How Far is Traffic from User Equilibrium?

ATRF 2017, Auckland, New Zealand

Mehmet Yildirimoglu[1] and Osman Kahraman[2]

[1] *University of Queensland, Australia*

[2] *University of Southern California*

Abstract

The interplay between road infrastructure, traffic conditions and travel choices is in the core of any infrastructure initiative, traffic control strategy or policy change. While the traditional traffic analysis framework assumes full awareness of travel costs and relies on the definition of user equilibrium state, there are few studies in the literature that test the existence of such equilibrium state. This study evaluates the widely applied shortest path assumption and so user equilibrium state from two aspects; (i) *user perspective*: similarity between actual and shortest paths, (ii) *network perspective*: node loads that result from actual and shortest paths. This study uses the GPS trajectory data set of approximately 20,000 taxis from Shenzhen, China to estimate the travel costs and reveal the actual routes followed in the network. It also considers two types of shortest paths that are based on free flow and estimated travel time. *User perspective* analysis concludes that most travellers do not choose the shortest path based on neither free flow nor estimated travel time. On the other hand, *network perspective* examination demonstrates similar network-wide patterns with actual routes and shortest paths based on estimated travel time.

INTRODUCTION

Understanding the complex interaction between road infrastructure, traffic conditions and travel choices (e.g. route choice, departure time) has been a long-standing challenge to understand mobility patterns in cities. The path travellers follow in the complex traffic networks has been arguably the most influential decision that dynamically redefines the relation between the supply provided by road infrastructure and the demand generated by travel plans of drivers. Nearly all techniques that aim for congestion alleviation build on accurate estimations of travel demand and thus on underlying route choices. The mainstream understanding of this complex interplay builds on the definition of user equilibrium (UE) state, namely Wardrop's first principle: "the journey times in all routes actually used are equal and less than those which would be experienced by a single vehicle on any unused route" (1).

While UE assumes that travellers have the perfect knowledge of travel costs along the network, and choose the routes that minimize their travel costs, stochastic user equilibrium (SUE) states that travellers might not be fully informed about network conditions, and therefore choose the routes that minimize their perceived travel costs. Shortest path algorithm is the straightforward way to establish UE conditions; however, there is a vast literature of discrete choice models that could be exploited to reach SUE conditions. The most appreciated models are Multinomial

Logit, C-Logit (2), Path Size Logit (3), Link-Nested Logit (4), Multinomial Probit (5) and Error-Component model (6). Route choice models are essential to estimate travel demand under hypothetical scenarios and predict traffic conditions in real-time. Nevertheless, modelling route choice is a challenging task given the complexity of human behaviour and uncertainty about travellers' perceptions. Additionally, calibration of the above discrete choice models is questionable because of computational complexity, lack of detailed data and occasionally over-fitting of the parameters. These models also require explicit enumeration of all alternative routes (i.e. choice set). In route choice analysis, alternative paths cannot be easily monitored, enumerated or extracted from the complex networks where the routes are hidden. While the literature of route choice models is vast, a detailed description is beyond the scope of this paper. The interested reader can refer to (7).

While congestion seems an unavoidable sign of a vibrant city with economic growth and social interactions, drivers increasingly take advantage of real-time information through GPS devices and smart phones. With everyone being easily monitored by the new sensors, researchers can start testing the assumptions behind the equilibrium state on network traffic patterns. There are few studies that test the empirical existence of Wardrop's first principle. (8) evaluate the habitual routes reported by 188 respondents. Assuming 90% overlapping is required to define two routes as the same, 37% of respondents follow a shortest time path (travel times are estimated using a traffic assignment model), and 22% follow a shortest distance path. Similarly, (9) design a Web-based survey and evaluate 236 routes between 182 OD pairs. Again, assuming 90% threshold for two routes to be considered the same, 26.7% of respondents choose the shortest distance path, while 17.8% choose the shortest time path. (10) collect GPS records from 143 participants and evaluate their route choice decisions in a period of 13 weeks where there is a disruptive event of bridge reopening. Their analysis concludes that about 40% of trips follow the shortest time path using a 10% overlapping threshold. Note that all studies are conducted with limited amount of data and test the equilibrium assumptions from a user perspective, i.e. similarity between actual and shortest path at the individual level.

In this work, we address this issue by coupling travel demand and traffic conditions, and analyse whether actual vehicle movements across cities are in line with equilibrium assumptions. To generate travel demand, we begin by mining massive GPS records from taxis. Using the taxi trips with passengers, we identify the origin and the destination locations within the city. Additionally, we apply a map-matching algorithm to identify the paths that vehicles actually follow in the network. We then parse the publicly available OpenStreetMap data and build the road network graph. OpenStreetMap provides the physical distance across the network and the category of roads which allows us to calculate free flow travel times. To determine actual traffic conditions, we process the taxi trajectories and estimate link travel times. We then explore the relationship between observed and shortest paths from two perspectives: (i) user perspective that compares individual path similarities and (ii) network perspective that is focused on traffic loads at the nodes. The findings from these two-layer comparison indicate a strong contrast; actual network state and equilibrium conditions are quite dissimilar from user perspective, while network-level analysis presents significant correspondence.

This paper is structured as follows. The next section describes the data set. The following section introduces the major findings of the user perspective analysis, and the one after that presents the results from network perspective inspection. The last section gives the discussion and the conclusions of the study.

DATA DESCRIPTION

The data set includes GPS tracks of around 20,000 taxis in a fast growing Chinese mega-city; Shenzhen. The rapid investment created one of the fastest-growing cities in the world with a population close to 11 million and, as expected, large congestion problems both in the urban and freeway system of the city. The network structure includes 10065 nodes and 23351 links. The data set consists of trips (on the same day) from taxis equipped with a GPS sensor that stores its location every 10-40 seconds. For every GPS point, it is also known whether the taxi carries a passenger or not, which allows us to distinguish between trips with and without passengers. Assuming that taxi passengers follow routes similar to regular cars in the network, we only focus on taxi trips with passengers. Even if taxi drivers might seek non-standard paths, we expect that speed estimates based on taxis with passengers are a good representation of all vehicles and that aggregated patterns are not influenced much by local low-level route choices.

In order to identify the paths that result from GPS records collected every 10-40 seconds, we apply a map-matching algorithm (11). This procedure results in the removal of 25% of the available trips due to the lack of feasible paths between successive pairs of GPS records, Despite this loss, we have around 190,000 trips in the data set which offers a wide network coverage to observe traffic conditions and travel patterns.

Figure 1 introduces the distribution of distance, average speed and duration for the successfully map-matched trips. Note that we do not observe any significant difference between map-matched and discarded trips in terms of distance and duration distribution, which implies that the process of map-matching does not create an additional bias in our analysis. From Figure 1, we see that most trips span a distance less than 10 km, and there are very few trips that cross more than 20 km. As this analysis only represents taxi trips with passengers, one can expect a different trip distance distribution for all vehicles. In particular, there may be more trips with longer distance in the actual distribution; however, as these long trips are expected to constitute a very small portion among all trips, we do not expect significant changes in our results. Figure 1 also shows that the average speed of a trip is distributed between 3km/h and 70 km/h. Considering the mixed nature of the study network (i.e. signalised arterial roads and freeways), no trip is able to exceed the average speed of 70 km/h. In addition, we see that most taxi trips take less than 15 min, and only few exceed 30 min.

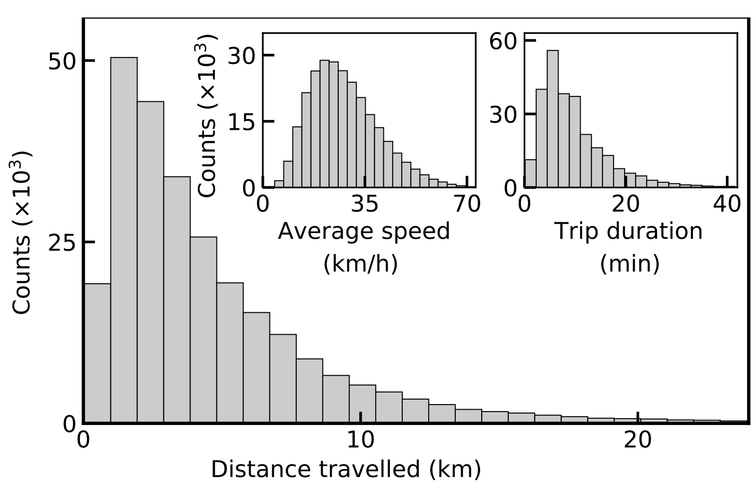


FIGURE 1: Trip distance, speed and duration distribution

In order to estimate the travel cost, we formally define the road network as a directed graph $G = (V, E)$ where V is the set of vertices or nodes, and E is the set of edges or links on the graph. For each road link indexed by $e \in E$, we call X_e the travel time and l_e length of the link. We also represent the set of link lengths as L. Denote N the number of GPS observations, and for each observation $n \in N$, the travel time between the start point and the end point is denoted x_n. We can represent the trajectory of observation n as a vector $\mathbf{w}_n$ of size E where $w_n^e = 1$, if the edge was taken during this observation and 0 otherwise. We use two different definitions to compute link travel times. The first definition of link travel time

$$X_e^{ff} = l_e/v_e \tag{1}$$

is based on free flow traffic conditions where we set free flow speed limit $v_e = 100$ km/h for freeways and $= 40$ km/h for urban roads. In the second definition, we estimate link travel time using the following formula.

$$X_e^{est} = \left(\sum_{n \in N} \mathbf{w}_n \cdot L\right) \cdot \left(\sum_{n \in N} \frac{l_e}{\mathbf{w}_n \cdot L} \cdot x_n\right)^{-1}. \tag{2}$$

The first term in Equation 2 refers to total distance travelled on link e, while the second term presents total time spent assuming vehicles have a constant speed between two successive GPS observations. This equation, therefore, provides an estimation of the space mean speed on link e.

USER PERSPECTIVE

In this section, we compare the actual routes and the shortest paths in terms of overlap percentages. In order to determine the shortest path, we use two types of travel cost; free flow travel time X_e^{ff} of Equation 1 that results from speed limits in road classes and estimated travel time X_e^{est} that follows Equation 2. Figure 2 presents route overlaps between the actual paths and the shortest paths based on two travel costs defined above. If two routes completely overlap, the difference should be 0. If they do not overlap at all, the difference should be 100%. Using the most strict overlap definition, we observe that around 29% of paths are in full compliance with shortest paths that rely on either one of the travel costs. Considering 5% threshold to consider two paths the same, this value goes up to 33% and 35% for free flow and estimated travel time, respectively. More importantly, even though estimated travel time consistently produces more paths with less than 30% difference, the overall distribution of values does not seem to be largely affected by the travel cost definition.

The number of alternative routes, and thereby route choice behavior is not the same for all origin and destination pairs. For example, when the trip covers a short distance, the entire path could consist of one straight road and even the second shortest path would be much longer. On the other hand, if the distance is very long, the common behavior would be to reach the nearest freeway connection and get off at the nearest interchange to the destination point, which would create a very high overlap percentage.

In order to investigate the relation between trip distance and overlap percentage, we have divided the data set into groups with respect to euclidean distance between origin and destination points. Figure 3 presents the percentage of routes that are equivalent to shortest paths based on both free flow and estimated travel time. As Figure 2 shows, 5% could be a natural cut to define two paths the same. Therefore, we classify routes that differentiate from the shortest path by less than 5% in length as those following the shortest time path. As expected, travelers are more likely

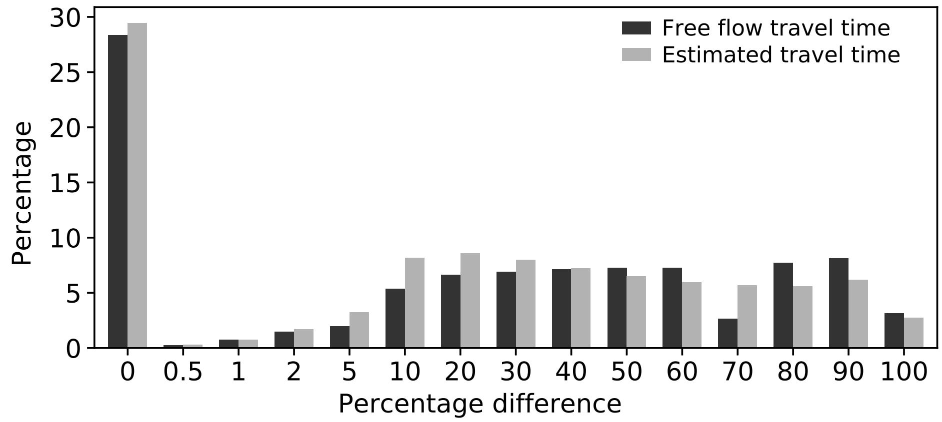


FIGURE 2: Difference between actual routes and shortest paths

to follow the shortest path when the trip is short. Interestingly, free flow travel time performs slightly better than estimated travel time in the shortest distance group. With free flow travel time, the percentage drops very fast as the trip distance increases; it reproduces only 5% of the actual routes in the longest distance group. On the other hand, despite the significant decrease in overlap percentage, estimated travel time seems to converge to a value around 17% with increasing distance. Our expectation regarding long distance trips seems to be valid only for a small portion of travelers and only for shortest paths that are based on estimated travel time.

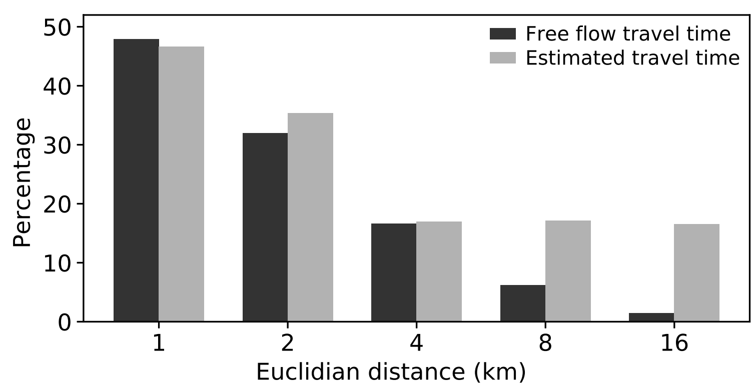


FIGURE 3: Percentage of same paths with respect to distance groups

In overall, the analysis from user perspective does not provide new insights into the empirical existence of user equilibrium in traffic networks. Our findings are in line with previous studies (8, 9, 10); even with an overlap tolerance, not more than 35% of trips follow the shortest path. Nevertheless, this observation does not fully explain "How far is traffic from user equilibrium?". There is a need for a novel approach from network perspective that investigates aggregated patterns resulting from the equilibrium assumptions. The next section intends to address this gap.

NETWORK PERSPECTIVE

As a result of the complex organization of mobility patterns in relation to road infrastructure, simple network representations of transportation systems are far from providing a comprehensive characterization of such systems within the framework of complex spatial networks (12). To disentangle the network of traffic flows from the underlying road infrastucture, here we follow Ref. (13) and use a two-layered network model with one physical and one logical layer. As illustrated in Figure 4(a), while the lower-layer physical graph $G^{\phi} = (V^{\phi}, E^{\phi})$ represents the physical road network with nodes $V^{\phi} \equiv V$ and edges $E^{\phi} \equiv E$, the upper-layer logical graph $G^{\lambda} = (V^{\lambda}, E^{\lambda})$ captures the traffic over the physical graph with nodes $V^{\lambda} \equiv V^{\phi}$ and edges $e^{\lambda} = (u, v)$ pairing the origin u and the destination v of traffic flows. In our case, both the physical and logical layers are directed and weighted graphs. Figure 4(b) shows node degrees $k^{\phi}_{\rm in}$ and $k^{\phi}_{\rm out}$ associated with, respectively, incoming and outgoing edges of the physical graph G^{ϕ}. Consistent with previous studies (14, 15), we find that most nodes have a degree ≤ 4, with a cumulative probability $P(k^{\phi} \geq 5) < 0.01$ and the average degree $\langle k^{\phi} \rangle \approx 2.32$ both for incoming and outgoing edges. Figure 4(c) shows that the distribution of node weights of the logical graph G^{λ} is highly right-skewed indicating that while a large portion of nodes has low weights (< 100), few nodes bear very large weights (> 1000).

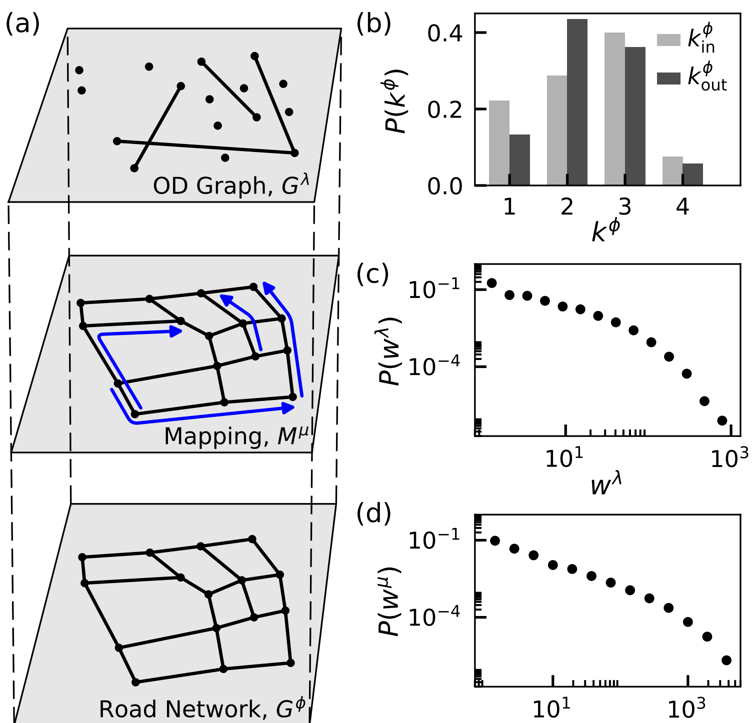
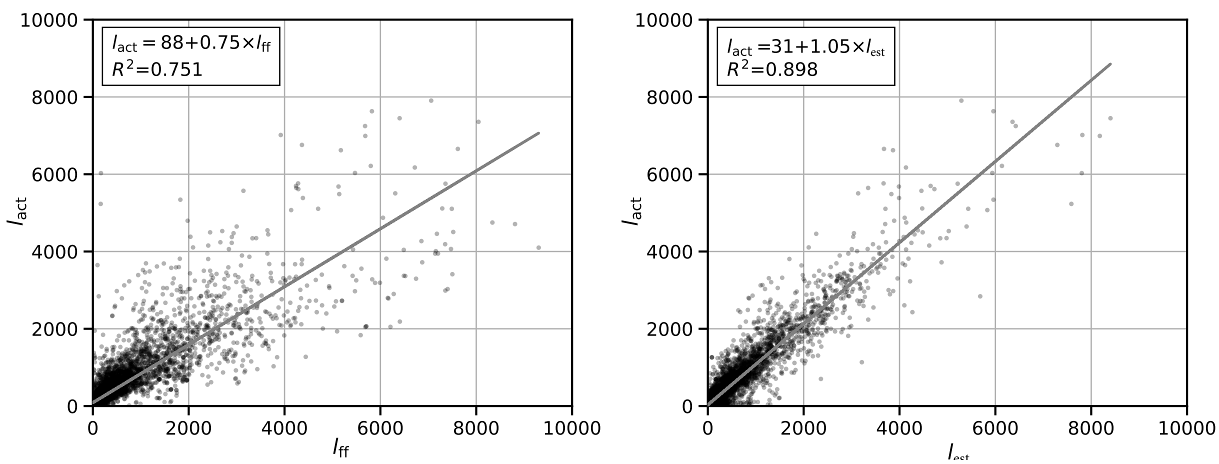


FIGURE 4: Description of the multilayered network approach. (a) Illustration of the two-layer network. (b) Node degree distribution of the physical layer. (c) Weight distribution of nodes in the logical layer G^{λ}. (d) Weight distribution of edges obtained from the mapping M^{μ} using actual taxi trajectories.

Each logical edge e^{λ} is mapped onto the physical graph as a path $M^{\mu}(e^{\lambda})$ navigating the traffic from u to v. This mapping may be given explicitly as the actual paths constructed from GPS tracking or may be defined implicitly through shortest paths. Figure 4(d) shows that edge weights obtained following the former scenario with actual paths obtained from GPS tracks have a right-skewed distribution similar to the distribution of logical weights shown in Figure 4(c). Finally, load l of a node u^{ϕ}, corresponding to the amount of traffic flowing through u^{ϕ}, is defined (13) as the sum of the weights of logical edges whose paths $M^{\mu}(e^{\lambda})$ cross u^{ϕ}.

$$l(u^{\phi}) = \sum_{e^{\lambda}:u^{\phi}\in M^{\mu}(e^{\lambda})} w(e^{\lambda}). \tag{3}$$

Figure 5(a) presents the node loads that result from actual routes as revealed by GPS tracks and map-matching implementation (i.e. l_{act}) and the node loads l_{ff} that are associated with shortest paths based on free flow travel time X_e^{ff}. There is a strong proportionality between two node load types; Pearson's linear correlation coefficient is 0.86. This is a very high score; however, we notice that scatter in the plot becomes more evident with increasing l_{act}, which indicates lower estimation quality for high load carrying components. Additionally, we fit a linear function of the form $l_{act} = a * l_{ff} + b$, where a and b are calculated as 0.75 and 88, respectively. Despite high correlation, linear regression results (a being far from 1) point out consistent overestimation of l_{act} and indicate strong bias in the estimation of traffic loads with free flow travel times.



(a)

(b)

FIGURE 5: Node loads with (a) free flow travel times, (b) estimated travel times.

Similarly, Figure 5(b) depicts the node loads from actual routes (i.e. l_{act}) and the node loads l_{est} that are associated with shortest paths based on estimated travel time X_e^{est}. Pearson's linear correlation coefficient is significantly higher than the previous case; it is 0.95. More importantly, scatter seems to be homogeneous over the range of actual node loads. Lastly, we fit a linear function of the form $l_{act} = a * l_{est} + b$, where a and b are calculated as 1.05 and 31, respectively. This indicates a considerable improvement over the free flow travel time, as the value of a is much closer to 1.

To have a better understanding of estimation quality, we calculate mean absolute error (MAE) and root mean squared error (RMSE) with the following formulas.

$$\text{MAE}_x = \left[\sum_V (l_{act}(v) - l_x(v))\right] /|V| \tag{4}$$

$$\text{RMSE}_x = \left[\left(\sum_V (l_{act}(v) - l_x(v))^2\right) /|V|\right]^{1/2} \tag{5}$$

where $l_x(v)$ is the estimated load at node or vertice v with either free flow or estimated travel time. With free flow travel times, MAE and RMSE are 149 and 405, respectively. Estimated travel times lead to much smaller error values; 99 and 232 for MAE and RMSE, respectively. This represents a 34% decrease in MAE and 43% reduction in RMSE. The findings of this analysis imply that drivers anticipate traffic conditions across the alternative routes, they do not make decisions based on free flow travel times, and equilibrium state provides a proper estimator of network-wide traffic patterns.

Figures 5(a) and 5(b) provide an aggregated comparison of actual and estimated traffic loads, while Figures 6 and 7 depict the distribution of estimation errors in the network and investigate the existence of hidden patterns that may lead to biased evaluations. To develop a relative assessment measure, we define the percentage estimation error: $(l_{act}(v) - l_x(v))/l_{act}(v) * 100$. Nevertheless, this approach may exaggerate the error measures for the nodes that carry low traffic load. Therefore, we define a tolerance level of 500 veh and classify the nodes that remain within the bounds. In other words, we choose the nodes for which the absolute difference between actual and estimated load is less than 500 veh; gray points in Figures 6 and 7 represent them. For the nodes outside the tolerance bounds, we calculate the percentage error. Green, yellow and red points illustrate the nodes for which the estimation error is less than 50%, between 50% and 100% and more than 100%, respectively. Most of the nodes remain in the tolerance bounds with both travel cost definitions; however, l_{ff} produces significantly more points that lie outside the interval than l_{est}. More importantly, the coloured points in Figure 6 (yellow and red points in particular) seem to compose major roadways that run along in the east-west axis, while the ones in Figure 7 are scattered around the network and do not indicate a pattern. This comparative observation implies that free flow travel times do not accurately represent the traffic conditions on these major roads, and drivers anticipate the actual conditions to a certain extent and make choices accordingly. Note that in both cases, most coloured points are located in the east of the city or CBD where there is a denser network of low-load roadway links.

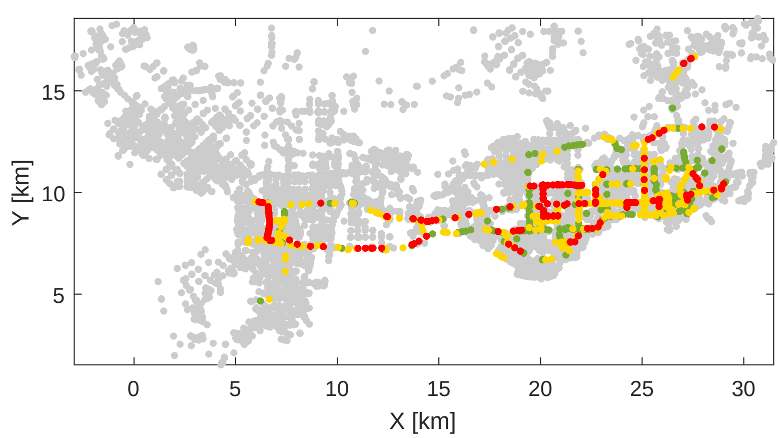


FIGURE 6: Distribution of estimation errors with l_{ff}

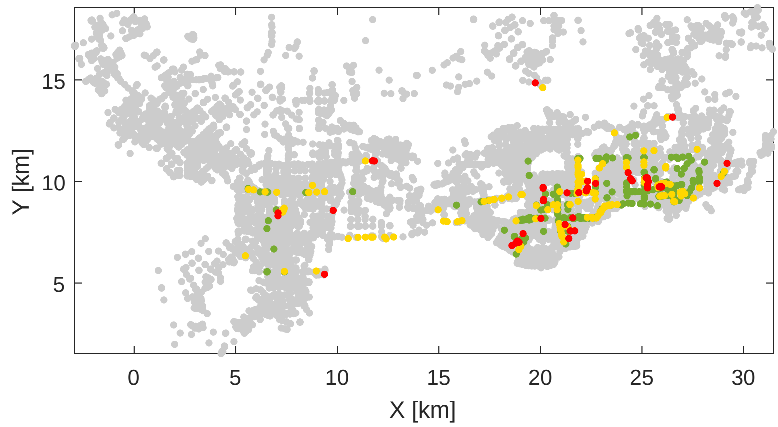


FIGURE 7: Distribution of estimation errors with l_{est}

CONCLUSION

This study empirically tests equilibrium assumptions from user and network perspective and using two travel cost definitions; free flow and estimated travel time. Free flow travel time is calculated using the distance of links and associated speed limit, while estimated travel time is computed with GPS observations and map-matching results. User perspective analysis does not indicate a significant difference between two travel cost definitions, and in both cases, it strongly rejects the assumption regarding widespread use of shortest paths. However, network perspective examination focuses on node loads and reveals significant differences between the two types of shortest path. Although free flow travel time does produce a strong correlation between actual and estimated node loads, results are strongly scattered, and evaluations of major roadways are biased. On the other hand, estimated travel time produces node loads that are very much in line with actual patterns.

Travellers clearly consider other alternatives when making route choice decisions, and all the observed alternatives cannot be captured by shortest paths or even advanced choice set generation algorithms (10). However, it is not clear how network-wide traffic patterns are affected by the mismatch of individual preferences. This paper is an attempt to fill this information gap. Despite the significant inconsistency of shortest paths at the individual level, the overall trend in the network loads seems to be captured by user equilibrium assumptions. Obviously, user equilibrium does not perfectly estimate the actual patterns observed in the network; estimations are not exactly positioned along a proportionality curve. However, it is not clear if and how well discrete choice models can estimate the network-wide patterns. This is a future research question to investigate.

REFERENCES

[1] Wardrop, J. G., Some Theoretical Aspects of Road Traffic Research. *Proceedings of the Institution of Civil Engineers*, Vol. 1, No. 3, 1952, pp. 325–362.

[2] Cascetta, E., A. Nuzzola, F. Russo, and A. Vitetta, A Modified Logit Route Choice Model Overcoming Path Overlapping Problems: Specification and Some Calibration Results for Interurban Networks. In *Proceedings of the 13th International Symposium on Transportation and Traffic Theory* (J. B. Lesort, ed.), Pergamon, Oxford, 1996, pp. 697–711.

[3] Ben-Akiva, M. E. and M. Bierlaire, Discrete Choice Methods and their Applications to Short-Term Travel Decisions. In *Handbook of Transportation Science* (R. Hall, ed.), Kluwer, Dordrecht, 1999, chap. 2, pp. 5–34.

[4] Vovsha, P. and S. Bekhor, Link-nested logit model of route choice: overcoming route overlapping problem. *Transportation Research Record: Journal of the Transportation Research Board*, , No. 1645, 1998, pp. 133–142.

[5] Daganzo, C. F., F. Bouthelier, and Y. Sheffi, Multinomial probit and qualitative choice: A computationally efficient algorithm. *Transportation Science*, Vol. 11, No. 4, 1977, pp. 338–358.

[6] Bekhor, S., M. Ben-Akiva, and M. Scott Ramming, Adaptation of logit kernel to route choice situation. *Transportation Research Record: Journal of the Transportation Research Board*, , No. 1805, 2002, pp. 78–85.

[7] Prato, C. G., Route choice modeling: past, present and future research directions. *Journal of choice modelling*, Vol. 2, No. 1, 2009, pp. 65–100.

[8] Bekhor, S., M. E. Ben-Akiva, and M. S. Ramming, Evaluation of choice set generation algorithms for route choice models. *Annals of Operations Research*, Vol. 144, No. 1, 2006, pp. 235–247.

[9] Prato, C. and S. Bekhor, Applying branch-and-bound technique to route choice set generation. *Transportation Research Record: Journal of the Transportation Research Board*, , No. 1985, 2006, pp. 19–28.

[10] Zhu, S. and D. Levinson, Do people use the shortest path? An empirical test of Wardrop's first principle. *PloS one*, Vol. 10, No. 8, 2015, p. e0134322.

[11] Lou, Y., C. Zhang, Y. Zheng, X. Xie, W. Wang, and Y. Huang, Map-matching for low-sampling-rate GPS trajectories. In *Proceedings of the 17th ACM SIGSPATIAL international conference on advances in geographic information systems*, ACM, 2009, pp. 352–361.

[12] Barthélemy, M., Spatial networks. *Physics Reports*, Vol. 499, No. 1, 2011, pp. 1–101.

[13] Kurant, M. and P. Thiran, Layered complex networks. *Physical Review Letters*, Vol. 96, No. 13, 2006, p. 138701.

[14] Buhl, J., J. Gautrais, N. Reeves, R. Solé, S. Valverde, P. Kuntz, and G. Theraulaz, Topological patterns in street networks of self-organized urban settlements. *The European Physical Journal B*, Vol. 49, No. 4, 2006, pp. 513–522.

[15] Cardillo, A., S. Scellato, V. Latora, and S. Porta, Structural properties of planar graphs of urban street patterns. *Physical Review E*, Vol. 73, No. 6, 2006, p. 066107.