



Article Extracting the Relationship and Evolutionary Rule Connecting Residents' Travel Demand and Traffic Supply Using Multisource Data

Zijia Wang ^{1,2}, Zhixiang Chen ¹, Youyin Shi ³ and Liping Huang ^{4,*}

- ¹ Department of Civil Engineering, Beijing Jiaotong University, Beijing 100044, China; zjwang@bjtu.edu.cn (Z.W.); 20121169@bjtu.edu.cn (Z.C.)
- ² Engineering and Technology Research Center of Rail Transit Line Safety and Disaster Prevention, No. 3 Shangyuancun, Haidian District, Beijing 100044, China
- ³ Department of Civil Engineering, Central South University, Changsha 410000, China; youyinshi@csu.edu.cn
- ⁴ School of Electrical and Electronic Engineering, Nanyang Technological University, 50 Nanyang Avenue, Singapore 639798, Singapore
- * Correspondence: liping.huang@ntu.edu.sg

Abstract: Urban rail transit (URT) systems are often regarded as the backbone of their respective city. The evolutionary features of URT systems have attracted much attention in recent years, but their evolution and their distinct function in contrast to other transit modes have seldom been investigated, especially quantitatively from the perspective of work-residence separation. Accordingly, we propose a framework for exploring the evolution of URT topological networks and demand-weighted networks, comparing the different impacts of all transit modes on work-residence separation. In this study, a URT passenger flow assignment model was formulated on the basis of travel cost function and an improved logit model was proposed that takes into account the heterogeneity of passengers. This model was used to generate a section load, which is regarded as a weight and able to reflect the residents' demand for travel by URT. Then, the fractal dimensions for a non-weighted network and demand-weighted network are proposed and their indications for transportation explained. Finally, the Beijing Subway System (BSS) is used as a case study by employing fifty years of network data and ten years of smart card data. Using fractal approaches, the different characteristics illustrated by the two networks were investigated and the reasons behind the observed patterns explained. In addition, the spatial features of the rail network, in terms of fractal indictors, were compared with population distribution and urban mobility for all modes, extracted from phone data as a proxy. Thus, the relationship between the residents' travel demand and traffic supply can be revealed to some extent. The main finding of this work is that demand must be taken into account when analyzing the fractal features of a transport network, lest the demand side be separated from the supply and important issues missed such as inconsistencies between demand and supply. Additionally, the role of rail transit in work-home imbalance can be investigated in the context of urban mobility for an entire city.

Keywords: urban rail transit; fractal approaches; improved traffic assignment model; spatiotemporal evolution; travel demand and traffic supply

1. Introduction

Public transportation plays an important and indeed indispensable role in daily commuting in almost all cities [1–3]. Due to the saturation of ground transport, urban rail transit (URT) systems are preferred by many metropolises to reduce traffic congestion because of their high speed, large capacity, energy efficiency, and use of urban non-surface space [4–6]. Moreover, URT systems can promote the development of a city and even reshape urban structure such as by enhancing or mitigating work–residence separation. The topological network of URT systems and their coverage can be seen as the transportation infrastructure supply from an urban planner's point of view, with the network traffic flow of URT systems



Citation: Wang, Z.; Chen, Z.; Shi, Y.; Huang, L. Extracting the Relationship and Evolutionary Rule Connecting Residents' Travel Demand and Traffic Supply Using Multisource Data. *Sensors* 2021, *21*, 2179. https:// doi.org/10.3390/s21062179

Academic Editor: Kyandoghere Kyamakya

Received: 13 February 2021 Accepted: 18 March 2021 Published: 20 March 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https:// creativecommons.org/licenses/by/ 4.0/). representing the demand side [7–9]. Increasing traffic demand drives the extension of the URT system, with the new rail lines generating still further demand. The different evolutionary features of topological networks and demand-weighted networks (topological networks loaded by station and section demand) can reflect the supply-demand matching level and the different stages of city growth. However, on a complex rail network, section load cannot be easily retrieved because many Origin-Destination (OD) pairs have more than one reasonable path, and different passengers may prefer different paths. To obtain a relatively accurate section demand between two adjacent stations, a traffic assignment model, often some type of multinominal logit model (MLM), is consistently adopted in the literature, however, passenger heterogeneity is seldom considered. Furthermore, the parameters used in the path choice model are normally calibrated against stated preferences derived from questionnaire data, which might not conform well to real life. Thus, we used Bayesian inference as a more reliable alternative to the traditional questionnaire survey method, with an Markov chain Monte Carlo (MCMC) algorithm used to improve calculation efficiency. In terms of estimated sectional passenger flow, URT development level has been well analyzed using fractal approaches.

There are several methods for evaluating the evolutionary features of a URT. As the fractal approach is an effective way of describing fragmented anthropogenic objects and obtaining scale-independent results [10,11], we used it to assess the evolutionary features of a non-weighted URT network and a transit-weighted network. Fractal concepts were proposed by Mandelbrot [10], and several definitions of fractal dimensions have been introduced to describe fractality in a quantified way, with application to an urban transport network following nearly a decade later. Benguigui [12] first identified the fractal features of an urban transport network and carried out case studies on the London Underground, Paris Metro, Moscow Metro, and Rhine Subway (Germany). Batty [13,14] linked transport analysis to urban geography on the basis of fractal dimensions, after which fractal approaches have been widely applied to different parts of cities including the city center [15], the whole scope of a city [16], and conurbations [11]. Different modes of transport have also been analyzed using fractal methods including traditional surface transport [15], subway [16], and the whole public transport network. Several urban issues are discussed in the research in relation to transport networks such as the complexity and space filling capacity of urban transport networks [17], city morphology [18], and the consistency of transportation and the built-up environment [19]. As this study goes beyond these considerations, with fractal characteristics unable to be properly described using only one fractal dimension, multifractals [19–21] were seen as an extension for complicated fractal structures. The existing literature is devoted chiefly to the structure of a transport network without demand on it, with no research addressing the fractal features of a transit network for its entire life from both a topological network perspective and a demand-weighted network perspective. In addition, fractal comparison has rarely been made of an all-mode transit network with respect to work-residence separation within a city. This study focused on the different evolutionary fractal features of a rail topological network and a rail demand-weighted network. Various conclusions have been suggested relating to the dynamic interactions between rail network development and urban expansion. Moreover, the fractal dimensions of URT mode and all modes in Beijing have been calculated to further explore the work-residence separation problem, which has also seldom been addressed in the literature.

Then, a precise and reasonable traffic assignment model is used as the basis for comprehensive analysis of fractal features. Traffic assignment is first introduced to the urban surface network [22]. However, as URT use has increased, traffic assignment of subway systems has drawn increasing attention. This study discusses improvements to passengers' generalized travel cost function, probability determination for path choices, and calculation efficiency. To capture the passengers' more complicated path choice behavior in the real world, various stochastic factors and uncertainties have been introduced such as passengers' preferences [23], crowding or service level [24], and transport network

familiarity. In recent years, research concentrating on the generalized cost function has turned to passengers' heterogeneity, with several assignment models proposed based on passenger clustering [25,26]. However, the socioeconomic attributes of passengers are hard to access, constraining further investigation of the stochastic assignment method. Fortunately, thanks to the widespread use of automatic fare collection (AFC) systems in URT systems, abundant travel information recorded in smart card data (SCD) can be collected. Furthermore, passengers' spatial and temporal attributes can be derived by mining SCD. Accordingly, an improved traffic assignment model can be devised. Thus, topic model was chosen to distinguish the passengers' heterogeneity, with K-means used to acquire passenger clustering for calculation of the path-chosen probability. As dynamic traffic assignment for the URT system requires high levels of accuracy and efficiency, an effective machine mining method, Bayesian inference [27,28], was adopted to calibrate the parameters of the utility function. Moreover, the passenger attributes already discussed could be used as the input data to differentiate parameter calibration for each passenger clustering. When assessing data of large scale and computational complexity, the MCMC algorithm is selected for its efficiency and accuracy [29].

A public transportation system, especially in big cities, always contains multiple modes. Thus, a single trip could comprise several modes of transport trip-leg. In this study, we further investigated the ways in which URT affects city structure and the different roles that URT systems play in thee spatial imbalance of workplaces and residences, compared with the overall mobility of a whole city. Distinct characteristics of different modes determine the attraction of different passengers to each mode and also reveal or even affect the city's urban structure in different ways. However, because quantitative measures are rarely devoted to investigating this issue, it should be enlightening to compare the travel features of rail riders using the fractal dimensions of a demand-weighted network, in contrast to that of urban mobility as a whole. To this end, mobile phone base station data were used as a proxy for population distribution and whole-city mobility, serving as a benchmark against which to reveal the distinct features of the rail transit network and its effects on urban development. The main contributions of this study are as follows: (i) An improved traffic assignment model that takes into account the spatiotemporal characteristics of passengers is introduced. (ii) A fifty-year time span fractal analysis of a rail topological network and a ten-year time span fractal analysis of a demand-weighted rail network are carried out, revealing long-term evaluation patterns of a rail network with or without passenger load. (iii) A comprehensive investigation of fractal features for rail transit in Beijing from the perspective of work-residence separation is provided, allowing assessment of the dynamic interaction between transportation infrastructure supply and urban growth.

The remainder of this paper is organized as follows. In Section 2, the proposed method is formulated. In Section 3, the materials and methods for this study are described, a case study presented, and results and findings analyzed and discussed. Conclusions and future work are discussed in Section 4.

2. Materials and Methods

In this section, the AFC system and mobile phone data are briefly introduced, and the fractal dimensions are presented in detail. In particular, the URT passenger flow assignment model was formulated to classify URT passengers, revealing thee residents' travel demand by URT and pointing to the evolutionary rule of BBS.

2.1. Automatic Fare Collection (AFC) Systems

The automatic fare collection (AFC) systems are an important part of URT, which can not only provide passengers with convenient, humanized services such as a quick and easy ticket checking process, but also providing operators with an automatic management platform containing ticket production, ticket sales, ticket inspection, financial, statistical analysis, audit and other functions. As displayed in Figure 1, a typical AFC system consists of five hierarchical levels: AFC cleaning center (ACC), line computer (LC), station computer (SC), station level equipment (SLE), and smart cards and tickets.



Figure 1. Five hierarchical levels of automatic fare collection (AFC) system.

The smart ticket and card contain an integrated circuit (IC) clip (a type of microsensor) inside, so if a passenger touches a smart ticket or card to a turnstile when boarding or alighting, the sensor in the turnstile will record and respond some necessary information such as card code, boarding/alighting station, boarding/alighting time, and some other useful information. Then, this information will be transmitted to the SC, LC, and finally to the CC. Meanwhile, the AFC data do not contain the passenger's private information such as passenger's name, age, occupation, etc. Therefore, this does not invade people's privacy.

The AFC system can not only be employed by operators to calculate the fares from passengers directly, but the massive data collected by AFC system can also help researchers to detect the urban mobility structure and travel demand of residents.

2.2. Mobile Phone Data

The mobile phone data applied in this paper contained two types of datasets. The first, which presents ODs between different geographic coordinate points at different times for a single day, contains detailed information about the individual passengers' trips including time, latitude and longitude of origin, latitude and longitude of destination, and ODs. The second presents the numbers of people per geographic points at different times on a single day including accurate time, latitude and longitude of geographic coordinate points, and numbers of people.

2.3. Fractal Dimensions

2.3.1. Volume Dimension

Along with economic and urban development, urban mobility demand fluctuates around a certain value in a period [30]. In a circular area of radius r, load level in a topological network and in a demand-weighted network can be defined, respectively, by

$$S(r) = C_S * r^{D_S} \tag{1}$$

$$V(r) = C_V * r^{D_V} \tag{2}$$

where S(r) and V(r) represent the number of stations and their total demand, respectively, within the area of radius r; C_S and C_V are constant coefficients; D_S and D_V are the station dimension and volume dimension using the fractural method, respectively.

The volume dimension D_V depicts the supply level change from the city center to the surrounding areas. Using the derivative transform, the spatial attenuation formula for demand-weighted network volume is obtained by

$$\rho(r) \propto r^{D_V - d} \tag{3}$$

where *d* is the Euclidean dimension d = 2. If $D_V < 2$, the density distribution of transportation supply decreases from the city center to the suburbs, with capacity decreasing gradually. If $D_V = 2$, the distribution of capacity is homogeneous. If $D_V > 2$, the network has a heterogeneous density distribution, which is common in multicenter cities.

For a detailed reflection of demand density distribution, inbound passenger flow and outbound passenger flow are used to define the inflow volume dimension D_V^{in} and outflow volume dimension D_V^{out} , respectively,

$$V^{in}(r) = C_V^{in} * r^{D_V^{in}} \tag{4}$$

$$V^{out}(r) = C_V^{out} * r^{D_V^{out}}$$
⁽⁵⁾

where $V^{in}(r)$ and $V^{out}(r)$ are the cumulative flow volume in each circular area with radius r of inflow and outflow, respectively.

2.3.2. Traffic Impedance Dimension

The sectional passenger flow and interstation distance can reflect the efficiency and accessibility of a transport network. The distance d_{ij} between two stations *i* and *j* is denoted as the real distance between two stations. The traffic impedance dimension in terms of interstation distance D_I^d is defined in Equation (6), and the impedance dimension represented by sectional flow D_I^f is defined in Equation (7),

$$\mathbf{f}^d(\mathbf{r}) = C_I^d * \mathbf{r}^{D_I^d} \tag{6}$$

$$I^f(r) = C^f_I * r^{D^f_I} \tag{7}$$

where $I^{d}(r)$ is the cumulative interstation distance in each circular area with radius *r* and $I^{f}(r)$ is the cumulative sectional passenger flow in each circular area of radius *r*.

Rail lines are normally bidirectional: upstream and downstream, both predefined. For detailed description of flow direction, upstream passenger flow and downstream passenger flow are used to define the upstream volume dimension and the downstream volume dimension, respectively,

$$V^{up}(r) = C_V^{up} * r^{D_V^{up}}$$
(8)

$$V^{down}(r) = C_V^{down} * r^{D_V^{down}}$$
⁽⁹⁾

where $V^{up}(r)$ and $V^{down}(r)$ is the cumulative flow volume in each circular area with radius r of upstream and downstream, respectively.

2.3.3. Branch Dimension

Volume dimension D_V represents the URT network's capacity and the traffic impedance dimension's accessibility. The connection relationship is fully represented by these two dimensions' indicators, but because the development level of the network structure itself is not captured, the branch dimension is introduced to describe the complexity of the network structure. The subway lines can be divided into branches, with branch dimension D_B defined as

$$B(r) = C_B * r^{D_B} \tag{10}$$

where B(r) is the number of branches in the circular area described by radius r. The larger B(r), the more branches of lines and the better the URT network supply.

2.3.4. Fractal Dimension Consistency Index

Passenger flow and the topological network interact in a URT system. To examine the interrelation of topological structure and passenger flow, the fractal dimension consistency index γ is chosen for quantitative analysis [31],

$$\gamma = \begin{cases} 1 - \frac{|D_S - D_V^{in}|}{D_V^{in}}, & D_S \le 2D_V^{in} \\ 0, & D_S > 2D_V^{in} \end{cases}$$
(11)

where D_S is the station dimension and D_V^{in} is the inflow volume dimension. The larger the value of γ , the greater the consistency between the network supply and the demand of passenger flow.

2.4. Subway Passenger Flow Assignment

To acquire reliable sectional passenger flow data, an effective dynamic traffic assignment method is proposed that takes into account the passengers' heterogeneity.

2.4.1. Passenger Clustering

The temporal and spatial characteristics of passengers are indicators of passenger heterogeneity and can be considered for passenger clustering. The temporal features are represented by the travel days within a period for as long as possible considering data accessibility—at least a month. Regular passengers can be selected by trip frequency. The spatial features are characterized by spatial consistency based on the diversity of the station visited. The proportion of the most frequent origins for each passenger can be calculated by

$$PR = \frac{m_{i,\max}}{\sum_{i=1}^{n} m_i} \tag{12}$$

where $m_{i,\max}$ is the travel time of the most frequently visiting origin station; n is the total travel time of all origin stations for one passenger; and m_i is the travel time for origin station O_i . We took passengers with $PR > \alpha(\alpha > 0)$ as regular travel passengers for the following study.

For the regular travel passengers, a topic model and *K*-means were adopted for further clustering. Like the irregular passengers, they were classified as one clustering, with application of the clustering method not needed. First, the topic model was used to calculate the probability distribution of travel regularity for each passenger. In this study, each passenger can be regarded as a document containing a large number of words, which reflect the characteristics of the passengers' time of station entry such as Friday at 10 a.m. The number of trips of passengers at different times indicates the corresponding topic,

representing the probability of passengers' preferring to travel at different times, whether morning, midday, or evening. The formulas for the topic model are described as

$$z \sim M(1,\pi) \tag{13}$$

$$u|z \sim M(D, \beta_z) \tag{14}$$

$$p(u) = \sum_{z} p(z) \prod_{d=1}^{D} p(u_d | z)$$
(15)

where *z* denotes the topic; π the proportion of topic distribution; *M* the polynomial distribution; *D* total travel frequency for one topic; and β weekly travel regularity distribution for one topic; Equation (15) expresses the Expectation-Maximum (EM) algorithm and allows calculation of π , β .

Based on the topic model, the probability distributions of each passenger for different topics can be calculated, reflecting the probability distributions that correspond to each passenger's various travel time patterns. Then, the probability distribution of different travel patterns of each passenger is taken as its eigenvalue, with the *K*-means algorithm used to cluster passengers.

2.4.2. Passenger Travel Generalized Cost Function

Passenger path choices are affected by various factors, so that full consideration of impact factors, and reasonable quantification of them, is essential to the path choice model. We chose traveling time cost, interchange time cost, URT network familiarity, and degree of congestion as the key factors. For one OD pair *rs*, for a passenger in class *n*, the set of possible paths is K_n^{rs} , with the total cost of path *k* expressed as

$$D_{k,n}^{rs} = C_{k,n}^{rs} + \varepsilon_{k,n}^{rs} = \gamma \sum_{l} \overline{T}_{rs,k}^{l} + \sum_{i} \overline{E}_{i,n}^{l,m} + \varepsilon_{k,n}^{rs} = \gamma \sum_{l} \sum T_{i,j}^{l} * \delta_{ij,k}^{rs} * (1 + Y(x_{w})) + \sum_{i} \alpha_{n} * (e_{i,k}^{rs})^{\beta_{n}} * (w_{i,n}^{l,m} + 0.5 * f_{m}) * \varphi_{i,k}^{rs} * \eta_{l,k}^{rs} * \eta_{m,k}^{rs} + \varepsilon_{k,n}^{rs}$$
(16)

where $C_{k,n}^{rs}$ is the quantitative cost, which consists of in-vehicle time $\sum_{l} \overline{T}_{rs,k}^{l}$ and interchange time $\sum_{i} \overline{E}_{i,n}^{l,m}$; $\varepsilon_{k,n}^{rs}$ is the stochastic cost error; and interchange time $\sum_{i} \overline{E}_{i,n}^{l,m}$ comprises walking time $w_{i,n}^{l,m}$ and waiting time $0.5 * f_m$ (f_m is the headway, and assuming most of the service are regular). Walking time is obtained through field research, which is affected by station type, differences in transfer mode, transfer distance, and transfer height, whereas waiting time is calculated by headway, as obtained from the Beijing subway map. $e_{i,k}^{rs}$ is the accumulative interchange time of OD pair rs for path k at station i; α_n and β_n are the parameters needing estimation; $\delta_{ij,k}^{rs}$, $\eta_{l,k}^{rs}$, $\eta_{m,k}^{rs}$ represent the affiliation relationship between sections or interchange stations or lines and path k, with a value of 1 indicating that the foregoing segments belong to path k; and $Y(x_w)$ denotes the degree of congestion, producing the expression

$$Y(x_{w}) = \begin{cases} 0, & x_{w} \leq p_{n} \\ A\left(\frac{x_{w}-p_{n}}{p_{n}}\right)^{\eta}, & p_{n} < x_{w} < p_{c} \\ A\left(\frac{p_{c}-p_{n}}{p_{n}}\right)^{\eta} + B\left(\frac{x_{w}-p_{c}}{p_{n}}\right)^{\psi}, & x_{w} \geq p_{c} \end{cases}$$
(17)

where x_w is the actual load of each vehicle; p_n is the number of seats; p_c is the rated load factor of each vehicle, when x_w is smaller than p_n , which means there are seats for all passengers and the influence of congestion degree is 0; when x_w is larger than p_n and smaller than p_c , it means that the degree of congestion is a little high; when x_w is larger

than p_c , it means that the degree of congestion is very high; *A* and η are the coefficients for average level congestion (second situation), *B* and ψ are the coefficients for serious congestion (third situation).

Perceived interchange time and path preference differ for different classes of passengers. Combined with the subway operating parameters **C**, the time-allowable deviation coefficient σ , the time coefficient in vehicle γ , the transfer time coefficient α , β and the joint distribution of prior probability can be denoted as $\pi(\mathbf{C}, \sigma, \gamma, \alpha, \beta)$, the conditional probability as $p(\Gamma | \mathbf{C}, \sigma, \gamma, \alpha, \beta)$ and the posterior probability distribution of each parameter as $\pi(\mathbf{C}, \sigma, \gamma, \alpha, \beta | \Gamma)$. By comparing actual time set Γ with the theoretical time for each class of passenger, the parameters were calibrated step by step. The theory of Bayesian inference [18,19] is denoted as

$$\pi(\mathbf{C}, \sigma, \gamma, \alpha, \beta | \Gamma) \propto p(\Gamma | \mathbf{C}, \sigma, \gamma, \alpha, \beta) \pi(\mathbf{C}, \sigma, \gamma, \alpha, \beta)$$
(18)

Assuming that all parameters are independent of each other, the posterior distribution theoretical solution of each parameter could be represented as

$$\prod_{w \in W} \left(\prod_{\Gamma \in T_w} \left(\sum_{r \in R_w} h(\Gamma | r, \mathbf{C}, \sigma, \gamma, \alpha, \beta) f_w(r | \mathbf{C}, \sigma, \gamma, \alpha, \beta) \right) \right) \times \pi(\mathbf{C}) \pi(\sigma) \pi(\gamma) \pi(\alpha) \pi(\beta)$$
(19)

With regard to the arithmetic solution of the posterior distribution, the large volume of passenger flow data must be processed and several parameters calibrated. Accordingly, the algorithm MCMC [20] was chosen as a time-saving and computerized capacity feasible method. The MCMC algorithm comprises three stages: Metropolis–Hasting sampling, Markov chain analysis, and error analysis.

2.4.3. Acquisition of Sectional Passenger Flow

Using the passenger clustering and parameter estimation method, passengers' generalized travel cost function for each class of passengers was determined. Adopting MLM [32], the choice of each effective path was determined as

$$p_{k,n}^{rs} = \frac{\exp\left(-\theta C_{k,n}^{rs}/C_{\min}\right)}{\sum_{p \in K_{rs}^{n}} \left(-\theta C_{p,n}^{rs}/C_{\min}\right)} \quad k \in K_{rs}^{n}, n \in N$$
(20)

where C_{\min} denotes the least cost among all the effective paths' costs and θ is the network familiarity level of the passengers.

Then, the passenger flow of each path $f_{k,n}^{rs}$ is estimated:

$$f_{k,n}^{rs} = q_n^{rs} \cdot p_{k,n}^{rs} \tag{21}$$

Finally, the cumulative $f_{k,n}^{rs}$ is defined as the sectional passenger volume x_a :

$$x_a = \sum_r \sum_s \sum_k f_k^{rs} \cdot \delta_{a,k}^{rs} \left(1\right)$$
(22)

2.5. The Hybrid Model

The model combination of subway passenger flow assignment and fractal dimensions was proposed to extract the relationship and evolutionary rule connecting the residents' travel demand and traffic supply. The flow chart of this hybrid model is displayed in Figure 2 and all variables are summarized in Table A1 (see Appendix A).



Figure 2. The flow chart of the combination of subway flow assignment and fractal dimensions.

3. Empirical Study

3.1. Data Collection

In this study, we used BSS smart card data to extract the urban mobility demand of BSS. In addition, mobile phone location data were used to derive the proxy travel demand for the entire city. Each of the datasets is described as follows.

3.1.1. BSS Smart Card Data

Six weeks of BSS smart card data recorded by automatic fee collection (AFC) systems were collected for each year during 2009 to 2018. The data included smart card ID, entry time, exit time, entry station, and exit station. Data were collected each day for the period from 5 a.m. to 11 p.m. The corresponding record numbers for each year are presented in Table 1.

Year	Date	Records
2009	14 December~20 December	33,589,115
2013	4 March~10 March	34,224,238
2015	21 December~27 December	33,458,698
2016	19 December~25 December	33,581,774
2017	18 December~24 December	32,478,975
2018	18 September~24 September	32,558,423

Table 1. BSS data description.

The AFC data utilized in this paper only contains passengers who use smart card and does not contain other payments such as paper-ticket, NFC solutions, APP, etc. However, the percentage of passengers using smart cards is different in different lines and different stations, according to the data analysis results, which shows that the average percentage of passengers using smart cards was almost 70%, which means that the AFC data can reflect the general passenger flow of Beijing URT. Meanwhile, adopting AFC data may underestimate passenger demand due to fare evasion in some cities [33], however, there are

security and gate systems for each station in Beijing URT systems, and it is well-staffed, so the number of fare evaders is low and cannot make a big difference to the research results.

3.1.2. Mobile Phone Data

BBS demand does not give the whole picture of urban mobility in Beijing. It is difficult, if not impossible, to collect the data for all transport modes, but mobile phone data provides an alternative. Mobile phone data for 10 August 2016, were collected and used in this study to extract the entire travel demand in Beijing.

3.2. BSS Data Processing

The geographical subway lines and subway stations were acquired via the Baidu Map API. The BSS network was visualized in the GIS platform via embedded vector plotting. When processing AFC data, we found that some AFC data had errors such as missing some necessary data, however, the number of these incorrect data was very small and could not influence the research results. Therefore, we deleted these wrong data in the BSS data processing. Then, boarding and alighting demand for each station were retrieved directly from AFC data. Using the traffic assignment method, sectional passenger flow was also obtained.

Taking the BSS travel data from 29 February to 3 April 2016 as an example, passengers who traveled more than six days in five weeks were regarded as regular travelers, who accounted for nearly 80% of total passengers on weekdays and 60% on weekends. For the spatial feature analysis, we took passengers with PR > 0.3 to be regular passengers (accounting for 87.6%). Using the topic model, we calculated the probability distributions of each passenger for nine topics, obtaining the results shown in Figure 3a. According to Figure 3a, fewer than 1% of passengers covered only one topic, whereas the passengers covering four topics were the most numerous; passenger travel was diverse and related to multiple topics. Passenger clustering results using the *K*-means algorithm are shown in Figure 3b, indicating that all passengers can be divided into seven clusters by their probability distribution of different travel time patterns. Parameter estimations are presented in Table 2. Using MLM, the demand for each section can be generated to allow for application of the fractal model.

The geographical distribution of the sectional passenger flow is visualized in Figure 4a, with fractal dimensions calculated using GIS and Oracle. Considering the BSS as an undirected- and weighted-network, we applied the Dijkstra algorithm to calculate the network's shortest path length between each pair of stations from 2016 to 2018. Among all stations, the lowest average shortest path length appeared at Tiananmen West Station and was chosen as the center of all rings, as shown in Figure 4b.



Figure 3. Percentages of passengers by topic model and shares of different classes: (**a**) percentages of passengers in different topics and (**b**) percentages of passenger clusters.

Passenger Class	γ	α_n	β_n	θ
Class 1	0.90	1.53	3.97	4.45
Class 2	1.33	3.68	3.01	3.87
Class 3	1.16	0.79	1.72	3.51
Class 4	1.19	0.84	1.64	3.12
Class 5	1.21	0.42	1.85	6.85
Class 6	1.11	3.98	3.77	2.16
Class 7	1.15	1.87	2.66	1.88

Table 2. Parameter results.



Figure 4. (a) Upstream flow in 2016 and (b) fractal units.

3.3. Fractal Analysis of BSS

3.3.1. Station and Volume

For the BSS, station dimensions, total volume, entry volume, and exit volume were respectively calculated using Equations (1)–(10). Figure 5 shows $\ln(r)$ and $\ln(S(r))$ for different years, with the blue dotted lines denoting Beijing's six ring roads.

This shows that the non-scale area is expanding from the inside of the third ring road to the fifth ring road (points scattered along a straight line), consistent with increases in station dimension D_S (slope of the straight line) for the BSS network.

Figure 6 summarizes the volume dimension D_V . All volume dimensions for each year were less than 2, indicating that demand density decreased from the city center to the suburb area, with demand commensurately lower in the suburbs than in the city center. This points out that suburb area could become a focus of government urban planning in the future such as tourist industry, breeding industry, etc.

The volume dimension D_V in BSS was a bit lower than that of the London Underground and Paris Metro (around 1.7) [12], indicating that they are at different stages of development. We also observed a fluctuation in the indicators of the demand-weighted network from 2009 to 2018, and the non-scale area within the fourth ring underlines a gap between URT topology and rail transit demand.



Figure 5. Fitting plots of station dimension for different years.



Figure 6. Volume dimension for each year.

3.3.2. Traffic Impedance

Using the subway traffic assignment model and the OD matrix, we acquired the section load of each link between each adjacent station pair including the upstream and downstream flows. These sectional loads were incorporated in the dimension calculation. Figure 7 shows the interstation distance D_I^f .

An increasing trend of interstation distance without section load could be observed, indicating that the accessibility of the whole network increases with time. Note that this indicator was close to 2 within the fifth ring road, implying that the accessibility in the city center was well developed in its network structure. The interstation distance dimensions D_I^f were all below 2 out of the fifth ring road, indicating that the density of the BSS network decreased from the center to the suburbs.



Figure 7. Interstation distance for select years.

Taking into account the section load on the link produced the results shown in Figure 8, which demonstrated the same trend but a lower dimension than without demand. Such a result is common for a mono-centered city, in which the majority of demand is concentrated in the center area.



Figure 8. Impedance dimension with demand as a weight.

3.3.3. Branch Dimension

The branch dimension D_B is used to evaluate the URT network's development level. The higher the branch dimension, the more complicated the network and the greater the accessibility. Figure 9 shows the results for various years.

As revealed by $\ln(B(r))$, increasing amounts of urban area are being covered by the BSS, and the fractal characteristic is becoming more and more obvious with time, having grown from 1.53 in 1987 to 1.86 in 2018.



Figure 9. Fitting plots of branch dimension.

3.3.4. Fractal Dimension Consistency Index in BSS

Using Equation (11), Figure 10 shows the fractal dimension consistency index γ for each year. The consistency index γ was close to 1, indicating that the supply of transportation is strongly related to the demand of passenger flow. Furthermore, the index increased monotonically before 2015, peaking in 2015 before hitting its low point in 2016. The locations of newly built subway lines contributed to this pattern. Before 2015, construction of URT lines emphasized more than the city center. As the network matured in the city center, the focus began shifting toward remote suburban areas, with as many as three suburban lines opening at the end of 2015. The implications are twofold. First, these lines attracted very low demand, with a large train headway, especially at their opening. The low load level generates inconsistency between supply and demand. Second, these lines increase the suburban accessibility, attracting more residents to downtown areas, thereby increasing the spatial heterogeneity of demand. Thus, the consistency indicator decreased in 2016. As newly opened suburban lines cultivated demand, and thanks to the implementation of decentralizing policies in Beijing, residents even traveled to suburban areas for work. These inconsistencies were mitigated as time passed.



Figure 10. Fractal dimension consistency index.

3.4. Analysis of Urban Travel Demand from Mobile Phone Data

Rail demand was used in the foregoing analysis, but is only part of the overall picture of urban mobility. To further investigate the role of rail in the residents' daily transport, we must consider the whole picture. As collecting demand for all modes is quite difficult, we used demand extracted from mobile phone data as a proxy. Taking 2016 as an example, we estimated the population distribution of the whole city and derived travel demand using mobile phone data.

As shown in Figure 11a, the urban population is distributed mainly outside the fourth ring road at the start of morning peak, scattered along the subway lines. In contrast, at the start of the evening peak, as shown in Figure 11b, the population is concentrated inside the fourth ring road and distributed along the subway lines. Such observations indirectly demonstrate that the BSS plays a major role in work–residence separation and has a strong gathering and guiding effect on passenger flow.



Figure 11. Population distribution (a) from 7 a.m. to 8 a.m. and (b) from 5 p.m. to 6 p.m.

Based on mobile phone data, Figure 12 shows $\ln(V(r))$ for two peak periods in the day, the morning (7–10 a.m.) and evening (5–8 p.m.) peaks. The volume dimension D_V from 8 a.m. to 9 a.m. was almost the same as from 9 a.m. to 10 a.m. The change rate of the volume in the morning peak was faster than that of the evening peak, demonstrating that the commuting period in the morning peak was more concentrated than that in the evening peak. For the evening peak, $\ln(V(r))$ increased gradually, demonstrating that the residents returned to their dwellings after work within a longer time span.

The ODs derived from the mobile phone data could help directly demonstrate spatial travel demand, revealing an evident trend of traveling from the outside city to the center of the city. Figure 13 shows the corresponding geographical OD distributions for the morning peak (7–10 a.m.).



Figure 12. Volume dimension of population distribution deprived from mobile phone data.



Figure 13. OD distribution in the morning peak, for (a) origin and (b) destination.

This shows that the origins were concentrated mainly in the suburb areas, whereas the destinations were distributed mainly in the center of the city during the morning peak.

According to Figure 14, for the morning peak origin and evening peak destination, the volume was similar to that of the population distribution. It can be inferred that the morning mobility origins of most urban residents are in the home, regardless of their occupation. Much as for the morning–evening comparison results of population distribution, it can be concluded that the imbalance between employment and housing is dramatic in Beijing.

We extracted urban resident distribution and traffic demand from mobile phone data. Different transport modes play different roles in transportation supply, and their volume dimensions should be compared from the perspective of fractal analysis.

8

0

2

ln(r)

3



Figure 14. Volume dimension of mobile phone ODs.

8

3

2

ln(r)

10

3.5. Comparing BSS with Travel Demand Based on Mobile Phone Data

Figure 15 shows an obvious commuting pattern for subway passengers. Some typical residential areas are evident in Figure 15a such as TianTongYuan, ShengMingKeXueYuan, and XiErQi, and some work locations are identifiable in Figure 15b such as GuoMao, DaZhiMen, and XiZhiMen.

2

ln(r)

3

8

0

2

ln(r)

3



Figure 15. The density distribution of metro passengers (morning peak): (a) origin and (b) destination.

As Figure 16 shows, the volume dimension D_V for morning peak origin and evening peak destination in BSS was larger than that of the population and the total OD pairs extracted from phone data. Furthermore, the volume dimension of the morning peak destination and the evening peak origin was lower than that of the population and total OD pairs. It can be concluded that subway-related mobility demand is more unbalanced than the average level of all modes, as reflected in the phone data. Due to the rail service's speed, residents can choose to live far from their workplace, so that the subway system actually increases work–home imbalance.



Figure 16. Volume dimension of BSS demand.

4. Conclusions and Discussion

In this study, we proposed a fractal model with which to extract the relationship and evolutionary rule connecting resident travel demand and traffic supply using multisource data. We first presented a fractal model for addressing both the topologic network and the transportation demand on it. These methods can be used to reveal the level of development, volume capacity, accessibility, and consistency of the morphology network and the demand network. To acquire the sectional demand that is the input for fractal analysis, we further proposed a stochastic traffic assignment model based on passenger clustering. Using public transportation network data for Beijing, China, we presented a detailed analysis based on the proposed model. Furthermore, to investigate the role of BSS on overall urban mobility in Beijing, we extracted the population distribution and OD data from mobile phone data. We compared the fractal characteristics of a rail network with demand as a weight and the fractal features of proxy urban mobility based on phone data. Our results show that the transit network reveals more information than the topological network alone, especially in the context of network development. Comparison of the results of BSS and of using phone data as a proxy illustrates the role of rail transit in urban mobility. The main conclusions of this study thus include the following:

- Passenger flow is considered for comprehensive investigation of URT effectiveness. Demand-weighted network analysis reveals more information about the evolution of URT and its correspondence to urban mobility than a purely topological network.
- ii. Non-scale area and self-similarity of the URT network in both the topological and the demand aspects is seen in Beijing's public transportation systems, along with a nonlinear area in the log coordinates—namely, fractal degradation in the early years of URT growth. In addition, for the topological network, nearly all of the fractal dimensions is increasing. However, when demand is taken into account, irregular fluctuations can be seen for some fractal dimensions, revealing inconsistencies in URT demand and supply that offer network work planning insights unable to be captured by pure physical network analysis.
- iii. Using phone-based population distribution data as a benchmark, the travel pattern of rail transit and its implications for urban mobility can be investigated and compared through fractal analysis. Different transit modes have markedly different roles in catering to commuting demand. URT normally plays a leading role because of its large capacity and exclusive right of way, and it actually increases job–work separation.

Some other aspects could be further investigated in future studies including the following: (1) The volume dimensions mentioned in this paper cannot depict the different development levels of a network in various directions within one fractal unit. The box-count dimension [34] could provide complementary descriptions for directional analysis of URT and urban expansion. (2) The fractal dimension calculated at the city scale of each transport mode is not accurate enough to provide detailed information or allow further comparison. In this paper, we used the passenger flow density distribution of each transportation analysis zone (TAZ) as supplementary information to offer additional detail,

but the information provided by the density distribution is limited to use for comparison with the fractal dimensions. Thus, calculation of fractal dimensions at the TAZ level could be conducted in future studies.

Author Contributions: Conceptualization Z.W. and Z.C.; Methodology Z.W. and L.H.; Data analysis Z.C.; Writing/original draft preparation Z.C.; Writing/review and editing L.H. and Y.S.; Visualization Z.C.; Funding acquisition Z.W. and L.H. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the general projects of the National Natural Science Foundation of China, grant numbers 51978044.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data sharing is not applicable to this article.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A

Table A1. Summary of all variables.

Variables	Meanings	
S(r)	Number of stations	
V(r)	Cumulative flow volume	
D_S	Station dimension	
D_V	Volume dimension	
D_V^{in}	Inflow volume dimension	
D_V^{out}	Outflow volume dimension	
$V^{in}(r)$	Cumulative inflow volume in each circular area with radius <i>r</i>	
$V^{out}(r)$	Cumulative outflow volume in each circular area with radius r	
d _{ij}	Distance between two stations <i>i</i> and <i>j</i>	
D_I^d	Traffic impedance dimension in terms of interstation distance	
D_I^f	Impedance dimension represented by sectional flow	
$I^d(r)$	Cumulative interstation distance in each circular area with radius <i>r</i>	
$I^{f}(r)$	Cumulative sectional passenger flow in each circular area of radius r	
$V^{up}(r)$	Cumulative flow volume in each circular area with radius <i>r</i> of upstream	
$V^{down}(r)$	Cumulative flow volume in each circular area with radius r of downstream	
D_V^{up}	Upstream volume dimension	
D_V^{down}	Downstream volume dimension	
D_B	Branch dimension	
B(r)	Number of branches in the circular area described by radius r	
γ	Fractal dimension consistency index	
$C_{k,n}^{rs}$	Quantitative cost	
$\sum_{l} \overline{T}_{rs,k}^{l}$	In-vehicle time	
$\sum_{i} \overline{E}_{i,n}^{l,m}$	Interchange time	

Variables	Meanings	
$\varepsilon_{k,n}^{rs}$	Stochastic cost error	
$w_{i,n}^{l,m}$	Walking time	
f_m	Headway	
$e_{i,k}^{rs}$	Accumulative interchange time of OD pair rs for path k at station i	
$Y(x_w)$	Degree of congestion	
x_w	Actual load of each vehicle	
p_n	Number of seats	
p_c	Rated load factor of each vehicle	
Г	Actual time set	

Table A1. Conts.

References

- Talib, N.H.; Hasnan, K.B.; Nawawi, A.B.; Abdullah, H.B.; Elewe, A.M. Monitoring large-scale rail transit systems based on an analytic hierarchy process/gradient-based cuckoo search algorithm (GBCS) scheme. Urban Rail Transit 2020, 6, 132–144. [CrossRef]
- 2. Obsie, A.; Woldeamanuel, M.; Woldetensae, B. Service Quality of Addis Ababa Light Rail Transit: Passengers' Views and Perspectives. *Urban Rail Transit* 2020, *6*, 231–243. [CrossRef]
- 3. Abutaleb, A.; McDougall, K.; Basson, M.; Hassan, R.; Mahmood, M.N. The Impact of Transit-Oriented Shopping Mall Developments (TOSMDs) on Metro Station Ridership: Dubai Metro Redline. *Urban Rail Transit* 2020, *6*, 157–170. [CrossRef]
- 4. Screen, D.; Parkinson, J.; Shilton, C.; Rjabovs, A.; Marinov, M. Data Analysis to Study Sub-threshold Delays Incurred by Tyne and Wear Metro Trains. *Urban Rail Transit* 2020, *6*, 93–103. [CrossRef]
- 5. Weerawat, W.; Samitiwantikul, L.; Torpanya, R. Operational Challenges of the Bangkok Airport Rail Link. *Urban Rail Transit* 2020, *6*, 42–55. [CrossRef]
- 6. Sönmez, H.Y.; Öztürk, Z. Effects of traffic loads and track parameters on rail wear: A case study for Yenikapi–Ataturk Airport Light Rail Transit Line. *Urban Rail Transit* 2020, *6*, 244–264. [CrossRef]
- 7. Potti, P.; Marinov, M. Evaluation of Actual Timetables and Utilization Levels of West Midlands Metro Using Event-Based Simulations. *Urban Rail Transit* 2020, *6*, 28–41. [CrossRef]
- 8. Guerrieri, M. Correction to: Tramways in Urban Areas: An Overview on Safety at Road Intersections. *Urban Rail Transit* 2020, *6*, 185–186. [CrossRef]
- 9. Christoforou, Z.; Chandakas, E.; Kaparias, I. Investigating the Impact of Dwell Time on the Reliability of Urban Light Rail Operations. *Urban Rail Transit* 2020, *6*, 116–131. [CrossRef]
- 10. Mandelbrot, B.B. The Fractal Geometry of Nature; WH Freeman: New York, NY, USA, 1983.
- 11. Pavón-Domínguez, P.; Rincón-Casado, A.; Ruiz, P.; Camacho-Magriñán, P. Multifractal approach for comparing road transport network geometry: The case of Spain. *Phys. A Stat. Mech. Appl.* **2018**, *510*, 678–690. [CrossRef]
- 12. Benguigui, L. The fractal dimension of some railway networks. J. Phys. I 1992, 2, 385–388. [CrossRef]
- 13. Batty, M.; Longley, P.A. Fractal Cities: A Geometry of Form and Function; Academic Press Inc.: San Diego, CA, USA, 1994.
- 14. Batty, M. The Size, Scale, and Shape of Cities. Science 2008, 319, 769–771. [CrossRef]
- 15. Wang, H.; Luo, S.; Luo, T. Fractal characteristics of urban surface transit and road networks: Case study of Strasbourg, France. *Adv. Mech. Eng.* **2017**, *9*, 1687814017692289. [CrossRef]
- 16. Valério, D.; Lopes, A.M.; Machado, J.A.T. Entropy Analysis of a Railway Network's Complexity. Entropy 2016, 18, 388. [CrossRef]
- 17. 2019 Beijing Transport Annual Report [EB/OL]. 2020, p. 31. Available online: http://www.bjtrc.org.cn/List/index/cid/7.html (accessed on 19 March 2021). (In Chinese).
- 18. Chen, Y. Multi-scaling allometric analysis for urban and regional development. Phys. A Stat. Mech. Appl. 2017, 465, 673–689. [CrossRef]
- 19. Lu, Z.; Zhang, H.; Southworth, F.; Crittenden, J. Fractal dimensions of metropolitan area road networks and the impacts on the urban built environment. *Ecol. Indic.* **2016**, *70*, 285–296. [CrossRef]
- 20. Liu, J.-L.; Yu, Z.-G.; Anh, V. Determination of multifractal dimensions of complex networks by means of the sandbox algorithm. *Chaos: Interdiscip. J. Nonlinear Sci.* **2015**, *25*, 023103. [CrossRef]
- Jalan, S.; Yadav, A.; Sarkar, C.; Boccaletti, S. Unveiling the multi-fractal structure of complex networks. *Chaos Solitons Fractals* 2017, 97, 11–14. [CrossRef]
- 22. Chriqui, C.; Robillard, P. Common Bus Lines. Transp. Sci. 1975, 9, 115–121. [CrossRef]
- 23. Hassan, M.N.; Rashidi, T.H.; Nassir, N. Consideration of different travel strategies and choice set sizes in transit path choice modelling. *Transportation* **2019**, 2019, 1–24. [CrossRef]
- 24. Kim, K.M.; Hong, S.-P.; Ko, S.-J.; Kim, D. Does crowding affect the path choice of metro passengers? *Transp. Res. Part A Policy Pract.* **2015**, *77*, 292–304. [CrossRef]

- Zhu, W.; Fan, W.-L.; Wahaballa, A.M.; Wei, J. Calibrating travel time thresholds with cluster analysis and AFC data for passenger reasonable route generation on an urban rail transit network. *Transportation* 2020, 47, 3069–3090. [CrossRef]
- 26. Zhu, W.; Zhou, F.; Huang, J.; Xu, R. Validating rail traffic assignment models with cluster analysis and automatic fare collection data. *Transp. Res. Rec.* 2015, 2526, 10–18. [CrossRef]
- 27. Rahbar, M.; Hickman, M.; Mesbah, M.; Tavassoli, A. Calibrating a Bayesian traffic assignment model using smart card data. *IEEE Trans. Intell. Transp. Syst.* 2018, 20, 1574–1583. [CrossRef]
- 28. Rahbar, M.; Hickman, M.; Mesbah, M.; Tavassoli, A. Determining effective sample size to calibrate a traffic assignment model: A Bayesian perspective. *Transp. Res. Rec.* 2018, 2672, 849–858. [CrossRef]
- 29. Xu, X.; Xie, L.; Li, H.; Qin, L. Learning the route choice behavior of subway passengers from AFC data. *Expert Syst. Appl.* **2018**, *95*, 324–332. [CrossRef]
- 30. Beijing Subway. Subway Query 8 January 2019. Available online: http://www.bjsubway.com/en/ (accessed on 15 August 2019).
- 31. Wang, Y.P.; Chen, K.M.; Ma, C.Q. Quantitative analysis of coordination between rail transit network configuration and urban form. *J. Railw. Eng. Soc.* **2008**, *11*, 11–15.
- 32. Chen, Z.; Kuo, L. A Note on the Estimation of the Multinomial Logit Model With Random Effects. Am. Stat. 2001, 55, 89–95. [CrossRef]
- Barabino, B.; Lai, C.; Olivo, A. Fare evasion in public transport systems: A review of the literature. *Public Transp.* 2020, 12, 27–88.
 [CrossRef]
- 34. Frankhauser, P. Aspects fractals des structures urbaines. L'Espace Géographique 1990, 19, 45–69. [CrossRef]