THE VALUE OF ROUTE PLANNING DATA FOR ORIGIN-DESTINATION ESTIMATION AND PREDICTION

Cheng ZHANG [a], Chengyuan MA [b], Xiaoguang YANG [c], William H.K. LAM [d], Yuelong SU [e], Zhenning DONG [f]

[a] *The Key Laboratory of Road and Traffic Engineering, Ministry of Education, Tongji University, Shanghai, China;*
Department of Civil and Environmental Engineering, The Hong Kong Polytechnic University, Hong Kong, China
Email: cheng.zhang@tongji.edu.cn; cheng1.zhang@polyu.edu.hk
[b] *The Key Laboratory of Road and Traffic Engineering, Ministry of Education, Tongji University, Shanghai, China*
Email: machengyuan@tongji.edu.cn
c The Key Laboratory of Road and Traffic Engineering, Ministry of Education, Tongji University, Shanghai, China
Email: yangxg@tongji.edu.cn
[d] *Department of Civil and Environmental Engineering, The Hong Kong Polytechnic University, Hong Kong, China*
Email: william.lam@polyu.edu.hk
[e] *Traffic Management Solution Division, AutoNavi Software Co, Beijing, China*
Email: yuelong.syl@alibaba-inc.com
[f] *AutoNavi Software Co, Beijing, China*
Email: 1537446246@qq.com

ABSTRACT

The origin-destination (OD) estimation and prediction have received lots of attention over the last several decades. However, few approaches considered route planning data. The data come from mobile map applications where some travelers plan their trips before departure. To some extents, they imply future OD demand. Motivated by this, the paper compares the volumes of planned trips and actually occurred trips extracted from trajectory data. Strong correlation (more than 0.97) exists between them. Planned trips are generally more than actual trips, especially for long-distance travel (64% more). For morning and evening peak hours, their volumes fluctuate more. These conclusions show the value of route planning data for OD estimation and prediction. Specific approaches can be studied on the basis of this preliminary analysis, but they are not included in this paper.

Keywords: route planning data, origin-destination, estimation, prediction, mobile phone

1. INTRODUCTION

The OD demand is important for traffic management and control. However, it is difficult to achieve its true value due to the uncertainty and dynamics of traffic systems. The OD demand essentially origins from people's activities. Traditionally, trip surveys were conducted to obtain samples of the OD demand. The result can reflect general trends but cannot represent real-time changes.

The development of sensor technology enables to collect real-time traffic dynamics. Counting sensors are the most popular kinds of sensors, including loop detectors, microwave detectors, infrared sensors, etc. They collect traffic states like volume, occupancy, and speed at the aggregate level. Many approaches were developed to estimate OD based on them, such as the gravity model (Nihan 1982), Bayesian statistical approach (Maher 1983) and the generalized least square model (Cascetta 1984). Their objectives are to minimize the error between estimated traffic counts and the ground truth.

Later, individual vehicles can be distinguished via identification (ID). Vehicle ID can be regularly returned by mobile sensors (e.g. GPS devices and mobile phones) or collected by fixed sensors when vehicles pass through (e.g. Bluetooth sniffers and license plate readers). Individual vehicle movements provide additional information for OD estimation.

As for GPS data, Ásmundsdóttir (2008) used them for validation, while Parry and Hazelton (2012) took them as supplemental input for OD estimation. Yang, Lu, and Hao (2017) presented both types of models and compared their performance under different conditions. Meanwhile, mobile phones have become necessities in people's daily life. They record a series of positions as GPS data do (J. Ma et al. 2013). Iqbal et al. (2014) used mobile phone call data to track individual movements in a large spatial scale. Zhang et al. (2019) extracted OD from mobile navigation data and analyzed the influence of penetration rates to estimation performance. Also, Barcelö et al. (2010) incorporated Bluetooth data in their dynamic estimation via Kalman filtering. Rao et al. (2018) reconstructed vehicle paths based on automatic license plate recognition and then estimated OD. These estimation methods generated a deterministic result.

Some estimation methods also considered stochastic features of the OD demand. Shao et al. (2014) estimated not only mean but also covariance of OD from day-to-day traffic counts. Ma and Qian (2018) proposed a data-driven approach to estimate day-to-day and hour-to-hour dynamic OD. Both mean and variance of OD are estimated and analyzed, which led to a better understanding of OD evolution.

For proactive traffic management and control, real-time OD estimation is still late and OD prediction is more important. On the basis of estimation, prediction approaches have been developed. Ashok and Ben-Akiva (2000) utilized filtering approach to combine historical and estimated O–D information in their OD prediction. The prediction showed significant advantage compared to historical values. Zhou and Mahmassani (2007) formulated the OD prediction as a time series problem. Djukic, Van Lint, and Hoogendoorn (2012) applied principal component analysis to significantly decrease the time series of OD.

With the development of mobile internet technology, more types of data can be considered into OD estimation and prediction. So far, there are only researches on how to plan routes with real-time information. The data generated from the route planning process actually represent real-time travel demand. More importantly, the demand can be captured before people enter road networks. Hence, it can lead to more accurate OD prediction. This paper preliminarily analyzes the value of route planning data. Section 2 describes the route planning data. Section 3 compared the number of planned trips and the ones actually taking place. Section 4 concludes this paper.

2. DATA DESCRIPTION

Route planning data are generated when travellers use mobile navigation applications such as Google Map and Amap for route suggestion. To plan routes, travelers input origins and destinations of the trips. The applications then provide several alternative routes as shown in Figure 1, together with estimated travel time and trip distance.

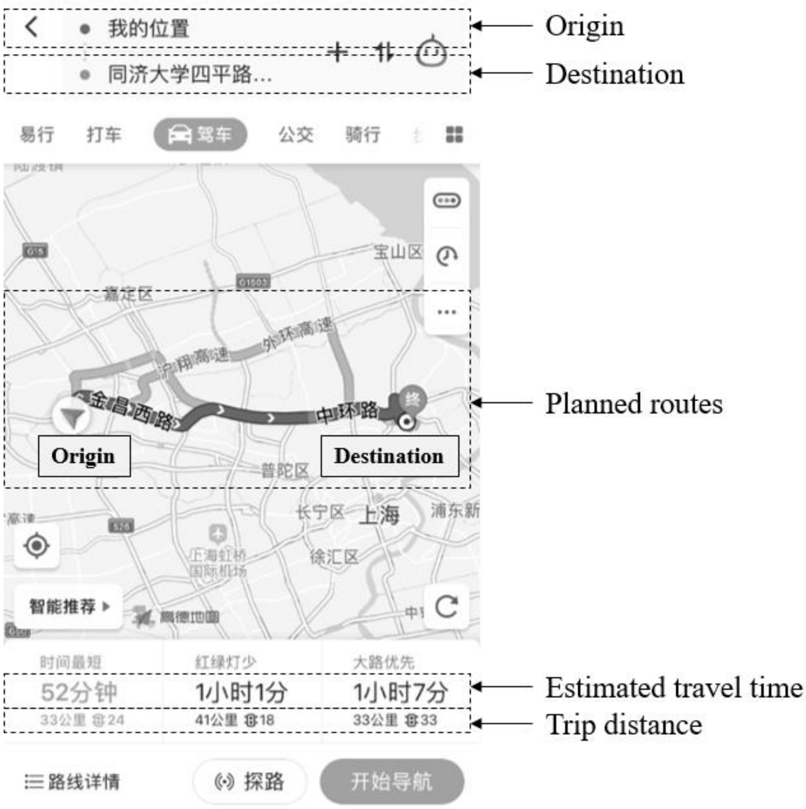

Figure 1. The interface of route planning

Table 1 shows an example of route planning data. Key information includes user ID, departure time, the locations of origins and destinations in terms of longitude and latitude. These indicators constitute travel demand. Also, information of planned routes is included such as trip distance and estimated travel time and estimated time of arrival.

Table 1. An excerpt of route planning data

User ID	Time of departure	Longitude of origin	Latitude of origin	Longitude of destination	Latitude of destination	Estimated time of arrival	Trip distance	Estimated travel time
36e092***	2017/10/23 10:47:46	121.440894	31.313185	121.488701	31.292775	2017/10/23 11:06:46	6842|9511|7272	19|17|20
dc115b***	2017/10/23 10:47:50	121.377724	31.210949	121.547286	31.308036	2017/10/23 11:20:50	24799|23369|23019	33|42|46
778615***	2017/10/23 10:53:48	121.498084	31.345008	121.476746	31.293263	2017/10/23 11:07:48	7014|9191|7070	14|13|14
6a6d3e***	2017/10/23 10:59:11	121.425193	31.286574	121.230771	31.311462	2017/10/23 11:49:11	22007|23209|25762	50|48|42

3. DATA ANALYSIS

3.1 A case in Shanghai

This paper chooses the road network in Shanghai for case study. As illustrated in Figure 2, the city is divided into region A and region B. Region A is about 90 square kilometers located in downtown. It is further divided into 6 subregions (A1, A2, A3, A4, A5, A6). The rest part is region B.

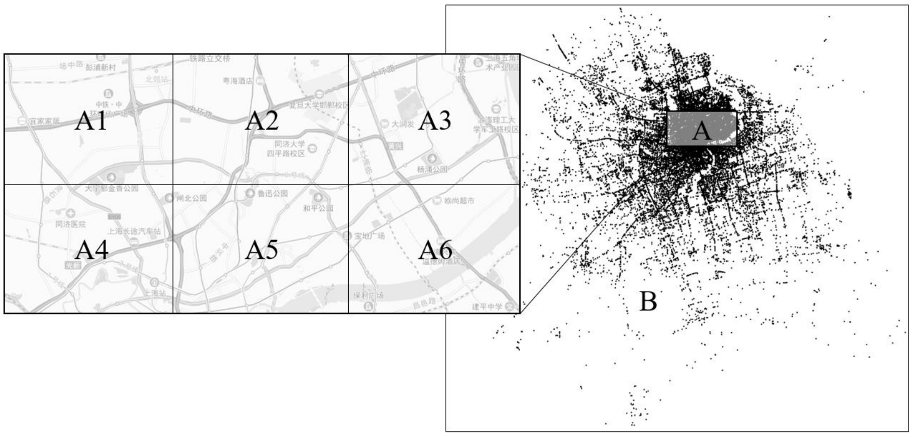

Figure 2. The map of the network in this case

Route planning data referred to region A is collected, i.e. either the origin or the destination is in region A. The time duration is from 23 October 2017 to 29 October 2017, a typical week, including about 4.2 million records.

3.2 Data processing

Besides route planning data, trajectory data are used to extract trips actually occurred in the network following the procedure in (Zhang et al. 2019). Then, OD pairs are aggregated as the OD volume for each hour. Let $X = \{x_1, \dots, x_n\}$ be the OD volume captured by route planning data and $Y = \{y_1, \dots, y_n\}$ be the actually occurred OD volume, where n is the size of the statistics, which equals to 168, the number of hours in a week. Pearson correlation coefficient ρ is utilized to evaluate the linear correlation between X (planned) and Y (actual).

$$\rho = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}\sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2}} \tag{1}$$

3.3 Temporal trends of the OD volume

Figure 3 compares the trip volumes within the same region (region A). The planned volumes fluctuate closely to the actual value. The correlation coefficient ρ equals 0.9957.

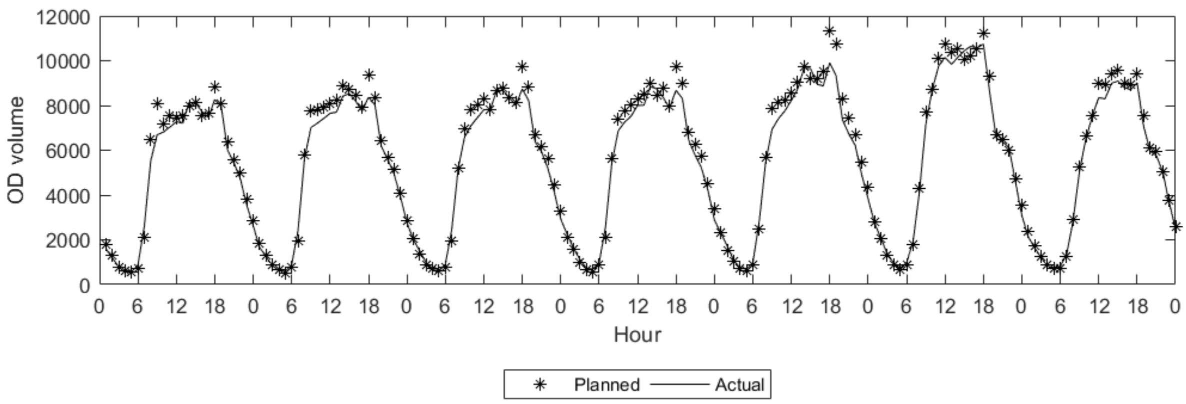

Figure 3. OD volume of A-A: planned vs. actual

Figure 4 compares the inter-region trip volumes. The planned volumes are averagely 64% larger than the actual value, but they still show strong correlation (ρ equals 0.9745). The reason may be that travelers tend to be unfamiliar for long-distance trips and they are more likely to use route planning service.

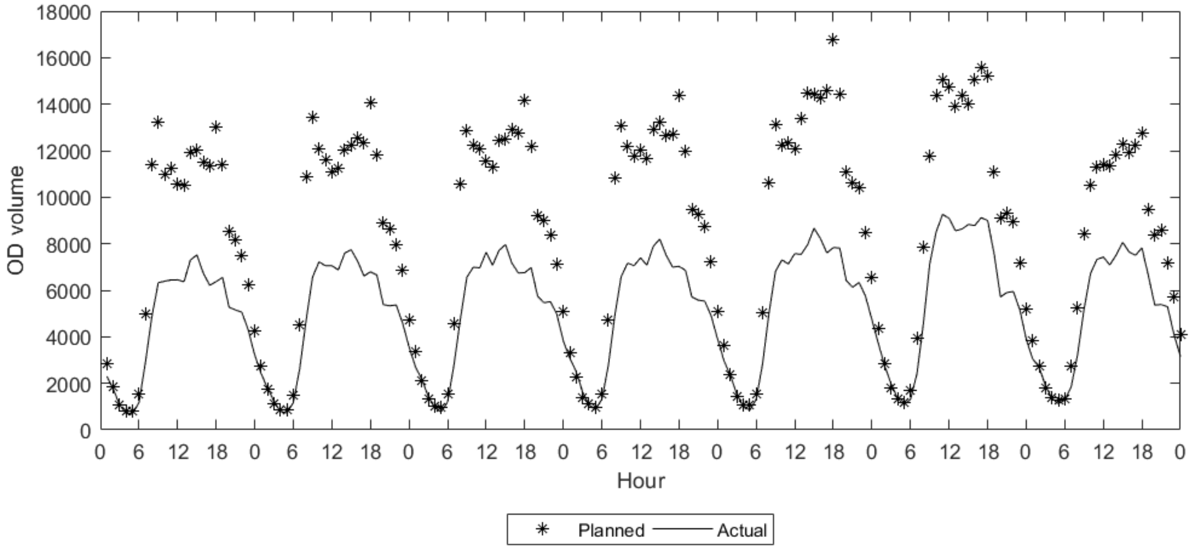

Figure 4. OD volume of A-B: planned vs. actual

3.4 Mean and variation of the OD matrix

OD matrix within region A in the morning peak hour (8:00 a.m. – 9:00 a.m.) is achieved for seven consecutive days. Table 2 compares the mean of the planned and the actual OD matrix. Though some OD volumes show big difference, such as A1-A5 and A2-A6 bold in the tables, there exists strong relationships between two matrices in general (ρ equals 0.9771).

Table 2. The mean of the OD matrix in the morning peak: planned vs. actual

(a) Planned

O \ D	A1	A2	A3	A4	A5	A6
A1	432	148	69	253	**213**	56
A2	112	697	204	130	357	**152**
A3	44	170	451	58	114	191
A4	143	105	64	626	262	51
A5	71	184	84	199	843	135
A6	22	71	108	49	117	282

(b) Actual

O \ D	A1	A2	A3	A4	A5	A6
A1	465	190	61	221	**118**	19
A2	126	576	187	103	346	**72**
A3	41	181	395	33	161	137
A4	125	89	31	571	201	20
A5	56	208	99	131	894	128
A6	12	62	111	18	177	234

The variation of OD volumes is also considered. Coefficient of Variation (CV) is used as index. Table 3 manifests that the result from route planning data has larger variation than the actual records. On average, the CV of planned volumes is 0.2216 compared to 0.1947 of actual ones.

Table 3. The CV of the OD matrix in the morning peak: planned vs. actual

(a) Planned

O \ D	A1	A2	A3	A4	A5	A6
A1	0.2598	0.2274	0.2977	0.3223	0.3263	0.4248
A2	0.2472	0.2118	0.1475	0.1452	0.2024	0.3706
A3	0.1632	0.2639	0.1750	0.2696	0.1360	0.1895
A4	0.1495	0.1335	0.1303	0.1688	0.1717	0.3346
A5	0.2169	0.2225	0.2449	0.1480	0.1542	0.2133
A6	0.2488	0.1911	0.1805	0.2272	0.2713	0.1910

(b) Actual

O \ D	A1	A2	A3	A4	A5	A6
A1	0.2259	0.2220	0.2235	0.2170	0.2554	0.3955
A2	0.2305	0.1589	0.1704	0.1317	0.1174	0.2802
A3	0.3565	0.2405	0.0696	0.2424	0.1394	0.1721
A4	0.1818	0.1778	0.1942	0.1130	0.1889	0.2899
A5	0.1690	0.1959	0.1393	0.0819	0.1054	0.1379
A6	0.1060	0.1946	0.1447	0.4272	0.1445	0.1668

Similarly, the mean and the CV is calculated for the evening peak hour (17:00 a.m. – 18:00 a.m.) as Table 4 and Table 5 show. The correlation coefficient ρ of two matrices is 0.9876. The average CV of planned volumes is 0.1638 and actual ones 0.1529.

Table 4. The mean of the OD matrix in the evening peak: planned vs. actual

(a) Planned

O \ D	A1	A2	A3	A4	A5	A6
A1	679	143	77	252	149	36
A2	154	809	309	154	342	107
A3	71	230	818	77	150	179
A4	237	123	87	848	309	61
A5	164	334	189	308	1462	180
A6	32	90	182	51	188	358

(b) Actual

O \ D	A1	A2	A3	A4	A5	A6
A1	639	178	55	241	112	17
A2	196	710	234	116	344	72
A3	72	231	610	43	177	140
A4	243	114	40	790	237	27
A5	126	378	171	272	1501	187
A6	22	87	172	28	220	287

Table 5. The CV of the OD matrix in the evening peak: planned vs. actual

(a) Planned

O \ D	A1	A2	A3	A4	A5	A6
A1	0.0884	0.1607	0.2311	0.1258	0.1278	0.2518
A2	0.1270	0.1016	0.1840	0.1967	0.1545	0.1161
A3	0.2458	0.1891	0.2066	0.1548	0.2215	0.1748
A4	0.1846	0.1077	0.1905	0.0684	0.0807	0.1813
A5	0.1847	0.1409	0.1783	0.1568	0.0959	0.1359
A6	0.2420	0.1942	0.1733	0.2325	0.2277	0.0652

(b) Actual

O \ D	A1	A2	A3	A4	A5	A6
A1	0.0671	0.1336	0.2642	0.1297	0.1360	0.2038
A2	0.1008	0.1034	0.2186	0.2088	0.2062	0.0985
A3	0.2574	0.1408	0.1928	0.3039	0.2129	0.1447
A4	0.1118	0.1074	0.2110	0.0578	0.1549	0.1255
A5	0.1457	0.0862	0.1690	0.1114	0.0971	0.0743
A6	0.3436	0.1446	0.1359	0.1688	0.0606	0.0749

4. CONCLUSION

This paper introduces potential value of route planning data for OD estimation and prediction. First, this new data type is described with an illustration of a map application interface. Then, OD volumes from route planning data and actual trajectory data are compared to study their correlation. Mean and variation of OD volumes are also analyzed using one-week data. Results show:

1) Route planning data can represent actual OD demand in a near future (with all correlation coefficients larger than 0.97);
2) Planned trips are generally more than actually occurred trips, especially for long-distance trips (64% more);
3) Compared to real OD volumes, the volumes of planned trips fluctuate more (with larger coefficients of variation).

This research reveals strong correlation between planned trips and actually occurred trips, i.e., some OD demand can be captured by route planning data in advance. This shows the value of route planning data for OD estimation and prediction. However, detailed methodologies are not included in this paper. Further research can be conducted on the basis of this preliminary analysis.

5. ACKNOWLEDGMENTS

This research was supported by the Research Grants Council of the Hong Kong Special Administrative Region, China (Project No. 152628/16E), National Natural Science Foundation of China (Grant No. 61773293) and Joint Laboratory for Future Transport and Urban Computing of Amap.

6. REFERENCES

Ashok, K., and M. E. Ben-Akiva. 2000. “Alternative Approaches for Real-Time Estimation and Prediction of Time-Dependent Origin–Destination Flows.” *Transportation Science* 34 (1): 21–36.

doi:10.1287/trsc.34.1.21.12282.

Ásmundsdóttir, Runa. 2008. "Dynamic OD Matrix Estimation Using Floating Car Data." Delft University of Technology. http://its-edulab.nl/files/MSc Thesis Runa Asmundsdottir.pdf.

Barcelö, Jaume, Lidin Montero, Laura Marqués, and Carlos Carmona. 2010. "Travel Time Forecasting and Dynamic Origin-Destination Estimation for Freeways Based on Bluetooth Traffic Monitoring." *Transportation Research Record: Journal of the Transportation Research Board* 2175 (1): 19–27. doi:10.3141/2175-03.

Cascetta, Ennio. 1984. "Estimation of Trip Matrices from Traffic Counts and Survey Data: A Generalized Least Squares Estimator." *Transportation Research Part B* 18 (4–5): 289–299. doi:10.1016/0191-2615(84)90012-2.

Djukic, Tamara, J. W. C. Van Lint, and S. P. Hoogendoorn. 2012. "Application of Principal Component Analysis to Predict Dynamic Origin–Destination Matrices." *Transportation Research Record: Journal of the Transportation Research Board* 2283 (1): 81–89. doi:10.3141/2283-09.

Iqbal, Md Shahadat, Charisma F. Choudhury, Pu Wang, and Marta C. González. 2014. "Development of Origin-Destination Matrices Using Mobile Phone Call Data." *Transportation Research Part C: Emerging Technologies* 40. Elsevier Ltd: 63–74. doi:10.1016/j.trc.2014.01.002.

Ma, Jingtao, Huan Li, Fang Yuan, and Thomas Bauer. 2013. "Deriving Operational Origin-Destination Matrices From Large Scale Mobile Phone Data." *International Journal of Transportation Science and Technology* 2 (3). Tongji University and Tongji University Press: 183–204. doi:10.1260/2046-0430.2.3.183.

Ma, Wei, and Zhen (Sean) Qian. 2018. "Estimating Multi-Year 24/7 Origin-Destination Demand Using High-Granular Multi-Source Traffic Data." *Transportation Research Part C: Emerging Technologies* 96 (September). Elsevier: 96–121. doi:10.1016/j.trc.2018.09.002.

Maher, Michael J. 1983. "Inferences on Trip Matrices from Observations on Link Volumes: A Bayesian Statistical Approach." *Transportation Research Part B* 17 (6): 435–447. doi:10.1016/0191-2615(83)90030-9.

Nihan, Nancy L. 1982. "Procedure for Estimating Freeway Trip Tables." *Transportation Research Record* d (1): 1–5.

Parry, Katharina, and Martin L. Hazelton. 2012. "Estimation of Origin-Destination Matrices from Link Counts and Sporadic Routing Data." *Transportation Research Part B: Methodological* 46 (1). Elsevier Ltd: 175–188. doi:10.1016/j.trb.2011.09.009.

Rao, Wenming, Yao-jan Wu, Jingxin Xia, Jishun Ou, and Robert Kluger. 2018. "Origin-Destination Pattern Estimation Based on Trajectory Reconstruction Using Automatic License Plate Recognition Data." *Transportation Research Part C* 95 (June). Elsevier: 29–46. doi:10.1016/j.trc.2018.07.002.

Shao, Hu, William H.K. Lam, Agachai Sumalee, Anthony Chen, and Martin L. Hazelton. 2014. "Estimation of Mean and Covariance of Peak Hour Origin-Destination Demands from Day-to-Day Traffic Counts." *Transportation Research Part B: Methodological* 68. Elsevier Ltd: 52–75. doi:10.1016/j.trb.2014.06.002.

Yang, Xianfeng, Yang Lu, and Wei Hao. 2017. "Origin-Destination Estimation Using Probe Vehicle Trajectory and Link Counts." *Journal of Advanced Transportation* 2017 (1). doi:10.1155/2017/4341532.

Zhang, Cheng, Jiawen Wang, Jintao Lai, Xiaoguang Yang, Yuelong Su, and Zhenning Dong. 2019. "Extracting Origin-Destination with Vehicle Trajectory Data and Applying to Coordinated Ramp Metering." *Journal of Advanced Transportation* 2019: 1–15. doi:10.1155/2019/8469316.

Zhou, Xuesong, and Hani S. Mahmassani. 2007. "A Structural State Space Model for Real-Time Traffic Origin-Destination Demand Estimation and Prediction in a Day-to-Day Learning Framework." *Transportation Research Part B: Methodological* 41 (8): 823–840. doi:10.1016/j.trb.2007.02.004.