

# Understanding User Equilibrium States of Road Networks using Big Trajectory Data

Bi Yu Chen<sup>1\*</sup>, Xuan-Yan Chen<sup>1\*</sup>, Hui-Ping Chen<sup>2\*</sup>

<sup>1</sup>State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, Wuhan 430079, China - (chen.biyu, xuanyanc)@whu.edu.cn

<sup>2</sup>School of Management, Huazhong University of Science and Technology, Wuhan 430074, China – chenhuiping@hust.edu.cn

**KEY WORDS:** User equilibrium, route choice behavior, big trajectory data, transport planning

## ABSTRACT:

User equilibrium (UE) has long been regarded as the cornerstone of transport planning studies. Despite its fundamental importance, our understanding of the actual UE state of road networks has remained surprisingly incomplete. Using big datasets of taxi trajectories, this study investigates the UE states of road networks in Wuhan. Effective indicators, namely relative gaps, are introduced to quantify how actual traffic states deviate from theoretical UE states. Advanced machine learning techniques, including XGBoost and SHAP values, are employed to analyze nonlinear relationships between network disequilibrium states and seven influencing factors extracted from trajectory data. The results reveal significant gaps between actual traffic states and the theoretical UE states at various times of the day during both weekdays and weekends. The XGBoost analysis shows that differences in travel distances, travel speeds, and signalized intersection numbers among alternative routes are the primary causes of road network disequilibrium. The results of this study could have several important methodological and policy implications for using the UE models in transport applications.

## 1. INTRODUCTION

User Equilibrium (UE) theory (Wardrop, 1952) has long been regarded as the cornerstone of transport planning studies. It assumes that all travelers, given perfect traffic knowledge, select their routes for minimal travel times. As a result, travel times of all used routes are equal and minimum, and travel times of all unused routes are greater than or equal to the travel times of used routes. Under UE state, travelers can not reduce the travel time by unilaterally changing the route between the origin and destination (OD) pair. UE is widely used in various transport planning applications, such as transport network designs, disaster evacuating planning, etc. (Sheffi, 1985; Yang and Bell, 1998; Chen et al., 2012; Fu et al., 2022).

Although UE model is elegant, its two behavioral assumptions have been recognized too strong to some extent (Garcia-Sierra et al., 2015; Havlíčková and Zámečník, 2020; Giannotti et al., 2011). It is generally difficult for travelers to acquire the completed traffic conditions of the road network due to their limited spatial knowledge and reasoning abilities. Numerous empirical studies using survey data have shown that several other criteria could be considered in travelers' route choice decisions, such as shorter travel distance, higher reliability and more major roads use.

In view of this, several traffic assignment models have been proposed to model travelers' fine-grained route-choice behaviors by incorporating travel time perception errors, psychological and behavioral mechanisms. State-of-the-art models include stochastic user equilibrium model (Daganzo and Sheffi, 1977; Sheffi and Powell 1982), random utility maximization model (Bowman and Ben-Akiva, 2001; Habib 2011), reliability-based model (Lam et al., 2008; Chen et al., 2011), regret minimization model (Chorus, 2012), rational choice theory model (Mahmassani and Chang, 1987; Han et al., 2015; Lou et al., 2010; Di and Liu, 2016; González Ramírez et al., 2021), prospect theory model (Ben-Elia and Shiftan, 2010;

Gao et al., 2010; Xu et al., 2011; Kahneman and Tversky, 1979) and stochastic network equilibrium with routing inertia (Xie et al. 2014). Despite their sophistication, how well these models capture the actual traffic state remains unclear. In practice, the UE model still serves as a dominated traffic assignment model in various transport applications.

Technological advancements have made it possible to collect abundant vehicle trajectories, particular taxi trajectories (Yildirimoglu and Kahraman, 2018; Calabrese et al., 2013; Rayle et al., 2016; Chen et al., 2023). Taxi trajectories have provided an excellent opportunity to examine real-world route choice behaviors in large-scale transport networks. Many empirical studies have utilized taxi trajectory data to analyze individual-level route choice behaviors and indicated the necessity to revisit the basics of how routes are chosen. For example, Ma et al. (2020) estimated network disequilibrium levels using big trajectory data in Chengdu and Pittsburgh and introduced a traffic management strategy for optimal routing (Ma et al. 2020). Manley et al., (2015) utilized a large dataset of nearly 700,000 taxi routes in London to observe the route choice behaviors. They found that travelers prefer to choose anchor-based routes (i.e., major roads, roads with well-known places, and etc.) rather than the shortest distance routes. In addition, a study based on the GPS traces of 20,000 taxis collected in Shenzhen implied that travelers do not substantially anticipate the existing traffic conditions when making their route choice decisions (Yildirimoglu and Kahraman, 2018). Through analysis of 496 participants recorded 5,535 choices over 41 OD pairs in Lyon (González Ramírez et al., 2021), the results confirmed that travelers evaluate relative rather than absolute differences in travel times of different routes. However, there has been little attention in the literature on using taxi trajectory data to investigate the collective patterns of individual route choice behaviors, i.e., the UE state of road networks.

The research objectives of this study are twofold: to quantify how much the actual traffic states of road networks deviate from

\* Corresponding authors: chen.biyu@whu.edu.cn (B.Y. Chen), xuanyanc@whu.edu.cn (X.-Y. Chen), chenhuiping@hust.edu.cn (H.-P. Chen)

the theoretical UE state; and to examine what factors influence the disequilibrium states of road networks. To fulfill these research objectives, we collect big datasets of one-month taxi trajectories in Wuhan, a Chinese mega-city. Using the collected dataset, travel times of all used routes by taxis between numerous OD pairs are exacted. Effective indicators, namely relative gaps, are introduced to quantify the degrees of disequilibrium states at two different levels, i.e., the OD level and the network level. The evolutions of disequilibrium states at different times of the day for both weekdays and weekends are examined. Seven key factors affecting disequilibrium states of road networks are identified. Advanced machine learning techniques, i.e., XGBoost and Shapley value, are employed to analyze nonlinear relationships between disequilibrium states and seven influencing factors. The results will deepen our understandings of UE states in real road networks and provide methodological and policy implications in transport applications.

## 2. STUDY AREA AND DATA COLLECTION

Wuhan, the largest city in central China, is selected as the study areas. Because there are numerous lakes and rivers (e.g., the Yangtze River) pass through Wuhan, the road network structure is complex with many bridges and tunnels. By the end of 2009, the vehicle parc in Wuhan was approximately 0.9 million.

The taxi trajectory dataset from Wuhan offers a comprehensive insight into the city's dynamism and complexity of the urban mobility patterns. The characteristics of the dataset in Wuhan are summarized in Table 1.

**Table 1.** The characteristics of taxi trajectory dataset in Wuhan.

No. of days	No. of taxis	No. of GPS points	Average sampling time
29	10,790	380 million	47 seconds

## 3. METHOD

### 3.1. Relative Gap Indicators

Relative gap was widely used as an indicator of whether the traffic assignment model converges to an equilibrium solution (Rose et al. 1988, Chen et al. 2011, Patil et al. 2021). In this study, the relative gap is introduced to quantify the UE state of a road network using trips data extracted from taxi trajectories.

Let  $o$  and  $d$  be the origin and destination nodes respective. Between each OD pair, there has a set of alternative routes  $R_{od} = \{\dots, r_{od}^i, \dots\}$  used by taxi drivers. Let  $t_{od}^i$  and  $f_{od}^i$  be average travel time and taxi flows of route  $r_{od}^i$  during the  $\tau$ th time interval. Both  $t_{od}^i$  and  $f_{od}^i$  are directly extracted from taxi trajectory datasets. Let  $t_{od}^{min}$  be the least travel time between the OD pair during the  $\tau$ th time interval. It can be calculated in the road network by using the shortest path algorithm. The detailed procedure for calculating  $t_{od}^i$ ,  $f_{od}^i$  and  $t_{od}^{min}$  are described in Section 3.2.2. Then, the relative gap at the OD level, denoted by  $GAP_{od}$ , describes the relative difference between used routes and the least travel time route during the  $\tau$ th time interval and it can be expressed as:

$$GAP_{od} = (\sum_{\forall i} f_{od}^i (t_{od}^i - t_{od}^{min})) / f_{od} t_{od}^{min} \quad (1)$$

where  $f_{od}$  is total taxi flows between the OD pair during the  $\tau$ th time interval. It can be calculated by:

$$f_{od} = \sum_{\forall i} f_{od}^i \quad (2)$$

Given the set of OD pairs, the relative gap for the whole network, denoted by  $GAP_{net}$ , describes the traffic state deviates from the theoretical UE state at snapshot  $\tau$  and it can be expressed as

$$GAP_{net} = (\sum_{\forall od} f_{od} t_{od}^{min} GAP_{od}) / (\sum_{\forall od} f_{od} t_{od}^{min}) \quad (3)$$

The value of  $GAP_{od}$  and  $GAP_{net}$  both range in  $[0, +\infty)$ . The value equal to 0 indicates the traffic state under the perfect UE state: taxi drivers between all OD pairs choose the least travel time routes and no taxi driver could improve their travel times by switching the routes. A larger value implies that the more observed traffic state (i.e., route choice behaviors of all taxi drivers) deviates from the theoretical UE state.

### 3.2. Data Analysis Procedure

#### 3.2.1. Step 1: Extracting Representative Routes Used by Taxis between OD Pairs

This process involves two stages. Firstly, taxi trajectory data is cleaned: abnormal GPS jumps and erroneous points are removed, and a map matching algorithm (Chen et al. 2014) is applied to correct positioning errors and reconstruct trajectories. The result is a continuous trajectory on the road network.

The second stage involves identifying commonly used taxi routes between origin-destination (OD) pairs. Trajectories are divided into OD trips using passenger boarding information. Then, representative routes are identified by clustering similar trips between the same OD pair using dynamic time warping and DBSCAN algorithm (Lima et al., 2016), resulting in valid OD pairs, their routes, and corresponding trips, as illustrated in Table 2.

**Table 2.** The characteristics of processed taxi trajectory data.

No. of OD pairs	No. of Routes	No. of Trips
6,253	121,031	1048575

#### 3.2.2 Step 2: Quantifying User Equilibrium States of Road Networks

This step is to quantify UE states at different time intervals of both two road networks using the introduced relative gap indicators. Based on the results of Step 1, it performs a two-phase procedure for each road network during every time interval. In this study, one hour is used as time interval  $\Delta$  for both road networks.

The first phase is to determine route travel time (i.e.,  $t_{od}^i$ ) and route taxi flows (i.e.,  $f_{od}^i$ ) for each route  $r_{od}^i \in R_{od}$  between the OD pair during the hourly time interval. To determine these  $t_{od}^i$  and  $f_{od}^i$ , we group taxi trips during the same OD pair during the same time interval according to their departure times. After this phase, we can determine route travel time  $t_{od}^i$  and route taxi flows  $f_{od}^i$  for all routes  $\forall r_{od}^i \in R_{od}$  between each OD pair during the hourly time interval. It is worth noting that we set  $f_{od}^j = 0$  for an unused route  $r_{od}^j \in R_{od}$  and it is not used in the relative gap calculation.

The second phase is to calculate the least travel time (i.e.,  $t_{od}^{min}$ ) between every OD pair during the time interval. We firstly

utilize map-matched trajectories of all taxis during the hourly time interval to estimate traffic conditions of the road network. The method of estimating hourly link travel times using taxi trajectories can be referred to Shi et al. (2017). We then employ the shortest path algorithm (Li et al., 2015) to calculate  $t_{od}^{min}$  for each OD pair. After performed this phase, we can calculate the UE state of a road network, i.e.,  $GAP_{net}$ , during the time interval using Eqs. (1-3). Consequently, we quantify the UE states of Wuhan networks for 540 hourly time intervals during one month, i.e., 18 hours/day \* 30 days.

### 3.2.3 Step 3: Analysis of Factors Affecting User Equilibrium States of Road Networks

This step is to utilize the advanced machine learning technique to examine factors affecting UE states of road networks by using extracted  $GAP_{net}$  for 540 hourly snapshots. First, seven factors in terms of OD pairs are extracted from the collected taxi trajectories at each snapshot  $\tau$ . Afterwards, according to taxi flow  $f_{od}$  between each OD, weighted to obtain the average value of each factor within the entire network range at each snapshot  $\tau$ . The results are (1) average alternative route numbers (denoted by  $n_{od}$ ); (2) average travel distance (denoted by  $\bar{d}_{net}$ ); (3) coefficient of variation (CV) of travel distances (denoted by  $cv_{net}^d$ ); (4) average travel speeds (denoted by  $\bar{v}_{net}$ ); (5) CV of travel speeds (denoted by  $cv_{net}^v$ ); (6) average signalized intersection numbers (denoted by  $\bar{s}_{net}$ ); and (7) CV of signalized intersection numbers (denoted by  $cv_{net}^{sg}$ ).

After extracting these factors, we examine their influences on  $GAP_{net}$  at 540 hourly snapshots as

$$GAP_{net} = F(X) \quad (4)$$

where  $X = (n_{net}, \bar{d}_{net}, cv_{net}^d, \bar{v}_{net}, cv_{net}^v, \bar{s}_{net}, cv_{net}^{sg})$  is the set of seven factors. Because these factors could be correlated and have nonlinear relationships with  $GAP_{net}$ , we cannot simply use the traditional multivariable linear regression method but utilize a powerful machine learning technique, namely eXtreme Gradient Boosting (XGBoost) (Chen and Guestrin, 2016; Ma et al., 2017; Li, 2022; Ji et al., 2022). Apart from a high regression accuracy, the XGBoost exhibits strong interpretative power by integrating with the SHapley Additive exPlanations (SHAP) (Shapley, 1953).

To calculate the SHAP relative importance  $\theta_i$  of each factor and the interaction term as:

$$\theta_i = \text{Mean}(|\phi_i|) / \sum \text{Mean}(|\phi_i|) \quad (5)$$

Where  $\phi_i$  is the SHAP value of each factor/interaction term. Thus, the value of the relative importance of each factor/interaction term is between 0 and 1, and the sum of the relative importance of each factor/interaction term is 1. The result represents the percentage of marginal contribution of each factor/interaction term to the relative gap.

## 4. RESULTS

### 4.1. Disequilibrium States of Road Networks

This section reports the network equilibrium states of Wuhan. We first examined the detailed equilibrium states at the OD level, i.e.,  $GAP_{od}$ , at a selected peak hour, i.e., 18:00-19:00 on the first Monday out of one month data. As shown in Figure 1, the red line links represent extreme disequilibrium OD pairs with

$GAP_{od} > 0.5$ , the blue line connects moderate disequilibrium OD pairs with  $GAP_{od} \in [0.05, 0.5]$ , and the green line indicates the equilibrium OD pairs with  $GAP_{od} < 0.05$ . It is easy to find that OD pairs at extreme and moderate disequilibrium states are predominant, while only a few OD pairs are at equilibrium states.

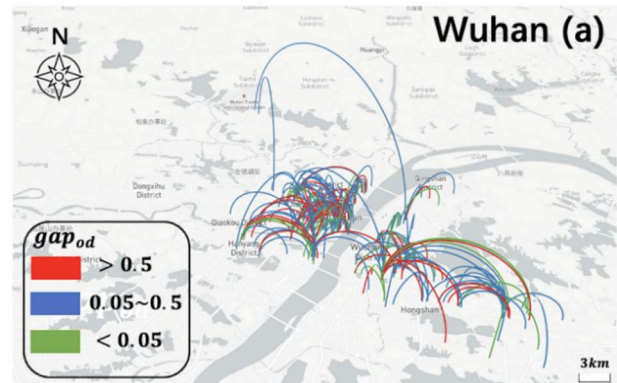


Figure 1. Network user equilibrium state in Wuhan

To quantify the  $GAP_{od}$  distribution for all OD pairs, we fitted the  $GAP_{od}$  probability density distribution into five pre-given types of distributions, including exponential, gamma, beta, lognormal and Pareto distributions. The goodness-of-fit of each distribution type was evaluated using the K-S test. As shown in Figure 2(b), the majority of time periods in Wuhan (82.5%) road networks obeyed the exponential distributions. As shown in Figure 2(a), during a typical peak hour, 12.3% of OD pairs reached equilibrium (i.e.,  $GAP_{od} < 0.05$ ); 28.6% of OD pairs were in slightly disequilibrium ( $GAP_{od} \in [0.05, 0.2]$ ); 32.3% were in moderate disequilibrium ( $GAP_{od} \in [0.2, 0.5]$ ); and 26.8% of OD pairs were in extreme disequilibrium ( $GAP_{od} > 0.5$ ). The overall relative gap for Wuhan network was  $GAP_{net} = 0.34$ , indicating a moderate disequilibrium state.

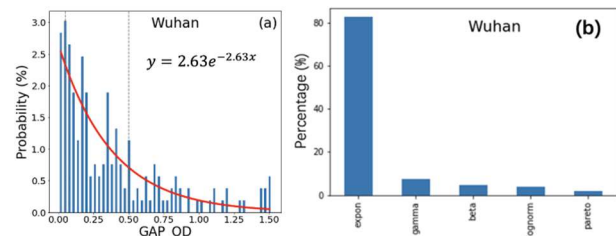


Figure 2. Distribution of  $GAP_{od}$  in Wuhan

We then investigated the equilibrium states at the network level (i.e.,  $GAP_{net}$ ) during different times of the day. Figure 4 reports the temporal patterns of road network for weekdays and weekends by calculating the average  $GAP_{net}$  as well as its 95% confidence intervals at the same time of the day. As shown in Figure 3(a), the  $GAP_{net}$  95% confidence intervals of the road network in weekdays ranged from 0.287 to 0.378 and fluctuated at different times of the day. The peak of  $GAP_{net}$  value occurred at the evening peak hour (17:00-18:00). The  $GAP_{net}$  value significantly raised at morning and afternoon. Between these two timeslots, the  $GAP_{net}$  was relatively stable. After 18:00,  $GAP_{net}$  value kept decreasing until the mid-night. As shown in Figure 3(b), the  $GAP_{net}$  value of Wuhan road network in weekends had a distinct pattern compared to that in weekdays. It ranged at a larger range of 95% confidence interval from 0.233 to 0.413. The  $GAP_{net}$  value gradually increased from the morning until the noon, and then fluctuated at a high level of disequilibrium state until the mid-night. This result highlights that Wuhan road

network kept moderate disequilibrium states at various times of the day for both weekdays and weekends.

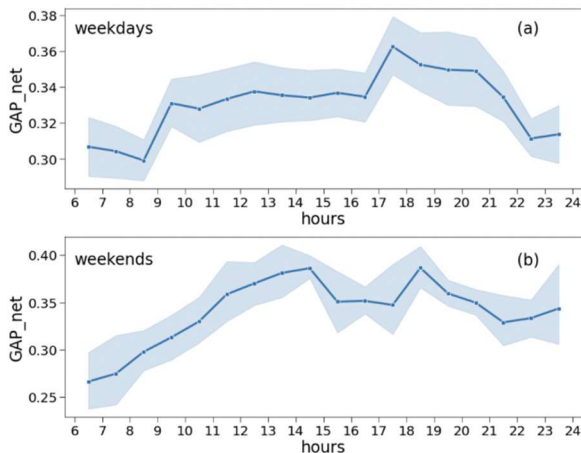


Figure 3. The relative gap pattern at different times of the day in Wuhan network.

#### 4.2. Factors Influencing Disequilibrium States of Road Networks

Table 3. Relative importance of influencing factors and interaction terms to  $GAP_{net}$

Features	Wuhan network	
	Relative Importance (%)	Ranking
<b>Influencing factors</b>	<b>90.6</b>	
DistMean ( $\bar{d}_{net}$ )	3.3	7
DistCV ( $cv_{net}^d$ )	40.7	1
SpeedMean ( $\bar{v}_{net}$ )	5.9	5
SpeedCV ( $cv_{net}^v$ )	16.7	2
CrossNum ( $\bar{s}g_{net}$ )	6.8	4
CrossCV ( $cv_{net}^{sg}$ )	11.7	3
RouteNum ( $n_{net}$ )	5.5	6
<b>Interaction terms</b>	<b>9.4</b>	
$R^2$	0.559	

$n_{net}$ : average alternative route numbers;  $\bar{d}_{net}$ : average travel distances;  $cv_{net}^d$ : CV of travel distances;  $\bar{v}_{net}$ : average travel speeds;  $cv_{net}^v$ : CV of travel speeds;  $\bar{s}g_{net}$ : average signalized intersection numbers;  $cv_{net}^{sg}$ : CV of signalized intersection numbers.

Figure 4 gives SHAP dependence plots to examine detailed relationships between influencing factors and  $GAP_{net}$  in Wuhan. Top 9 features, including seven factors ( $n_{net}$ ,  $\bar{d}_{net}$ ,  $cv_{net}^d$ ,  $\bar{v}_{net}$ ,  $cv_{net}^v$ ,  $\bar{s}g_{net}$ , and  $cv_{net}^{sg}$ ) and two interaction terms (DistCV\*SpeedCV and SpeedMean\*SpeedCV) are plotted, since they accounted for about 93% of total relative importance. In each SHAP dependence plot, the x-axis represents the corresponding feature value, and the y-axis gives the SHAP value indicating how much this feature impacts the prediction of  $GAP_{net}$ . Each dot represents a feature value of a  $GAP_{net}$  value. For interaction terms, the color of each dot represents the value of the other feature.

Based on the figure, following observations can be found:

Using XGBoost and SHAP techniques, we then examined how seven factors influencing network disequilibrium states. Table 3 presents the relative importance of seven influencing factors and their interaction terms. This relative importance ( $\theta_i$ ) represents the percentage of a factor's marginal contribution to  $GAP_{net}$ . The higher  $\theta_i$  value, the larger contribution to the prediction of  $GAP_{net}$ .

Overall, the XGBoost performed well for the Wuhan dataset with  $R^2 = 0.559$ . Among all factors and interaction terms in Wuhan city, the CV of travel distances  $cv_{net}^d$  made the largest contribution with  $\theta_i = 40.7\%$ . It was followed by the CV of travel speeds  $cv_{net}^v$  with  $\theta_i = 16.7\%$ , the CV of signalized intersection numbers  $cv_{net}^{sg}$  with  $\theta_i = 11.7\%$ , average signalized intersection numbers  $\bar{s}g_{net}$  with  $\theta_i = 6.8\%$ , average travel speeds  $\bar{v}_{net}$  with  $\theta_i = 5.9\%$ , average alternative route number  $n_{net}$  with  $\theta_i = 5.5\%$ , and the average travel distances  $\bar{d}_{net}$  with  $\theta_i = 3.3\%$ . All seven influencing factors predicted 90.6% of  $GAP_{net}$ , while other interaction terms predicted only the rest 9.4%.

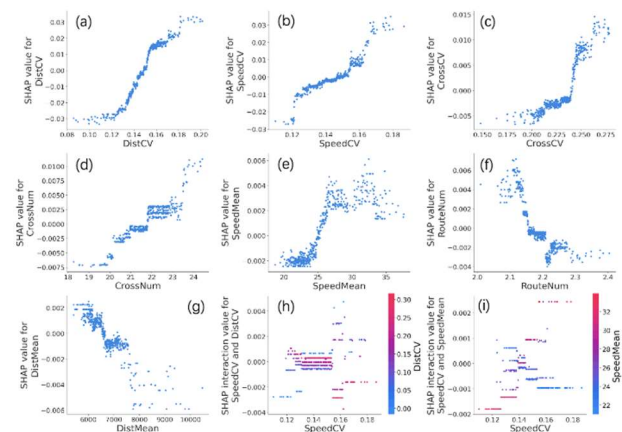


Figure 4. SHAP dependence plots of top nine features in Wuhan network.

- (a) When CV of travel distances ( $cv_{net}^d$ ) is below 0.15, its contribution to  $GAP_{net}$  is negative, but above this threshold it contributes positively.
- (b) The CV of travel speeds ( $cv_{net}^v$ ) behaves similarly; under 0.15 it reduces  $GAP_{net}$ , while above this, it increases it.
- (c) The CV of signalized intersection numbers ( $cv_{net}^{sg}$ ) negatively impacts  $GAP_{net}$  until 0.24, after which it positively affects it.
- (d) The average signalized intersection number ( $\bar{sg}_{net}$ ) decreases  $GAP_{net}$  under 21 but increases it beyond.
- (e) For average travel speeds ( $\bar{v}_{net}$ , under 25 km/h they reduce  $GAP_{net}$ , while over this speed they increase it, sharply until 27 km/h and then fluctuates at high level.
- (f) The average alternative route numbers ( $n_{net}$ ) contribute positively to  $GAP_{net}$  until 2.18, after which they contribute negatively.
- (g) Average travel distance ( $\bar{d}_{net}$ ) below 7km increases  $GAP_{net}$ , while above 7km it decreases it.
- (h) The interaction of  $cv_{net}^v$  and  $cv_{net}^d$  shows a complex interplay of similar travel distances and differences in travel speeds affecting the disequilibrium states.
- (i) The influence of  $cv_{net}^v$  and  $\bar{v}_{net}$  together on  $GAP_{net}$  depends on whether the average travel speed is less or more than 25 km/h, with each combination contributing differently to the disequilibrium states.

## 5. CONCLUSION AND DISCUSSION

This study investigated user equilibrium states of road networks in Wuhan using taxi trajectories. The user equilibrium states were explicitly evaluated by relative gap indicators at both network and OD levels, i.e.,  $GAP_{net}$  and  $GAP_{OD}$ . The advanced regression techniques, i.e., XGboost and SHAP, were employed to investigate the nonlinear relationships between seven factors and  $GAP_{net}$  values. Results found that road networks kept moderate and extreme disequilibrium states with  $GAP_{net} > 0.2$  at various times of the day for weekdays and weekends. Results also showed that  $GAP_{OD}$  obeyed exponential distributions during most time periods. Regression analysis found that the nonlinear relationships between seven influencing factors and  $GAP_{net}$  values can be well established by using the XGBoost method. CV factors ( $cv_{net}^v$ ,  $cv_{net}^d$ , and  $cv_{net}^{sg}$ ) were top three contributors to predict  $GAP_{net}$  values.

The results of network disequilibrium states in Wuhan provided several new insights on the route choice behavior field. Firstly, taxi trajectory data mining allowed the quantification of UE states on road networks. The results from Wuhan suggested that road networks tend to maintain moderate or extreme imbalance states at various times of the day, with different patterns appearing during weekdays and weekends. Therefore, this study provided strong empirical evidence to support the previous assertion that Wardrop's user equilibrium state is difficult to reach in real road networks at various times of the day during both weekdays and weekends (Daganzo and Sheffi 1977, Sheffi and Powell 1982, Lam et al. 2008).

Secondly, this work measures user equilibrium states at an intricate OD level, i.e.,  $GAP_{od}$ . The findings show that the majority of OD pairs were slightly or moderately imbalanced, which resonates with prior studies (Papinski and Scott 2011; Zhu and Levinson, 2015; Yildirimoglu Kahraman, 2018). This research extends those studies by exploring the distribution characteristics of  $GAP_{od}$  values.

Thirdly, we used XGBoost and SHAP to study non-linear relationships between seven influencing factors and  $GAP_{net}$  values. XGBoost's performance in Wuhan demonstrates the impact of these factors and their interaction terms in predicting  $GAP_{net}$ . Interestingly, CV factors ( $cv_{net}^v$ ,  $cv_{net}^d$ , and  $cv_{net}^{sg}$ ) were found to be more influential than mean factors ( $\bar{v}_{net}$ ,  $\bar{d}_{net}$  and  $\bar{sg}_{net}$ ) in predicting  $GAP_{net}$  (Train and Wilson, 2008; Manley et al. 2015; Yang et al., 2017).

Based on the findings, several methodological and policy implications can be derived. Transport planners and policymakers should be aware that Wardrop's user equilibrium state is hard to achieve in reality. The use of UE-based models can lead to significant bias in policy evaluation, especially during peak hours.

Despite the valuable insights, this study, being one of the first to investigate actual user imbalance states of road networks using taxi trajectory mining techniques, has its limitations. It used trajectories from over 10,000 taxis, which only form a small fraction of total daily travels. Also, taxi drivers may exhibit different driving behavior than other drivers. More data from private cars (Xiao et al. 2020) and ride-hailing services (Tirachini 2020) could enhance the evaluation accuracy.

Several future research directions can be proposed. Extending this study to multi-mode transport networks, investigating factors contributing to disequilibrium levels in different cities, or exploring equilibrium state at the OD pair level using the massive data available for individual OD pairs are all valuable next steps. Examining the consistency between theoretical equilibrium state and actual state from vehicle trajectory data can provide a more comprehensive understanding of traffic equilibria.

## ACKNOWLEDGMENTS

The work described in this paper was jointly supported by the National Key Research and Development Program (No. 2021YFB3900900), National Natural Science Foundation of China (No. 42271473), the Fundamental Research Funds for the Central Universities (No. 2042022kf1199), the Natural Science Foundation of Hubei Province (No. 2020CFA054), and LIESMARS special Research Funding.

## REFERENCES

- Ben-Elia, E. and Shiftan, Y. 2010. Which road do I take? A learning-based model of route-choice behavior with real-time information. *Transportation Research Part A*, 44, 249-264.
- Bowman, J. L. and Ben-Akiva, M. E. 2001. Activity-based disaggregate travel demand model system with activity schedules. *Transportation Research Part A*, 35(1), 1-28.
- Calabrese, F., et al. 2013. Understanding individual mobility patterns from urban sensing data: A mobile phone trace example. *Transportation Research Part C*, 26, 301-313.

- Chen, B.Y., Lam, W.H.K., Sumalee, A. and Shao, H., 2011. An efficient solution algorithm for solving multi-class reliability-based traffic assignment problem. *Mathematical and Computer Modelling*, 54, 1428-1439.
- Chen, B.Y., Lam, W.H.K., Sumalee, A., Li, Q.Q. and Li, Z.C., 2012. Vulnerability analysis for large-scale and congested road networks with demand uncertainty. *Transportation Research Part A*, 46, 501-516.
- Chen, B. Y., et al. 2014. Map-matching algorithm for large-scale low-frequency floating car data. *International Journal of Geographical Information Science*, 28(1), 22-38.
- Chen, B.Y., Luo, Y.-B., Jia, T., Chen, H.-P., Chen, X.-Y., Gong, J. and Li, Q., 2023. A spatiotemporal data model and an index structure for computational time geography. *International Journal of Geographical Information Science*, 37, pp. 550-583.
- Chen, T. and Guestrin, C., 2016. Xgboost: A scalable tree boosting system. In: *Proceedings of the 22nd ACM SigKDD international conference on knowledge discovery and data mining*, pp. 785-794.
- Chorus, C. G. 2012. Random regret minimization: An overview of model properties and empirical evidence. *Transport Reviews*, 32, 75 - 92.
- Daganzo, C. F. and Sheffi, Y. 1977. On stochastic models of traffic assignment. *Transportation Science*, 11(3), 253-274.
- Di, X. and Liu, H. X. 2016. Boundedly rational route choice behavior: A review of models and methodologies. *Transportation Research Part B*, 85, 142-179.
- Fu, X., Lam, W.H.K., Chen, B.Y. and Liu, Z., 2022. Maximizing space-time accessibility in multi-modal transit networks: an activity-based approach. *Transportmetrica A*, 18, pp. 192-220.
- Gao, S., Frejinger, E. and Ben-Akiva, M. E. 2010. Adaptive route choices in risky traffic networks: A prospect theory approach. *Transportation Research Part C*, 18, 727-740.
- Garcia-Sierra, M., Bergh, J. v. d. and Miralles-Guasch, C. 2015. Behavioural economics, travel behaviour and environmental-transport policy. *Transportation Research Part D*, 41, 288-305.
- Giannotti, F., et al. 2011. Unveiling the complexity of human mobility by querying and mining massive trajectory data. *VLDB Journal*, 20(5), 695-719.
- González Ramírez, H., et al. 2021. Travel time and bounded rationality in travellers' route choice behaviour: A computer route choice experiment. *Travel Behaviour and Society*, 22, 59-83.
- Habib, K. M. N. 2011. A random utility maximization (RUM) based dynamic activity scheduling model: Application in weekend activity scheduling. *Transportation*, 38(1), 123-151.
- Han, K., Szeto, W. Y. and Friesz, T. L. 2015. Formulation, existence, and computation of boundedly rational dynamic user equilibrium with fixed or endogenous user tolerance. *Transportation Research Part B*, 79, 16-49.
- Havlicková, D. and Zámečník, P. 2020. Considering habit in research on travel mode choice: a literature review with a two-level methodology. *Transactions on Transport Sciences*, 11, 18-32.
- Ji, S., Wang, X., Lyu, T., Liu, X., Wang, Y., Heinen, E. and Sun, Z., 2022. Understanding cycling distance according to the prediction of the XGBoost and the interpretation of SHAP: A non-linear and interaction effect analysis *Journal of Transport Geography*, 103, 103414.
- Kahneman, D. and Tversky, A. 1979. Prospect theory: analysis of decision under risk. *Econometrica*, 47, 263-291.
- Lam, W.H.K., Shao, H. and Sumalee, A., 2008. Modeling impacts of adverse weather conditions on a road network with uncertainties in demand and supply. *Transportation Research Part B*, 42, 890-910.
- Li, Q.Q., Chen, B.Y., Wang, Y. and Lam, W.H.K., 2015. A hybrid link-node approach for finding shortest paths in road networks with turn restrictions. *Transactions in GIS*, 19, pp. 915-929.
- Li, Z. 2022. Extracting spatial effects from machine learning model using local interpretation method: An example of SHAP and XGBoost. *Computers, Environment and Urban Systems*, 96, 101845.
- Lima, A., et al. 2016. Understanding individual routing behaviour. *Journal of The Royal Society Interface*, 13.
- Lou, Y., Yin, Y. and Lawphongpanich, S. 2010. Robust congestion pricing under boundedly rational user equilibrium. *Transportation Research Part B*, 44, 15-28.
- Ma, W., Quan, S. Measuring and reducing the disequilibrium levels of dynamic networks with ride-sourcing vehicle data. *Transportation Research Part C*, 110, 222-246, 2020.
- Ma, X., et al. 2017. Prioritizing influential factors for freeway incident clearance time prediction using the gradient boosting decision trees method. *IEEE Transactions on Intelligent Transportation Systems*, 18(9), 2303-2310.
- Mahmassani, H. S. and Chang, G.-L. 1987. On boundedly rational user equilibrium in transportation systems. *Transportation Science*, 21, 89-99.
- Manley, E., Addison, J. D. and Cheng, T. 2015. Shortest path or anchor-based route choice: A large-scale empirical analysis of minicab routing in London. *Journal of Transport Geography*, 43, 123-139.
- Papinski, D. and Scott, D. M. 2011. A GIS-based toolkit for route choice analysis. *Journal of Transport Geography*, 19, 434-442.
- Patil, P. N., Ross, K. C. and Boyles, S. D. 2021. Convergence behavior for traffic assignment characterization metrics. *Transportmetrica A*, 17(4), 1244-1271.
- Rayle, L., et al. 2016. Just a better taxi? A survey-based comparison of taxis, transit, and ridesourcing services in San Francisco. *Transport Policy*, 45, 168-178.
- Rose, G., Daskin, M. S. and Koppelman, F. S. 1988. An examination of convergence error in equilibrium traffic

- assignment models. *Transportation Research Part B*, 22, 261-274.
- Shapley, L.S., 1953. A value for n-person games. In: Kuhn, H. and Tucker, A., Eds., *Contributions to the Theory of Games II*, Princeton University Press, Princeton, pp. 307-317.
- Sheffi, Y., 1985, *Urban Transportation Networks: Equilibrium Analysis With Mathematical Programming Methods*. Prentice-Hall. New Jersey.
- Sheffi, Y. and Powell, W. B. 1982. An algorithm for the equilibrium assignment problem with random link times. *Networks*, 12(2), 191-207.
- Shi, C., Chen, B.Y. and Li, Q., 2017, Estimation of travel time distributions in urban road networks using low-frequency floating car data. *ISPRS International Journal of Geo-Information*, 6, p. 253.
- Sun, S. and Szeto, W. Y. 2021. Multi-class stochastic user equilibrium assignment model with ridesharing: Formulation and policy implications. *Transportation Research Part A*, 145, 203-227.
- Tirachini, A. 2020. Ride-hailing, travel behaviour and sustainable mobility: An international review. *Transportation*, 47(4), 2011-2047.
- Train, K. and Wilson, W.W., 2008, Estimation on stated-preference experiments constructed from revealed-preference choices. *Transportation Research Part B*, 42, pp. 191-203.
- Wardrop, J. G. 1952. Road paper: some theoretical aspects of road traffic research. *Proceedings of the Institution of Civil Engineers*, 1(3), 325-362.
- Xiao, Z., et al. 2020. On extracting regular travel behavior of private cars based on trajectory data analysis. *IEEE Transactions on Vehicular Technology*, 69(12), 14537-14549.
- Xie, C. and Liu, Z.L. On the stochastic network equilibrium with heterogeneous routing inertia. *Transportation Research Part B*, 66, 90-109, 2014.
- Xu, H.L., Lou, Y.Y., Yin, Y.F. and Zhou, J., 2011, A prospect-based user equilibrium model with endogenous reference points and its application in congestion pricing. *Transportation Research Part B*, 45, pp. 311-328.
- Yang, H. and Bell, M.G., 1998, Models and algorithms for road network design: a review and some new developments. *Transport Reviews*, 18, pp. 257-278.
- Yang, L., Kwan, M.-P., Pan, X., Wan, B. and Zhou, S., 2017, Scalable space-time trajectory cube for path-finding: A study using big taxi trajectory data. *Transportation Research Part B*, 101, pp. 1-27.
- Yildirimoglu, M. and Kahraman, O. 2018. Searching for empirical evidence on traffic equilibrium. *PLoS ONE*, 13.
- Zhu, S. and Levinson, D., 2015, Do people use the shortest path? An empirical test of Wardrop's first principle. *Plos One*, 10, p. e0134322.