Journal of Urban Mobility 1 (2021) 100004

Contents lists available at ScienceDirect

Journal of Urban Mobility

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

Extracting spatiotemporal commuting patterns from public transit data

Trivik Verma[a,*], Mikhail Sirenko[a], Itto Kornecki[b], Scott Cunningham[c], Nuno A.M. Araújo[d,e]

[a] Faculty of Technology, Policy and Management, Delft University of Technology, Delft 2628BX, the Netherlands
[b] ETH Zürich, Universitätstrasse 16, Zürich 8092, Switzerland
[c] Faculty of Humanities and Social Science, University of Strathclyde, 18 Richmond Street, Glasgow G1 1XQ, Scotland, United Kingdom
[d] Departamento de Física, Faculdade de Ciências, Universidade de Lisboa, Lisboa P-1749-016, Portugal
[e] Centro de Física Teórica e Computacional, Faculdade de Ciências, Universidade de Lisboa, Lisboa 1749-016, Portugal

ARTICLE INFO

Keywords:
Smart card data
Mixture models
Clustering
Demand forecasting
Public transit

ABSTRACT

Public transit networks in cities are crucial in addressing the transforming mobility needs of citizens for work, services and leisure. The rapid changes in urban demographics pose several challenges for the efficient management of transit services. To forecast transit demand, planners often resort to sociological investigations, modelling or population data that are either difficult to obtain, inaccurate or outdated. How can we then estimate the variable demand for mobility? We propose a simple method to identify the spatiotemporal demand for public transit in a city. Using a Gaussian mixture model, we decompose empirical ridership data into a set of temporal demand profiles representative of ridership over any given day. A case of ≈ 4.6 million daily transit traces of the primary mode of underground services from the Greater London region reveals distinct commuting profiles. We find that a weighted mixture of these profiles can generate any station traffic remarkably well, uncovering spatially concentric clusters of mobility needs. Our results also suggest that heavily used stations that exhibit mixed-use commuting patterns are generally located in the cluster of the central business district and stations away from the centre of the city are largely single use residential areas. Overall, identifying mixed temporal and spatial use of stations diverging from macro mobility patterns in public transit indicates that our approach may be useful in a detailed understanding of integrated transit planning for heterogeneous needs of travellers.

1. Introduction

Public Transit (PT) networks play a significant Rodrigue et al. (2013) and challenging role Zhang et al. (2019) in serving citizens' business, industrial, social, cultural, educational and recreational needs Alonso (1960); Ceder (2016). Development around transit stations promotes mixed-use of urban areas by encouraging people to live near clusters of amenities and services, built around transit stations. Such development enhances accessibility of the city for a wider group of citizens Tomer et al. (2011), improves pedestrian access Dittmar and Ohland (2012), and reduces pollution and congestion by discouraging reliance on private vehicles Calthorpe (1993). As cities are continually growing Bettencourt et al. (2007), commuting patterns show diversity and relate closely with land-use Gan et al. (2020). Even though much scholarly work has revealed the estimation of commuting behaviour in public transit Briand et al. (2017), there is a gap in our understanding of how diversity in our mobility needs relate to land-use. Moreover, decision makers often end up under or overestimating varying transit demand for substantial infrastructure investments Flyvbjerg et al. (2005) and often such projects threaten to create segregated or sprawling urban areas Dawkins and Moeckel (2016); Henig (1980). To address the complex demand for mobility, it is important to identify the variability in commuting needs that change with space and time De Domenico et al. (2014).

Due to the changing transit needs and developments in urban land-use Smith and Hall (2013), the planning and management of transit services pose a huge challenge in addressing complex and competing priorities of urban demand Byrne (2003). There are at least three main approaches to analysing demand for PT. First, scholars use census-based population statistics and usage estimates from transit authorities to analyse transportation demand Ceder (2016). While such methods were quite prevalent among transportation planners before Teodorović and Janić (2017), the changes in urban fabric Beaverstock and Smith (1996); Darling (2017) make derived estimates less reliable Flyvbjerg et al. (2005). Second, researchers emphasise qualitative methods to estimate transport demand, involving direct observation of urban populations Gutiérrez et al. (2011); Raudenbush and Sampson (1999); Taylor et al. (2009). Due to the costs associated with surveys, these methods analyse transport analysis zones where activities are aggregated by zones that are classified as origins or destinations Teodorović and Janić (2017) and are often only classified as residential or work places. Third, there are studies that focus on revealing

* Corresponding author.
E-mail address: t.verma@tudelft.nl (T. Verma).

https://doi.org/10.1016/j.urbmob.2021.100004
Received 23 November 2020; Received in revised form 13 July 2021; Accepted 16 July 2021

macroscopic urban structures Anas et al. (1998); Barthélemy (2011); Burgess (2008) by measuring aggregated Origin-Destination (OD) trajectories of people using mobile phone Calabrese et al. (2011); Louail et al. (2015, 2014); Noulas et al. (2013) and social media data McNeill et al. (2017). While there is evidence that urban mobility patterns are reproducible using aggregated statistics of populations derived from mobile data González et al. (2008), this data usually only accounts for a subset of the entire population. Approximately 95% of the population is missing from such an analysis: people who may not be able to afford mobile services or provide their data Louail et al. (2015). What is more, recent work suggests methods based on incomplete statistical data underestimate important trips, especially in larger cities Chico et al. (2019). In addition, data from 210 transport projects across 14 nations show that forecasts for demand are often inaccurate Flyvbjerg et al. (2005) leading to substantial financial and economic risks. The general framework of estimating transit demand aims to identify universal demand without going into the heterogeneity of travel behaviours.

Over the past decade, digital services like Automatic Fare Collection (AFC) through the use of smart cards have been introduced into transit networks worldwide. There is important literature on flow estimation that extracts detailed and complete OD trajectories Long and Thill (2015); Park et al. (2008); Roth et al. (2011); Zhong et al. (2014) from AFC data. Using high resolution data that is more granular, these studies provide aggregated instances of mobility flows between large hotspots in a city, mainly focusing on recovering morphological characteristics of the urban structure Anas et al. (1998); Barthélemy (2011). Several scholars have also designed new methods of studying smart card data that generate a lot of knowledge in multiple transportation domains Pelletier et al. (2011) and shown its usefulness in demand forecasting Briand et al. (2017). Though such studies reveal complex characteristics of typical commuting behaviours, we do not yet understand the heterogeneity in complex commuting patterns and their relationship with space or the transit network, where citizens' needs are beyond regular work-home commutes Ceder (2016), and differs by use in the network context Jun et al. (2015).

We propose a method to understand the varying commuting patterns in public transportation in space and time using anonymous, privacy-preserving, granular and open-source entry-only ridership data. Our case uses the data of ≈ 4.6 million daily commutes in the Transport for London (TfL) services in the Greater London Underground network. We find that the daily traffic through this PT network is a mixture of *six* demand profiles. Using the weights of these profiles, our model is able to reproduce individual network station traffic throughout the day and cluster stations into *six* categories. Upon mapping these categories, we find how these profiles can identify the spatial distribution of varying mobility demand for the PT infrastructure. We discuss how the temporal nature of complex urban demand reveals the spatial structure of the city consisting of central Murphy (2017), polycentric Louf et al. (2013); Roth et al. (2011) and concentric Burgess (2008); Hoyt (1939) zones of development. We expect our method could also be useful for granular transportation demand analysis of PT infrastructures in any region in the world where transport planners can strengthen efforts in evaluating variable use of urban spaces.

2. Related work

Comprehensive data about the travel patterns of passengers supports the strategic, tactical and operational management of transport services throughout cities. Much of this work can be accomplished using general mathematical models of urban mobility González et al. (2008). Traditionally, and in the absence of plentiful digitised evidence of travel demand, planners created origin-destination matrices. Demand may be estimated by direct observation, or through travel surveys - this is expensive and time-consuming Munizaga and Palma (2012). Increasingly however, cities are implementing smart card systems, thereby enabling comprehensive capture of travel patterns by individuals, by hour and weekday, and by station Pelletier et al. (2011). Usage patterns of stations are affected by the network of stations, as well as by the surrounding catchment of the station. The following review examines theory and prior work. The first section concerns networks and their catchments, the second concerns urban populations and land usage, and the third section concerns the diversity of temporal patterns evidenced by urban travellers.

Travel Patterns by Network and Catchment. The surrounding catchment of urban population is a significant determinant of subway station ridership, with declining demand for station usage as the station is located more distantly from the central business district Jun et al. (2015). This declining gradient of demand is caused by lower population and lower population densities through the periphery of the city Cervero and Kockelman (1997). Previous studies have also demonstrated an association between ridership and the available employment opportunities, which also decline at the periphery of the city Thompson et al. (2012). The most typical means of forecasting demand using catchment characteristics is by means of an ordinary least squares regression analysis, using a variety of different land use features, and a fixed but feasible walking catchment area to the station Guerra et al. (2012); Jun et al. (2015).

Previous work clusters subways by their similarity of passenger type, or by their similarity of temporal usage. Kim et al. Kim et al. (2017) cluster stations by usage patterns, revealing predominant demographic predictors of usage. El Mahrsi et al. Mahrsi et al. (2017) cluster subways in the city of Rennes, where station similarity is judged by temporal patterns of station usage. Eleven clusters are identified, of which many indicate heavy usage during rush hour peaks. Other clusters are active throughout the day, and are located in the centre of Rennes. Other clusters show important activity during the weekend, perhaps indicating leisure and sporting activity. In contrast, other clusters are inactive during weekend, perhaps indicating business districts lacking other features of interest.

Another predictor of subway traffic is the extended networked context of the subway and its respective rail lines Jun et al. (2015). For some passengers a rail station may be neither an origin or a destination, but a stopover on an extended journey through the city. The travel may entail exiting the station, for instance to engage in other modes of travel. Or the passenger may be switching lines, since a direct route to their final destination is not available. The networked context of subway rail operations emerges naturally from efforts to plan or schedule more efficient or more equitable terms of service Shang et al. (2018). Some stations within the network may have a primary role as mobility hubs, and may therefore witness unusual traffic volumes, or daily profiles of use Sohn and Shim (2010). In truth, and as will be discussed later in this review, the most effective clustering procedures are generative and multi-levelled, reliant on multiple attributes of passenger, trip, time and station.

Travel Patterns by Population and Land Use. The understanding of the spatial variation of subway demand within the city requires a deeper understanding of population and employment patterns throughout the city. A highly generalisable model of urban land competition is available in the literature based on economic competition over competing uses of the scarce urban quantity of land Alonso (1964). In the most basic form of this model employers crowd into the central business district of the city, where they mutually benefit from a range of amenities, including business services, exchange of know-how, and a pool of available workers Quigley (1998). Employees make trade-offs between the costs of property, the time spent on travel, and the availability of land, resulting in expensive and high-density inhabitation at the core of the city. As with any such stylised model, there are notable empirical departures from the ideal. Specifically, the formation of multiple peripheral hubs results in multiple hubs of activity manifesting as the polycentric city Anas et al. (1998). Descriptive models such as these have been invaluable in establishing demand patterns based on popu-

lation, employment and location, and have been useful in establishing the functional properties of specific districts in light of their patterns of transport.

A variety of transport and mobility sources are being used to discover the dynamic attributes of urban districts and their access and usage Terroso-Sáenz et al. (2021). The functional uses of urban districts are reflected by the volume and modality of incoming and outgoing traffic to the district. Previous work has used data and traces from taxi trips, bike trips, subways and truck travel as sources for land use discovery Gan et al. (2020); Terroso-Sáenz et al. (2021); Zhai et al. (2019); Zhang et al. (2018). Other work characterises origin-destinations as a complex, weighted network, revealing common dynamical patterns of functional land use Saberi et al. (2017). The use of mobility data provides a valuable dynamic component to understanding functional land use given the fact that the traditional sources of data such as satellite imagery or administrative statistics is primarily static in character.

Travel Patterns and the Dynamics of Individual Mobility. One of the complexities of clustering smart card data stems from the heterogeneity of potential passenger types. Previous work clusters passengers according to similarities in smart card usage. Any specific passenger is likely a mixture drawn from multiple ideal passenger types. The resultant clusters demonstrate stability over time, as well as distinct patterns of fare and card usage. This demonstrates the validity of the clustering, and the resultant effectiveness of the classification. While the initial work focuses on smart-card data from the public transport system of Rennes Metropole, related later work applies similar models to smart card data from the city of Gatineau, Canada Briand et al. (2015, 2017).

Another complexity of temporal clustering is analysing the continuously varying patterns of transport usage, seen across the day and across the week. There are distinct temporal characteristics of traffic, reflecting multiple and inter-connected rhythms, repeating once and twice daily, and over weekly and yearly duration. Despite commonalities across such traffic patterns, there remains considerable heterogeneity across passenger types regarding their station access and subway usage. Recent work analyses the Nanjing subway system, finding seven distinct temporal patterns. These patterns correlate significantly with local patterns of land usage Gan et al. (2020). Other work identifies temporal patterns of metro usage in Shanghai and Shenzhen Duan et al. (2018). Work on the Shenzhen subway system demonstrates the relevance of geographical variables and land use in establishing the dynamics of ridership in stations He et al. (2020). Another relevant paper examined daily and yearly fluctuations in public transport in the city of Riga Pavlyuk et al. (2020). Simultaneous clustering of passenger and time profiles can be performed using two-level generative models, where the first level assigns passengers to usage types and the second level assigns types to various time-dependent travel profiles Briand et al. (2015, 2017). Other multi-level models are possible including models of population aggregates rather than individuals, and models clustering stations using surrounding land as covariates.

Opportunities for New Work. A comprehensive analysis of subway traffic, coupled with administrative statistics about population, density, rail use and land use can assist both public transport planners as well as land use planners. While there is certainly a diurnal rhythm of work and home-life, there are more complex patterns of transportation activity in the city, created by multiple functional uses of urban districts, including polycenters, multi-modal transport hubs, shopping, entertainment, and tourism. Understanding of this urban land use context can help manage the strategic positioning of subway stations within the city, as well as the day to day scheduling or operations of stations. A research gap remains in understanding the joint dynamics of population-level ridership given local land use and available subway connectivity.

The contribution of this paper is to enable an improved understanding of the context and use of metropolitan subway, metro or underground stations. This is accomplished by clustering user trajectories over the work week, and establishing a characteristic mixture of user classes by station. These mixtures are further positioned within the network context of surrounding stations, and grounded in the attraction basins for city regions proximal to the station. The results demonstrate a complex profile based on the spatial network context as well as a population profile of temporal use of stations.

3. Data and methods

3.1. Transport data

We use the London Underground Passenger Count dataset as a proxy for ridership, which is provided freely by TfL London (2018). This data was collected using AFC systems through the Oyster smart card setup in Greater London, UK. Enabled by AFC, travellers in the system are required to check-in and check-out when entering the station or departing it, respectively. It is important to note that the passenger count dataset used in this study does not have passenger trajectories following individual riders on their journeys. Instead, we only use counts of passengers checking in and out at every station in the network. The dataset describes the average number of *entrances* and *exits* to and from each station in the Underground Network, represented as a discrete time series spanning 24 hours. The time series is aggregated at 15-minute intervals, resulting in 96 data points per station for a total of 264 stations. We remove instances of erroneous or zero counts for every station between the times of $02:00am$ and $05:00am$ where the PT network is not in function. This data represents an average of all days in the month of November 2017, separated into weekdays and weekends. We carry out our analysis only on the weekday data. The data description provided by TfL claims that November 2017 illustrates a typical sample of winter travelling behaviour in the year and has been adjusted for any disruptions in the Underground service (such as related to weather, malfunctions and accidents or large community events). For details on potential sources of error in the ridership data, see Supplementary Note 4.

Beyond traditional transportation patterns. We use ≈ 4.6 million geolocated entrance observations of daily TfL passengers. To understand overall system ridership behaviour in PT we analyse the passenger entrance counts, $P_i(t)$, entering a station i at time interval t ($\forall t \in 1, 2, \ldots, m$) where each interval is a 15-minute observation window in which the data is collected with $m = 96$ intervals per day. The variable P_i is a proxy for the ridership behaviour, an indication of the usage of every station at different times in a day.

Fig. 1 a represents the average ridership for the PT system across all stations in the network throughout the day with the quantile interval showing variations across all stations in the city (see Supplementary Figure 1 for an overview of the entire system traffic). As it is evident from Fig. 1b, aggregate station ridership is symmetric: data from a day shows excellent correlation between the total number of entrances and exits for every station. To understand how station traffic is related to population statistics, we visualise the relationship between the number of total station entrances on a given weekday, $\sum_t P_i(t)$, and the working adult population of every zone associated with the station (see Supplementary Note 1 for details on estimating population sizes for station zones). Fig. 1c illustrates a very weak relationship between the number of people residing in a zone and the entrances at the corresponding station. The weak correlation suggests population statistics are not a good proxy for identifying commuting patterns. If people are not using the station closest to them, that may be due to some stations not fulfilling the potential for accessing opportunities in a city. Even though aggregate counts of entrances and exits are matched well (Fig. 1b), an indication of the complexity of urban mobility can be witnessed in asymmetric correlations between trips made in opposite directions illustrating the increasing use of stations for other activities than work and residential (see Supplementary Note 2 for details). Considering this evidence, census-based population surveys are not accurate enough for understanding commuting patterns Batty (1976) and are often found to underestimate importance of regular home-work trips Chico et al. (2019). Thus, to further understand the complexities of varying ridership behaviour in PT

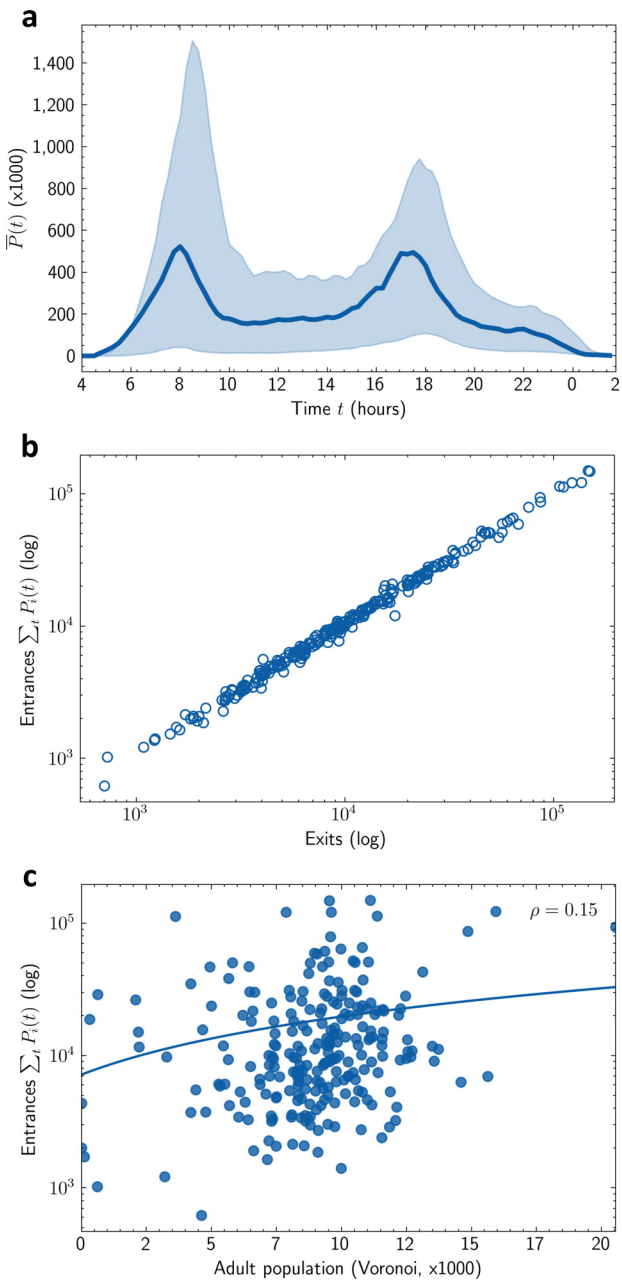

Fig. 1. Describing the TfL data of station traffic over a representative day. (a). Average Passenger counts $\overline{P}(t)$ for the PT system showing entrances for every 15-minute intervals. The confidence intervals show the quantile interval (10-90% of data) and the blue line is the average traffic for each cluster **(b).** Relationship between the total entry vs exit counts for every station in the system. **(c).** Relationship between population and ridership (entry counts) for Voronoi cells that are attributed to every station (see Supplementary Information for estimating population counts for station zones). The number of stations in this figure are less than the total number in the dataset because some stations are outside Greater London for which population estimates were not available. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

we focus on identifying more detailed and distinct commuting patterns in a day.

3.2. Methodology for analysis of commuting patterns

Transportation demand analysis helps transport planners in understanding the use of the system across space and time. A much more detailed understanding of the peak and off-peak hours of commuting patterns across the system in both space and time will help planners improve transport services for all groups of people, not just during peak hours. Our model formulation is based on a three-step approach using a generative model that describes the temporal commuting patterns of the public transit system. First, using smart card data specifying entries of passengers in the transit system, we identify the general profiles of commuting patterns over an average weekday. Second, using the specific parameters for each of the general commuting patterns, we reconstruct individual station traffic to understand how each station traffic relates to the overall system traffic. This allows us to estimate both in-sample and out-of-sample traffic at individual stations. Third, we cluster the stations based on their generated traffic data to understand the location and use of specific archetypes of stations in space. This methodology aids us in understanding the commuting patterns, and thus demand for public transit, in both time and space through a generative model representation.

Identifying Transport Commuting Patterns. To identify the composition of general profiles of commuting patterns in the ridership data, we formulate a Gaussian Mixture Model (GMM). GMMs are formed so subpopulations can be automatically learned from a large dataset without annotating any data points in advance with user-defined labels. A formulation of this type constitutes a class of unsupervised learning algorithms and allows us to not define specific kinds of commuting patterns beforehand. The foundation of these models is built upon a mixture of several normal distributions. In the case of a passenger count dataset of a public transit system, the underlying distribution of the overall traffic at a station per day follows the sum of multiple scaled normal distributions (Fig. 2) with their means at different times representing local peaks of different commuting patterns. Translating this mixture onto a standard GMM means that each distribution learned by the model represents one specific commuting pattern. It is useful to observe that a data point in any one normal distribution does not necessarily classify a person entering a station as belonging to a specific commuting pattern. A GMM is probabilistic in nature that associates to each data point a likelihood of belonging to a specific normal distribution (a specific type of commuting pattern).

A GMM is characterised by three parameters: individual mixture weights ϕ_k, mixture means μ_k and variances σ_k for all mixtures C_k. Because of the discrete nature of the ridership data, a GMM is a natural choice for representing temporal commuting patterns of passengers in a transit system and helps in estimating the mean and variance around each of the typical commuting patterns Briand et al. (2015). For the case of TfL data, every passenger checking into the system leaves a smart-card trace behind. Let's name this observation x. Every observation x thus represents the smart-card log of the trip's time rounded to every 15 min interval within an hour and has a likelihood of belonging to a particular type of commuting pattern in a day (for instance, morning or evening traffic).

Fig. 1 suggests that there are substantial peaks in the overall traffic in the system outside of the regular home-work commute as well. Selecting more than 2 gaussians provides a clearer representation of this off-peak behaviour. To identify the most suitable value of k and calibrate our model, we vary the number of gaussian mixtures describing commuting patterns between $k = 5$ to $k = 7$ (see Supplementary Figure 6 for effects of variations in k). We use a set of 4 measures to guide us with our selection criterion (Akaike Information Criterion, Bayesian Information Criterion, Calinski-Harabasz and Davies-Bouldin Steinley and Brusco (2011)). Using the information criteria suggested above, we find that using anywhere between 4–7 gaussians will represent the data well and minimise information losses (see Supplementary Figure 7 for the information criterion). The higher the value of k, the longer will be the computational efforts required to find a stable solution. Seeing that there are different commuter profiles within the Greater London region (work, home nighttime workers, tourism, etc.) London (2018), is it also important to contextualise the computational time a model takes within

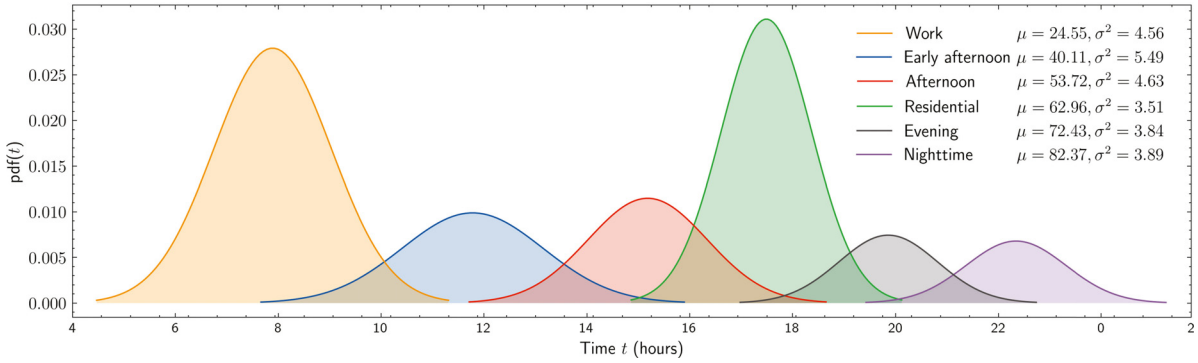

Fig. 2. Temporal commuting patterns over the representative day. Each normal distribution refers to a particular kind of traffic in time and the density estimates $pdf(t)$ measures the likelihood that an individual entering the system belongs to a certain commuting pattern. $\sum_k pdf(t) = 1$.

the broader suitability to a particular set of behaviours seen in a transit system. Hence we selected $k = 6$.

Given our input data in the form of a discrete time-series and the number of mixtures we specify, the formulated GMM first estimates a-posteriori the unknown parameters $(\phi_k, \mu_k, \sigma_k)$ using an expectation-maximisation algorithm Dempster et al. (1977). The probability distribution of the data points x in a GMM is given by,

$$p(x) = \sum_{k=1}^{K} \phi_k \, \mathcal{N}(x \mid \mu_k, \sigma_k), \tag{1}$$

$$\mathcal{N}(x \mid \mu_k, \sigma_k) = \frac{1}{\sigma_k \sqrt{2\pi}} \exp\left(-\frac{(x-\mu_k)^2}{2\sigma_k^2}\right), and \tag{2}$$

$$\sum_{k=1}^{K} \phi_k = 1. \tag{3}$$

Eq. 3 shows that each mixture C_k is weighted such that the total probability distribution normalises to 1.

Generation of individual station traffic. To generate individual station traffic, we need to first have the estimates of the parameters of the GMM. As described in the previous section, Ref. Dempster et al. (1977) defines the expectation-maximisation algorithm used for converging to reasonable values of ϕ_k, μ_k, σ_k. The process of estimating the densities at every station involves two steps. First, we sample the Gaussian component according to the distribution of commuting profiles defined by $p(C_k) = \phi_k$. Second, we sample each data point belonging to the station from the distribution of mixture component C_k using $p(x \mid C_k) = \mathcal{N}(x \mid \mu_k, \sigma_k)$. Though we estimate densities at each station for observations that belong to the dataset, traffic for a new station that records out-of-sample observations could also be estimated using the parameters for our formulation (as traffic at a new station may not change the overall system commuting patterns significantly, especially in big cities like Greater London, the UK).

Clustering stations. As the discrete time series data of passenger counts is aggregated by 15-minute intervals, we implicitly map the generated station traffic into a matrix where each row can be interpreted as station traffic over different times in a day. $P_i(t_j)$ is the measure of use of station i at time-step t_j (Table 1). To curb the skewing effects of larger stations that also witness incoming and outgoing traffic to and from other cities, we normalise the passenger counts such that $\sum_i P_i(t) = 1 \; \forall$ stations i. Columns showing traffic over all stations at every time-step t_j have long-tail distributions and each station traffic vector $[t_1, t_2, t_3, \ldots, t_n]$ is a multi-modal (not related to *modes* of transportation) distribution (Fig. 1a). Given this description, we formulate a multivariate GMM for clustering stations into *six* characteristic station types. To arrive at the conclusion of using six station clusters, we use a set of 3 measures to guide us with our selection criterion (Silhouette Score, Calinski-Harabasz and Davies-Bouldin Steinley and Brusco (2011)). Supplementary Figure 8 illustrates that the ratio of variance starts to degrade beyond 8 clusters. Since our goal is to focus on the significant differences in commuting behaviours and not on the absolute number of station types, we choose the value of 6 such that our models also converge faster.

Table 1
Data schema for the feature matrix used for clustering. Each station has n features t_j where every column (feature) represents the passenger traffic count $\vec{P}(t_j)$ at each time interval of 15 minutes. As an example, $P_3(t_3)$ is the passenger traffic entering station 3 at time interval $05:30am$ - $05:45am$, generated using our modelling approach.

Feature Matrix					
Station	t_1	t_2	t_3	$t_{...}$	t_n
1	.	.	.	.	.
2	.	.	.	.	.
3	.	.	$P_3(t_3)$	.	.
...	.	.	.	.	.
m	.	.	.	.	.

To model the multivariate GMM case, we use the formulation,

$$p(\vec{x}) = \sum_{k=1}^{K} \phi_k \, \mathcal{N}(\vec{x} \mid \vec{\mu}_k, \Sigma_k), \tag{4}$$

$$\mathcal{N}(\vec{x} \mid \vec{\mu}_k, \Sigma_k) = \frac{1}{\sigma_k \sqrt{2\pi}} \exp\left(-\frac{(\vec{x}-\vec{\mu}_k)^2}{2\sigma_k^2}\right), \tag{5}$$

where Eq. 5 is the probability density function of the multivariate normal distribution, $\vec{\mu}_k$ represents the means and Σ_k the covariance matrices Eirola and Lendasse (2013).

To cluster station traffic that follow similar trends over a day, we utilise the expectation step of the expectation-maximisation algorithm Dempster et al. (1977) which forms the basis of a GMM. Using the estimated model parameters for the multivariate distributions, we find the likelihood that a station traffic pattern $(\vec{x})$ belongs to a mixture C_k by calculating,

$$p(C_k \mid \vec{x}) = \frac{\phi_k \, \mathcal{N}(\vec{x} \mid \vec{\mu}_k, \Sigma_k)}{\sum_{k=1}^{K} \phi_k \, \mathcal{N}(\vec{x} \mid \vec{\mu}_k, \Sigma_k)}. \tag{6}$$

4. Results and discussion

Multiple temporal commuting patterns. The time series of entrances at stations represents an aggregation of many different commuting patterns. Understanding the distribution of these patterns can be very useful in inferring demand for public transit. To identify and interpret the different temporal commuting patterns within the PT system, we formulate a simple univariate GMM (see the Data and Methods

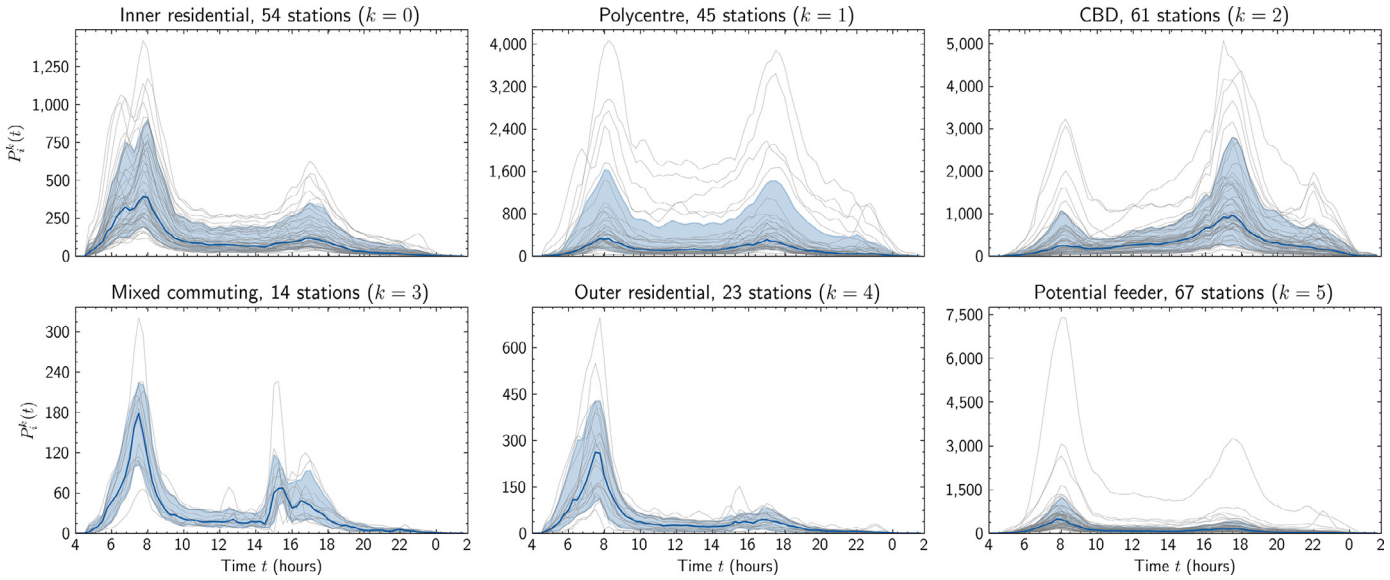

Fig. 3. Clusters of generated station traffic. *Six* clusters of station traffic showing different peculiar patterns of station use. Each sub plot is the time series of generated station traffic $P_i^k(t)$. The confidence intervals show the quantile interval (10 – 90% of data) and the blue line is the average traffic for each cluster. The thin grey lines show all station traffic belonging to a cluster. All 264 stations of the Greater London region are shown here.

section for details) modelling the time series of $\sum_i P_i(t)$. The model represents the different commuting patterns as gaussian distributions with varying mean (μ_k), variance (σ_k) and mixture weights (ϕ_k) for each commuting pattern C_k. We identify *six* commuting patterns that all represent the characteristic patterns of mobility in the system (see Supplementary Figure 6 for variation in the number of commuting patterns in a day). Each Gaussian distribution in Fig. 2 illustrates one typical commuting pattern, see Supplementary Table 1 for details of the parameters (μ, σ *and* ϕ.)

In addition to the measurements of goodness of clustering (see Supplementary Information for details) between different number of k values, we also qualitatively analysed the profiles of commuting trips in the Greater London region London (2018); Smith and Hall (2013). Our interpretation is the following,

1. Work (W): Morning trips for work and education;
2. Early Afternoon (LM): Flexible workers, tourists, shoppers and miscellaneous activities;
3. Afternoon (A): School and lunchtime traffic, flexible workers, tourists, and shoppers;
4. Residential (R): Evening trips returning from work;
5. Evening (E): Late night workers returning home and dinnertime traffic;
6. Nighttime (N): Service industry (restaurant, bars, healthcare) workers, and traffic from entertainment districts.

The typical commuting profiles mentioned above are generalisations of traffic but there is mixed usage throughout the day: tourists travel at all times and night workers come home in the morning as well. In this work, we do not categorise individual traces into any types.

Clustering stations by temporal commuting patterns. The commuting patterns we identify using the GMM are expressed using three parameters, a mixture component weight, mean and variance. Using this set of values we reconstruct traffic at every station as a linear combination of different types of generalised commuting profiles and classifying the station based on the relative value of the weights and the three estimated parameters of the GMM. The estimated probability density reveals the subscription of each station to every mixture. Next, using a multivariate GMM (see the Methods section for details on generating station traffic and clustering), we identify six characteristic station types.

We analyse the different clusters of stations in Fig. 3. Each cluster has a particular multi-modal distribution of average daily traffic entering the station:

1. *Central Business District (CBD)* stations show a higher number of entrances in the evening compared to the rest of the day. People usually move to the district for using services and business all day and return home at night;
2. *Polycentre* stations witness similar amount of workbound traffic in the morning and residential traffic in the afternoon, and a high amount of activity throughout the day compared to other clusters. These are large secondary hubs Louail et al. (2015) that have mixed use for residences, workplaces and services;
3. *Potential feeder* stations that are in a zone of transition Hoyt (1939) with changing land-use from a compact and busy CBD to wider residential regions with self-sufficient services. A peak in workbound traffic and a declining residential traffic pattern suggests middle-class housing workers residing in this region possibly feed into the city for work;
4. *Inner residential* stations may be serving working-class groups but are further away from the CBD (see Supplementary Figure 11). These stations differ from the feeders because of a striking peak in the early morning traffic just before the workbound ridership peaks;
5. *Outer residential* stations have lesser evening traffic and are much further from the CBD compared to inner residential stations. The lower evening volumes point to the region's greater residential nature;
6. Finally, *commuter* stations have a mix of suburban or satellite traffic in the morning and residential traffic in the evening, pointing to some work locations in the vicinity as is expected from clusters of suburban areas.

Combining the temporal commuting patterns and station clusters, Fig. 4a reports two scenarios. In the first one (left frame), we show how station clusters show variations among typical (W and R), midday (LM and N) and nighttime (EN and LN) commuting patterns. Note that the CBD is skewing the distribution toward nighttime traffic, possibly because entertainment centres are located close to business districts. The plots also reveal that the PT system in London is used much less in the midday hours in comparison to the regular home-work traffic at other times of the day. The second scenario (Fig. 4a - right frame) shows a prevalence of work (W) and other (LM, N, EN and LN) commuting pro-

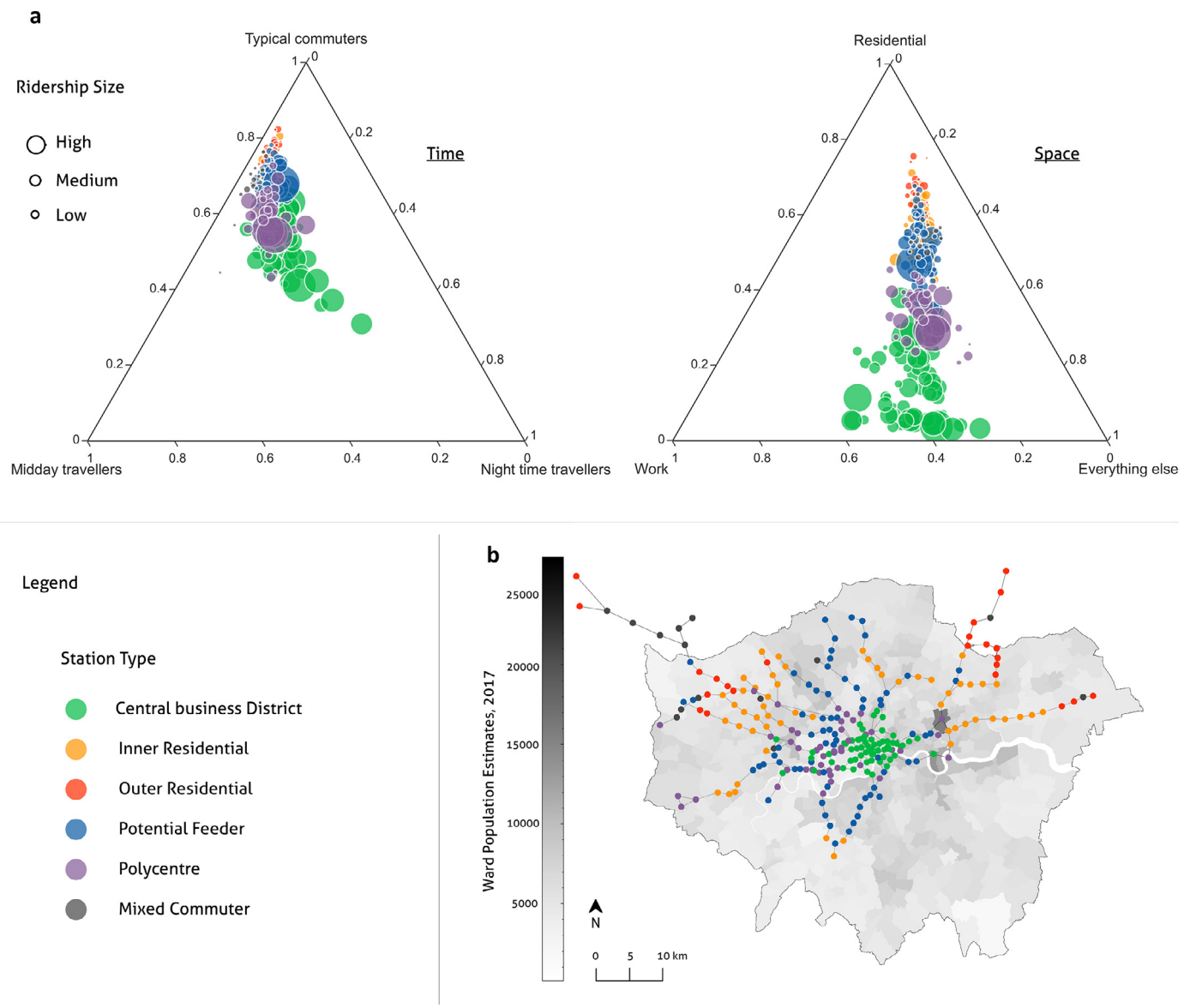

Fig. 4. The spatiotemporal geography of the London public transit system. (a). The two ternary plots show scenarios of time and space where commuting patterns are aggregated by time: typical (W and R), midday (A and EA) and after-work (E and N), and space: work (W), residential (R) and other (A, EA, E and N) **(b).** The station clusters mapped onto the Greater London region with locations as the coordinates of the underground service. The gradient in grey is a population map showing the adult working population of the Wards from a 2017 intercensal estimate.

files than residential (R) in the CBD stations. It is possible that many residents in the CBD region have improved access to bus services and therefore do not use underground transit. Additionally, these residents may be affluent and own a car, live within walking / biking distance of their place of work or are older, non-working citizens. Also, there are more residential and typical stations in the city and these stations have less traffic than the business district. This finding is consistent with the "many-to-few" characterisation of a typical city Roth et al. (2011), where many residential areas feed a small number of polycentres and the CBD.

Visualising the station clusters in space in Fig. 4b, mapped onto the public transport system, reveals how the city is organised in concentric circles as argued and proposed by Burgess *et al.* Burgess (2008). The spatial distribution of clusters describes a complex urban structure wherein public transit links the CBD to outer residential spaces through distinct linkage patterns, revealing a monocentric structure with respect to the primary means of underground transportation. Primarily, the users enter the system from the periphery and advance to the CBD for work or other activities. This indicates that most residential areas are spread out at the outskirts of the city, while the stations close to work centres are clustered together in the CBD. The map thus reveals that there is a dense urban core that is the City of London, surrounded by a residential periphery. This is supported by measuring the average travelling distance between the stations, where we find that the stations in work districts are much closer together than ones in residential districts (see Supplementary Figure 10 for details). Our analysis confirms both concentric Burgess (2008) and polycentric nature Roth et al. (2011) of the city. Polycentres also include some National Rail connections, such as Waterloo, Brixton, and Stratford stations. Though some of these outliers appear as important polycentric hubs for the city (for example, Waterloo), the stations themselves may not have many residents in the vicinity using the system. They are important connections which collect residents coming from other cities via the national train network.

To understand the complex nature of urban flows we disentangled station usage into various clusters (Fig. 3). Next, we individually examine the correlations between ridership and the adult working population

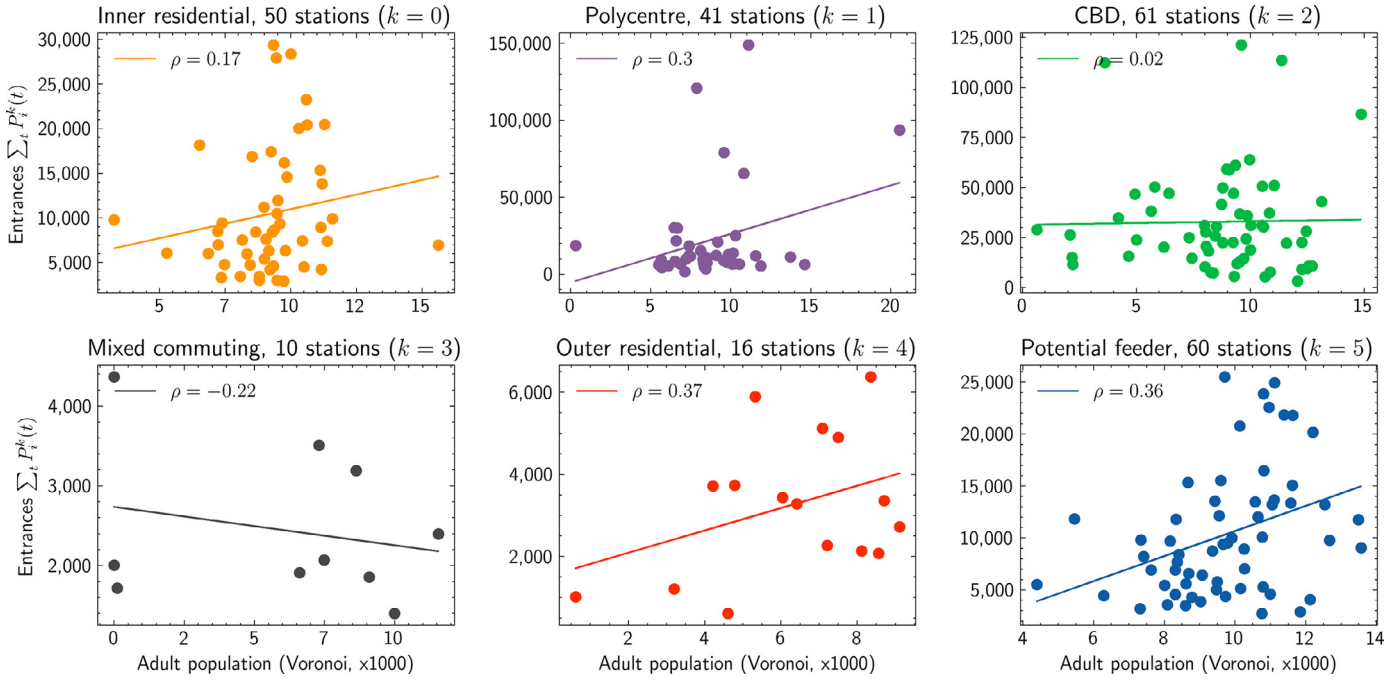

Fig. 5. Cluster specific scatter-plot representation of ridership versus adult working population. Relationship between population and ridership (entrance counts) for Voronoi cells that are attributed to every station (see Supplementary Note 1 for estimating population counts for station zones). The number of stations in this figure are less than the total number in the dataset (238 as opposed to 264) because some stations are outside Greater London for which population estimates were not available.

of each zone separated by station clusters (recall Fig. 1c for this discussion). Fig. 5 illustrates that there is a stronger correlation between the ridership at a station and the population within the station's zone for outer-*residential*, inner-*residential* and polycentric clusters. This is intuitive: the number of people taking the train in the morning from a residential station is proportional to the number of people living within the proximity of that station where other activity is also minimal (see Supplementary Note 1 for a theoretical proposition on accessibility analysis). Polycentres by definition are secondary hubs that attract people owing to their business/services/residential mix. As expected, the relationship for other clusters, the CBD, mixed commuting and feeder stations, is weak and prone to outliers.

In addition to understanding the complex use of public transit in space and time, our method can reveal other nuanced details about areas in the city. For example, we observe White City Station as being commercial, which corresponds to the concentration of several large businesses, including the British Broadcasting Corporation (BBC) offices and the Westfield Mall, the largest mall in Europe. Since our method is normalised for traffic volume, it also reveals areas which, though relatively low in ridership, are heavily business-oriented. For example, there is a strong work-like ridership pattern in Canary Wharf Station, corresponding to the large financial district in the Isle of Dogs area. Thus, our modelling has the potential to analyse mixed-use of stations which has implications for estimating demand and in improving services (see Supplementary Note 3 for a use case explanation).

5. Conclusions

Digitisation has enabled an unprecedented amount of anonymous and location-based transit ridership data that is both complete (representative of the entire population using the service) and easily tractable. Our method leverages such information to extract mobility demand profiles across a city over the course of the day. It is independent of trajectories of individuals, thus preserving their privacy. Upon clustering the station traffic generated using these demand profiles, we are able to extract significant information about urban structures of a city. We have applied this method to the Greater London region using a dataset consisting of ≈ 4.6 million transit traces. The empirical results show three key findings.

First, the aggregate usage of a public transit network can be decomposed into distinct temporal demand profiles that represent various clusters of daily ridership for work, services, leisure and other combinations of use Ahas et al. (2010); Gordon et al. (1986); Lenormand (2015). Second, stations clustered by their traffic patterns suggest that there are concentric zones of development in Greater London, identifying polycentres, entertainment and tourist locations, residential and highly specialised business districts. Third, larger station show mixed-use demand and stations farther away from the centre of the city are likely to exhibit a prevalent residential ridership. While there is evidence of declining populations and traffic from the central business district Tobler (1970), there are rhythms to human activity Smith and Hall (2013) and matching the various demands will lead to efficient transport utilisation across the entire network.

Transportation demand analysis in large metropolitan areas is an important problem that is relevant for urban planning Byrne (2003). Researchers either investigate detailed sociological data Gutiérrez et al. (2011); Raudenbush and Sampson (1999); Taylor et al. (2009) or extract macroscopic urban structures Calabrese et al. (2011); Louail et al. (2015, 2014); Noulas et al. (2013) as proxy for demand patterns. However, as transit needs evolve, intra-urban mobility structures Anas et al. (1998); Barthélemy (2011); Burgess (2008) also transform. Incomplete data Louail et al. (2015) from digital sources such as social media McNeill et al. (2017) prove inaccurate for transportation demand analysis Chico et al. (2019). Our empirical framework highlights an important finding that using digital ridership data we are able to extract a set of complete and accurate microscopic demand profiles which are also determinants of macroscopic urban structures in cities.

Though the data reveals a lot of information about the temporal and spatial profiles of traffic in a city, there are certain limiting factors to consider for future research. The model does not achieve high accuracy for predicting traffic (see Supplementary Note 4 for details). Even

though our work does not focus on the accuracy of prediction, the data-driven method is able to highlight candidate station locations which suffer from poor ridership (both under and over represented), which in turn could be indicative of an unsatisfactory transport service that is a good case for improvement. Thus, the spatiotemporal geography we have presented is an important framework for assessing spatial use of a city with respect to its transportation infrastructure. Larger cities are increasingly interested in providing safe and secure travel options for nighttime workers and preserving or enhancing their nightlife as a cultural amenity and source of economic revenue. A report published by the Greater London Authority London (2018) notes that fully a third of London workers work evening and nights, and two-thirds are actively engaging in nighttime activities. In addition, citizens may have very different expectations about how their districts should be used across time Pinkster and Boterman (2017) and cities are acknowledging those needs London (2018). Our results indeed confirm multiple uses of space over time and highlight the very specific districts where different kinds of activities occur, or might be enhanced with appropriate intervention. Our work can be used as a methodology for analysing and repurposing transport data for studies of cohesion, safety and growth.

Without requiring detailed origin-destination trajectories, which has become a common tool in urban studies, important information about urban structures could very well prompt studies in the direction of sustainable transit-oriented development Papa and Bertolini (2015); Zhang et al. (2019) and their potential negative impacts on displacing communities Dawkins and Moeckel (2016), promoting urbanisation Kasraian et al. (2019) and exacerbating sprawl Bertaud and Malpezzi (2003); Dieleman and Wegener (2004). Our quantitative analysis of transportation demand has the potential to initiate new developments in extracting precise micro-scale OD matrices useful for urban planning on a city level by studying distributions of amenities around generalised station clusters (see Supplementary Note 5 for details on a simple analysis of amenities' distribution). Since our method is generative, given mixed use of space, new stations can be planned or old ones re-designed, to match traffic demand for new technology hubs, social housing neighbourhoods or secondary or tertiary centres of tourism. Our method can be straightforwardly expanded to other transit networks, including multi-modal systems, and can therefore become a critical tool for urban and transit planners.

Data Availability

All datasets that support the findings of this study are publicly available (as cited in the references) and a version used in the study can be collected, requested or directly downloaded from the following link: https://github.com/mikhailsirenko/spacetimegeo

Code Availability

The GMM representation together with all the necessary functions for running the model are available in python at the following link on github: https://github.com/mikhailsirenko/spacetimegeo

Author contributions

All authors designed the study. TV, IK, MS, SC and NA evaluated the data. TV, MS and SC developed the model. All authors analyzed the results and wrote the manuscript.

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.

Supplementary material

Supplementary material associated with this article can be found, in the online version, at 10.1016/j.urbmob.2021.100004

References

Ahas, R., Aasa, A., Silm, S., & Tiru, M. (2010). Daily rhythms of suburban commuters' movements in the tallinn metropolitan area: Case study with mobile positioning data. *Transportation Research Part C: Emerging Technologies, 18*(1), 45–54. February

Alonso, W. (1960). A theory of the urban land market. *Papers in Regional Science, 6*(1), 149–157.

Alonso, W. (1964). The historic and the structural theories of urban form: Their implications for urban renewal. *Land economics, 40*(2), 227–231. Publisher: [Board of Regents of the University of Wisconsin System, University of Wisconsin Press]

Anas, A., Arnott, R., & Small, K. A. (1998). Urban spatial structure. publisher: american economic association. *Journal of economic literature, 36*(3), 1426–1464.

Barthélemy, M. (2011). Spatial networks. *Physics reports, 499*(1–3), 1–101. ArXiv: 1010.0302 Publisher: Elsevier B.V. ISBN: 0370–1573

Batty, M. (1976). *Urban modelling. algorithms, calibrations, predictions*. Cambridge University Press.

Beaverstock, J. V., & Smith, J. (1996). Lending jobs to global cities: Skilled international labour migration, investment banking and the city of london. *Urban Studies, 33*(8), 1377–1394. October

Bertaud, A., & Malpezzi, S. (2003). The spatial distribution of population in 48 world cities: Implications for economies in transition. *Center for Urban Land Economics Research, University of Wisconsin, 32*(1), 54–55.

Bettencourt, L. M. A., Lobo, J., Helbing, D., Kühnert, C., & West, G. B. (2007). Growth, innovation, scaling, and the pace of life in cities. *Proceedings of the National Academy of Sciences, 104*(17), 7301–7306. Publisher: National Academy of Sciences Section: Social Sciences

Briand, A., Côme, E., Mahrsi, M. K. E., & Oukhellou, L. (2015). A mixture model clustering approach for temporal passenger pattern characterization in public transport. In *2015 IEEE International Conference on Data Science and Advanced Analytics (DSAA)* (pp. 1–10). October

Briand, A.-S., Côme, E., Trépanier, M., & Oukhellou, L. (2017). Analyzing year-to-year changes in public transport passenger behaviour using smart card data. *Transportation Research Part C: Emerging Technologies, 79*, 274–289. June

Burgess, E. W. (2008). The growth of the city: An introduction to a research project. In J. M. Marzluff, E. Shulenberger, W. Endlicher, M. Alberti, G. Bradley, C. Ryan, … C. ZumBrunnen (Eds.), *Urban Ecology: An International Perspective on the Interaction Between Humans and Nature* (pp. 71–78). Boston, MA: Springer US.

Byrne, D. (2003). Complexity theory and planning theory: A necessary encounter. *Planning Theory, 2*(3), 171–178. November

Calabrese, F., Lorenzo, G. D., Liu, L., & Ratti, C. (2011). Estimating origin-destination flows using mobile phone location data. *IEEE Pervasive Computing, 10*(4), 36–44. April

Calthorpe, P. (1993). *The next american metropolis: Ecology, community, and the american dream*. Princeton Architectural Press.

Ceder, A. (2016). Public transit planning and operation: modeling, practice and behavior. *Second edition*. CRC Press. March

Cervero, R., & Kockelman, K. (1997). Travel demand and the 3Ds: Density, diversity, and design. *Transportation Research Part D: Transport and Environment, 2*(3), 199–219. September

Chico, Q., Camargo, Bright, J., & Hale, S. A. (2019). Diagnosing the performance of human mobility models at small spatial scales using volunteered geographic information. *arXiv:1905.07964 [physics]*. ArXiv: 1905.07964

Darling, J. (2017). Forced migration and the city: Irregularity, informality, and the politics of presence. *Progress in human geography, 41*(2), 178–198. April

Dawkins, C., & Moeckel, R. (2016). Transit-induced gentrification: Who will stay, and who will go? *Housing policy debate, 26*(4–5), 801–818. 10.1080/10511482.2016.1138986. Publisher: Routledge _eprint:

De Domenico, M., Solé-Ribalta, A., Gómez, S., & Alex, A. (2014). Navigability of interconnected networks under random failures. *Proceedings of the National Academy of Sciences, 111*(23), 8351–8356. June

Dempster, A. P., Laird, N. M., & Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society. Series B (Methodological), 39*(1), 1–38. Publisher: [Royal Statistical Society, Wiley]

Dieleman, F., & Wegener, M. (2004). Compact city and urban sprawl. *Built Environment (1978-), 30*(4), 308–323.

Dittmar, H., & Ohland, G. (2012). *The new transit town: Best practices in transit-oriented development*. Island Press. June

Duan, Z., Lei, Z., Zhang, M., Li, H., & Yang, D. (2018). Understanding multiple days' metro travel demand at aggregate level. *IET Intelligent Transport Systems, 13*(5), 756–763. December, Publisher: IET Digital Library

Eirola, E., & Lendasse, A. (2013). Gaussian mixture models for time series modelling, forecasting, and interpolation. In A. Tucker, F. Höppner, A. Siebes, & S. Swift (Eds.), *Advances in Intelligent Data Analysis XII, lecture Notes in Computer Science* (pp. 162–173). Berlin, Heidelberg: Springer.

Flyvbjerg, B., Holm, M. K. S., & Buhl, S. L. (2005). How (in)accurate are demand forecasts in public works projects?: The case of transportation. *Journal of the American Planning Association, 71*(2), 131–146. 10.1080/01944360508976688. Publisher: Routledge _eprint:

Gan, Z., Yang, M., Feng, T., & Timmermans, H. (2020). Understanding urban mobility patterns from a spatiotemporal perspective: Daily ridership profiles of metro stations. *Transportation, 47*(1), 315–336. February

González, M. C., Hidalgo, C. A., & Barabási, A.-L. (2008). Understanding individual human mobility patterns. *Nature, 453*(7196), 779–782. June

Gordon, P., Richardson, H. W., & Wong, H. L. (1986). The distribution of population and employment in a polycentric city: The case of los angeles. *Environment and Planning A: Economy and Space, 18*(2), 161–173. February

Guerra, E., Cervero, R., & Tischler, D. (2012). Half-mile circle: Does it best represent transit station catchments? *Transportation research record, 2276*(1), 101–109. Publisher: SAGE Publications Inc

Gutiérrez, J., Cardozo, O. D., & García-Palomares, J. C. (2011). Transit ridership forecasting at station level: An approach based on distance-decay weighted regression. *Journal of transport geography, 19*(6), 1081–1092. November

He, Y., Zhao, Y., & Tsui, K. L. (2020). An adapted geographically weighted LASSO (ada-GWL) model for predicting subway ridership. *Transportation*. February

Henig, J. R. (1980). Gentrification and displacement within cities: a comparitive analysis. *Social science quarterly, 61*(3/4), 638–652.

Hoyt, H. (1939). The structure and growth of residential neighborhoods in american cities. *U.s. government printing office*. Google-Books-ID: VtjZdGSOWhgC

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. *Journal of transport geography, 48*, 30–40. October

Kasraian, D., Maat, K., & van Wee, B. (2019). The impact of urban proximity, transport accessibility and policy on urban growth: A longitudinal analysis over five decades. *Environment and Planning B: Urban Analytics and City Science, 46*(6), 1000–1017. July

Kim, M.-K., Kim, S.-P., Heo, J., & Sohn, H.-G. (2017). Ridership patterns at subway stations of seoul capital area and characteristics of station influence area. *KSCE Journal of Civil Engineering, 21*(3), 964–975. March

Lenormand, M., Picornell, M., CantúRos, O. G., Louail, T., Herranz, R., Barthelemy, M., Frías-Martínez, E., Miguel, M. S., & Ramasco, J. J. (2015). Comparing and modelling land use organization in cities. *Royal Society open science, 2*(12), 150449. December

London (b). Underground passenger counts data.

London (2018). at night - an evidence base for a 24-hour city. Library Catalog: www.london.gov.uk.

Long, Y., & Thill, J.-C. (2015). Combining smart card data and household travel survey to analyze jobs-housing relationships in beijing. *Computers, environment and urban systems, 53*, 19–35. September

Louail, T., Lenormand, M., Picornell, M., Cantu, O. G., Herranz, R., Frias-Martinez, E., Ramasco, J. J., & Barthelemy, M. (2015). Uncovering the spatial structure of mobility networks. *Nature communications, 6*. ArXiv: 1501.05269 ISBN: 2041-1723 (Electronic)\r2041-1723 (Linking)

Louail, T., Lenormand, M., Ros, O. G. C., Picornell, M., Herranz, R., Frias-Martinez, E., Ramasco, J. J., & Barthelemy, M. (2014). From mobile phone data to the spatial structure of cities. *Scientific reports, 4*(1), 5276. ArXiv: 1401.4540v1 Publisher: Nature Publishing Group

Louf, R., Jensen, P., & Barthelemy, M. (2013). Emergence of hierarchy in cost-driven growth of spatial networks. *P Natl Acad Sci USA, 110*, 8824–8829.

Mahrsi, M. K. E., Côme, E., Oukhellou, L., & Verleysen, M. (2017). Clustering smart card data for urban mobility analysis. *IEEE Transactions on Intelligent Transportation Systems, 18*(3), 712–728. Conference Name: IEEE Transactions on Intelligent Transportation Systems

McNeill, G., Bright, J., & Hale, S. A. (2017). Estimating local commuting patterns from geolocated twitter data. *EPJ Data Science, 6*(1), 1–16. December

Munizaga, M. A., & Palma, C. (2012). Estimation of a disaggregate multimodal public transport origin-destination matrix from passive smartcard data from santiago, chile. *Transportation Research Part C: Emerging Technologies, 24*, 9–18. October

Murphy, R. E. (2017). *The central business district: A study in urban geography*. Routledge. Google-Books-ID: fDwrDwAAQBAJ

Noulas, A., Mascolo, C., & Frias-Martinez, E. (2013). Exploiting foursquare and cellular data to infer user activity in urban environments. In *2013 IEEE 14th International Conference on Mobile Data Management, volume 1* (pp. 167–176). June

Papa, E., & Bertolini, L. (2015). Accessibility and transit-oriented development in european metropolitan areas. *Journal of transport geography, 47*, 70–83. July

Park, J., Kim, D.-J., & Lim, Y. (2008). Use of smart card data to define public transit use in seoul, south korea. *Transportation Research Record: Journal of the Transportation Research Board, 2063*(1), 3–9. Publisher: SAGE PublicationsSage CA: Los Angeles, CA ISBN: 0361–1981

Pavlyuk, D., Spiridovska, N., & (Jackiva), I. Y. (2020). Spatiotemporal dynamics of public transport demand: Acase study of riga. *Transport, 35*(6), 576–587. Number: 6

Pelletier, M.-P., Trépanier, M., & Morency, C. (2011). Smart card data use in public transit: A literature review. *Transportation Research Part C: Emerging Technologies, 19*(4), 557–568. August

Pinkster, F. M., & Boterman, W. R. (2017). When the spell is broken: Gentrification, urban tourism and privileged discontent in the amsterdam canal district. *cultural geographies, 24*(3), 457–472. July, Publisher: SAGE Publications Ltd

Quigley, J. M. (1998). Urban diversity and economic growth. *Journal of Economic Perspectives, 12*(2), 127–138. June

Raudenbush, S. W., & Sampson, R. J. (1999). Ecometrics: Toward a science of assessing ecological settings, with application to the systematic social observation of neighborhoods. *Sociological methodology, 29*, 1–41. August, Publisher: John Wiley & Sons, Ltd (10.1111)

Rodrigue, J.-P., Comtois, C., & Slack, B. (2013). The geography of transport systems. *third edition edition*. London; New York: Routledge.

Roth, C., Kang, S. M., Batty, M., & Barthélemy, M. (2011). Structure of urban movements: Polycentric activity and entangled hierarchical flows. *PloS one, 6*(1), e15923.

Saberi, M., Mahmassani, H. S., Brockmann, D., & Hosseini, A. (2017). A complex network perspective for characterizing urban travel demand patterns: Graph theoretical analysis of large-scale origin-destination demand networks. *Transportation, 44*(6), 1383–1402. November

Shang, P., Li, R., Liu, Z., Yang, L., & Wang, Y. (2018). Equity-oriented skip-stopping schedule optimization in an oversaturated urban rail transit network. *Transportation Research Part C: Emerging Technologies, 89*, 321–343. April

Smith, R. J., & Hall, T. (2013). No time out: Mobility, rhythmicity and urban patrol in the twenty-four hour city:. *The Sociological review*. June

Sohn, K., & Shim, H. (2010). Factors generating boardings at metro stations in the seoul metropolitan area. *Cities (London, England), 27*(5), 358–368. October

Steinley, D., & Brusco, M. J. (2011). Evaluating mixture modeling for clustering: recommendations and cautions. *Psychological methods, 16*(1), 63–79. March

Taylor, B. D., Miller, D., Iseki, H., & Fink, C. (2009). Nature and/or nurture? analyzing the determinants of transit ridership across US urbanized areas. *Transportation Research Part A: Policy and Practice, 43*(1), 60–77. January

Teodorović, D., & Janić, M. (2017). Chapter 8- transportation demand analysis. In D. Teodorović, & M. Janić (Eds.), *Transportation Engineering* (pp. 495–568). Butterworth-Heinemann. January

Terroso-Sáenz, F., Muñoz, A., & Arcas, F. (2021). Land use dynamic discovery based on heterogeneous mobility sources. *International Journal of Intelligent Systems*.

Thompson, G., Brown, J., & Bhattacharya, T. (2012). What really matters for increasing transit ridership: Understanding the determinants of transit ridership demand in broward county, florida. *Urban Studies, 49*(15), 3327–3345. Publisher: SAGE Publications Ltd

Tobler, W. R. (1970). A computer movie simulating urban growth in the detroit region. *Economic geography, 46*, 234–240. June

Tomer, A., Kneebone, E., Puentes, R., & Berube, A. (2011). Missed opportunity: Transit and jobs in metropolitan america. *Technical report*. May

Zhai, W., Bai, X., Shi, Y., Han, Y., Peng, Z.-R., & Gu, C. (2019). Beyond word2vec: An approach for urban functional region extraction and identification by combining place2vec and POIs. *Computers, environment and urban systems, 74*, 1–12. March

Zhang, X., Li, W., Zhang, F., Liu, R., & Du, Z. (2018). Identifying urban functional zones using public bicycle rental records and point-of-interest data. *ISPRS international journal of geo-information, 7*(12), 459. Number: 12 Publisher: Multidisciplinary Digital Publishing Institute

Zhang, Y., Marshall, S., & Manley, E. (2019). Network criticality and the node-place-design model: classifying metro station areas in greater london. *Journal of transport geography, 79*, 102485. July

Zhong, C., Arisona, S. M., Huang, X., Batty, M., & Schmitt, G. (2014). Detecting the dynamics of urban structure through spatial network analysis. *International Journal of Geographical Information Science, 28*(11), 2178–2199.