was generated by adding Gaussian noise to the historical demand, using a deviance coefficient equal to 0.15. The full dataset was split into training and test sets (keeping 20% samples in the test set). Predictions for the test set were compared with the results for the inverse optimization problem-solving algorithm (reshaping the results of the last one into the euclidean distance between demands). Fig. 7 shows the distribution of the percentage error between the prediction and the estimated demand (ED) obtained from Section 2.3 ($(prediction - ED)/ED$).

Despite significant percentage errors, the model is a valuable filter from a practical point of view. Having the system manager a threshold level that determines when a change in the demand triggers an action over the system, predicted values under this threshold (minus an error percentage) are safe to indicate non-significant changes.

4. Discussion

An improvement for the flow assignment algorithm was proposed, replacing MSA with a self-regulated method. This change reduces the computation times, which becomes crucial when solving the demand estimation problem without changing the results. In addition, the bilevel model for bus demand estimation has been improved compared to the previous version. Proportionality factors and a regularization parameter have been incorporated, allowing for managing the difference between the amount of collected information and reducing numerical errors by avoiding working with small orders of magnitude. Finally, an ANN was designed and calibrated to detect if the changes in the system are significant enough to assume that there is a change in demand. This



(a) Bus stops and centroids. (b) Traffic analyzing zones.

Fig. 5. Urban areas served by public transport.

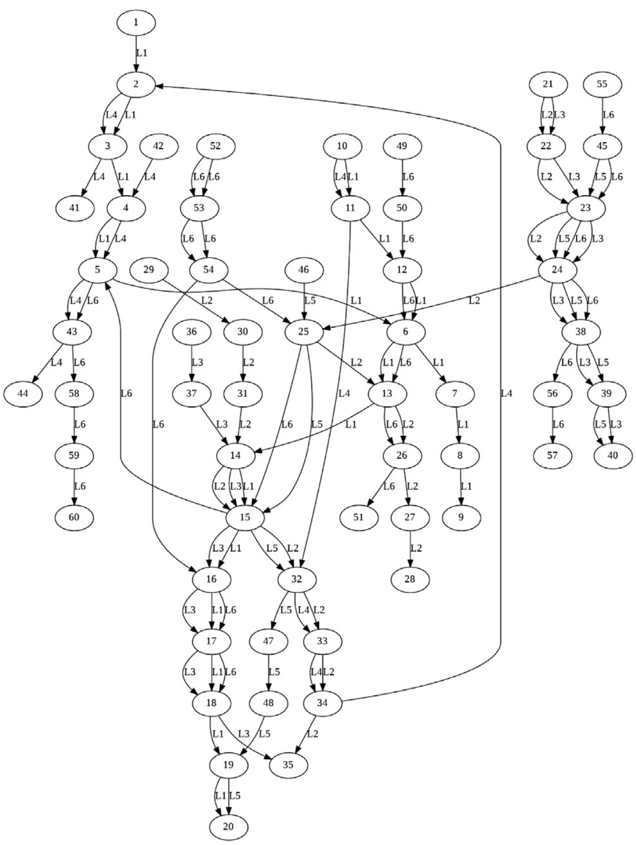


Fig. 6. Graph without boarding, alight and walking arcs for example 1.

approach provides a quick tool to determine if it is necessary to run the demand estimation algorithm or if the changes in the system are not significant enough to ensure that the demand has changed.

The main limitation of the proposed bilevel model is the long computation time that its resolution requires when applied to large networks. However, this has not impeded the performing of the experiments of this research, which allowed testing both the numerical model and the approach with ANN. Future works that would solve this limitation are proposed in the next section.

5. Conclusions

In this work, the bilevel model for demand estimation and the flow assignment algorithm involved during the estimation have been modified. The results show that these modifications reduce calculation times and numerical errors from working with small orders of magnitude. The proposed bilevel problem can calculate the associated demand by knowing historical demand and surveying real frequencies. The model was tested with theoretical examples from the literature and a real-world example from the Tandil transport network, which is presented in this work. The results vary depending on the value assigned to the regularization parameter but showed relatively low estimation error by keeping said parameter within a specific range. Through tests on different examples, this would allow establishing an appropriate range for this value and then using it to estimate demand on the entire network. An advantage of the proposed model is that it can be used with any flow assignment model, not only with the one chosen in this work.

However, the results show that computation times are still considerably high, even in medium-sized networks. To address that limitation, we have proposed a machine learning approach, using an ANN as a tool that detects whether changes in the system are significant enough to evidence a change in demand. This tool is very useful since traffic authorities can use it to rule out false positives and only run the demand estimation algorithm when they are certain that the demand has changed. As mentioned, the models proposed in this paper could be adapted to estimate the demand for any product or service besides public transportation.

For future works, we are already exploring several research lines:

- Estimate the demand on larger sub-networks.
- Improve the current algorithm implementation to reduce the calculation times significantly. There are several approaches to work with: improve data classes and procedures, move to a parallel-based schema, and work with a lower-level programming language/paradigm.
- Improve the machine learning model, reducing errors and allowing full demand prediction (instead of the current usage as a filter). Moving from a conventional neural network to a graph-based one seems like a promissory path.
- Mix the current algorithm with a machine learning approach, for example, replacing some of the steps, in some iterations, with estimations.

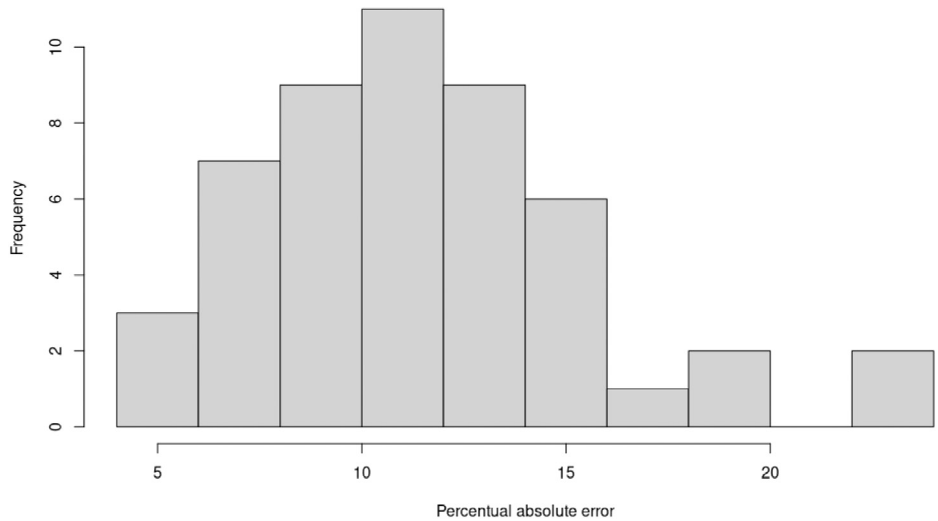


Fig. 7. Distribution of the percentage error.

CRediT authorship contribution statement

Victoria M. Orlando: Conceptualization, Methodology, Software, Investigation, Writing – original draft. **Enrique G. Baquela:** Conceptualization, Methodology, Software, Investigation, Writing – original draft. **Neila Bhouri:** Conceptualization, Writing – review & editing, Funding acquisition. **Pablo A. Lotito:** Conceptualization, Methodology, Formal analysis, Investigation, Supervision.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgments

The research of V. M. Orlando and P. A. Lotito is supported by CONICET Argentina (PIP-11220200100815CO) and Agencia Nacional de Promoción de la Investigación, el Desarrollo Tecnológico y la Innovación, Argentina (PICT-2019-02886). The research of E. G. Baquela is supported by Universidad Tecnológica Nacional, Argentina (PIDUTN8252). Neila Bhouri's research is supported by the French Ministry of Higher Education and Research (MESRI). The authors thank the Municipality of Tandil for providing the data of the transport smart card used in this work.

References

Ahern, Z., Paz, A., Corry, P., 2022. Approximate multi-objective optimization for integrated bus route design and service frequency setting. Transp. Res. B 155, 1–25. http://dx.doi.org/10.1016/j.trb.2021.10.007, URL: https://www.sciencedirect.com/science/article/pii/S0191261521001910.

Asmael, N., Waheed, M., 2018. Demand estimation of bus as a public transport based on gravity model. MATEC Web Conf. 162, 01038. http://dx.doi.org/10.1051/matecconf/201816201038.

Bagloee, S.A., Asadi, M., Sarvi, M., Patriksson, M., 2018. A hybrid machine-learning and optimization method to solve bi-level problems. Expert Syst. Appl. 95, 142–152. http://dx.doi.org/10.1016/j.eswa.2017.11.039.

Bajari, P., Nekipelov, D., Ryan, S.P., Yang, M., 2015. Machine learning methods for demand estimation. Amer. Econ. Rev. 105 (5), 481–485. http://dx.doi.org/10.1257/aer.p20151021.

Barcelo, J., Montero, L., Bullejos, M., Serch, O., Carmona, C., 2012. Dynamic OD matrix estimation exploiting bluetooth data in urban networks. pp. 116–121.

Barry, J.J., Newhouser, R.J., Rahbee, A., Sayeda, S., 2002. Origin and destination estimation in New York city with automated fare system data. Transp. Res. Rec. 1817, 183–187. http://dx.doi.org/10.3141/1817-24.

Bezanson, J., Edelman, A., Karpinski, S., Shah, V.B., 2017. Julia: A fresh approach to numerical computing. SIAM Rev. 59 (1), 65–98. http://dx.doi.org/10.1137/141000671.

Bhouri, N., Lebacque, J.P., Lotito, P.A., Orlando, V.M., 2021. A bilevel model for public transport demand estimation. Transp. Res. Procedia 52, 669–676. http://dx.doi.org/10.1016/j.trpro.2021.01.080, 23rd EURO Working Group on Transportation Meeting, EWGT 2020, 16-18 September 2020, Paphos, Cyprus.

Cepeda, M., Cominetti, R., Florian, M., 2006. A frequency-based assignment model for congested transit networks with strict capacity constraints: characterization and computation of equilibria. Transp. Res. B 40 (6), 437–459. http://dx.doi.org/10.1016/j.trb.2005.05.006.

Cipriani, E., Gemma, A., Mannini, L., Carrese, S., Crisalli, U., 2021. Traffic demand estimation using path information from bluetooth data. Transp. Res. C 133, 103443. http://dx.doi.org/10.1016/j.trc.2021.103443.

Codina, E., 2013. A variational inequality reformulation of a congested transit assignment model by Cominetti, Correa, Cepeda, and Florian. Transp. Sci. 47 (2), 231–246. http://dx.doi.org/10.1287/trsc.1120.0427.

Codina, E., Rosell, F., 2017. A heuristic method for a congested capacitated transit assignment model with strategies. Transp. Res. B 106, 293–320. http://dx.doi.org/10.1016/j.trb.2017.07.008.

Cominetti, R., Correa, J., 2001. Common-lines and passenger assignment in congested transit networks. Transp. Sci. 35 (3), 250–267. http://dx.doi.org/10.1287/trsc.35.3.250.10154.

De Cea, J., Fernández, E., 1993. Transit assignment for congested public transport systems: An equilibrium model. Transp. Sci. 27 (2), 133–147. http://dx.doi.org/10.1287/trsc.27.2.133.

Dick, A.A., 2008. Demand estimation and consumer welfare in the banking industry. J. Bank. Financ. 32 (8), 1661–1676. http://dx.doi.org/10.1016/j.jbankfin.2007.12.005.

Echaniz, E., Cordera, R., Rodriguez, A., Nogués, S., Coppola, P., dell'Olio, L., 2022. Spatial and temporal variation of user satisfaction in public transport systems. Transp. Policy 117, 88–97. http://dx.doi.org/10.1016/j.tranpol.2022.01.003.

Innes, M., 2018. Flux: Elegant machine learning with Julia. J. Open Source Softw. http://dx.doi.org/10.21105/joss.00602.

Jaber, A., Al-Sahili, K., Juhász, J., 2022. Demand-responsive users' travel behavior and satisfaction analysis in small cities: Case study of the public transportation system in palestine. Period. Polytech. Transp. Eng. http://dx.doi.org/10.3311/PPtr.19914.

Lagarias, J.C., Reeds, J.A., Wright, M.H., Wright, P.E., 1998. Convergence properties of the Nelder-Mead simplex method in low dimensions. SIAM J. Optim. 9 (1), 112–147. http://dx.doi.org/10.1137/S1052623496303470.

Lera-López, F., Sánchez, M., Faulin, J., Cacciolatti, L., 2014. Rural environment stakeholders and policy making: Willingness to pay to reduce road transportation pollution impact in the Western Pyrenees. Transp. Res. D 32, 129–142. http://dx.doi.org/10.1016/j.trd.2014.07.003.

Liu, H., He, X.S., He, B.S., 2009. Method of successive weighted averages (MSWA) and self-regulated averaging schemes for solving stochastic user equilibrium problem. Netw. Spat. Econ. 9, 485–503. http://dx.doi.org/10.1007/s11067-007-9023-x.

Montero, L., Barcelo, J., Codina, E., 2012. Adapting a dynamic OD matrix estimation approach for private traffic based on bluetooth data to passenger OD matrices. pp. 41–49.

Nguyen, S., Pallottino, S., 1988. Equilibrium traffic assignment for large scale transit networks. European J. Oper. Res. 37 (2), 176–186. http://dx.doi.org/10.1016/0377-2217(88)90327-X.

Oliker, N., Bekhor, S., 2020. A frequency based transit assignment model that considers online information and strict capacity constraints. EURO J. Transp. Logist. 9 (1), 100005. http://dx.doi.org/10.1016/j.ejtl.2020.100005.

Orlando, V.M., Baquela, E.G., Bhouri, N., Lotito, P.A., 2021. An inverse bilevel equilibrium problem for public transport demand estimation. IFAC-PapersOnLine 54 (2), 13–18. http://dx.doi.org/10.1016/j.ifacol.2021.06.002, 16th IFAC Symposium on Control in Transportation Systems CTS 2021.

Orlando, V.M., Baquela, E.G., Bhouri, N., Lotito, P.A., 2022. Analytical and simulation-based estimation of public transport demand. Transp. Res. Procedia 62, 727–734. http://dx.doi.org/10.1016/j.trpro.2022.02.090, 24th Euro Working Group on Transportation Meeting.

Sánchez-García, M., Zouaghi, F., Lera-López, F., Faulin, J., 2021. An extended behavior model for explaining the willingness to pay to reduce the air pollution in road transportation. J. Clean. Prod. 314, 128134. http://dx.doi.org/10.1016/j.jclepro.2021.128134.

Sawik, B., Faulin, J., Serrano-Hernandez, A., Ballano, A., 2022a. A simulation-optimization model for automated parcel lockers network design in urban scenarios in pamplona (Spain), Zakopane, and Krakow (Poland). In: 2022 Winter Simulation Conference. WSC, pp. 1648–1659. http://dx.doi.org/10.1109/WSC57314.2022.10015458.

Sawik, B., Serrano-Hernandez, A., Muro, A., Faulin, J., 2022b. Multi-criteria simulation-optimization analysis of usage of automated parcel lockers: A practical approach. Mathematics 10 (23), http://dx.doi.org/10.3390/math10234423.

Spiess, H., Florian, M., 1989. Optimal strategies: A new assignment model for transit networks. Transp. Res. B 23 (2), 83–102. http://dx.doi.org/10.1016/0191-2615(89)90034-9.

Spinner, S., Casale, G., Brosig, F., Kounev, S., 2015. Evaluating approaches to resource demand estimation. Perform. Eval. 92, 51–71. http://dx.doi.org/10.1016/j.peva.2015.07.005.

Sun, W., Schmöcker, J.D., 2021. Demand estimation for public transport network planning. In: Mulley, C., Nelson, J.D., Ison, S. (Eds.), The Routledge Handbook of Public Transport. Routledge, pp. 289–305. http://dx.doi.org/10.4324/9780367816698.

Tirachini, A., Godachevich, J., Cats, O., Muñoz, J., Soza-Parra, J., 2021. Headway variability in public transport: a review of metrics, determinants, effects for quality of service and control strategies. Transp. Rev. 42, 1–25. http://dx.doi.org/10.1080/01441647.2021.1977415.

Van Rossum, G., Drake, F.L., 2009. Python 3 Reference Manual. CreateSpace, Scotts Valley, CA.

Virtanen, P., Gommers, R., Oliphant, T.E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., van der Walt, S.J., Brett, M., Wilson, J., Millman, K.J., Mayorov, N., Nelson, A.R.J., Jones, E., Kern, R., Larson, E., Carey, C.J., Polat, İ., Feng, Y., Moore, E.W., VanderPlas, J., Laxalde, D., Perktold, J., Cimrman, R., Henriksen, I., Quintero, E.A., Harris, C.R., Archibald, A.M., Ribeiro, A.H., Pedregosa, F., van Mulbregt, P., SciPy 1.0 Contributors, 2020. SciPy 1.0: Fundamental algorithms for scientific computing in python. Nature Methods 17, 261–272. http://dx.doi.org/10.1038/s41592-019-0686-2.

Walpen, J., Mancinelli, E.M., Lotito, P.A., 2015. A heuristic for the OD matrix adjustment problem in a congested transport network. European J. Oper. Res. 242 (3), 807–819. http://dx.doi.org/10.1016/j.ejor.2014.11.009.