Zefreh, Mohammad Maghrour ; Saif, Atiullah; Esztergár-Kiss, Domokos ; Torok, Adam

A data-driven decision support tool for public transport service analysis and provision

Published in:
Transport Policy

DOI:
10.1016/j.tranpol.2023.01.015

Published: 01/05/2023

Document Version
Publisher's PDF, also known as Version of record

Transport Policy 135 (2023) 82–90

Contents lists available at ScienceDirect

Transport Policy

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

A data-driven decision support tool for public transport service analysis and provision

Mohammad Maghrour Zefreh [a,*], Muhammad Atiullah Saif [b], Domokos Esztergár-Kiss [c], Adam Torok [c,d]

[a] Division of Transport Planning, KTH Royal Institute of Technology, Brinellvägen 23, Stockholm SE-100 44, Sweden
[b] Department of Built Environment, Aalto University, 4 Otakaari, Espoo 02150, Finland
[c] Department of Transport Technology and Economics, Budapest University of Technology and Economics, Muegyetem rkp 3, Budapest 1111, Hungary
[d] KTI - Institute for Transport Sciences, Department for Transport Policy and Economics, Budapest, 1119, Hungary

ARTICLE INFO

Keywords:
Public transport service
Policy action
High-dimensional supervised classification
Data-driven
Jenks algorithm
Mahalanobis TOPSIS

ABSTRACT

Public transport service (PTS) analysis and provision is an important and challenging issue for public transport agencies. The results of the PTS analysis help transport planners to identify the areas in need of PTS improvement. Furthermore, relevant policy actions need to be determined for service provision to reach the desired level of PTS improvement in the identified areas. Without an appropriate decision support tool, planners need to apply several blind trials to find a policy action which improves the PTS in the examined areas. This paper introduces a data-driven decision support tool for PTS analysis and provision. The proposed framework combines a potentially large number of PTS measures while taking the correlation among the investigated measures into account and develops high-dimensional supervised classification models that predict the PTS levels for different policy actions. With this approach, planners can identify and prioritize the areas in need of PTS improvement, determine what policy actions should be targeted to improve the PTS in the identified areas, and predict the PTS impacts of these policy actions in the examined areas. The application of the proposed framework is demonstrated in detail through a case study of Budapest, Hungary, which is followed by a hypothetical policy implementation. The results show that mostly outskirts are in need of PTS improvement. Furthermore, the underlying reasons behind the areas with poor overall PTS are studied to target the relevant policy actions that improve the PTS in the identified areas. The PTS impacts of the targeted policy actions are studied by using the developed high-dimensional supervised classification models.

1. Introduction

Public transport service (PTS) provision is considered as an important element of the overall transport planning and management (Murray, 2001). However, the analysis and provision of PTS is a challenging issue for public transport agencies. The results of the PTS analysis help transport planners to identify the areas in need of PTS improvement (Mavoa et al., 2012; Wu and Hine, 2003; Fransen et al., 2015; Currie, 2010). Moreover, relevant policy actions need to be determined to reach the desired level of PTS improvement in the examined areas. Without an appropriate decision tool, planners need to conduct several blind trials to find a policy action that improves the PTS in the identified areas.

The public transport agencies primarily use bespoke database software tools to investigate the impacts of various policy actions on the PTS. Some examples of such tools, among others, are Amelia (Mackett et al., 2008), Accession (Preston and Rajé, 2007), and Snapta (Karou and Hull, 2014). These tools are useful to examine the effects of various policy actions. However, these are unavailable to the wider academic public, on top of the functionalities embedded in the software, planners/analysts are unable to develop their own procedures since the tools are not flexible enough (Fransen et al., 2015).

Regarding the PTS analysis, the importance and need of combining multiple PTS measures into one comprehensive measure is highlighted in the literature (Aman and Smith-Colin, 2020; Klumpenhouwer and Huang, 2021). Current paper introduces a data-driven decision support tool for the analysis and provision of PTS. The proposed framework combines a potentially large number of PTS measures while taking the correlation among the investigated measures into account. Additionally, high-dimensional supervised classification models that predict the PTS levels for various policy actions are developed. With this

* Corresponding author.
E-mail address: momz@kth.se (M. Maghrour Zefreh).

https://doi.org/10.1016/j.tranpol.2023.01.015
Received 15 February 2021; Received in revised form 29 November 2022; Accepted 23 January 2023
Available online 22 February 2023

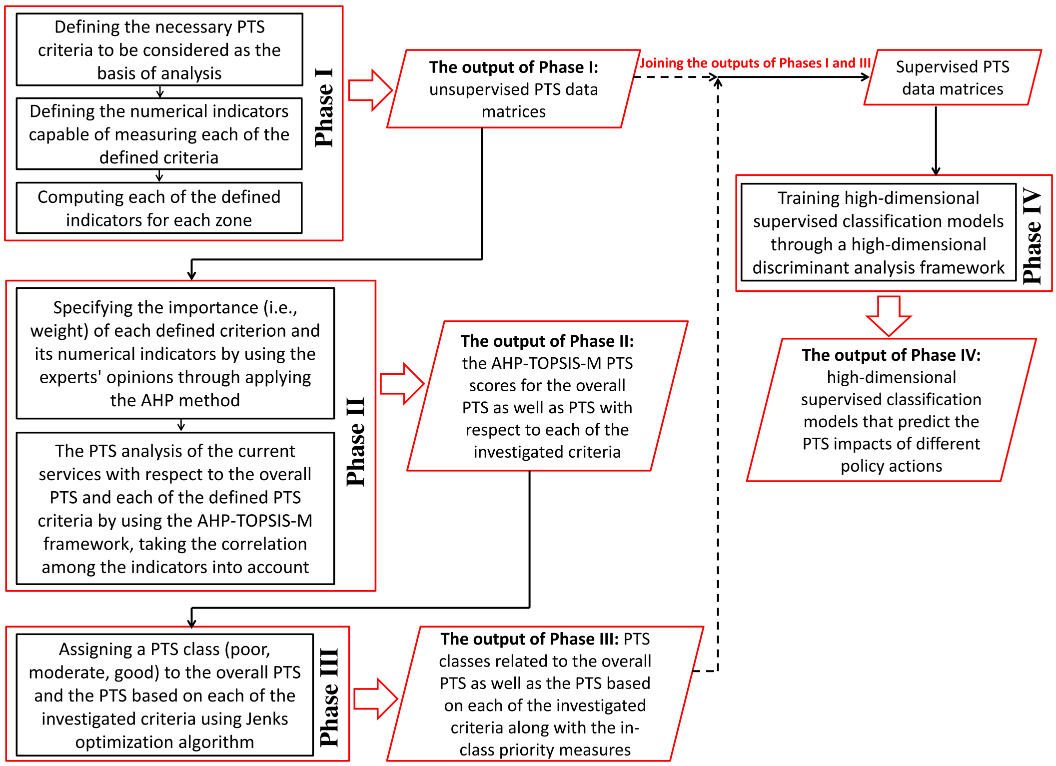

Fig. 1. Framework of the proposed decision support tool for the PTS analysis and provision.

approach, planners can identify and prioritize the areas in need of PTS improvement, determine what policy actions should be targeted to improve the PTS in the identified areas, and predict the impacts of these policy actions and service provisions on the PTS in the examined areas. The introduced framework is flexible. Thus, planners/analysts are not restricted to some predetermined PTS measures, and in this framework, they can study any relevant measure of interest according to the scope of the study and the availability of the data. Moreover, analysts have the opportunity to incorporate a potentially large number of PTS measures not merely in the evaluation of the existing PTS but in the proposed prediction framework, as well, to predict the PTS levels of various policy actions.

This paper is structured as follows. After the introduction in Section 1, the details of the proposed approach are discussed in Section 2. Section 3 demonstrates the application of the developed approach through a case study, which is followed by a hypothetical policy implementation along with a discussion of the results. Some limitations and suggested future works are presented in Section 4. Finally, Section 5 provides the conclusion of the paper.

2. Methodology

In this paper, the framework of the proposed decision support tool contains four phases, as shown in Fig. 1. The first three phases evaluate the existing PTS for each zone – by combining various PTS measures into comprehensive measures while taking the correlation among them into account – and create supervised PTS data matrices. The Phase IV applies the obtained supervised PTS data matrices to build high-dimensional supervised classification models that predict the PTS levels of different policy actions.

2.1. The evaluation of the existing PTS

To evaluate the existing PTS in the study area, planners often need to study multiple service measures/criteria and indicators, each of which has different units. In Phase I, following the scope of the analysis, relevant PTS criteria and indicators are set as the basis of the analysis. Afterward, the defined PTS measures can be computed at the level of the desired unit of analysis (e.g., traffic analysis zones (TAZs), census tracts, districts). The output of Phase I is an unsupervised overall PTS data matrix, which is the following:

$$M_1 = \begin{array}{c} \\ UoA_1 \\ UoA_2 \\ \vdots \\ UoA_z \end{array} \begin{array}{c} \begin{array}{cccc} i_1 & i_2 & \cdots & i_n \end{array} \\ \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{z1} & a_{z2} & \cdots & a_{zn} \end{pmatrix} \end{array}$$

Where the desired unit of the analysis (UoA) in the investigated area is denoted by $UoA = \{UoA_Z | Z = 1, \ldots, z\}$, $i = \{i_N | N = 1, \ldots, n\}$ stands for the PTS indicators, and a_{zn} represents the value of the nth PTS indicator for UoA z. It is worth highlighting that the overall PTS data matrix M_1 could be further split into multiple, criterion-specific data matrices. For illustration, see the overall PTS data matrix M_1 and the criterion-specific data matrices $M_{1.1}$, $M_{1.2}$, and $M_{1.3}$ as the calculated data matrices of the case study shown in Table 2.

In Phase II, the various PTS measures computed in Phase I are combined to evaluate the PTS of the existing services. Combining multiple PTS measures into a comprehensive measure requires an approach which weights the relative importance of each PTS measure. Some studies in the literature simply assume the same importance for all of their investigated measures (Aman and Smith-Colin, 2020; Mavoa et al., 2012). Other studies use extensive survey data as their underlying structure in the weighting process (McNeil, 2011; Zheng et al., 2019). Using the experts' opinions in the weighting process as a flexible approach that could be easily applied to different PTS measures has been recently highlighted in the literature (Klumpenhouwer and Huang, 2021; Hawas et al., 2016). Current paper weights the relative importance of the defined PTS criteria and indicators by using the experts' opinions through applying the analytic hierarchy process (AHP) method proposed by Saaty (1980). AHP is a structured approach that utilizes the experts' opinions to quantify the weights of the decision criteria (i.e., PTS measures) through pairwise comparisons by using a comparison scale proposed by Saaty (1980). In the AHP method, the decision criteria are compared to each other to compute their relative importance. Furthermore, within each criterion, the sub-criteria (i.e., the indicators representing each PTS criterion) are compared with each other to assess their relative importance in forming the criterion. The overall relative importance of each sub-criterion is calculated by multiplying their relative importance in their own criterion and the relative importance of their criterion compared to the other criteria. Therefore, the sub-criteria (i.e., PTS indicators) have one relative importance (i.e., weight) within their own criterion (i.e., indicator weight in the criterion) and one overall relative importance (i.e., overall indicator weight) compared to the other sub-criteria representing other PTS measures. For illustration, see Table 3 presenting the results of the weights of the criteria and the indicators in the case study. Having computed the relative weights of the PTS criteria and the indicators with the AHP method, the PTS in the UoA can be evaluated based on its relative closeness to the ideal situation by using the technique for order preference by similarity to ideal solution (TOPSIS) proposed by Hwang and Yoon (1981). In the traditional TOPSIS, each alternative (e.g., the UoA in the PTS analysis) is globally evaluated following its Euclidean distances to the positive and negative ideal solutions assuming that the evaluation criteria are independent (Vega et al., 2014). Within the context of the PTS analysis, several PTS measures are expected to correlate with each other. For instance, having more public transport stops in the UoA may result in more service area, more opportunity of direct origin-destination (O-D) coverage, and potentially, more overall service frequency in the UoA. Therefore, it is of great importance to take the correlation among the PTS indicators into account in the evaluation framework.

TOPSIS-M is one of the approaches that has been widely used in the literature (Wang and Wang, 2014; Liu et al., 2019; Sheikh et al., 2019) to address the dependent criteria by applying a decorrelation procedure through incorporating the Mahalanobis distance measures into the traditional TOPSIS method. Thus, TOPSIS-M method is used in Phase II of the proposed tool to evaluate the PTS in the UoA. In this process, the unsupervised PTS data matrix M_1 computed in Phase I is normalized and further used to calculate the positive and negative ideal solutions taking the benefit (i.e., the more is better, such as the area of the service) and cost (i.e., the less is better, such as waiting time) nature of the PTS indicators into account. The Mahalanobis distances from both the positive (DM_z^+) and negative (DM_z^-) ideal solutions are computed by using Eq. (1).

$$\begin{aligned} DM_z^+ &= \sqrt{(U_{z,n} - U^+)^T \ \Delta^T \ S_U^{-1} \ \Delta \ (U_{z,n} - U^+)} \\ DM_z^- &= \sqrt{(U_{z,n} - U^-)^T \ \Delta^T \ S_U^{-1} \ \Delta \ (U_{z,n} - U^-)} \end{aligned} \tag{1}$$

Where $U_{z,n}$ stands for the normalized value of a_{zn} (i.e., PTS indicator values in M_1), U^+ and U^- are the positive and negative ideal solutions, Δ^T represents the transposed diagonal matrix of the squared root of the weights obtained through the AHP method, and S_U^{-1} is the inverse covariance matrix of the normalized unsupervised PTS data matrix M_1. Finally, the relative closeness measures (i.e., TOPSIS-M PTS scores) are calculated in a way that the UoA with the higher scores are the ones with the better PTS. Further details about the AHP and TOPSIS methods as well as TOPSIS-M method can be read in Tzeng and Huang (2011) and Shih and Olson (2022), respectively.

It should be highlighted that the PTS in each UoA can be evaluated based on the overall PTS (i.e., by using the overall PTS data matrix M_1 and the overall indicator weight values) as well as based on each of the defined PTS criteria (i.e., by using the criterion-specific data matrices and the indicator weight in the criterion values). Hence, for each UoA, there is one TOPSIS-M PTS score for the overall PTS and one for each of the defined PTS criteria.

Furthermore, by taking the resulted scores into account, the UoA can be ranked with respect to their overall PTS as well as each of the investigated PTS criteria in a way that the one with the lowest score is considered to be in the worst situation thus in higher priority for PTS improvement. For illustration, see the computed TOPSIS-M PTS scores of the case study shown in Fig. 4.

While Phase II assesses and ranks the UoA based on the computed TOPSIS-M PTS scores ($y_z^* \in [0,1]$), Phase III assigns a PTS class (i.e., poor, moderate, or good) to the UoA in the examined area according to the related scores by using Jenks natural breaks optimization algorithm (Jenks, 1967), which is proved to be a reliable classification scheme in the literature (Lu et al., 2021). The Jenks algorithm attempts to minimize the distance between the TOPSIS-M PTS scores and the center of the clusters (i.e., poor, moderate, or good) they belong to while maximizing the difference between the cluster centers by minimizing the following cost function (Khan, 2012):

$$J = \sum_{\substack{1 \le i \le z \\ 1 < j < k}} dist(y_z^*, c_j) - \sum_{1 \le j \le (k-1)} dist(c_{j+1}, c_j) \tag{2}$$

Where z represents the number of the UoA, k denotes the number of the clusters, $dist(y_z^*, c_j)$ shows the distance between the TOPSIS-M PTS score of UoA_z computed in Phase II and its nearest cluster center c_j.

It is worth mentioning that the PTS class assignment could be done based on the overall PTS (i.e., by using TOPSIS-M PTS scores representing the overall PTS) as well as based on each of the defined PTS criteria (i.e., by using the scores representing each of the criteria). The outputs of Phase III are the accessibility classes related to the overall PTS as well as the PTS based on each of the investigated criteria. To assess in-class priorities (e.g., which UoA has a higher priority for PTS improvement in poor class zones), rankings obtained in Phase II are normalized following a Min–Max normalization approach. Thus, the UoA with higher rankings (i.e., worse PTS) have higher priority compared to others with lower rankings (see Fig. 6).

Joining the assigned overall PTS classes (i.e., the outputs of Phase III) of each UoA to the unsupervised PTS data matrix M_1 (i.e., the outputs of Phase I) creates a supervised PTS data matrix M_2, as follows:

$$M_2 = \begin{array}{c c} & \begin{array}{ccccc} i_1 & i_2 & & i_n & c_{UoA} \end{array} \\ \begin{array}{c} UoA_1 \\ UoA_2 \\ \vdots \\ UoA_z \end{array} & \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} & c_{UoA_1} \\ a_{21} & a_{22} & \cdots & a_{2n} & c_{UoA_2} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ a_{z1} & a_{z2} & \cdots & a_{zn} & c_{UoA_z} \end{pmatrix} \end{array}$$

Where c_{UoA_z} represents the overall PTS class of the zth unit of the analysis. Likewise, joining the PTS classes of each defined criterion to the criterion-specific data matrices (i.e., $M_{1.1}$, $M_{1.2}$, and $M_{1.3}$ matrices in the case study) creates supervised, criterion-specific PTS data matrices (i.e., $M_{2.1}$, $M_{2.2}$, and $M_{2.3}$).

2.2. Predicting the PTS levels for different policy actions

In the first three phases of the proposed framework, areas in need of PTS improvement (i.e., areas with poor overall PTS) are identified and ranked (i.e., prioritized for PTS improvement). Furthermore, the underlying reasons (i.e., which PTS criteria result in poor overall PTS) behind the UoA with poor PTS are studied to target the relevant policy actions improving PTS in the identified areas. In Phase IV, by using the obtained supervised data matrices, high-dimensional supervised classification models are trained through a high-dimensional discriminant analysis (HDDA) framework proposed by Bouveyron et al. (2007) to predict the impacts of the different policy actions on PTS. It is proven that the prediction performance of the HDDA models does not sink by increasing the number of the dimensions (i.e., PTS measures) (Bergé et al., 2012). Therefore, in the proposed framework, planners have the opportunity of incorporating a potentially large number of PTS criteria and indicators (i.e., represented by the number of dimensions)

Table 1
The investigated HDDA models (Bergé et al., 2012).

No.	HDDA model
1	$A_{kj}B_kQ_kD_k$
2	$A_{kj}BQ_kD_k$
3	$A_kB_kQ_kD_k$
4	$AB_kQ_kD_k$
5	$A_kBQ_kD_k$
6	ABQ_kD_k
7	$A_{kj}B_kQ_kD$
8	$A_{kj}BQ_kD$
9	$A_kB_kQ_kD$
10	AB_kQ_kD
11	A_kBQ_kD
12	ABQ_kD
13	A_jBQD
14	$ABQD$

not solely in the PTS evaluation stage (i.e., Phases I to III) but in the PTS prediction stage, as well.

This prediction framework is made of a learning step, in which model parameters are estimated from a set of learning observations, and a classification step, which aims to predict the class belonging to the new unlabeled observations (Bergé et al., 2012). In current paper, the HDDA model development is conducted by using the R package *HDclassif* (Bergé et al., 2012). In the training process, 14 HDDA models proposed by Bergé et al. (2012), as shown in Table 1, are compared to each other, and the best model is selected according to the BIC values.

In the models presented in Table 1, A_{kj} ($j = 1, \ldots, D_k$) models the variance of the data of the kth class, B_k models the variance of the noise, Q and Q_k represent the orientation matrices, and D and D_k represent the intrinsic dimensions of the classes.

It is worth mentioning that the supervised PTS data matrices need to be centered and scaled (i.e., mean = 0, standard deviation = 1) in the model development phase because the values of some PTS indicators might be much larger than the values of others. For illustration, see the trained HDDA models of the case study in Tables 4, 5, 6, and 7.

3. The case study

The application of the proposed framework is demonstrated in detail through a case study of Budapest, Hungary. Budapest has a population of approximately 1.7 million inhabitants in an area of 525 km^2, which is geographically divided into 23 districts. These districts are separated by political boundaries and administered by separate administrative units, i.e., each has its own municipal government.

General transit feed specification (GTFS) data along with the census data of Budapest in the TAZs (KSH (Hungarian Central Statistical Office), 2017) are the main data sources applied in this case study. This information is used to analyze a series of PTS criteria and their numerical indicators at the level of the TAZs in Budapest. The study area along with the location of the public transport routes and stops as well as the population density (i.e., population per square kilometer) of the TAZs in Budapest are shown in Fig. 2.

According to the availability of data, three PTS criteria are investigated in this case study. The three criteria are the followings: the relative value of the spatial spread and the population exposure of the PTS (i.e., PTS coverage), the extent of the PTS available in a TAZ and the magnitude of the available public transport infrastructures (i.e., PTS supply), and the sprawl of the PTS (i.e., PTS diversity) in a TAZ. The PTS coverage is studied by taking the area of the service concept into account. The area of the service is the catchment area of public transport, which serves as the actual area with access to the location of the public transport stops with a threshold limit of distance. To estimate the area of the service in Budapest, a 400 meter walking distance threshold to the public transport stops is set, as suggested in the literature (Kimpel et al., 2007). The estimated area of the service in Budapest is illustrated in Fig. 3. The service coverage is measured based on the following two indicators: the magnitude of the service area per inhabitant in each TAZ (i.e., population coverage i_1) and the percentage of the land area of each TAZ within its area of service (i.e., land coverage i_2). The PTS supply can be represented by a number of numerical indicators, such as vehicle-kilometers, passenger trips per capita, and passenger trips per hour Dargay and Hanly (2002), Holmgren (2008), Webster and Bly (1981) and Meyer (2000). In current study, the PTS supply is measured based on the following three indicators: the PTS availability of the TAZ based on the demand and the area of the TAZ (i.e., trips per day per square kilometer area i_3), the relative service frequency based on the area of the TAZ (i.e., frequency of service per square kilometer area, i_4), the provision of public transport infrastructure and the area of the TAZ (i.e., stops per square kilometer area i_5). The PTS diversity can be assessed by investigating every possible service between an origin and a destination (Frappier et al., 2018). In current research work, the PTS diversity refers to the concept of more connections to more destinations (i.e., TAZs), which represents a better availability of PTS, an indicator of service effectiveness (Hawas et al., 2016). In this case study, the public transport alternatives are defined as a sequence of public transport routes and transfers. This criterion of the evaluation of the public transport system denotes the number of the possible routes available between any origin and any destination pair. The PTS diversity aims to find the empirical evaluation of the access to and from a TAZ to another by using public transport. In current research, PTS diversity is measured by the following two indicators: the percentage of the total TAZs accessible from a TAZ by making a single public transport journey (i.e., direct O-D coverage i_6) and the public transport pair routes available to reach all other remaining possible TAZs from an origin district (i.e., O-D pair routes i_7).

Having computed all of the defined PTS indicators (i.e., seven indicators representing three PTS criteria) separately for each TAZ in the case study, the overall unsupervised PTS data matrix (M_1) is obtained, as shown in Table 2. The PTS data matrix M_1 can be further split into $M_{1.1}$, $M_{1.2}$, and $M_{1.3}$ data matrices representing the PTS coverage, the PTS supply, and the PTS diversity data matrices, respectively (see Table 2).

The existing PTS is further evaluated through an AHP-TOPSIS-M framework, which results in one TOPSIS-M PTS score representing the overall PTS in the TAZs and one score for each of the defined PTS criteria. Furthermore, the obtained scores are applied to rank the TAZs based on their overall PTS and each of the investigated criteria. As mentioned before, in this multicriteria framework, the relative importance (i.e., weights) of the criteria and the indicators are quantified through the AHP method, the TAZs are the alternatives, the PTS criteria with the different units are the decision criteria, and the indicators are the sub-criteria. In current research, the relative weights of the criteria are quantified through the AHP method and equally distributed among their respective indicators. The obtained weights and TOPSIS-M PTS scores are presented in Table 3 and Fig. 4, respectively. As expected, by taking the overall PTS into account, the results show that the inner-city zones have better PTS compared to the outskirts. This difference is most obvious in the PTS supply and diversity criteria while the PTS coverage criterion has somewhat more equal distribution among the TAZs (i.e., once compared to the PTS supply and diversity). Regarding the PTS supply, most of the TAZs have low scores, except for the densely-built inner city areas. Overall, the computed TOPSIS-M PTS scores show that the studied TAZs are more in need of PTS supply and diversity improvement than the development of PTS coverage.

Furthermore, Jenks natural breaks optimization algorithm is used to assign a PTS class (i.e., poor, moderate, or good) to the studied TAZs based on their TOPSIS-M PTS overall scores and their PTS scores with respect to each of the three investigated criteria, as shown in Fig. 5. Furthermore, the obtained rankings are normalized through a

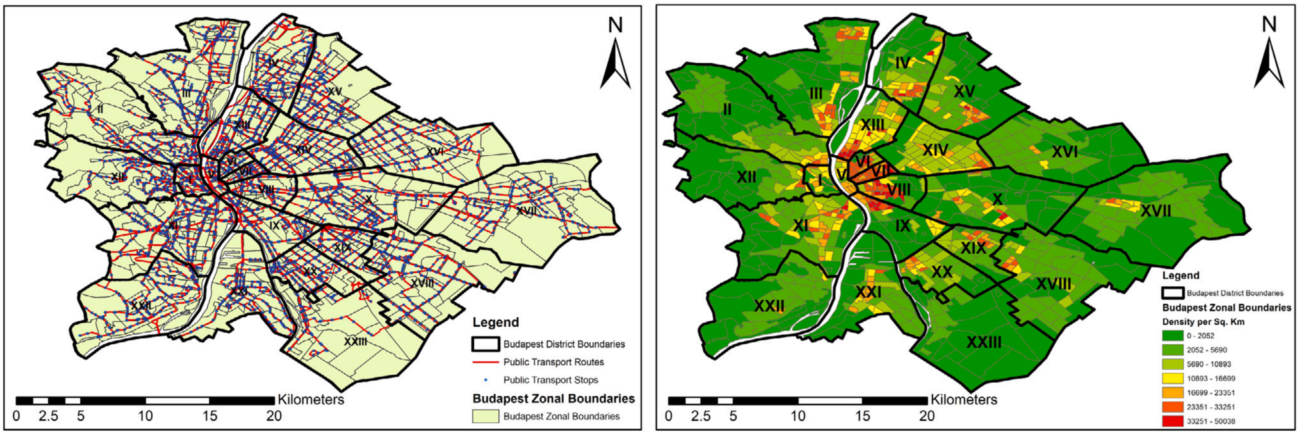

Fig. 2. The study area along with its population density (KSH (Hungarian Central Statistical Office), 2017).

Table 2
The values of the PTS indicators in the case study (M_1).

District No.	Zone ID	PTS coverage ($M_{1.1}$)		PTS supply ($M_{1.2}$)			PTS diversity ($M_{1.3}$)	
		i_1: population coverage (m^2/inh.)	i_2: land coverage (%)	i_3: trips per day per square kilometer area	i_4: frequency of service per square kilometer area	i_5: stops per square kilometer area	i_6: direct O-D coverage (%)	i_7: O-D pair routes
	1001	1054.7	99.7	40220.1	10044.9	21.32	12	150
District 1	1002	1691.3	99.7	25587.4	39256.2	94.8	22	342
	1003	890.1	99.7	14753.9	27844.6	49.9	21.9	306
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
	8228	541.9	99.72	49602.7	30984.3	26.03	16.48	318
District 8	8229	406.1	99.73	62605.01	33907.7	31.1	13.73	288
	8230	321.01	99.7	23326.7	30152.1	15.35	5.57	120
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
	15516	2318.8	99.6	1192.7	5746.7	18.04	12.5	204
District 15	15517	1605.4	99.7	1663.4	7541.1	21.6	15.1	264
	15518	2606.02	99.7	2041.1	12967.1	20.08	18.4	362
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
	23734	1590.5	99.6	877.1	3860.6	15.7	5.25	64
District 23	23735	999.08	84.7	339.04	795.8	6.26	3.46	38
	23736	1663.2	98.1	1123.2	2493.4	14.3	5	62
	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮

Table 3
The weights of the criteria and the indicators.

Criterion	The weight of the criterion	Indicator	Indicator weight in the criterion	Overall indicator weight
PTS coverage	0.38	i_1: population coverage	0.5	0.19
		i_2: land coverage	0.5	0.19
PTS supply	0.33	i_3: trips per day per square kilometer area	0.33	0.11
		i_4: frequency of service per square kilometer area	0.33	0.11
		i_5: stops per square kilometer area	0.33	0.11
PTS diversity	0.29	i_6: direct O-D coverage	0.5	0.145
		i_7: O-D pair routes	0.5	0.145

Min–Max normalization within each class to represent in-class priorities. This would help to prioritize the zones in need of improvement (i.e., the zones with poor overall PTS). Thus, the zones in need of PTS improvement (i.e., mostly outskirts) are identified and prioritized.

Moreover, the underlying reasons (i.e., which PTS criteria resulted in poor overall PTS) behind the TAZs with poor overall PTS are studied to target the relevant policy actions improving the PTS in the identified zones. For example, zone 17594 in District 17 has poor overall PTS with a considerable priority, and it is in need of improvement for PTS coverage and supply (see Fig. 6). The resulted class belongings as well as in-class priorities are presented in Fig. 6 (note for the sake of clarity merely poor-class priorities are illustrated in the figure).

As mentioned before, joining the assigned PTS classes to the obtained unsupervised PTS data matrices (i.e., M_1, $M_{1.1}$, $M_{1.2}$, and $M_{1.3}$)

Table 4
HDDA model for the overall PTS class prediction, model type: $A_{kj}B_kQ_kD_k$, out-of-sample correct classification rate: 93.9%.

Class	Prior prob.	Intrinsic dimen.	A_{kj}				B_k
			a_1	a_2	a_3	a_4	
Poor	0.044	2	1.88	0.56	–	–	0.057
Moderate	0.91	4	2.41	1.14	0.66	0.41	0.1
Good	0.039	2	16.05	4.89	–	–	0.67

Table 5
HDDA model for PTS coverage class prediction, model type: $A_{kj}B_kQ_kD$, out-of-sample correct classification rate: 95.2%.

Class	Prior prob.	Intrinsic dimen.	A_{kj}	B_k
			a_1	
Poor	0.04	1	1.86	0.37
Moderate	0.93	1	0.5	0.15
Good	0.02	1	2.8	0.37

Table 6
HDDA model for PTS supply class prediction, model type: $A_{kj}B_kQ_kD$, out-of-sample correct classification rate: 93.9%.

Class	Prior prob.	Intrinsic dimen.	A_{kj}		B_k
			a_1	a_2	
Poor	0.62	2	0.13	0.04	0.37
Moderate	0.26	2	0.76	0.43	0.15
Good	0.1	2	4.57	2.02	0.37

Table 7
HDDA model for PTS diversity class prediction, model type: $A_{kj}B_kQ_kD$, out-of-sample correct classification rate: 98.6%.

Class	Prior prob.	Intrinsic dimen.	A_{kj}	B_k
			a_1	
Poor	0.36	1	0.13	0.005
Moderate	0.52	1	0.46	0.029
Good	0.11	1	1.54	0.1

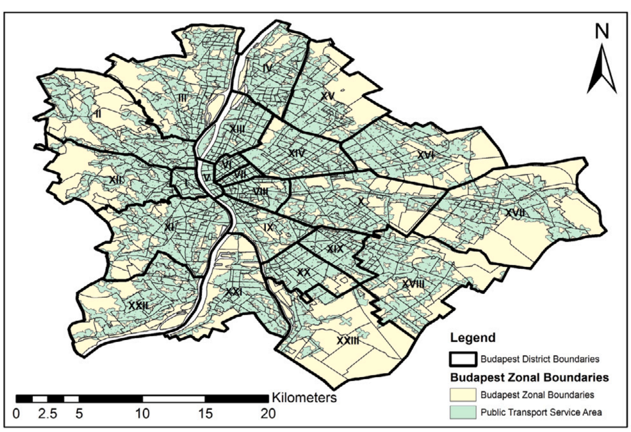

Fig. 3. The service area in Budapest.

creates supervised PTS data matrices. The obtained supervised PTS data matrices are further used to train the HDDA models (i.e., one model for the overall PTS and one for the PTS with respect to the coverage, supply, and diversity) to predict the PTS impacts of the different policy actions. The trained HDDA models are presented in Tables 4, 5, 6, and 7. It is worth highlighting that the HDDA models are validated via the out-of-sample validation method (i.e., 80%–20% training–testing).

3.1. Hypothetical policy implementation

To demonstrate the functionality of the trained HDDA models within the framework of the proposed decision support tool, a two-step hypothetical policy is set aiming to develop the PTS of the zones in need of improvement (i.e., zones with poor overall PTS) hereinafter called target zones. In Step 1, the target criteria (i.e., criteria with poor PTS class) in the target zones are increased by 25%. For instance, the PTS coverage and supply are the target criteria for the target zone 17594 in District 17 (see Fig. 6). The impacts of this increment on those criteria and the overall PTS of the target districts are predicted by the trained HDDA models. The predictions made in Step 1 are expected to update the target zones and their target criteria. In Step 2, the target criteria of the remaining target zones are increased by another 25%, which is followed by predicting the impacts of this increment on those criteria and the overall PTS of the updated target zones. The predictions made in Step 2 are expected to update the results obtained in Step 1.

The results of this hypothetical, two-step policy are shown in Fig. 7. As mentioned before, the target zones are identified, and the underlying reasons (i.e., target criteria) behind their poor overall PTS are studied. The leftmost map in Fig. 7 shows the existing 68 target zones with poor overall PTS. Following the rules set in Step 1 of the hypothetical policy, the numerical indicators of the target criteria (see the PTS coverage, supply, and diversity classes of the target zones in Fig. 6) of the 68 target zones are increased by 25%.

The impacts of the increment applied in Step 1 on the target criteria and the overall PTS of the target zones are predicted by the trained HDDA models, as shown in Fig. 7 (note for the sake of clarity, merely the predicted overall PTS is illustrated). The predictions made in Step 1 update the target zones and their target criteria as the target zones become 57.

As mentioned before, in Step 2 of this hypothetical policy implementation, the target criteria of the remaining target zones are increased by another 25%. The impacts of this increment on PTS are predicted by using the trained HDDA models, as shown in Fig. 7. The results of the prediction show that the increment applied in Step 2 results in the decrease of the target zones from 57 to 44 zones.

4. Limitations and future works

This paper and the introduced framework have some limitations. Firstly, cost implications are not included in the framework presented in this study. The framework might be extended to an optimization problem to find the best combination of the targeted policy actions and to improve the PTS of the identified areas with minimized implementation cost to be minimized. Secondly, a wide range of PTS measures could be incorporated in the case study to demonstrate the flexibility of the introduced approach to a greater extent. However, due to the data availability issues at the time this research is conducted, introductory measures are used in the case study. Future works might demonstrate the flexibility of the approach by using more varied PTS measures.

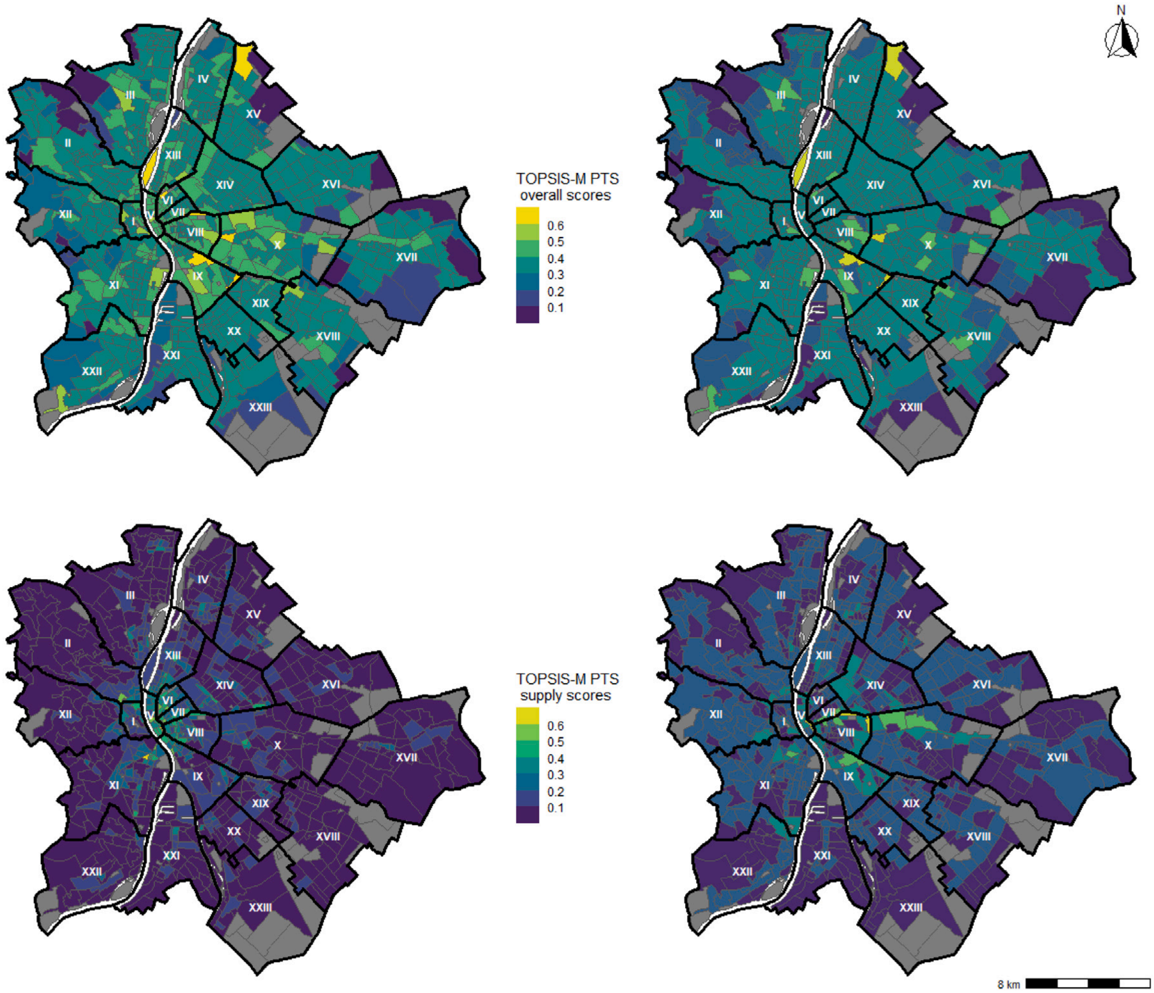

Fig. 4. PTS evaluation: the TOPSIS-M PTS overall, coverage, supply, and diversity scores.

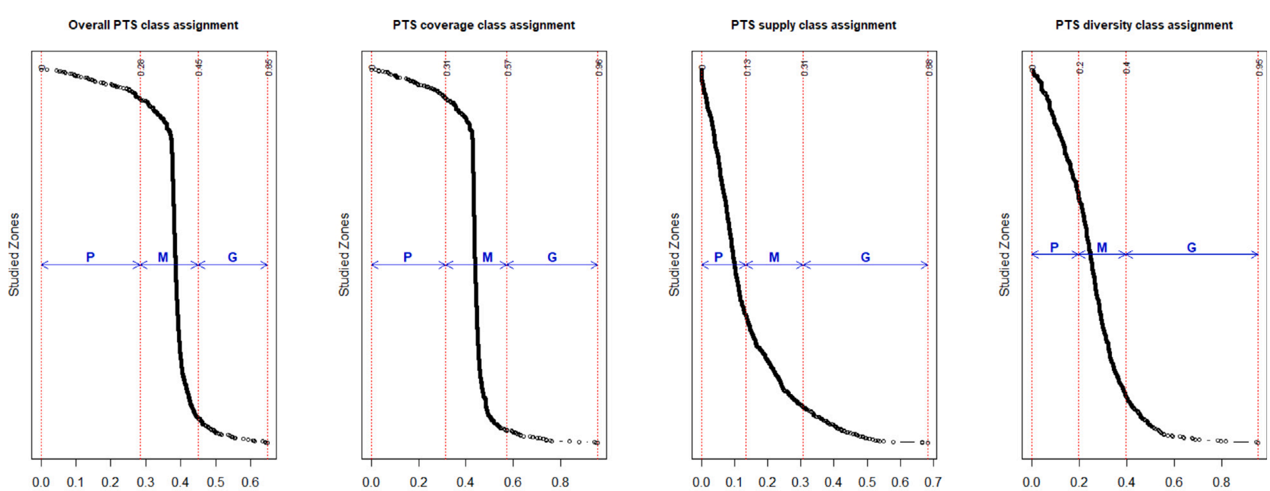

Fig. 5. PTS class assignment by using Jenks natural breaks optimization algorithm. P:poor, M:moderate, G:good.

5. Conclusion

Current paper proposes a comprehensive approach to combine a potentially large number of PTS measures and to build high-dimensional supervised classification models that predict the PTS levels for various policy actions. With this data-driven decision support tool, planners can identify and rank the areas in need of PTS improvement, determine what policy actions should be targeted to improve the PTS in the

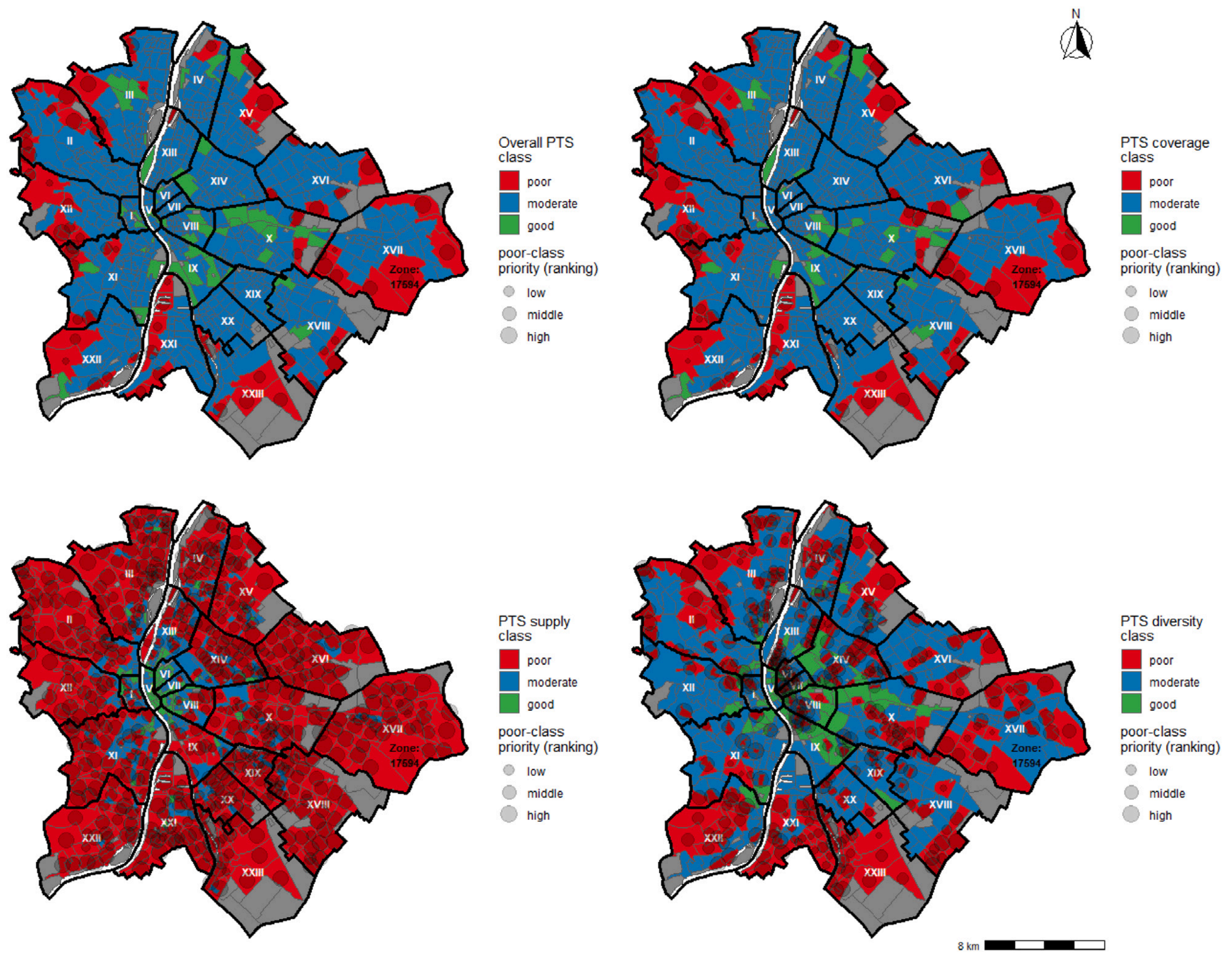

Fig. 6. The results of the PTS class assignment along with in-class priorities (i.e., existing services/current situation).

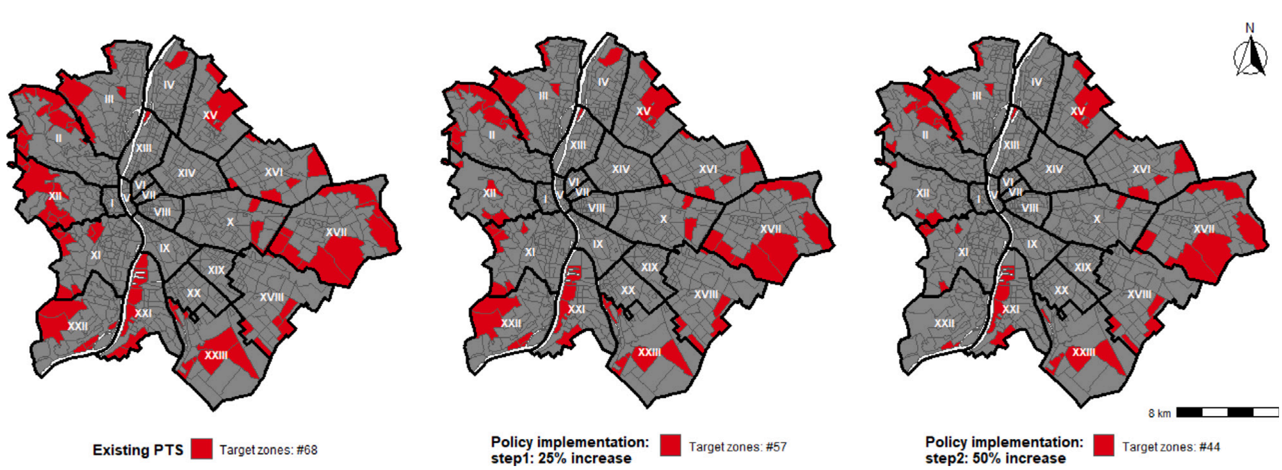

Fig. 7. The results of the policy implementation.

identified areas, and predict the impacts of these policy actions on the PTS in the identified areas. The proposed framework is flexible in a way that the planners/analysts are not restricted by some predetermined PTS measures and can study any relevant measure of interest in this framework according to the scope of the analysis and the availability of data. Moreover, a potentially large number of PTS measures can be incorporated not solely in the evaluation of the existing PTS but in the prediction of the PTS levels for different policy actions, as well.

The application of the proposed framework is demonstrated in detail through a case study, which is followed by a hypothetical policy implementation. In this case study, three PTS criteria and seven indicators are assessed to evaluate the PTS of the TAZs in the city of Budapest, Hungary. Furthermore, the impacts of a two-step hypothetical policy implementation on the target zones and the targeted PTS criteria are predicted by using the trained HDDA models in the case study (i.e., four HDDA models for predicting the PTS classes of the overall PTS, the PTS coverage, supply, and diversity). It is worth highlighting that solely seven indicators are assessed in the case study; however, the proposed approach is capable of handling considerably more number of indicators depending on the objectives of the study and the availability of data.

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

The data used in the case study will be made available from the corresponding author upon reasonable request.

Acknowledgments

The research reported in this paper is part of project no. BME-NVA-02, implemented with the support provided by the Ministry of Innovation and Technology of Hungary from the National Research, Development and Innovation Fund, Hungary, financed under the TKP2021 funding scheme. The first author further acknowledges the support of the KTH Royal Institute of Technology, Sweden for providing the open access fund.

References

Aman, Javad Jomehpour Chahar, Smith-Colin, Janille, 2020. Transit Deserts: Equity analysis of public transit accessibility. J. Transp. Geogr. 89, 102869.
Bergé, Laurent, Bouveyron, Charles, Girard, Stéphane, 2012. HDclassif: An R package for model-based clustering and discriminant analysis of high-dimensional data.
Bouveyron, Charles, Girard, Stéphane, Schmid, Cordelia, 2007. High-dimensional discriminant analysis. Comm. Statist. Theory Methods 36 (14), 2607–2623.
Currie, Graham, 2010. Quantifying spatial gaps in public transport supply based on social needs. J. Transp. Geogr. 18 (1), 31–41.
Dargay, Joyce M., Hanly, Mark, 2002. The demand for local bus services in England. J. Transp. Econ. Policy 36 (1), 73–91.
Fransen, Koos, Neutens, Tijs, Farber, Steven, De Maeyer, Philippe, Deruyter, Greet, Witlox, Frank, 2015. Identifying public transport gaps using time-dependent accessibility levels. J. Transp. Geogr. 48, 176–187.
Frappier, Alexis, Morency, Catherine, Trépanier, Martin, 2018. Measuring the quality and diversity of transit alternatives. Transp. Policy 61, 51–59.
Hawas, Yaser E., Hassan, Mohammad Nurul, Abulibdeh, Ammar, 2016. A multi-criteria approach of assessing public transport accessibility at a strategic level. J. Transp. Geogr. 57, 19–34.
Holmgren, Johan, 2008. Studies in Local Public Transport Demand (Ph.D. thesis). Linköping University Electronic Press.
Hwang, Ching-Lai, Yoon, Kwangsun, 1981. Methods for multiple attribute decision making. In: Multiple Attribute Decision Making. Springer, pp. 58–191.
Jenks, George F., 1967. The data model concept in statistical mapping. Int. Yearb. Cartogr. 7, 186–190.
Karou, Saleem, Hull, Angela, 2014. Accessibility modelling: predicting the impact of planned transport infrastructure on accessibility patterns in Edinburgh, UK. J. Transp. Geogr. 35, 1–11.
Khan, Fouad, 2012. An initial seed selection algorithm for k-means clustering of georeferenced data to improve replicability of cluster assignments for mapping application. Appl. Soft Comput. 12 (11), 3698–3700.
Kimpel, Thomas J., Dueker, Kenneth J., El-Geneidy, Ahmed M., 2007. Using GIS to measure the effect of overlapping service areas on passenger boardings at bus stops. Urban Reg. Inf. Syst. Assoc. J. 19 (1).
Klumpenhouwer, Willem, Huang, Wei, 2021. A flexible framework for measuring accessibility with destination bundling. J. Transp. Geogr. 91, 102949.
KSH (Hungarian Central Statistical Office), 2017. Statistical Yearbook of Budapest, 2016. Hungarian Central Statistical Office.
Liu, Dong, Qi, Xiaochen, Li, Mo, Zhu, Weifeng, Zhang, Liangliang, Faiz, Muhammad Abrar, Khan, Muhammad Imran, Li, Tianxiao, Cui, Song, et al., 2019. A resilience evaluation method for a combined regional agricultural water and soil resource system based on Weighted Mahalanobis distance and a Gray-TOPSIS model. J. Clean. Prod. 229, 667–679.
Lu, Yingshuang, He, Tong, Xu, Xinliang, Qiao, Zhi, 2021. Investigation the robustness of standard classification methods for defining urban heat islands. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 14, 11386–11394.
Mackett, Roger L., Achuthan, Kamalasudhan, Titheridge, Helena, 2008. AMELIA: A tool to make transport policies more socially inclusive. Transp. Policy 15 (6), 372–378.
Mavoa, Suzanne, Witten, Karen, McCreanor, Tim, O'Sullivan, David, 2012. GIS based destination accessibility via public transit and walking in Auckland, New Zealand. J. Transp. Geogr. 20 (1), 15–22.
McNeil, Nathan, 2011. Bikeability and the 20-min neighborhood: How infrastructure and destinations influence bicycle accessibility. Transp. Res. Rec. 2247 (1), 53–63.
Meyer, Michael D., 2000. Transport planning for urban areas: A retrospective look and future prospects. J. Adv. Transp. 34 (1), 143–171.
Murray, Alan T., 2001. Strategic analysis of public transport coverage. Socio-Econ. Plann. Sci. 35 (3), 175–188.
Preston, John, Rajé, Fiona, 2007. Accessibility, mobility and transport-related social exclusion. J. Transp. Geogr. 15 (3), 151–160.
Saaty, Thomas L., 1980. The Analytic Hierarchy Process. McGrow-Hill, New York.
Sheikh, Vahedberdi, Kornejady, Aiding, Ownegh, Majid, 2019. Application of the coupled TOPSIS–Mahalanobis distance for multi-hazard-based management of the target districts of the Golestan Province, Iran. Nat. Hazards 96 (3), 1335–1365.
Shih, Hsu-Shih, Olson, David L., 2022. TOPSIS and its Extensions: A Distance-Based MCDM Approach. Vol. 447, Springer Nature.
Tzeng, Gwo-Hshiung, Huang, Jih-Jeng, 2011. Multiple Attribute Decision Making: Methods and Applications. CRC Press.
Vega, Alberto, Aguarón, Juan, García-Alcaraz, Jorge, Moreno-Jiménez, José María, 2014. Notes on dependent attributes in TOPSIS. Procedia Comput. Sci. 31, 308–317.
Wang, Zheng-Xin, Wang, Yan-Yu, 2014. Evaluation of the provincial competitiveness of the Chinese high-tech industry using an improved TOPSIS method. Expert Syst. Appl. 41 (6), 2824–2831.
Webster, F.V., Bly, P.H., 1981. The demand for public transport: Part I. The changing environment in which public transport operates. Transp. Rev. 1 (4), 323–351.
Wu, Belinda M., Hine, Julian P., 2003. A PTAL approach to measuring changes in bus service accessibility. Transp. Policy 10 (4), 307–320.
Zheng, Lijuan, van Wee, Bert, Oeser, Markus, 2019. Combining accessibilities for different activity types. J. Transp. Land Use 12 (1), 853–872.