Case Studies on Transport Policy 18 (2024) 101282

Contents lists available at ScienceDirect

Case Studies on Transport Policy

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

The application of direct ridership models in the evaluation of the expansion of the Porto Light Rail Transit

Paulo Pinho *, Miguel Lopes, Marcelo Altieri, Frederico Moura e Sá, Cecília Silva, Ana Amante

CITTA - Research Centre for Territory, Transports and Environment, Faculty of Engineering, University of Porto, Portugal

A B S T R A C T

The main purpose of this paper is to show how a direct demand ride model coupled with a transfer model was able to support the choice of alternative routes for the expansion of the light rail system serving the Porto Metropolitan Area. After an overview of the literature on direct ridership models, emphasizing some key issues such as the need for a systematic assessment of their forecasting performance, the issues related to the definition of the pedestrian catchment area and the limitations of the simultaneous consideration of demand and supply effects, the paper moves into the case study, providing some background information on current occupation densities, land uses and mobility patterns, as well as on the performance of the existing LRT in the Metropolitan Area of Porto. The development of the direct ridership model, measuring the potential attractiveness of each station, and the transfer model, measuring the number of transfers at each station, are presented in detail. A justification is provided why, in this case, a two-step modelling approach was necessary. Further details about the statistical tests for model validation are also provided. After a brief characterization of the alternative routes under analysis for the expansion of the network, the modelling results are presented enabling a comparative assessment of the potential performance of each proposed route. The paper ends with a discussion of the relevance of this modelling results vis a vis the actual final decisions on investment priorities taken jointly by the Metropolitan Council and the Metro Company.

1. Introduction

The advantages of introducing light rail transit (LRT) in medium size cities and metropolis are relatively consensual in the scientific literature and seem to justify the continuous investments in new systems or in the expansion of existing ones. Indeed, the long term benefits such systems can generate go beyond the scope of the public transport sector, to include the fulfilment of wider objectives associated with urban planning and development, environmental policy, energy efficiency, and climate change.

The Metropolitan Area of Porto (MAP) with, approximately, 1,7 million inhabitants, is served by a LRT since 2002. By 2006, the so-called *first phase of the LRT investment project* was concluded. The rail network, made of five lines, had then a total extension of 59 km. Since the launching of the system its attractiveness sharply raised. In the first five years of operation, the annual passengers.km indicator, increased from 26,5 million to 245,9 million. In 2023 this indicator reached 425,3 million (source Metro do Porto Annual Reports). The overall impact in the metropolitan area was profound and justified the preparation of a strategic plan to guide the expansion of the system (Metro do Porto, 2007). The decision to go ahead with a second phase of investments was corroborated by the positive results of a systematic evaluation of the performance of the first phase of the LRT and its positive social, economic, and environmental impacts on the Porto metropolitan area (see Pinho and Vilares, 2009). However, the global financial crises of 2007–2008 and the subsequent national financial and economic crises of 2010–2014, resulted in the postponement of these investments. Later, in 2015, engineering studies on the possible routes for the expansion of the LRT were resumed. In the following year, the Metro do Porto Company commissioned the comparative assessment of the potential demand of those alternative routes. A first direct demand ride model was developed and applied in combination with a preliminary approach to estimate the total amount of transfers as a percentage of direct ridership at each station (Metro do Porto, 2017). The results were very satisfactory and encouraged the adoption of an improved methodology on a second and more ambitious study, commissioned in 2020, with identical objectives but now covering a larger number of alternative routes for the LRT expansion (Metro do Porto, 2021). This latter study also included the analysis of the likely demand of an associated metrobus system, which analysis, however, will not be dealt with in this paper.

The experience gained in the preparation of these two studies, and in particular in the more recent one, and the opportunity to feed and follow

* Corresponding author.
E-mail address: pcpinho@fe.up.pt (P. Pinho).

https://doi.org/10.1016/j.cstp.2024.101282
Received 5 January 2023; Received in revised form 15 July 2024; Accepted 19 August 2024
Available online 22 August 2024

the subsequent public debate to reach the final decision on the short list of selected routes for investment, constitute the basis on which this paper's main research question was formulated: can a direct demand ride model coupled with a transfer model be able to support the choice among alternative routes for the expansion of a light rail system?

After this introduction, direct ridership models are critically presented to illustrate their advantages and limitations concerning other more sophisticated transport demand models. Then follows the empirical sections of this paper with the development of the case study on the expansion of the Porto LRT. Emphasis is placed on the design of both the direct and the indirect ridership models, on the relevance of the combined results, and on how these results were able to influence the complex metropolitan decision-making process. The concluding section summarises the arguments on which the answer to the research question is built.

2. Direct ridership models

Forecasting ridership is key for public transport planning. Transport companies routinely assess how changes to their network affect ridership. From more complex models, like activity-based, tour-based, and agent-based models, to more simple ones, like trip-based models, have long been used to forecast public transport demand. All these models comprise the four main steps of travelling, namely, trip generation, trip distribution, modal choice and, finally, route assignment (see, for example, McNally, 2007). The practical application of these traditional models requires, however, extensive surveying, such as conventional household travel surveys, and complex computing, resulting in the consumption of massive amounts of technical and financial resources (Gutiérrez et al., 2011). Understandably, such demands are often difficult to accommodate within the resources available and time constraints of decision making, in particular in small and medium size metropolitan areas.

In this context, direct ridership models (DRM), also commonly termed direct demand models, have emerged as faster, cheaper, and yet reliable and scientifically sound alternatives to forecast ridership for a wide range of public transport services (e.g., Cervero et al., 2010; Duduta, 2013; Yang et al., 2016). These direct ridership models rely on linear econometric approaches, among which stand out, for their simplicity, the ordinary least square models (OLS) (e.g., Kim et al., 2016; Zhao et al., 2013) and the geographically weighted regression models (e.g., Cardozo et al., 2012; Jun et al., 2015), in this latter case due to their important capacity to incorporate spatial autocorrelations. More advanced econometric and geo-statistic models have been used, however, on a minor scale (see, for example, Zhang and Wang, 2014, for the use of Kriging models to accommodate prediction errors at unsampled locations). More recently DRM have also been applied by using machine learning methods (see Yan et al., 2020).

In contrast, DRM have been widely used by research exploring empirically the links between ridership, often at the station level, with a range of socio-economic factors, such as income, age, or gender, demographic factors, such as population and job densities, and accessibility-based indexes, such as land-use diversity and four-way intersections (Iseki et al., 2018; Jun et al., 2015; Kuby et al., 2004; Sohn and Shim, 2010).

The use of this type of models to forecast public transport ridership advances from the idea that once the econometric formulation establishes robust links between a variety of features that describe the surroundings of the stations with their respective ridership values, these links might be extended throughout the territory and, therefore, forecast ridership at areas not yet served by the network. As the estimation of each vector affecting ridership is highly sensitive to the full set of variables in the model, as well as to the particularities inherent to each case study, ridership forecasting should be designed on a case-by-case basis, rather than simply applying standard coefficients to each variable (Zhang and Wang, 2014).

2.1. Forecasting performance

Despite the increasing use of DRM to forecast ridership, few studies have assessed its performance, especially in comparison with the traditional four-step approaches. Mucci and Erhardt (2018) evaluated the ability of DRM to forecast medium-term changes for both bus and rail-based services in San Francisco. The research specified two distinct models, one for each transport mode, calibrated with data from 2009 to forecast ridership changes in 2016. The results found that the rail model correctly predicted the direction of change (ridership increase) but under-estimated the overall ridership by about 6 % of the true percentage change. On the other hand, the bus model failed to predict the direction of change (ridership increase, whereas it decreased), and consequently, over-estimated the overall ridership by about 15 % of the true percentage change. Another study by Gutiérrez et al. (2011) compared the estimations from a four-step model and a DRM with the actual boardings in Madrid's subway system, finding a slightly better correlation fit on the former approach (adjusted $R^2 = 0.811$) as compared to the latter approach (adjusted $R^2 = 0.753$).

2.2. Pedestrian catchment area

Another key issue of DRM is the definition of the station's pedestrian catchment area (PCA). Apart from ensuring that the model follows the appropriate statistical conditions and validations for forecasting purposes (e.g., homoscedasticity, multicollinearity, and other Gauss–Markov assumptions), the overall estimation accuracy, ultimately, also depends on the adopted methods and criteria to define the PCA. The definition can be roughly split into three components, namely, the method to draw the boundaries of the catchment area, the estimation of the influence threshold, and the incorporation of decaying effects.

Regarding the first component, the model accuracy may depend on the method deployed to define the station's PCA. In this respect past research has found inconsistent results. On the one hand, findings in Guerra et al. (2012) suggest that differences in the catchment area of radial-buffers, diamond-shaped polygons, or network-distance paths are largely irrelevant to change the overall model goodness-of-fit. On the other hand, findings in Gutiérrez et al. (2011) suggest that detailed network-distance paths tend to provide more accurate estimations compared to generalized radial-buffers. Furthermore, though hardly measurable, the second component deals with people's general willingness to walk to use public transport (Guerra et al., 2012). The standard walking threshold for rail-based services often ranges from 500 m (Choi et al., 2012; Kim et al., 2016; Sung and Oh, 2011) to 800 m (Cardozo et al., 2012; Guerra et al., 2012; Zhao et al., 2013). That is, the characteristics of the station's surroundings might find minimal, or even null, over a maximum walking threshold. This latter observation leads to the third component.

The influence of the independent variables, both in terms of the vector magnitude and statistical significance, might decay as the distance from the station increases rather than disappear right after crossing the maximum walking distance threshold. Gutiérrez et al. (2011) tested multiple regressions incorporating distance-decay weights based on 100 m bands to progressively fade the influence from population and employment. The authors found the model with 800 m walking threshold with the best goodness-of-fit. On the other hand, findings in Guerra et al. (2012), Sung et al. (2014) and Jun et al. (2015) suggest that each variable has a particular range of influence. For instance, Guerra et al. (2012) found that population and jobs affect ridership in particular ways. According to their results, the number of jobs found a clear decaying effect, whereas the population found no clear decaying effect, influencing ridership to a greater extent in the 0.25-to-0.50 miles band. Sung et al. (2014) found that while the influence of residential density has quite a consistent decay pattern, roughly disappearing after the 1 km threshold, the influence of large-scale commercial density tends to be concentrated on the 250–500 m band,

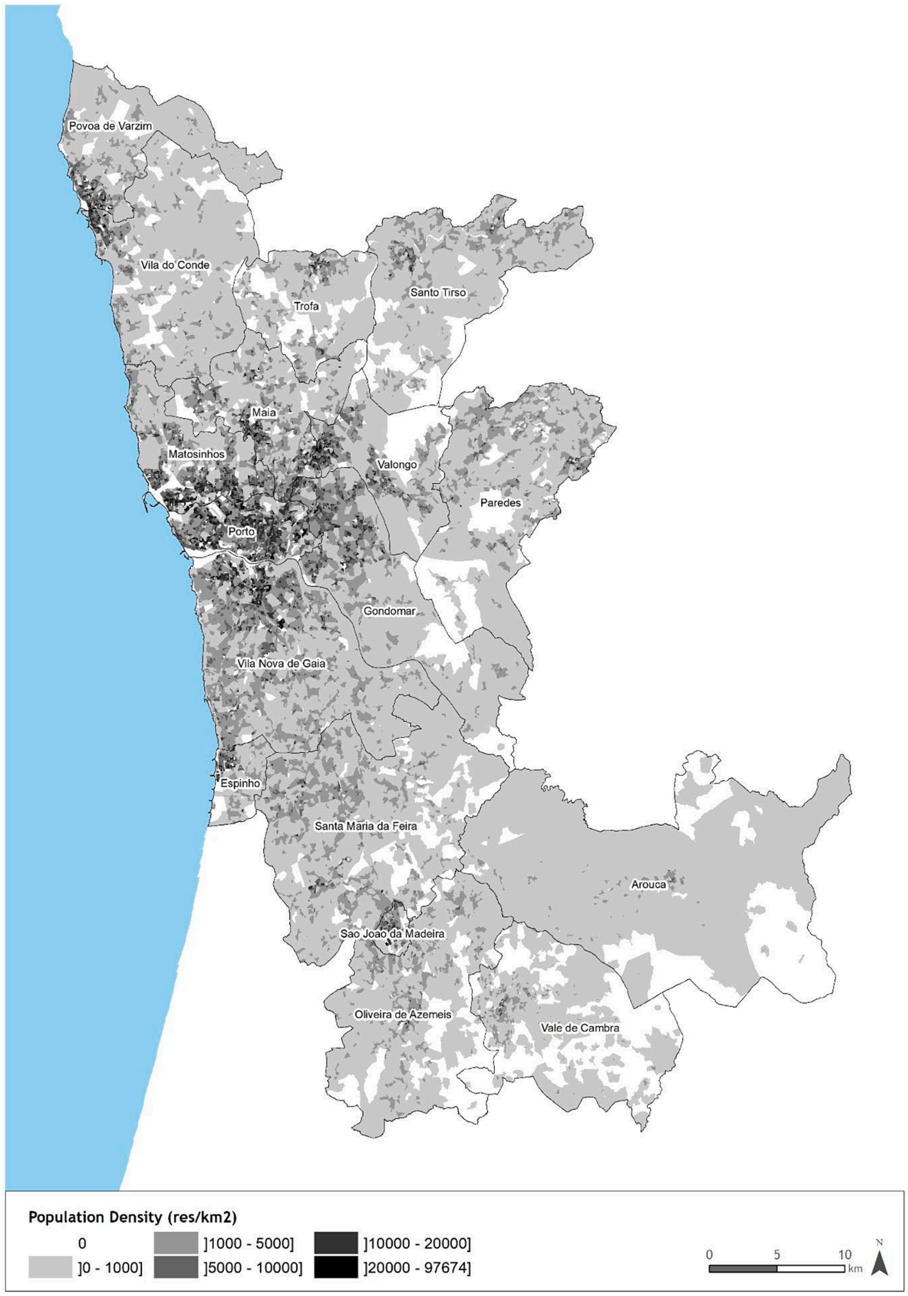

Fig. 1. Metropolitan and municipal boundaries and population densities in the Metropolitan Area of Porto (MAP). In the centre of the map, the so-called Greater Porto, including the municipalities of Porto, Matosinhos, Maia, Valongo, Gondomar and Gaia, concentrate the highest population densities.

whereas large-scale public service density on the 750–1000 m band. Likewise, Jun et al. (2015) found that while the influence of employment density tends to follow a consistent decay pattern, i.e., the coefficient decreases as the distance from stations increases, the influence of the population density finds greater influence over the 300–600 m range.

2.3. Demand and supply simultaneity

Ridership, which represents the demand for public transport and, therefore, the dependent variable on DRM, is typically represented by the average number of boardings (entries) per station/stop over a period, such as daily (e.g., Jun et al., 2013; Zhao et al., 2013), monthly (e.g., Cardozo et al., 2012; Gutiérrez et al., 2011), or yearly (e.g., Taylor et al., 2009). Few public transport services with distance-based fare systems, which require checking out at the destination, like Tokyo and Seoul, provide richer data accounting for boardings and alightings (Sung and Oh, 2011), as well as the complete journey information for a comprehensive origin–destination matrix (Choi et al., 2012; Lee et al., 2019). The further disaggregation into weekdays/weekends (e.g., Sung and Oh, 2011) or into different time periods during a single day (e.g., Choi et al., 2012) are less commonly found in the literature.

However, these data often aggregate two distinct types of boardings: the direct ridership (riders from within the PCA) and indirect ridership (riders from outside the PCA reaching via transfers). This distinction is important because DRM are designed to forecast the direct ridership, finding severe shortcomings to forecast the indirect ridership. The inability to disaggregate ridership into these two groups might lead DRM to misestimations of the coefficients and hence biased ridership forecasting. This is because, differently from the direct ridership, indirect ridership essentially depends on the relationships between the characteristics of the territories outside the stations/stops PCA, the level of supply and/or performance of feeder systems, and the broader connectivity that each station/stop provides to these territories. The inclusion of variables accounting for the level of service ultimately makes DRM sensitive to the simultaneity between supply and demand.

Put it simply, there is a feedback effect; transit operators, either from mass/trunk or feeder systems, are expected to adjust the supply of services in response to changes in demand, which in turn further influences public transport use (Peng et al., 1997). Thus, as evidenced by Taylor et al. (2009), the combination of endogenous variables into a single equation violates one of the conditions of OLS leading, potentially, to biased and inconsistent forecasts. Nonetheless, many studies still incorporate endogenous variables without considering the inevitable violation of the OLS, among which the number of feeder buses at the station's surroundings is the most common (e.g., Cardozo et al., 2012; Choi et al., 2012; Lee et al., 2013).

Considering these different kinds of limitations, some more difficult than others to be satisfactorily overcome, DRMs can provide satisfactory estimations at the station level, which is the reason why they are still often used in planning practice.

3. Case study – The expansion of the Porto light rail transit (LRT)

3.1. Background

The Metropolitan Area of Porto (MAP) is one of the NUTS 3 sub regions included in the Northern Region of Portugal (NUTS 2). The MAP is a polynucleated territory made of 17 municipalities with different socio-economic characteristics ranging from mostly urban in its core area, also called the Greater Porto (which includes six municipalities: Porto, Matosinhos, Maia, Valongo, Gondomar and Gaia), to mostly low density suburban and rural municipalities in the outskirts (see Fig. 1). The Greater Porto concentrates approximately one million inhabitants and, due to its higher population and job densities, as well as more intense commuting patterns, constitutes the privileged area to be served by the LRT.

In recent years, most MAP municipalities have revealed a stagnant or slight demographic loss, with a growing aging population. From 1991 to 2011, MAP population increased by 5,3% in the overall, whereas in the following decade, 2011 to 2021, decreased by −1,3% (INE, 2021). This decline was particularly felt in both the most central and the most peripheral municipalities, with an in-between ring still showing some signs of a modest demographic growth, which is in part associated with its capacity to attract some employment hubs and out-of-town shopping centres, decentralized higher education and professional schools, health centres and other public and private services, as well as housing investment. These emerging trends and changing spatial dynamics, strengthening municipal interdependences, have been considered in the design of the alternative routes for the expansion of the Porto LRT system.

Over the last twenty years, mobility patterns changed significantly in Portugal. While the importance of the public transport was continuously declining, the weight of the private car in the modal split, measured in percentage of trips, went from 45 % to 67 %, and the Porto Metropolitan Area was not an exception (see Fig. 2). It was in this context that the Porto LRT system emerged, as a counter-cyclical and decisive system to pursue the European agenda of decarbonization of mobility.

According to the last INE Travel Survey available (INE, 2017), regarding transport systems and travel patterns, the importance of the municipality of Porto and surrounding municipalities still stands out as a polarizing center for the entire metropolitan area.

In the Porto region (including the municipality of Póvoa do Varzim, since the LRT has a line that reaches this municipality), the motorization rate (per 1000 inhabitants) increased by almost 69 % between 2001 and 2017. In 2017, individual transport (private car) with an average occupancy rate of 1,56, accounted for around 68 % of all trips in MAP. The use of public transport (bus, company/school vehicle) reached 8 % (including a gain of 2 % in the Porto region in LRT trips from 2000 to 2017). Rail transport (heavy and light) corresponded to 5 % of total

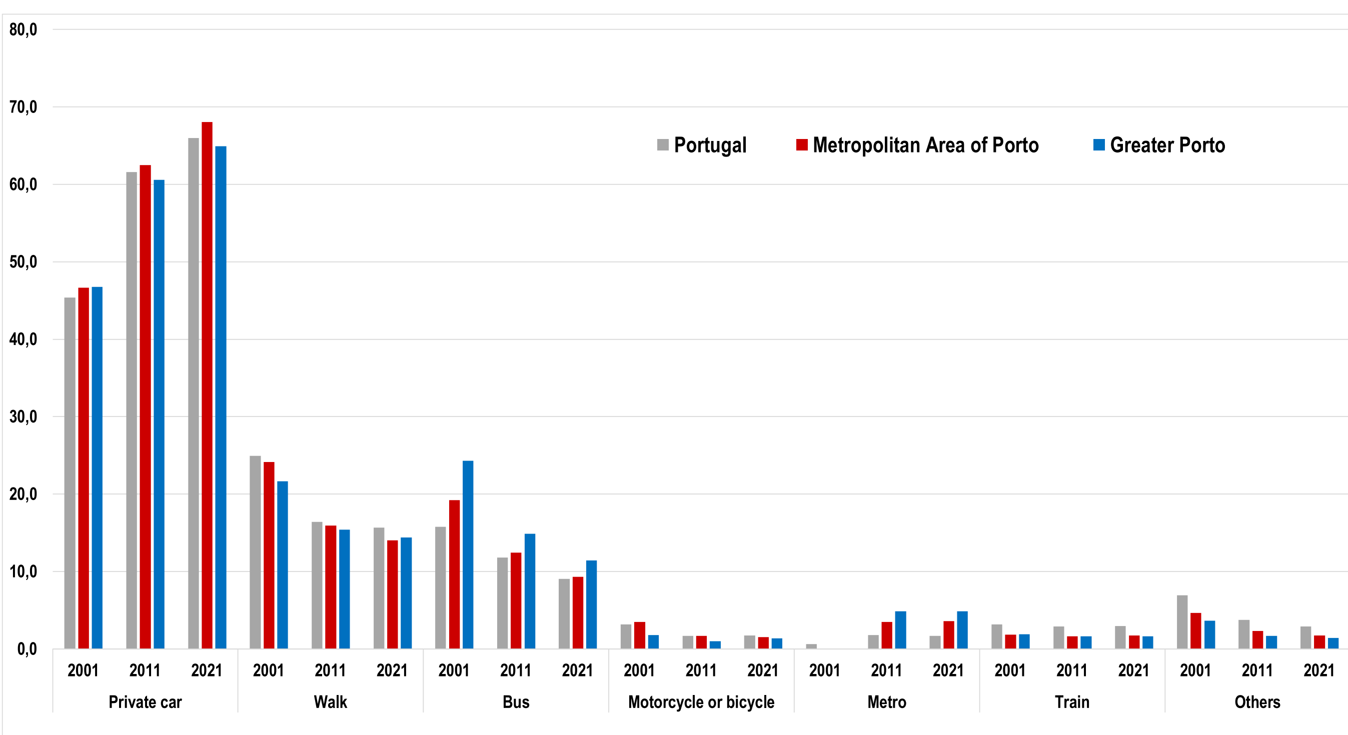

Fig. 2. Modal distribution by municipality in the last decades (according to Census data).

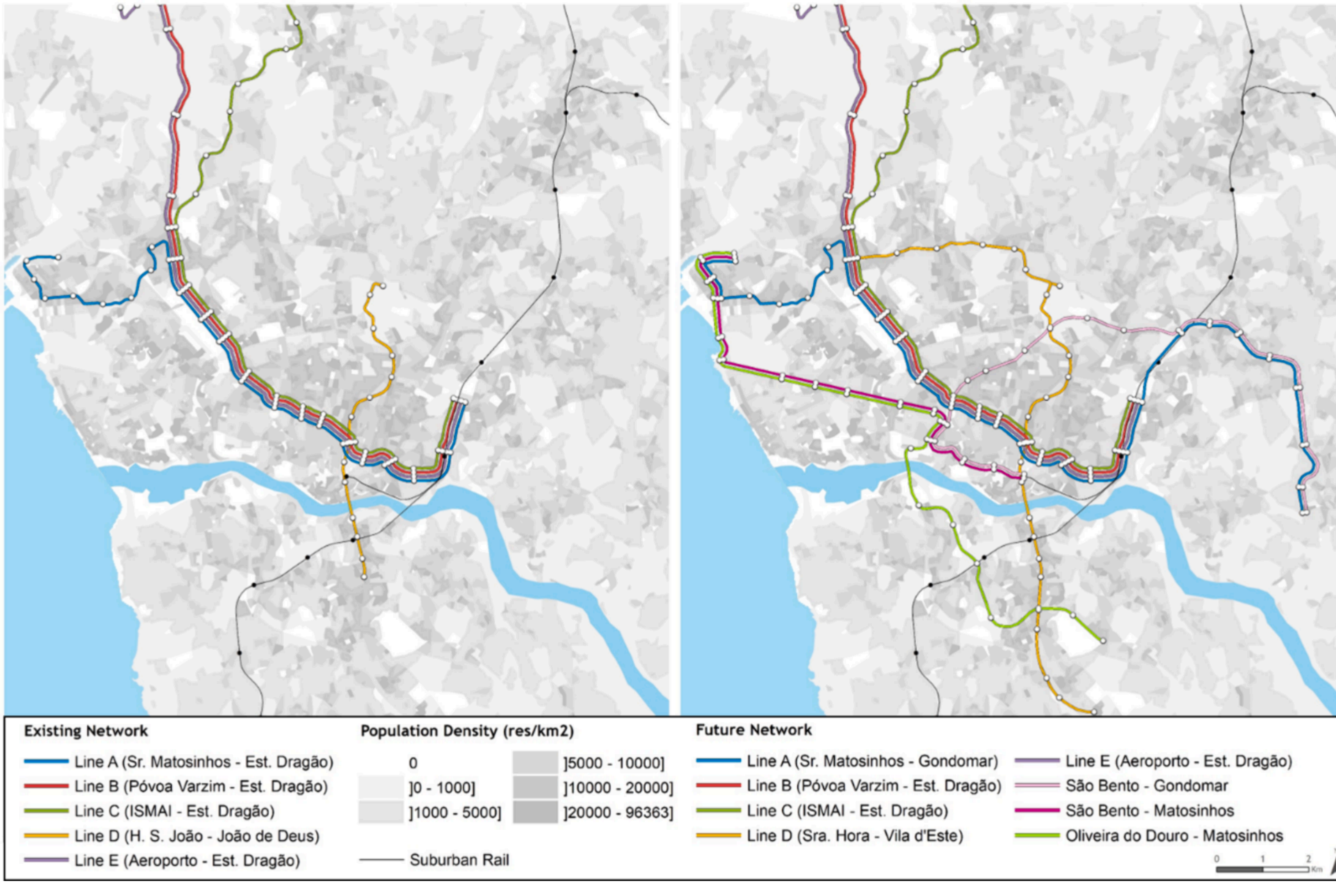

Fig. 3. On the left the Porto LRT network corresponding to the conclusion of the 1st investment phase and on the right the proposed network expansion according to the Strategic Plan (2007).

trips. Active/soft modes (walking and cycling) accounted for 19 % of the total. Based on the municipality of dwelling, car was less used by residents in areas of greater centrality and density, namely Porto and Espinho (just around 50 %), where shorter journeys were mostly made by active/soft modes. Inside MAP, intra-city trips accounted for 71 %, while 29 % were inter-city trips, indicating the importance of proximity between places and other factors related to the cost of travel as well as the optimization of time, comfort, and convenience.

The relative flows of origin and destination revealed a greater intensity of interactions between municipalities in MAP (29 %) compared to the municipality of Porto and between those contiguous, depending on the purpose of travel. About 40 % of commuting trips in MAP were between municipalities, with 61 % having the municipality of Porto as a destination. In addition, for commuting, shopping or leisure trips, the municipality of Porto was one of the three main destination municipalities in the MAP, among eleven, fifteen, and nine municipalities, respectively. Reinforcing what has been said before, although MAP is still dominantly polarized by the city of Porto its overall urban structure is strongly polycentric.

From a governance point of view, the mayors of the 17 municipalities, who are directly elected in nationwide municipal elections, constitute the so-called Metropolitan Council. Subsequently, through a Council's internal voting procedure, the President of the Metropolitan Council is then elected. Needless to say that the indirect nature of this presidency election tends to weaken his/her own political power and representativeness. In practice, the Council's President is always somehow divided in between the overall metropolitan perspective he/she is supposed to defend, and the municipal perspective he/she was directly elected to serve. This governance problem is not new or specific to Portuguese metropolitan areas. On the contrary, it is very common elsewhere (see, for instance, Vicuña, Elinbaum and Valenzuela-Montes, 2022).

The original proposal to build an LRT to serve MAP emerged in the late 1990 s. The first line was opened in 2002. Rapidly, the operation of the system introduced significant changes in the metropolitan mobility patterns and mode choices which opened the way to the development of an intermodal ticket system covering several public and private operators and run by the Intermodal Transport of Porto (ITP).

Despite the consensual recognition of its success, exponentiated by the architectural quality of the overall project and by the extensive qualification of the surrounding public spaces crossed on the surface by the network, the overall design of the Porto LRT was somehow victim of a *municipalization* process, the Metropolitan Council was not able or had not the political will to curb.

As the overall layout of the 1st phase of the LRT clearly illustrates (see Fig. 3 on the left), although the system's central station (Trindade) is geographically coincident with the metropolitan centre, some lines are very sinuous, covering areas with low densities and low urbanity levels. These options revealed how in some cases narrow municipal interests prevailed over a more comprehensive and strategic metropolitan perspective, penalizing in some cases the overall levels of demand which, otherwise, could even be higher.

It was in this context that one of the first commissions to the CITTA-FEUP team emerged, the preparation of the so-called *Strategic Programme for the Expansion of the Porto Metro System* (Metro do Porto, 2007). The prevalence of the metropolitan scale was then clearly assumed, recognizing, right from the outset, that an investment project with this dimension and impact is not limited to the sphere of public passenger transport. On the contrary, the Porto LRT could constitute a decisive instrument to guide the territorial restructuring of the

Table 1
Comparison of performance indicators between the LRT network after the conclusion of the 1st phase of investment and the introduction of a 2nd phase of investment as proposed by the Strategic Development Programme from 2007.

Porto Light Rail Transit	1st phase (existing)	1st + 2nd phase (proposed)
Total network length (km)	59	100
Population directly served (<500 m)	240,604	436,922
Population/km (network)	4055	4366
Employment served (<500 m)	186,042	296,192
Employment/km (network)	3136	2960
Large Travel Generators (LTG) directly served	42	82
Total LTG users	43,122	94,097

Metropolitan Area of Porto, promoting its economic competitiveness, its environmental sustainability, and the quality of life of its residents. From an essentially radial and monocentric network centred in the city of Porto, the system should evolve into a mix of radial and circular axis, expressing better and, indeed, reinforcing the polycentric nature of the metropolitan area. Fig. 3 shows the existing and the proposed new network reflecting this new perspective.

Table 1 presents several indicators to compare the performance of the existing network, corresponding to the project's 1st phase of investment, and the expansion proposal to be carried out in a future 2nd phase of investment. In the overall, the comparison of these two sets of performance indicators associated to demand potential seemed to clearly justify the decision to go ahead with the expansion of the network. Indeed, this decision was taken by the Metro company with the political support of the Metropolitan Council, but the subsequent studies

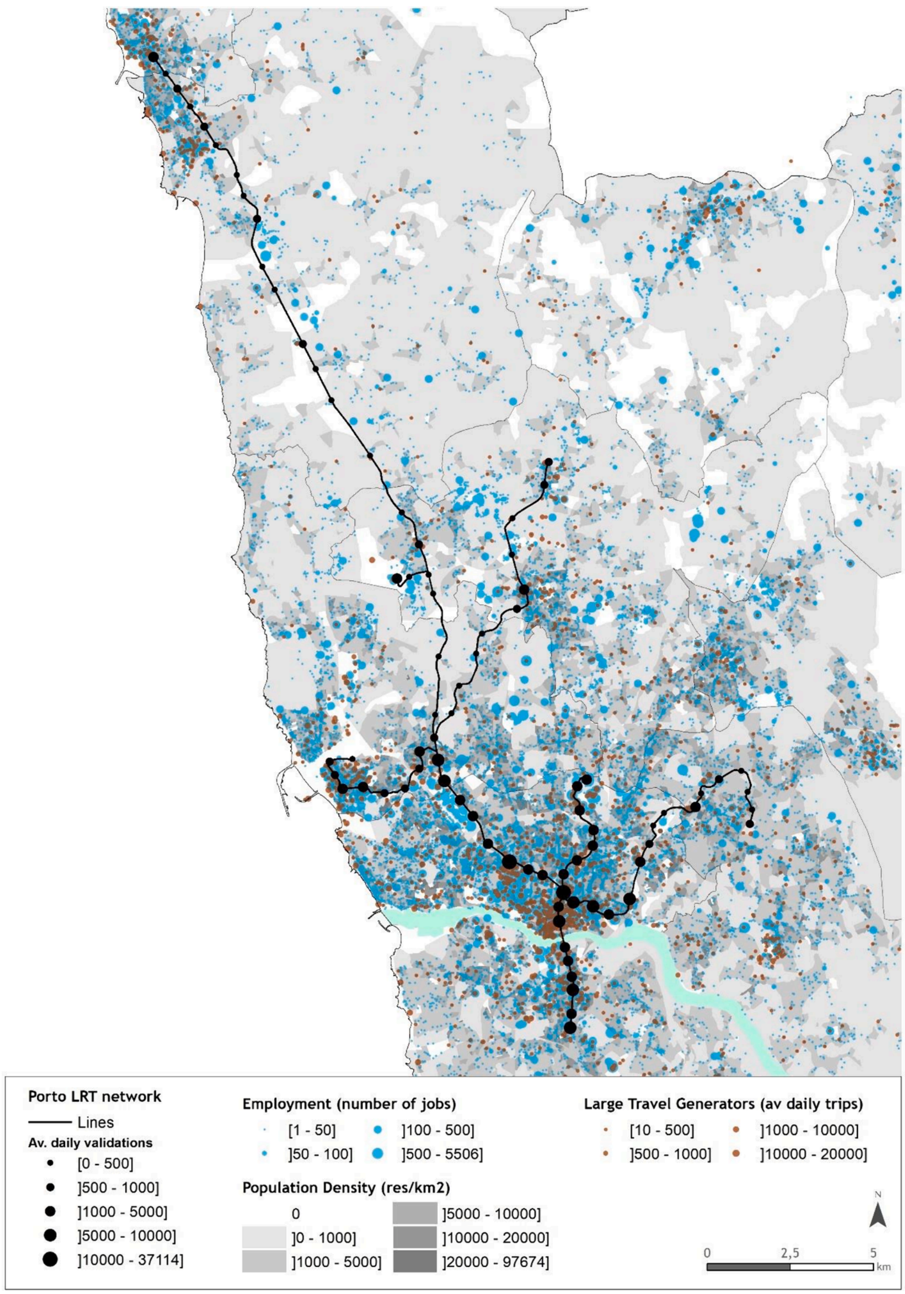

Fig. 4. GIS image with the existing Porto LRT network and all the main georeferenced datasets, including the resident population densities, the size and location of the public and private employment units and of all Large Travel Generators (LTG) in the study area.

Table 2
Descriptive statistics of tested variables in the direct demand model.

Parameters	Units	Min	Max	Mean	Median	Std. Deviation
Population	Residents @500 *m*	0	6,504.0	1,709.4	1,356.5	1,418.8
Employment	Workers @500 *m*	1	11,577.0	1,460.0	712.5	1,976.4
Large Travel Generator	Daily visitors @500 *m*	0	62,500.0	5,818.5	2,010.0	9,951.1
Accessible population	Residents @ 5.3 km	1,865	68,471.0	29,024,0	26,387.0	20,982.3
Accessible Employment	Workers @ 5.3 km	1,762	83,104.0	26,778.3	18,322.0	24,787.8
Distance to centre	Km	0	29.8	9.9	8.2	8.0
Number of Metro Lines	N°	1	6.0	1.7	1.0	1.4

had to be postponed following the economic and financial crises that, in the meantime, took place in the country, between 2010 and 2014, as already referred to in the introduction section.

3.2. Development of the direct ridership and transfer models

3.2.1. Approach and datasets

In line with a previous study in which a first DRM was developed (Metro do Porto, 2017), a large set of land use features was assembled to characterize the potential demand. These included the location of residents, jobs, and the so-called Large Travel Generators. This last variable seeks to account for non-commuting trips. It consists of the assigned number of average daily trips that a range of high-demand activities are likely to generate, such as health centres and hospitals, primary, secondary, and professional schools, university faculties, other major public services, and facilities, and hyperstores and shopping centres. This approach, though uncommon, might have some advantages compared to the often-used cumulative count of activities. It allows weighting facilities according to its attractiveness. Thus, facilities with the same activity may generate different average daily trips. Data on residential location was collected from the national Census at the so-called census tract level (the smallest statistical unit available, which is a good proxy to the block level in consolidated urban areas).

Two datasets were combined to build the employment database. The first to cover the public sector employment, from the Directorate-General for Administration and Public Employment, and the second to cover the private sector employment, from the Office for Strategy and Planning of the Ministry of Labour and Social Security. Both provided the exact address and the number of employees at each location of the respective employment unit, whether public or private. To respect data protection requirements, the name of the companies, the respective economic activity classification, and other related commercial information was omitted. While it is important to note that these two databases do not include self-employed persons and one person companies, the comparison with the active resident population figures, provided by the population census, make us conclude that over 90 % of all employment within the Porto Metropolitan Area was indeed considered in the development of the DRM.

Finally, on Large Travel Generators (LTG) belonging to both the public and the private sector, data was collected from a cross-analysis of different data sources, specific field work campaigns, and previous databases developed for other R&D projects carried out in our research centre. Fig. 4 shows a GIS image with the existing LRT network and a rigorous representation of all these georeferenced datasets, including a highly disaggregated spatial distribution of resident population densities, the size and location of the existing public and private employment units, and the size and location of all the LTG in the study area.

The second step involved a comprehensive modelling of the entire transport infrastructure of the Porto Metropolitan Area. The location of bus stops, routes and hourly frequencies was provided by the metropolitan transport authority. All equivalent information of rail-based transport, namely metro and suburban rail, was already available on open access databases. Average daily direct and indirect (transfers) demand at each of the existing metro stations, an essential piece of information to develop the demand models, was provided by the metro operator.

While direct demand estimation for each station derives from a set of built environment parameters within its corresponding coverage area, transfers are commonplace in a multimodal transport system. These can include changes within the same transport mode, as it is the case of a metro line transfer, or from different modes of transport, such as bus to metro or train to metro. The dataset provided by the Metro company included two distinct variables at each station. The first measured all new ticket (and card) *validations* i.e., new entries in the system, representing direct demand at each station, while the second targeted all transfers from other public transport modes measured at each station. This was important as most of the routes under evaluation intersected existing stations.

The need to predict these two distinct variables pointed to the development of two separate models. The first model estimated direct ridership at each station, while the second model focused on the estimation of the number of total transfers at each station. This two-step modelling scheme was necessary, as it was found that the inclusion of variables that influence the transfer profile of each station would lead to an overestimation of total demand figures. On the other hand, excluding the variables that impact the number of transfers would lead to an underestimation of demand, especially when new transfer possibilities were created with the construction of new lines. This combined ridership model allows for a more precise estimation of demand at new stations but also for the changes at existing stations due to induced demand.

3.2.2. Direct ridership model

Direct demand was measured through the modelling of the "potential attractiveness" of each station. Several models were tested, using different combinations of variables in a series of OLS regressions and Generalized Linear Models to account for the potential bias of fractional response variables. Due to the gaussian distribution of the dependent variable (direct demand), a linear regression was used.

$$y = \alpha_0 + \alpha_1 x_1 + \alpha_2 x_2 + \cdots + \alpha_n x_n + \epsilon$$

Thus, the value of the dependent variable is estimated via the sum of the product of the coefficients with the independent variables, plus an error term.

For this, seven different variables were tested. The first two measured the population and employment coverage provided by each metro station, within a 500-meter catchment area (measured through the pedestrian network). A third variable measured, within the same catchment area, the impact of the Large Travel Generators (LTG). These were systematized in 9 categories (commerce, culture, education, leisure, health, hotel, services, sports, and parking), combined with their size, which impacted potential demand, ranging from 20 daily trips in small sports pavilions to 20,000 daily trips generated by large shopping centers.

The next variables took into consideration the impact of the *network effect* at each station, an essential component to evaluate the future impact of the introduction of new routes. They included the total accessible employment and population in all stations, after entering the

Table 3
Comparison of tested models.

Tested variables	1	2	3	4	5	6	7	8	9	10
Sample size			82					78		
Population	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Employment	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Large Travel Generator	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Accessible population			✓	✓				✓	✓	
Accessible Employment			✓		✓			✓		✓
Distance to centre		✓	✓				✓	✓		
Number of Metro Lines	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Adj. R^2	0.896	0.896	0.905	0.895	0.897	0.892	0.892	0.901	0.891	0.893
Durbin-Watson	2.319	2.290	2.273	2.334	2.190	2.408	2.384	2.344	2.420	2.333

Table 4
Regression model for estimation of direct demand.

Parameters	Coefficient	Std. error	Significance	95 % CI	VIF
Constant	−339.204				
Population	0.206	0.043	<0.001	[0.121—0.291]	1.246
Employment	0.109	0.034	0.002	[0.040—0.177]	1.506
Large Travel Generator	0.099	0.007	<0.001	[0.085—0.114]	1.692
Accessible Employment	0.004	0.003	0.098	[0—0.010]	2.191
Number of Metro Lines	343.840	48.151	<0.010	[247.95—439.761]	1.385
Adj. R^2	0.897				
N	82.000				
Durbin-Watson	2.190				

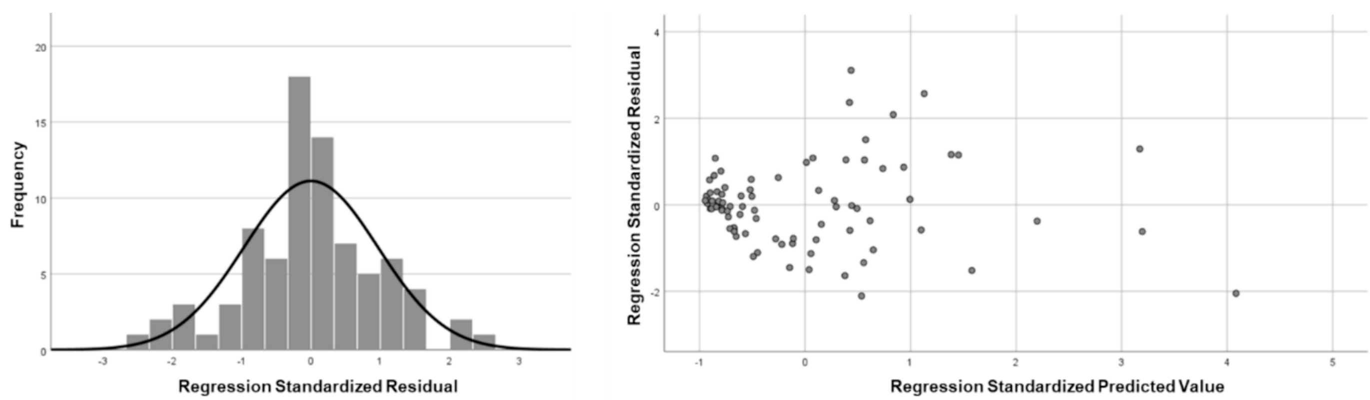

Fig. 5. Standardized residual plots for the direct demand model.

public transport system, within a 5,3km distance by the metro network, as such figure corresponds to the average travelled distance by metro passengers in the Porto metropolitan area. This corresponds, roughly, to a 25-minute trip. Other distances were tested but granted weaker outputs, especially at the stations outside the central urban core.

Also included in the pool of tested variables are the distance to the center of the city, located at Trindade station, as well as the number of number of lines at each station

In both models the calibration was tested both with the entire set and over a 95 % random sample of all existing stations (excluding some of the stations at both ends of the demand scale).

Table 2 presents the descriptive statistics of each of these variables and Table 3 the results of the different tests that had been performed with different combinations of variables and sample sizes.

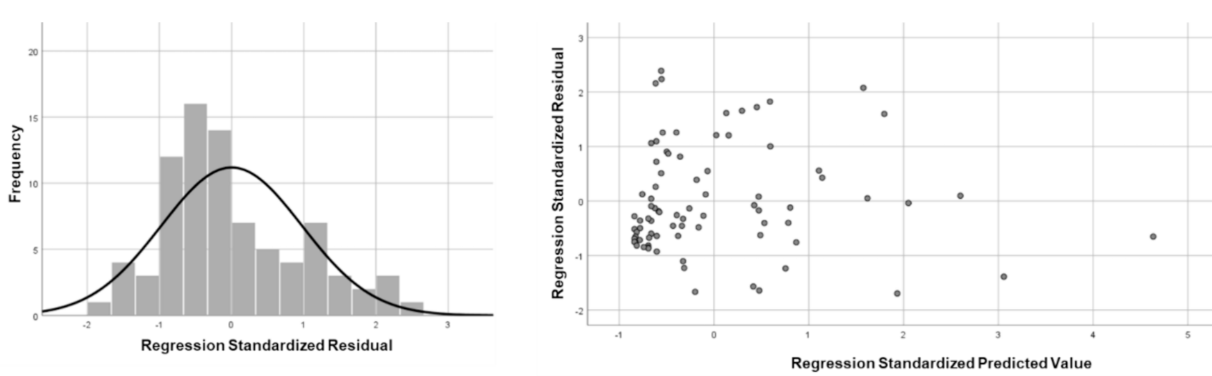

Fig. 6. Standardized residual plots for the transfer model.

Table 5
Descriptive statistics of tested variables in the transfers model.

Parameters	Units	Min	Max	Mean	Median	Std. Deviation
PT services at rush hour	N° services/ h	4	103.00	26.58	14.93	23.75
Terminal station	Number of service ends	0	1.00	0.10	0.00	0.30
Metro connection	Number of connections	0	1.00	0.05	0.00	0.22
Train connection	Number of connections	0	1.00	0.04	0.00	0.19

Table 6
Regression model for estimation of transfers (in percentage of average daily validations).

Parameters	Coefficient	Std. error	Significance	95 % CI	VIF
Constant	0.073	0.012	0.000	[0.049—0.098]	
Total PT services at rush hour	0.004	0.000	0.000	[0.003—0.005]	1.43
Terminal station	0.045	0.023	0.057	[-0.001—0.092]	1.10
Metro connection	0.283	0.039	0.000	[0.206—0.360]	1.11
Train connection	0.059	0.049	0.229	[-0.038—0.156]	1.34
Adj. R^2	0.759				
N	82.000				
Durbin-Watson	1.244				

Ten models were tested to determine the most suitable one for the task in hand. The goodness-of-fit, expressed by the r^2 value, ranging between 0.892 and 0.905, show that there were relatively small differences between them. As such, a decision was made to select model five, which showed the least amount of autocorrelation in the residual of the statistical model, measured by the Durbin-Watson, i.e. with values closer to 2.

Model number five was tested with 82 stations of the existing metro system and incorporating as independent variables the population, employment, and large travel generators at a 500 m distance from each station, the total employment at a 5,3 km trip and the number of metro lines at each station.

The adoption of a model with just five variables instead of a more complex one (such as model three with a higher r^2 value) reinforces its simplicity and the ease with which it could be explained to non-specialist audiences, which was the focus of the invitation of the Metro do Porto Company, as referred to before.

Table 4 and Fig. 5 resume, respectively, the adopted regression model and the corresponding residual plots (Fig. 6).

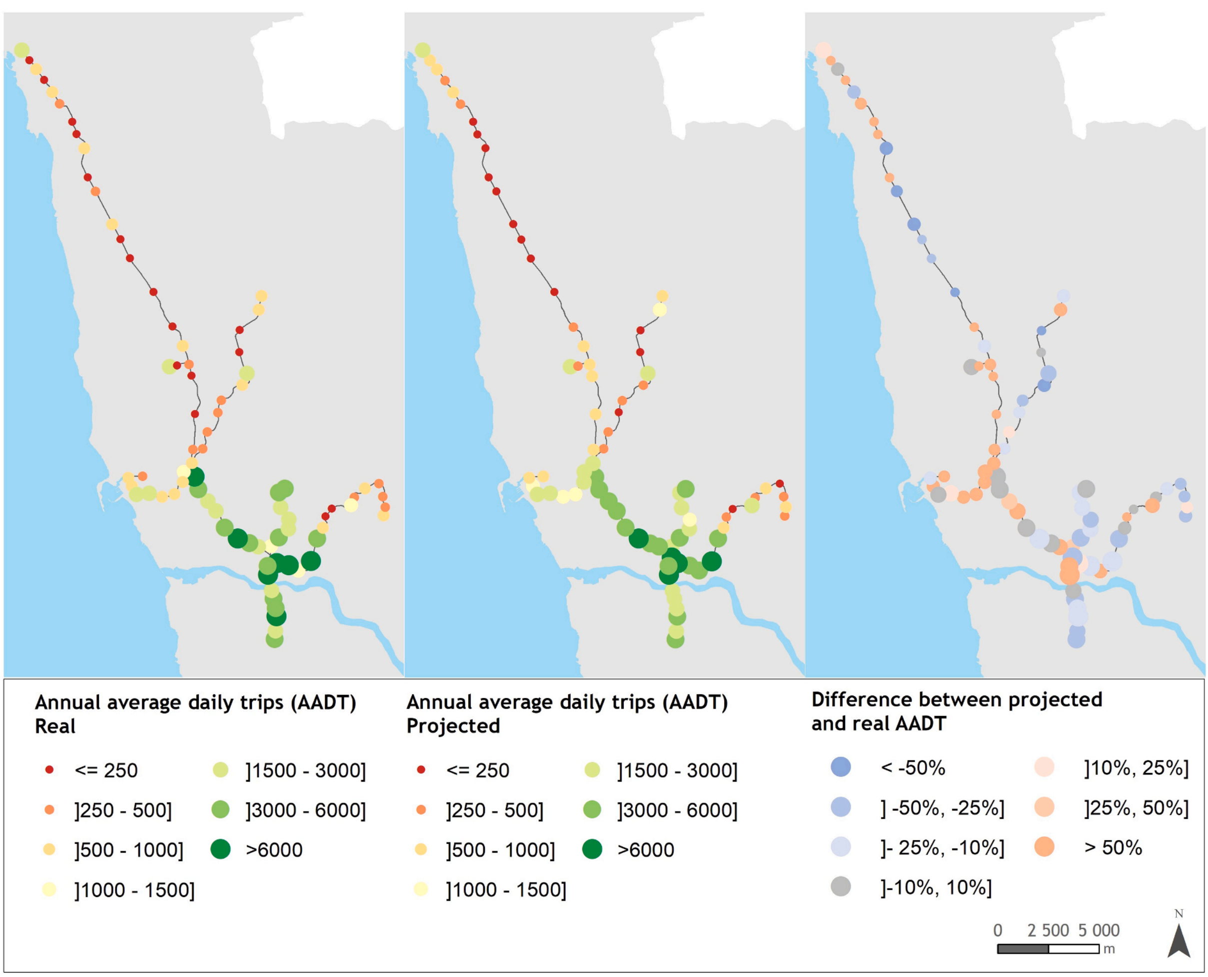

Fig. 7. Real (on the left) and estimated (on the centre) values expressed in Annual Average Daily Trips at each station of the Porto metro network. Percentual differences are presented on the right.

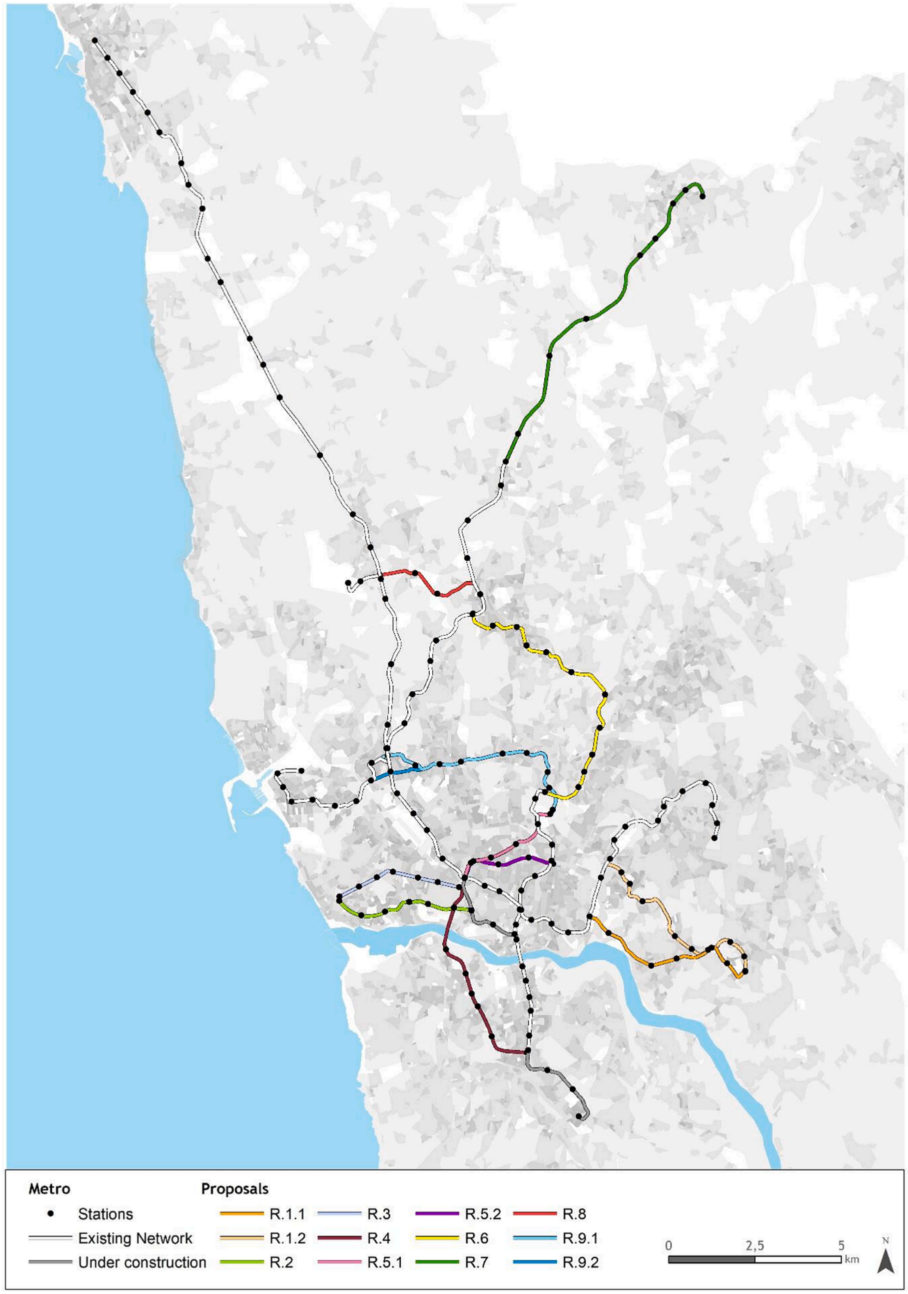

Fig. 8. The existing Porto LRT network and the new route proposals.

3.2.3. *Transfers model*

Having estimated the direct demand at each station, with a remarkably high level of goodness-of-fit, with the model explaining nearly 90 % of direct demand, the prediction of the number of transfers at each station followed. Similarly, to the first model, distinct arrangements of variables were tested in a series of OLS regressions.

To facilitate the interpretation of the results, a decision was made to express the number of transfers at each station as a percentage of the corresponding direct demand figures. The testing of several combinations of variables granted a higher explanatory power to those measuring the attractiveness of each station for transfers. The selected variables include the total hourly scheduled frequency at rush hour (considering bus, metro, and train) available within a 200 m radius from each station and three additional variables related to the likelihood of transferring to a different public transport: the presence of a bus terminal station, a metro connection with a different line, and a train connection.

Table 5 presents the descriptive statistics of the used variable and Table 6 a summary of the adopted regression model, including coefficients, standard errors, levels of significance and confidence intervals (CI) in addition to other statistical results. VIF values are within acceptable values, pointing to a relatively low probability of collinearity between variables. The results from the Durbin-Watson test also reinforce this conclusion.

3.2.4. *Bringing together the two models*

Finally, bringing together the two modelling exercises for direct demand and for transfers, Fig. 7 confronts the real values, expressed in Annual Average Daily Trips on the left, with the estimated values on the centre, at each station of the Porto metro network. The differences (in percentual terms) from the real with the estimated values, are presented on the right.

Table 7
Characteristics of the new route proposals.

New route proposals	Station characteristics			Route characteristics		Brief Description
	Number of new stations	Non-intermodal stations	Intermodal stations	Extension	New branch	
0R.1.1 – Campanhã – Souto	6	–	–		✓	Connection between Gondomar's' city centre and the existing network at AMP main interface (Intermodal Center of Campanhã).
0R.1.2 – Campanhã – Souto	9	–	–		✓	Similar to above – taking advantage of the existing MP channel between Estádio do Dragão and Campanhã).
0R.2 – Campo Alegre	6	1	–		✓	Connection to the existing network at the Galiza station. The route intends to cover the western part of Porto.
0R.3 – Boavista – Império	7	–	–		✓	Connection to the existing network at on the main stations of the existing network: Casa da Música. The route intends to cover the western part of Porto through the longest avenue of the city.
0R.4 – Casa da Música – Santo Ovídio	6	1	1		✓	Connections to the existing network at Casa da Música (important interface of the existing network) and Santo Ovídio (one of the most important stations in the southern end).
0R.5.1 – Circular Asprela	4	–	1	✓		Extension of the existing line G with connections to the existing network at Casa da Música and Polo Universitário stations.
0R.5.2 – Circular Combatentes	3	1	–	✓		Extension of the existing line G with connections to the existing network at Casa da Música and Combatentes stations.
0R.6 – Maia II (Hospital São João – Maia)	10	1	–	✓		Extension of the existing line D with connections to the existing network at Hospital São João and Parque Maia stations.
0R.7 – ISMAI – Trofa	8	–	–	✓		Extension of the existing line C with connection to the existing network at ISMAI station (end station of the current network).
0R.8 – Maia – Aeroporto	2	2	–		✓	With connections to the existing network at Fórum Maia and Verdes station (taking the existing channels Aeroporto – Verdes and Custió – Parque Maia).
0R.9.1 – São Mamede – via Fonte do Cuco	7	2	–	✓		With connections to the existing network at Polo Universitário, Fonte do Cuco and Hospital São João stations (after Fonte do Cuco station this route takes the existing line A).
0R.9.2 – São Mamede – via Senhora da Hora	7	2	–	✓		With connections to the existing network at Polo Universitário, Hospital São João and Senhora da Hora stations (after Senhora da Hora station this route takes line A as R.9.1.).

The very satisfactory goodness-of-fit across the network evidenced, coupled with the positive validation of the two models by statistical tests of adjustment and significance, open the way to the subsequent exercise of evaluating the alternative new routes for the expansion of the Porto metro network. In practice, these models clearly express how mobility patterns and transport systems are deeply dependent upon the characteristics of the territories they are supposed to serve.

3.3. Evaluation of the alternative routes for the expansion of the network

3.3.1. Brief characterization of the alternative routes under analysis

Fig. 8 and Table 7 present all the proposals for the expansion of the Porto LRT, nine new routes (1 to 9) in total, three of which including two different versions (1 and 2). To avoid confusion with the existing lines in operation, which are locally designated by letters complemented by a colour code, these new proposals are designated by numbers and are called routes (R).

With respect to the origin of these alternative route proposals for the enlargement of the Porto LRT network, some of the proposals follow with slight adjustments the original proposals included in the Strategic Development Programme from 2007 (Metro do Porto, 2007), e.g. R1.2, R.3, R.4, R.5.1, R.9.1 or R.9.2. The others resulted from municipal initiatives designed by the respective planning and transports departments with, usually, the involvement and the technical support of the Porto Metro Company.

3.3.2. Modelling results and preliminary findings

Each new route proposal was evaluated based on the prediction of its specific contribution to the total ridership of the entire metro network. This included both the estimations of the number of passengers captured at each of the new proposed stations, as well as on the existing ones due to the likely emergence of new accessible destinations and services. These estimations are expressed in Annual Average Daily Trips, AADT for short, and result from adding the outputs of the two models previously described, i.e., the direct ridership values and the transfer values at each new proposed station. It is worth remembering at this point that the results of any application of OLS indicate mere relationships and not any type of causality effect between the independent variables and the dependent variable.

As the characteristics of the routes under analysis differ considerably, particularly in length, the comparative performance of each new route or extension of an existing route is described not only in total net gains in the entire network expressed in AADT (see the third raw from below in Table 8), but also in AADT per length of the new route (AADT/km) and in AADT per number of new stations (AADT/station) (see, respectively, the second and the first raw from below in Table 8).

Considering first the proposals which present two slightly different versions, the ponderation of the three last indicators of Table 5 points to the election of routes R.1.2, R.5.1, and R.9.1 as the best solutions in comparative terms. In the case of the route R.1.2 there is an additional factor to take into consideration. This version circumvents the need to build a high and long bridge over the Campanhã valley associated to version R.1.1, which would significantly increase the overall cost of this route.

A first grading of the analysed routes can now be presented according to the three main indicators of potential demand:

· Firstly, the R.4 route, Casa da Música – St Ovídio, clearly stands out from the remaining routes. The global gains of AADT brought directly and indirectly to the MP system are the highest. In addition, the values of AADT/station, and of AADT/km are among the highest. From an operational point of view, Route R.4 has also the advantage of alleviating the overload that is currently felt in the existing Line D, the Yellow Line, so far, the only North – South line linking the most densely populated municipalities of Porto and Gaia.

· On a second level of potential demand, both the R.3 route, Boavista – Império, and the R.2 route along the Campo Alegre axis, emerge with

Table 8
Comparative performance of the new route proposals.

Proposed new routes / Models' results	0R.1.1 Campanhã – Souto	0R.1.2 Campanhã – Souto	0R.2 Campo Alegre	0R.3 Boavista – Império	0R.4 Casa Música St Ovídio	0R.5.1 Circular Asprela	0R.5.2 Circular Combatentes	0R.6 Maia II HSJ-Maia	0R.7 ISMAI – Trofa	0R.8 Maia – Aeroporto	0R.9.1 S. Mamede Fonte Cuco	0R.9.2 S. Mamede Sra. Hora
Total annual average daily trips (AADT)	299 369	309 166	305 471	303 513	324 876	285 829	278 017	286 108	279 339	281 982	307 465	307 115
Length of the new route (km)	5,68	6,77	4,14	3.70	6.77	3,53	2,97	9,31	10,73	3.10	7,92	7,80
Number of new proposed stations	6	9	7	7	9	4	3	10	8	2	7	7
Net gains in the new stations (AADT)	4 500	6 201	17 997	7 376	11 498	7 887	4 330	5 489	3 826	610	8 887	9 133
Net gains in the existing stations (AADT)	26 651	34 747	19 256	27 919	45 160	9 724	5 469	12 401	7 295	13 154	30 360	29 764
Total net gains in existing and new stations (AADT)	31 151	40 948	37 253	35 295	56 658	17 611	9 799	17 890	11 121	13 764	39 247	38 897
Total gains / length of new route (AADT/km)	5 489	6 046	8 998	9 539	8 369	4 995	3 302	1 921	1 037	4 440	4 955	4 987
Total gains / no. of new stations (AADT/stations)	5 192	4 550	5 322	5 042	9 443	4 403	3 266	1 789	1 390	6 882	5 607	5 557

very satisfactory results in all the three indicators. Although not initially proposed as mutual alternatives, routes R.3 and R.2 have in common the aim to serve the western side of Porto not yet covered by the existing lines in operation. As such it makes sense to point out that, in the overall, the R.3 route performs slightly better than the R.2 route. Another factor in favour of the R.3 route is the fact that it can be easily inserted in the central part of two existing boulevards, Av. da Boavista and Av. Marechal Gomes da Costa, reducing substantially all the respective construction costs.

On a third level, route R.1.2, Campanhã – Souto, route R.8, Maia – Aeroporto, route R.9.2, S. Mamede Sra. Hora, and route R.5.1, Circular – Asprela, stand out with intermediate levels of potential demand.

· At last, on a fourth level, with potential levels of performance significantly lower, appear the routes R.6, Maia II HSJ-Maia, and R.7, ISMAI-Trofa. These route proposals have in common the fact that they both cross low density suburban and rural areas, where demand tend to drop considerably.

3.4. *From modelling to practice*

At the time of writing (March 2024), the R.4 route and the R.3 route were already under construction, with the political decision unequivocally reflecting the action supported by the results of the models. Construction of R3 began in spring 2023 and the conclusion is estimated for the summer 2024. R4 only saw its construction start beginning of 2024 and is expected to be concluded by 2026. The routes which fell in the third level of potential demand, routes R.1.2, R.8, R.9.2 and R.5.1, returned to the design stage to fine-tune their layout. They are likely to be considered in a future investment envelope.

In these circumstances, it is fair to say that the political decision carried out by the Metropolitan Council has taken in due consideration the outputs of the comparative assessment of the potential demand of the different alternative routes for the expansion of the Porto Metro system.

Surely, the Council's final decision took also into consideration some other factors of a political nature, such as the historic record of previous decisions on the expansion of the Porto LRT, the balance among the different municipalities' political weight expressed by the size of the respective resident populations (i.e., voting populations), or the need to avoid decisions in which *all eggs would go into the same basket*.

Nevertheless, the robustness of the modelling exercise, and the clarity of the results summarised in Table 8, have certainly been two important contributing factors for the final decision. The technical team had the opportunity to make a general presentation of the assessment exercise to the Metropolitan Council, explaining in detail how the two models had been constructed and so why certain routes revealed lower levels of potential attractiveness as compared to others. Obviously, some of the Mayors integrating the Metropolitan Council were not pleased with the results after realizing that the proposed routes crossing their jurisdictions were unlikely to attract significant demand. Some of these Mayors asked for further explanations about how the modelling exercise was constructed and applied to the Porto metropolitan area, which the team readily offered. A *guided tour* to all the databases and modelling process development was provided to individual stakeholders at request, always leading to the acquiescence of the modelling outcomes. In this respect the GIS proved to be an excellent tool to make clear and perfectly visible how through the interaction between the layers of the different land use characteristics certain routes were abler than others to catch hold of the metropolitan territory.

Along the way the evaluation process took place, a sense of trust and respect was steadily built between the decision-makers and the technical team. The differences and the complementarities between the scientific and technical perspective on one side, and the political perspective on the other, were able to gradually emerge and be promptly addressed, contributing to the overall transparency and accountability of the decision-making process.

4. Conclusion

Based on the main outputs of the evaluation of alternative routes for the expansion of the Porto LRT, the research question framed in the introductory section of this paper can now be positively addressed in two sequential steps.

First, the case study illustrates that with accurate and sufficiently detailed socio-economic and land use datasets integrated into a GIS – and these are crucial preconditions – the development and application of direct ridership models, complemented with indirect (transfers) demand models, both validated by appropriate statistical tests of adjustment and significance, can produce very satisfactory results in terms of goodness-of-fit across the entire network.

Second, the case study also illustrates that when the above-mentioned models are applied to estimate the likely demand of new route proposals, the results can be easily communicated to and understood by the different stakeholders (including decision makers) as sufficiently accurate and unbiased to support the choice among alternative routes for the expansion of a light rail system, such as the Porto LRT.

Finally, it is important to emphasize that the results of any OLS application point to relationships and not to any sort of causality effects. In addition, even with an excellent database and a sophisticated GIS, the intrinsic limitations of direct ridership models are always there as pointed out in section 2. However, the thorough understanding and awareness of the nature of these theoretical limitations are key factors to increase the quality standards of applied research and professional practice.

CRediT authorship contribution statement

Paulo Pinho: Writing – review & editing, Supervision, Project administration, Funding acquisition, Conceptualization. **Miguel Lopes:** Writing – original draft, Investigation, Formal analysis. **Marcelo Altieri:** Writing – original draft, Methodology, Investigation. **Frederico Moura e Sá:** Writing – original draft, Methodology, Investigation. **Cecília Silva:** Methodology, Investigation, Formal analysis. **Ana Amante:** Writing – original draft, Formal analysis, Data curation.

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.

Acknowledgments

The authors would like to sincerely thank all the support, statistical information, and operational data kindly provided by the Porto Metro Company. The authors would also like to thank the valuable and constructive contributions of the three anonymous reviewers.

References

Cardozo, O.D., García-Palomares, J.C., Gutiérrez, J., 2012. Application of geographically weighted regression to the direct forecasting of transit ridership at station-level. Appl. Geogr. 34, 548–558.

Cervero, R., Murakami, J., Miller, M., 2010. Direct Ridership Model of Bus Rapid Transit in Los Angeles County, California. J. Transp. Res. Board 1–7.

Choi, J., Lee, Y.J., Kim, T., Sohn, K., 2012. An analysis of Metro ridership at the station-to-station level in Seoul. Transportation 39, 705–722.

Duduta, N., 2013. Direct ridership models of bus rapid transit and metro systems in Mexico City, Mexico. Transport. Res. Record: J. Transport. Res. Board 2394, 93–99.

Guerra, E., Cervero, R., Tischler, D., 2012. Half-Mile Circle. Transport. Res. Record: J. Transportat. Res. Board 2276, 101–109.

Gutiérrez, J., Cardozo, O.D., García-Palomares, J.C., 2011. Transit ridership forecasting at station level: an approach based on distance-decay weighted regression. J. Transport Geograp. 19, 1081–1092.

Iseki, H., Liu, C., Knaap, G., 2018. The determinants of travel demand between rail stations: A direct transit demand model using multilevel analysis for the Washington D.C. Metrorail system. Transp. Res. A Policy Pract. 116, 635–649.

Jun, M.J., Kim, J.I., Kwon, J.H., Jeong, J.E., 2013. The effects of high-density suburban development on commuter mode choices in Seoul, Korea. Cities 31, 230–238.

Jun, M.J., Choi, K., Jeong, J.E., Kwon, K.H., Kim, H.J., 2015. Land use characteristics of subway catchment areas and their influence on subway ridership in Seoul. J. Transp. Geogr. 48, 30–40.

Kim, M.K., Kim, S.P., Heo, J., Sohn, H.G., 2016. Ridership patterns at subway stations of seoul capital area and characteristics of station influence area. KSCE J. Civ. Eng. 00, 1–12.

Kuby, M., Barranda, A., Upchurch, C., 2004. Factors influencing light-rail station boardings in the United States. Transp. Res. A Policy Pract. 38, 223–247.

Lee, H., Park, H.C., Kho, S.Y., Kim, D.K., 2019. Assessing transit competitiveness in Seoul considering actual transit travel times based on smart card data. J. Transp. Geogr. 80, 102546.

Lee, S., Yi, C., Hong, S.P., 2013. Urban structural hierarchy and the relationship between the ridership of the Seoul Metropolitan Subway and the land-use pattern of the station areas. Cities 35, 69–77.

McNally, M.G., 2007. The Four-Step Model, in: Hensher, D.A., Button, K.J. (Eds.), Handbook of Transport Modelling. Emerald Group Publishing Limited, pp. 35–53.

Metro do Porto, 2007. Programa Estratégico de Desenvolvimento do Metro do Porto. Unpublished Report by FEUP-IC to Metro do Porto SA, Porto.

Metro do Porto, 2017. Avaliação da Procura Estrutural das Linhas Alternativas para Expansão da Rede do Metro do Porto. Unpublished Report by CITTA-IC to Metro do Porto AS, Porto.

Metro do Porto, 2021. Avaliação da Procura Potencial de Eixos para Consolidação do Sistema de Metro Ligeiro e Desenvolvimento de um Sistema de Metro Bus na Área Metropolitana do Porto. Unpublished Report by CITTA-IC to Metro do Porto AS, Porto.

Mucci, R.A., Erhardt, G.D., 2018. Evaluating the ability of transit direct ridership models to forecast medium-term ridership changes: evidence from San Francisco. Transportation Research Record 2672, 21–30.

Peng, Z., Dueker, K.J., Strathman, J., Hopper, J., 1997. A simultaneous route-level transit patronage model: demand, supply, and inter-route relationship. Transp. Res. A Policy Pract. 30, 70.

Pinho, P., Vilares, M., 2009. Avaliação das Grandes Obras Públicas: o caso do Metro do Porto. FEUP Press, Porto.

Sohn, K., Shim, H., 2010. Factors generating boardings at Metro stations in the Seoul metropolitan area. Cities 27, 358–368.

Sung, H., Choi, K., Lee, S., Cheon, S., 2014. Exploring the impacts of land use by service coverage and station-level accessibility on rail transit ridership. J. Transp. Geogr. 36, 134–140.

Sung, H., Oh, J.T., 2011. Transit-oriented development in a high-density city: Identifying its association with transit ridership in Seoul, Korea. Cities 28, 70–82.

Taylor, B., Miller, D., Iseki, H., Fink, C., 2009. Nature and/or nurture? analyzing the determinants of transit ridership across US urbanized areas. Transp. Res. A Policy Pract. 43, 60–77.

Vicuña, M., Elinbaum, P., Valenzuela-Montes, L.M., 2022. Planes metropolitanos en Iberoamérica: Culturas e instrumentos. RIL Press, Santiago do Chile.

Yan, X., Liu, X., Zhao, X., 2020. Using machine learning for direct demand modeling of ridesourcing services in Chicago. J. Transp. Geogr. 83, 102661.

Yang, J., Chen, J., Le, X., Zhang, Q., 2016. Density-oriented versus development-oriented transit investment: decoding metro station location selection in Shenzhen. Transport Policy 93–102.

Zhang, D., Wang, X.C., 2014. Transit ridership estimation with network Kriging: a case study of Second Avenue Subway, NYC. J. Transp. Geogr. 41, 107–115.

Zhao, J., Deng, W., Song, Y., Zhu, Y., 2013. What influences Metro station ridership in China? Insights from Nanjing. Cities 35, 114–124.