Frontiers in Traffic and Transportation Engineering (2024)
Clausius Scientific Press, Canada

DOI: 10.23977/ftte.2024.040102
ISSN 2616-2334 Vol. 4 Num. 1

Analysis of Traffic Flow Based on Toll Data

Xiangmin Zhu[1], Jian Luo[2,a,*], Meng Yang[2]

[1]Shandong Hi-speed Company Limited, Ji'nan, 250101, China
[2]China Merchants Chongqing Communications Technology Research & Design Institute CO., LTD., Chongqing, China
[a]luojian1@cmhk.com
[*]Corresponding author

Keywords: Expressway; Data mining; Traffic; OD

Abstract: Utilizing data mining technology, this study proposes a methodology for traffic flow analysis based on toll data. Initially, the OD information pertaining to the analyzed toll station is extracted from expressway toll data. Subsequently, traffic flow analysis is conducted employing time series methods over various temporal scales, including years, months, and days. Comparative assessment of traffic flow across different road sections within the expressway network elucidates variations in traffic flow status. Quantitative analysis serves as the cornerstone for this investigation. Taking the Xi'an Ringo Expressway Toll Station as a case study, this paper delves into the temporal and spatial distribution of traffic flow along the roadway. The deterministic, explosive, and growth-type evaluation indices for the road network provide quantitative metrics for assessing the operational and managerial efficacy of the expressway. Moreover, the establishment of a visualization platform for public service facilitates the dissemination of theoretical frameworks and data-backed insights concerning road traffic flow and road network operations.

1. Introduction

In recent years, the rapid acceleration of urbanization nationwide has been accompanied by significant advancements in highway infrastructure. The complex transportation network has enhanced travel convenience by presenting numerous alternative routes, thereby augmenting the array of options available to travelers aiming for the same destination. Consequently, the selection of an optimal route that balances convenience and efficiency has emerged as a central concern. In this context, the utilization of toll data to analyze the traffic flow across various segments of highways holds promise. By scrutinizing the spatiotemporal distribution patterns of traffic flow, it becomes feasible to forecast future traffic patterns within the road network, identify bottleneck segments, and furnish guidance for selecting the most efficient routes. Such endeavors not only enhance travel convenience but also furnish valuable insights for the strategic planning of the national highway system's future development.

Currently, the nationwide deployment of the Electronic Toll Collection (ETC) system is in effect. With the progression of computer technology and information technology, the methodology for gathering traffic data has evolved. Zhang Hongbin et al. [1] employed the traffic volume coefficient of variation to compensate for absent toll data, formulated Origin-Destination (OD) matrices, delineated path recognition and filtration principles based on the interconnectivity characteristics among neighboring toll stations within the road network, and devised allocation matrices for segment flow computation. Weng Jiancheng et al. [2] devised a velocity calculation model for highway journeys, leveraging ETC data and OD data from specific segments. Han Daqian et al. [3] introduced a computation technique grounded on swipe card timestamps, vehicle categorization, and geographical data extracted from toll records, ensuring precise and efficient estimation of traffic flow for any road segment. Hu Guixiu et al. [4] proposed corresponding conversion methodologies to transmute vehicle type classification data and highway toll mode data into traffic volume survey data, offering a framework for estimating highway section flow. Jiang Yiyue [5], by means of analysis and processing of highway toll data, conducted spatiotemporal characteristic analysis on highway traffic volumes, formulated a spatiotemporal correlated regression model for intercity OD travel volumes, and carried out an impact analysis on traffic travel volumes. Han Kunlin et al. [6], based on vehicle detector data and toll records, devised a methodology for detecting abnormal states on highways employing algorithmic voting fusion.

Currently, the scholarly exploration of expressway network traffic flow, utilizing networked toll data, predominantly concentrates on computing section traffic flow, thereby overlooking a comprehensive assessment of the spatiotemporal distribution of traffic flow within expressway networks. Addressing how to effectively leverage toll data sourced from expressways to derive insights into network-wide traffic flow dynamics across varying regions and temporal intervals stands as a paramount concern in scrutinizing the operational dynamics of expressway networks. Consequently, this paper advocates for an analytical investigation into expressway traffic flow patterns grounded in toll data, alongside the development of a corresponding road model. By extracting OD information from toll records and integrating it into the road model, this study aims to unravel the intricate traffic flow dynamics within expressway networks across diverse geographical zones and timeframes. Such an approach offers a novel methodology for comprehensively understanding the dynamics of expressway traffic flow analysis.

2. Acquiring Vehicle OD Matrix

2.1 OD Data Preprocessing

The initial dataset from toll stations on expressways captures pertinent details upon vehicle passage. These encompass particulars such as the exit toll station number, entry time, entry toll station number, exit time, vehicle license plate number, vehicle type, axle count, toll fees, among others. It is imperative to note that this raw dataset necessitates preprocessing before integration into existing road network models. For a comprehensive depiction, Table 1 showcases the raw dataset.

Table 1: Original Data

Entry Station Number	Entry Station Time	Exit Station Number	Exit Station Time	Amount	Vehicle Type
210103	2022-01-15 22:32:23	10002	2022-01-15 22:49:24	85.5	23
210104	2022-01-14 23:35:21	10021	2022-01-16 00:56:54	97.3	22
210105	2022-01-15 23:37:14	100103	2022-01-16 02:45:20	24.4	23
…	…	…	…	…	…
210206	2022-01-16 00:32:56	100201	2022-01-17 22:20:54	61.3	22

The original dataset encompasses a rich array of information beyond mere vehicle passage records. It comprises experimental data generated during toll system calibration, alongside disturbance data like cashier shift records. Furthermore, inherent to the data recording process are potential errors leading to instances of data loss, duplication, or inaccuracies, thereby yielding anomalous or incomplete datasets. Such data anomalies render them unsuitable for experimental research, warranting meticulous cleansing procedures.

(1) Exclude data pertaining to system testing and cashier shift transactions;

(2) Detect and eliminate duplicate entries;

(3) In cases where entry or exit toll station times exhibit anomalies, utilize adjacent entries to interpolate missing data while retaining the entries.

2.3 Correction of OD Matrix

By employing the aforementioned methodologies, we are able to derive an OD matrix for traffic volume based on exit records. At this juncture, the OD matrix reflects exit records within a defined temporal interval. The ensuing discourse elucidates the corrective measures and presents an approach for allocating traffic volume to vehicles persisting on the route (i.e., those yet to depart the specified segment) with regard to their entry and exit points.

Let $A(i)$ denote the inbound flow volume at the *i-th* toll station, $B(i)$ denote the outbound flow volume at the *i-th* toll station, $X(i, j)$ denote the number of vehicles entering the expressway at the *i-th* toll station and exiting at the *j-th* toll station, Y denote the number of vehicles remaining on the route, and $Y(i)$ denote the number of vehicles remaining inbound at the *i-th* toll station. Then we have:

Let $A(i)$ represent the inbound flow volume at the i-th toll station, $B(i)$ denote the outbound flow volume at the i-th toll station, $X(i, j)$ denote the count of vehicles entering the expressway at the i-th toll station and exiting at the j-th toll station, Y signify the tally of vehicles remaining on the route, and $Y(i)$ denote the count of vehicles lingering inbound at the i-th toll station. Thus, we have:

$$Y = \sum_{i=1}^{n} A(i) - \sum_{i=1}^{n} B(i) \tag{1}$$

$$Y = \sum_{i=1}^{n} Y(i) \tag{2}$$

Under the assumption that each vehicle remaining on the route at the front section of every toll station exits the expressway with equal probability, we denote $B'(i)$ as the probability of vehicles entering at the i-th toll station and exiting at the j-th toll station after adjustment.

$$P(i, j) = X(i, j) / B(j) \tag{3}$$

Therefore, the corrected $X'(i, j)$ should be:

$$X'(i, j) = X(i, j) + X(i, j) \sum_{i=1}^{n} Y(i) / B(j) \tag{4}$$

Through the above formula, the previously compiled OD matrix can be adjusted to obtain the traffic volume OD matrix for a specific time frame after correction.

3. Calculation Method for Section Traffic Volume

3.1 Model for Section Traffic Volume Calculation

Traffic flow holds significant importance in transportation research, facilitating the examination of road traffic operational dynamics through sectional analysis. The computation of traffic volume at road sections can be achieved by leveraging OD data collected between toll stations situated on expressways.

Consider a scenario where toll stations are positioned along a specific stretch of an expressway, as depicted in Figure 1, encompassing 2 mainline toll stations and 4 ramp toll stations within this segment.

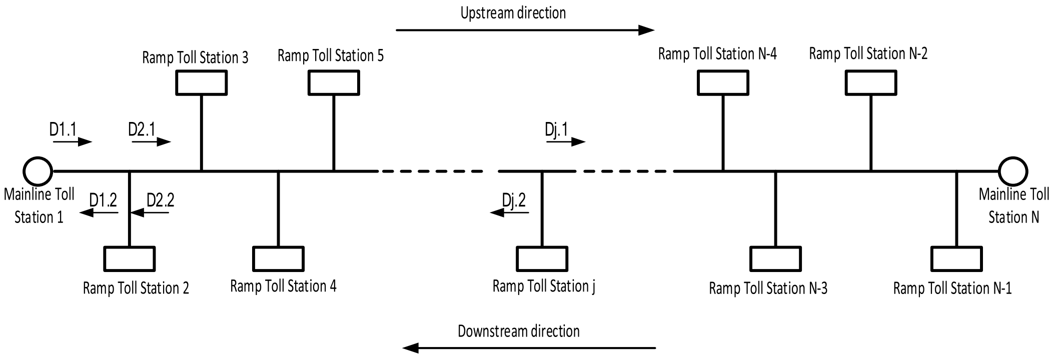

Figure 1: Toll station distribution map

In the provided illustration, the orientation from Mainline Toll Station 1 to Mainline Toll Station N signifies the upstream direction, whereas the opposite orientation is deemed the downstream direction. Traffic flow proceeding upstream is delineated as $D_{1.1}$, $D_{2.1}$,…,$D_{N.1}$, while traffic flow moving downstream is denoted as $D_{N.2}$, $D_{N-1.2}$,…, $D_{1.2}$.

The examined section lies adjacent to Mainline Toll Station 1. All traffic flow directed upstream enters the roadway via Mainline Toll Station 1, whereas all traffic flow traveling downstream exits the roadway through Mainline Toll Station 1. Consequently:

Upstream direction : $D_{1.1} = OD_{12} + OD_{13} + \cdots + OD_{1N}$

Downstream direction : $D_{1.2} = OD_{21} + OD_{31} + \cdots + OD_{N1}$

The section borders Ramp Toll Station 2. Regarding upstream traffic flow, with the exception of a fraction of vehicles departing the roadway from Ramp Toll Station 2, all other vehicles enter section $D_{1.1}$. Conversely, downstream traffic flow is derived by subtracting the number of vehicles entering from Ramp Toll Station 2 from the adjoining section to the Mainline Toll Station. Consequently:

Upstream direction : $D_{2.1} = D_{1.1} - OD_{12} + OD_{23} + OD_{24} + \cdots + OD_{2N}$

Downstream direction : $D_{2.2} = D_{1.2} - OD_{21} + OD_{32} + OD_{42} + \cdots + OD_{N2}$

Extending from this, we can formulate the general methodology for computing the traffic flow of the section.

For the segment adjoining the upstream direction and the mainline toll station:

$$D_{1.1} = \sum_{i=2}^{N} OD_{1i} \tag{5}$$

For the section not adjacent to the upstream direction and the mainline toll station:

$$Y = \sum_{i=1}^{n} Y(i) \quad D_{j.1} = D_{j-1.1} - \sum_{i=1}^{j-1} OD_{ij} + \sum_{i=j+1}^{N} OD_{ji} \quad (j \geq 2) \tag{6}$$

Similarly, for the section adjacent to the downstream direction and the mainline toll station:

$$D_{1.2} = \sum_{i=2}^{N} OD_{i1} \tag{7}$$

For the section not adjacent to the downstream direction and the mainline toll station:

$$D_{j.2} = D_{j-1.2} - \sum_{i=1}^{j-1} OD_{ji} + \sum_{i=j+1}^{N} OD_{ij} \quad (j \geq 2) \tag{8}$$

Based on the aforementioned analysis, the following conclusions can be derived:

(1) The upstream traffic flow on sections contiguous to mainline toll stations corresponds to the influx of traffic into the upstream direction of the mainline toll station.

(2) The downstream traffic flow on sections adjoining mainline toll stations equates to the egress of traffic from the downstream direction of the mainline toll station.

(3) The upstream traffic flow on sections neighboring ramp toll stations is determined by subtracting the traffic flow entering the ramp toll station from the traffic flow entering the upstream direction of the mainline toll station.

(4) The downstream traffic flow on sections adjacent to ramp toll stations is calculated by subtracting the traffic flow exiting the ramp toll station from the traffic flow exiting the downstream direction of the mainline toll station.

Through the application of traffic flow extraction techniques, diverse traffic volumes for distinct sections can be ascertained, encompassing traffic volumes for various vehicle categories, natural traffic volumes, and adjusted traffic volumes for sections.

3.2 Calculation Model of Road Network Traffic Volume

Traffic assignment represents a pivotal stage in the realm of transportation planning. It entails conducting OD surveys to compile matrices delineating the travel patterns of individuals, subsequently allocating these matrices onto the road network employing designated methodologies. This allocation process furnishes insights into the traffic volumes traversing various segments of the network. Depending on the objectives of the study, traffic assignment can be categorized into two principal types: static traffic assignment and dynamic traffic assignment. The former relies on pre-existing OD data and adheres to a predetermined allocation scheme, with the OD data remaining static over time. Conversely, dynamic traffic assignment involves the analysis of traffic flow in scenarios where traffic demand is known, incorporating traffic control and guidance measures to effectively manage traveler journeys, thereby enhancing the operational efficiency of the road network.

(1) All-or-Nothing Algorithm

The All-or-Nothing algorithm constitutes a simplistic yet fundamental method for allocation. It assigns the entirety of the traffic volume between any OD pair to the shortest path connecting said pair, while assigning zero traffic volume to alternate paths. While computationally straightforward, this approach overlooks the phenomenon wherein road impedance escalates with increased traffic volume, leading to diminished capacity. Consequently, the practical application of this algorithm is limited.

(2) Capacity Restraint Algorithm

The Capacity Restraint algorithm embodies an iterative approach to traffic assignment. It iteratively applies the All-or-Nothing algorithm while continuously updating road impedance to achieve equilibrium in allocation. This algorithm builds upon the principles of the shortest path assignment method, necessitating a considerable number of iterations (denoted by N) to attain equilibrium.

(3) Incremental Assignment Algorithm

The Incremental Assignment algorithm progressively allocates traffic volume onto the road network. Initially, it employs the All-or-Nothing algorithm to assign a specific proportion of traffic volume. Subsequently, it computes traffic volume, capacity, and road impedance for each road segment, informing subsequent allocations. Through incremental allocations, this method approximates equilibrium distribution models.

(4) Stochastic User Equilibrium (SUE) Algorithm

The Stochastic User Equilibrium (SUE) model represents a refined iteration of the User Equilibrium (UE) model. Unlike the UE model, which posits that travelers possess comprehensive traffic information and make optimal route choices, the SUE model acknowledges the uncertainty inherent in travelers' route selections owing to incomplete traffic information. It treats such uncertainty as stochastic variables, with SUE being achieved when individual travelers cannot decrease travel time by altering their route choice. SUE allocation presents a more realistic depiction than UE allocation, aligning more closely with actual traveler behavior.

4. Example Analysis

This study examines the Xi'an Ring Expressway located in Shaanxi Province, comprising nine toll stations: Hancheng Toll Station, Weiyangnan Toll Station, Zaoyuan Toll Station, Xinzhu Toll Station, Fangzhicheng Toll Station, Qujiang Toll Station, Gaoxin Zone Toll Station, Changan Toll Station, and Hechizhai Toll Station.

4.1 Data Description and OD Matrix Acquisition

The original toll station data records pertinent vehicle information as they traverse the expressway. Following data cleaning and integration, the focus narrows to the origin and destination points of vehicles, considering the multiple attributes within each raw data entry. This study aims to analyze the spatial distribution pattern of traffic flow on the Xi'an Ring Expressway in Shaanxi Province to predict future traffic flow on this road network.

Consequently, the OD data used herein pertains to travel between toll stations on the Xi'an Ring Expressway in Shaanxi Province. Extracting OD pairs between each toll station and computing the traffic volume between toll stations is essential. The consolidated data yield a many-to-many matrix, detailed in Table 2.

Table 2: The data integration results in a many-to-many matrix

O\D	Hechizhai	Gaoxin	Changan	Qujiang	Fangzhicheng	Xinzhu	Xingyuan	Weiyangnan	Hancheng
Hechizhai	--	1964	15780	40489	6545	3447	574	398	2823
Gaoxin	2678	--	51148	99594	17534	13293	2412	1631	5423
Changan	15462	55846	--	18079	11554	8400	1632	1661	5077
Qujiang	39122	114338	19173	--	2890	3817	575	614	3781
Fangzhicheng	**Xinzhu**	16262	10149	3864	--	2071	382	348	1073
Xinzhu	3483	15526	9928	4831	2554	--	740	312	1302
Xingyuan	456	2551	1900	1026	524	618	--	396	4745
Weiyangnan	387	1195	1389	731	406	277	998	--	4171
Hancheng	2537	4281	4342	4048	1272	1199	5201	4455	--

4.2 Analysis of Section Traffic Volume at Different Time Periods

The Annual Average Daily Traffic (AADT) serves as a metric indicating the average traffic volume observed throughout a year, thereby providing insights into the typical traffic patterns over extended periods. In road grading exercises, the anticipated daily traffic volume (considering both directions and standard passenger cars) for the designated future design year is often employed, denoted as the Annual Average Daily Traffic (AADT). This figure is derived by dividing the total traffic volume for a year by 365 days and subsequently adjusting it to suit the traffic demands of roads classified under varying grades.

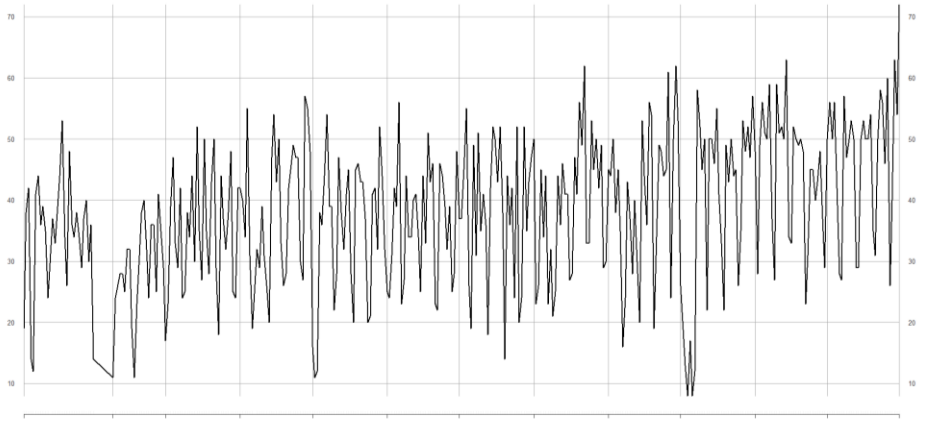

Figure 2: Annual Average Daily Traffic (AADT) Time Series

Based on the observations from Figure 2, the annual average daily traffic volume within section 5 predominantly fluctuates around 4000 vehicles per day. Notably, two conspicuous troughs are evident in early February and early October. Specifically, during early February 2022, coinciding with the Spring Festival holiday, there is a discernible reduction in vehicle traffic within section 5. This phenomenon suggests a minor decline in road traffic volume during holidays compared to typical days.

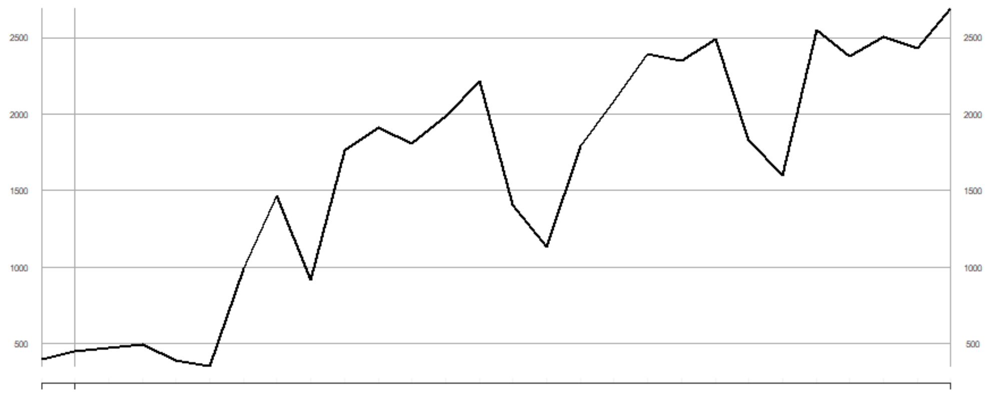

Figure 1: February Road Segment 5 Road Traffic Volume Time Series

The illustration above illustrates the traffic volume scenario on section 5 during February. As indicated, there is a consistent upward trend in vehicle traffic throughout February. At the onset of the month, the traffic volume on the road is below 500 vehicles per day, followed by a substantial increase, as shown in Figure 3.

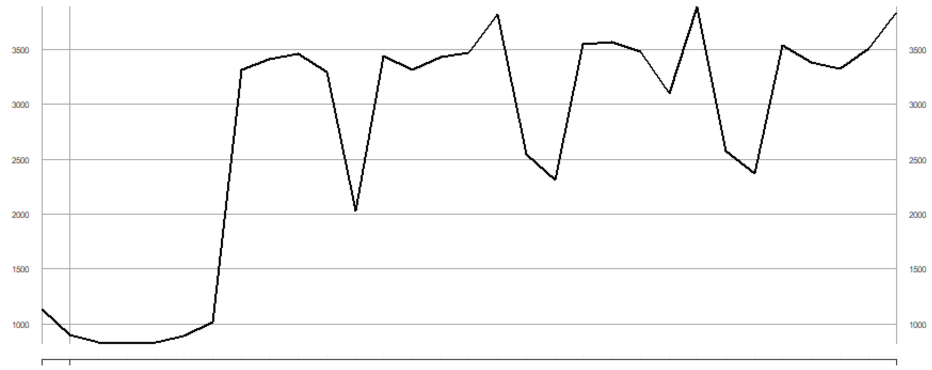

Figure 4: October Road Segment 5 Road Traffic Volume Time Series

The illustration above delineates the traffic volume dynamics observed on section 5 throughout October. It illustrates a consistent pattern in vehicle volumes during this month. Commencing October, the traffic flow on the road remains under 1000 vehicles per day initially, gradually escalating to approximately 3000 vehicles per day, and subsequently plateauing within a confidence interval. In contrast to the average period, the traffic flow on section 5 averages approximately 3000 vehicles, as shown in Figure 4.

4.3 Analysis of Traffic Volumes in Different Road Sections

The utilization of mosaic plots facilitates the examination of traffic flow across various road segments. These plots employ a color gradient, ranging from green to red, to signify different levels of congestion on the roads, enabling a comparative analysis of traffic volume, as shown in Figure 5.

O\D	210102	210103	210104	210105	210106	210107	210201	210202	210203
210102	251	1964	15780	40489	6545	3447	574	398	2823
210103	2678	269	51148	99594	17534	13293	2412	1631	5423
210104	15462	55846	1805	18079	11554	8400	1632	1661	5077
210105	39122	114338	19173	203	2890	3817	575	614	3781
210106	5430	16262	10149	3864	96	2071	382	348	1073
210107	3483	15526	9928	4831	2554	96	740	312	1302
210201	456	2551	1900	1026	524	618	275	396	4745
210202	387	1195	1389	731	406	277	998	55	4171
210203	2537	4281	4342	4048	1272	1199	5201	4455	127

(a) Traffic distribution on different road sections in February

O\D	210102	210103	210104	210105	210106	210107	210201	210202	210203
210102	20.9167	163.667	1315	3374.08	545.417	287.25	47.8333	33.1667	235.25
210103	223.167	22.4167	4262.33	8299.5	1461.17	1107.75	201	135.917	451.917
210104	1288.5	4653.83	150.417	1506.58	962.833	700	136	138.417	423.083
210105	3260.17	9528.17	1597.75	16.9167	240.833	318.083	47.9167	51.1667	315.083
210106	452.5	1355.17	845.75	322	8	172.583	31.8333	29	89.4167
210107	290.25	1293.83	827.333	402.583	212.833	8	61.6667	26	108.5
210201	38	212.583	158.333	85.5	43.6667	51.5	22.9167	33	395.417
210202	32.25	99.5833	115.75	60.9167	33.8333	23.0833	83.1667	4.58333	347.583
210203	211.417	356.75	361.833	337.333	106	99.9167	433.417	371.25	10.5833

(b) Traffic distribution on different road sections in Mayday

O\D	210102	210103	210104	210105	210106	210107	210201	210202	210203
210102	301.2	2356.8	20514	56684.6	9817.5	5515.2	975.8	716.4	5363.7
210103	3213.6	322.8	66492.4	139432	26301	21268.8	4100.4	2935.8	10303.7
210104	18554.4	67015.2	2346.5	25310.6	17331	13440	2774.4	2989.8	9646.3
210105	46946.4	137206	24924.9	284.2	4335	6107.2	977.5	1105.2	7183.9
210106	6516	19514.4	13193.7	5409.6	144	3313.6	649.4	626.4	2038.7
210107	4179.6	18631.2	12906.4	6763.4	3831	153.6	1258	561.6	2473.8
210201	547.2	3061.2	2470	1436.4	786	988.8	467.5	712.8	9015.5
210202	464.4	1434	1805.7	1023.4	609	443.2	1696.6	99	7924.9
210203	3044.4	5137.2	5644.6	5667.2	1908	1918.4	8841.7	8019	241.3

(c) Traffic distribution on different road sections in August

O\D	210102	210103	210104	210105	210106	210107	210201	210202	210203
210102	527.1	4320.8	36294	97173.6	16362.5	8962.2	1549.8	1114.4	8186.7
210103	5623.8	591.8	117640	239026	43835	34561.8	6512.4	4566.8	15726.7
210104	32470.2	122861	4151.5	43389.6	28885	21840	4406.4	4650.8	14723.3
210105	82156.2	251544	44097.9	487.2	7225	9924.2	1552.5	1719.2	10964.9
210106	11403	35776.4	23342.7	9273.6	240	5384.6	1031.4	974.4	3111.7
210107	7314.3	34157.2	22834.4	11594.4	6385	249.6	1998	873.6	3775.8
210201	957.6	5612.2	4370	2462.4	1310	1606.8	742.5	1108.8	13760.5
210202	812.7	2629	3194.7	1754.4	1015	720.2	2694.6	154	12095.9
210203	5327.7	9418.2	9986.6	9715.2	3180	3117.4	14042.7	12474	368.3

(d) Traffic distribution on different road sections in October

Figure 5: Traffic volume distribution on different road sections in different months

Based on the depicted data, it is evident that the traffic volume on the roads during February appears comparatively subdued when contrasted with other months. Notably, heightened traffic volumes are observed solely between 210103 (Chang'an West) and 210105 (Qujiang), as well as from 210105 (Qujiang) to 210103 (Chang'an West). Generally, the traffic flow exhibits a smooth progression. Conversely, in May, August, and October, the traffic volume on the roads follows predictable patterns. Analysis of different color distributions suggests a uniformity in traffic volumes across distinct roads within the same month. Furthermore, variations in traffic volumes along the same road segment correspond with the progression of months.

5. Conclusions

This paper presents an investigation into highway traffic flow leveraging toll data and establishes a comprehensive road model. It involves the extraction of OD information from toll data and the subsequent allocation of OD pairs onto the road model, enabling the identification of traffic flow patterns across highways categorized by both geographical regions and time periods. This approach introduces a novel methodology for analyzing highway traffic flow. Utilizing toll stations along the Xi'an Ring Expressway as a case study, the research delves into the temporal and spatial distribution of traffic flow along the roads, aiming to ascertain deterministic, explosive, and growth-type evaluation indices for the road network. These indices serve to facilitate the quantitative analysis of road traffic flow for highway operation and management entities, thereby laying the groundwork for the development of a theoretical framework and data infrastructure supporting the creation of a public service-centric road network operation visualization platform.

Acknowledgment

The work was supported by the Key Technology Projects in the transportation industry of China for 2021 (No. 2021-MS6-145).

References

[1] Zhang Hongbin, Ni Fujian, Yang Shunxin, et al. Calculation of highway section traffic volume based on toll data [J]. Journal of Southeast University: Natural Science Edition, 2013, 43(2): 320-324.

[2] Weng Jiancheng, Wang Yuanyuan, Rong Liang, et al. Expressway travel speed extraction model based on ETC transaction data [J]. Journal of Transport Information and Safety, 2017, 35(5):76-82.

[3] Han Daqian. Traffic flow analysis based on highway toll data [J]. Journal of Transportation Science and Engineering, 2015, 31(4):96-100.

[4] Hu Runxiu, Li Mengxue. Traffic volume survey based on highway toll data [J]. Journal of Shandong Jiaotong University, 2016, 24(1):21-26.

[5] Jiang Yiyue. Research on spatiotemporal correlation of traffic distribution based on highway toll data [D]. Beijing Jiaotong University, 2019.

[6] Han Kunlin. Research on highway abnormal state identification based on vehicle detector and toll data fusion [D]. Chongqing University, 2014.