ARTICLE

# Profiling mobility patterns and driving behaviors of individual drivers via trajectory trait

Yuhang Liu,[1,2] Zhipeng Gui,[2,3,4,5,*] Yang Xu,[6] Song Gao,[7] Anqi Zhao,[1] Fanhao Meng,[2,8] Dehua Peng,[2,3,4] Fa Li,[9] Lujia Bo,[2,10] Huayi Wu,[1,3,4] and Jianya Gong[1,2,3,4]

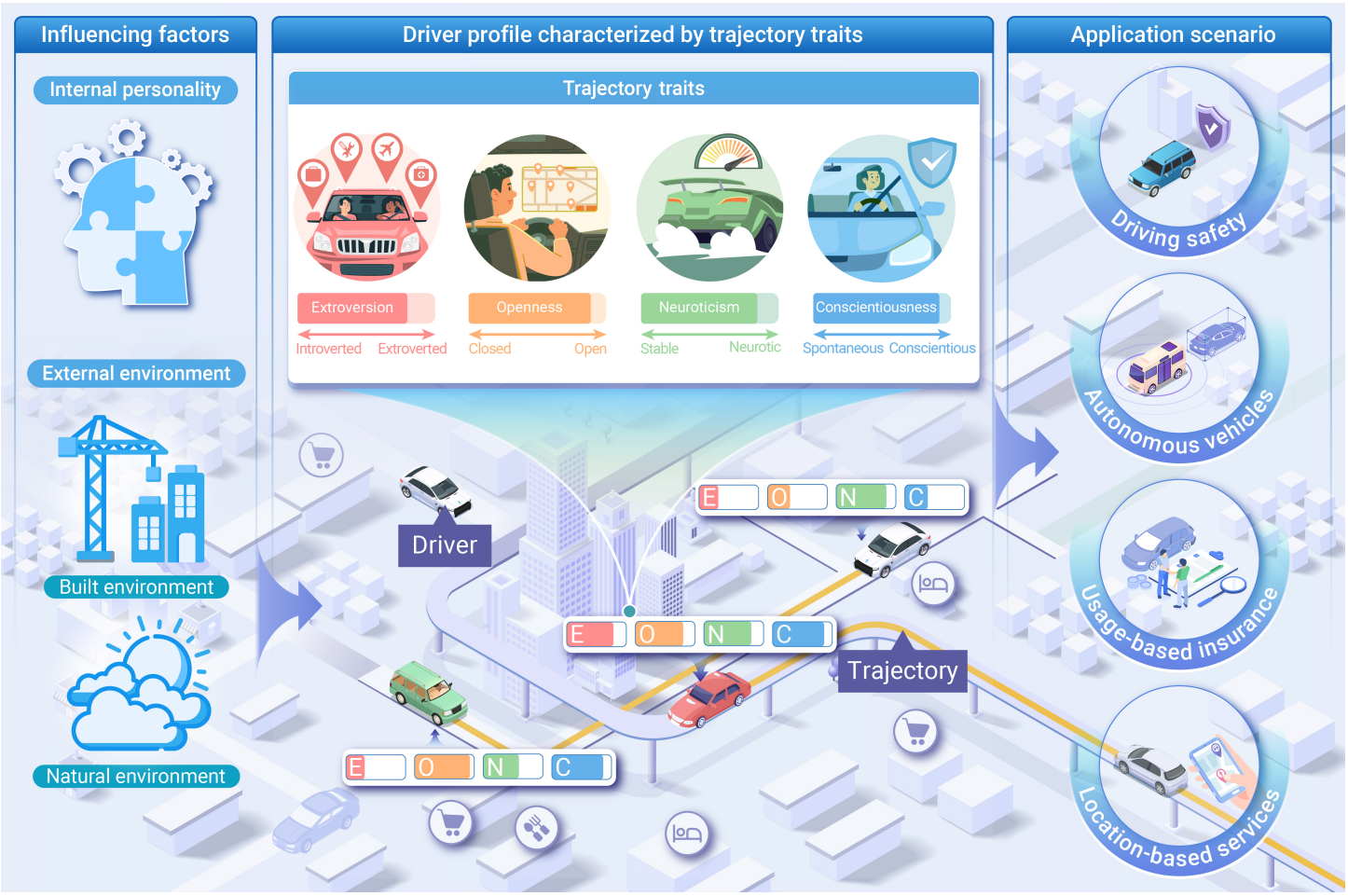
*Correspondence: zhipeng.gui@whu.edu.cn

Received: July 1, 2024; Accepted: December 11, 2024; Published Online: December 27, 2024; https://doi.org/10.59717/j.xinn-geo.2024.100114

© 2025 The Author(s). This is an open access article under the CC BY license (https://creativecommons.org/licenses/by/4.0/).

The Innovation Geoscience

## GRAPHICAL ABSTRACT




## PUBLIC SUMMARY

- Four trajectory traits are proposed to model driver profile: extroversion, openness, neuroticism, and conscientiousness.
- The consistency between the trajectory traits and external behavior observations is validated by experiments.
- Trajectory integrity, seasonal changes, and traffic conditions exert small but noteworthy impacts on profile stability.
- The proposed driver profile has broader potential for driving safety, advertising and marketing, sustainable transport.

ARTICLE

# Profiling mobility patterns and driving behaviors of individual drivers via trajectory trait

Yuhang Liu,[1,2] Zhipeng Gui,[2,3,4,5,*] Yang Xu,[6] Song Gao,[7] Anqi Zhao,[1] Fanhao Meng,[2,8] Dehua Peng,[2,3,4] Fa Li,[9] Lujia Bo,[2,10] Huayi Wu,[1,3,4] and Jianya Gong[1,2,3,4]

[1]State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, Wuhan 430079, China
[2]School of Remote Sensing and Information Engineering, Wuhan University, Wuhan 430079, China
[3]Collaborative Innovation Center of Geospatial Technology, Wuhan University, Wuhan 430079, China
[4]Hubei Luojia Laboratory, Wuhan 430079, China
[5]School of Geography and Planning, Ningxia University, Yinchuan 750021, China
[6]Department of Land Surveying and Geo-Informatics, The Hong Kong Polytechnic University, Hong Kong, China
[7]GeoDS Lab, Department of Geography, University of Wisconsin-Madison, Madison, Wisconsin 53715, USA
[8]Institute of Remote Sensing and Geographical Information Systems, Peking University, Beijing 100871, China
[9]Department of Earth System Science, Stanford University, Stanford, California 94305, USA
[10]School of Environment, Tsinghua University, Beijing 100084, China
*Correspondence: zhipeng.gui@whu.edu.cn

Received: July 1, 2024; Accepted: December 11, 2024; Published Online: December 27, 2024; https://doi.org/10.59717/j.xinn-geo.2024.100114

© 2025 The Author(s). This is an open access article under the CC BY license (https://creativecommons.org/licenses/by/4.0/).

Citation: Liu Y., Gui Z., Xu Y., et al. (2025). Profiling mobility patterns and driving behaviors of individual drivers via trajectory trait. *The Innovation Geoscience* **3**:100114.

Driver profiling can provide a human-centered approach to portraying individual travel behavior and revealing their motivation, objectives, and needs, thereby contributing to driving safety analysis, location-based service, and intelligent transportation. However, existing trajectory-based methods are limited to measuring low-level features, such as average speed and radius of gyration. Although these features can characterize specific observable behaviors, such as driving operation and movement range, they fail to depict stable traits underlying individual travel behavior. In this study, inspired by the Big Five Personality Traits, we model the driver profile through four fundamental trajectory traits: extroversion, openness, neuroticism, and conscientiousness, and quantify these traits by developing a Trajectory Trait Scale (TTS). Experiments on more than one million trajectories from 2,051 anonymized private vehicle volunteers over eight months demonstrate that our method can provide a valid representation of individual drivers' mobility patterns and driving behaviors. Specifically, we validate the consistency between trajectory traits and vehicle customer service records of drivers, including life rescue, navigation service, violation query, and fatigue companion. Besides, we find that trajectory integrity, seasonal changes, and traffic conditions exert small but non-negligible impacts on the stability of trajectory traits. These findings can enhance the understanding of human behavior in various spatiotemporal contexts, and illuminate the relations between trajectory traits and personality traits.

## INTRODUCTION

As an important part of urban traffic, billions of vehicles worldwide spend countless collective hours on roads every day.[1] Mining mobility patterns and driving behaviors of drivers is of immeasurable importance in alleviating urban traffic congestion and pollution,[2] designing urban infrastructure,[3] controlling disease spread,[4] recommending location-based services,[5] customizing usage-based insurance,[6] and promoting safe driving.[7] In particular, driver profiling can provide a human-centered approach to portraying individual travel, as well as revealing their motivation, objectives, and needs.[8] These insights are beneficial for predicting drivers' future behavior[9] and enhancing public services,[10] thereby broadening the application scenarios of trajectory mining.[11] For example, through characterizing the tendencies for impulsive driving and providing feedback, the drivers may consciously reduce such behavior, thereby improving driving safety.

However, constructing macroscopic and coherent driver profiles is challenging despite its importance. Driver profile is usually defined as a set of features describing a driver, including demographics, driving style, personality, and emotions.[12] Previous research on driver profiling can be categorized into self-report and passive behavioral-tracking approaches. The former relies on subjective responses obtained through questionnaires, such as the Aggressive Driving Behavior Scale,[13] Driving Attitude Questionnaire,[14] and Multidimensional Driving Style Inventory,[15] while the latter relies on in-car sensors, such as GPS devices,[16] monitoring cameras,[17] and Controller Area Network (CAN) bus data loggers.[18] Self-reporting can measure attributes that are not directly measurable, such as attitudes and tendencies. However, subjective assessment is susceptible to unintentional or intentional biases, stemming from drivers' inaccurate self-perception and concealment of real conditions.[19] These biases can result in inaccurate construction of driver profiles. With the availability of individual-level tracking data, researchers have shown an increased interest in constructing driver profiles based on passively tracking behavioral data. In particular, GPS trajectory data with high spatiotemporal resolution and long time series can reflect not only driving behaviors such as speed and acceleration, but also spatiotemporal regularities and contextual knowledge of travels.[20] If the trajectory data is collected from private vehicles, it can further capture drivers' travel preferences. Given these peculiarities, driving behavior can be depicted to assess driving risk by extracting vehicle operation features from trajectory data,[16,21] such as harsh acceleration and speeding. However, these approaches neglect individual spatiotemporal mobility patterns, which place more emphasis on the locations and durations of stays than on driving operational behaviors. This restricts the applications of driver profiles such as location-based service recommendations. Meanwhile, key mobility features (e.g., radius of gyration, travel entropy) are extracted[9,22-28] and their associations with individual attributes (e.g., occupation, mental illness) are explored[27,29-33] to interpret travel patterns,[23] predict travel behavior,[9] and infer the demographics of travelers.[27] These studies have offered essential and valuable observations on individual mobility. Nevertheless, these isolated and low-level vehicle operation and mobility features only link to specific observable behaviors, making them hard to be understood by various stakeholders, including drivers, service providers, and urban policymakers. By contrast, providing a concise and high-level representation of drivers' travel patterns that are consistent across situations and over time, such as weekdays and weekends, long and short-distance travel, can offer better insight into travel tendencies and driving preferences, and in turn facilitate multidisciplinary applications, like advertising and marketing, autonomous vehicles, and sustainable transport.

Therefore, we propose "trajectory traits" as descriptive domains of driver profile (called trajectory trait profile) to characterize individual mobility patterns and driving behaviors. Inspired by the Big Five Personality Traits (BFPT),[34] four fundamental trajectory traits are considered: extroversion, openness, neuroticism, and conscientiousness (Figure 1A). Extroversion characterizes travel vitality and the density of shopping and entertainment activities. Openness describes the irregularity and heterogeneity of destinations. Neuroticism portrays the tendency towards impulsive driving, while conscientiousness evaluates the sense of responsibility and punctuality in travel. Trajectory agreeableness is not considered in this study due to the lack of data involving driver interactions. To measure the four trajectory traits, we generate trajectory features as scale items from the perspectives of time, space, activity semantics, and driving behavior and develop a Trajectory Trait

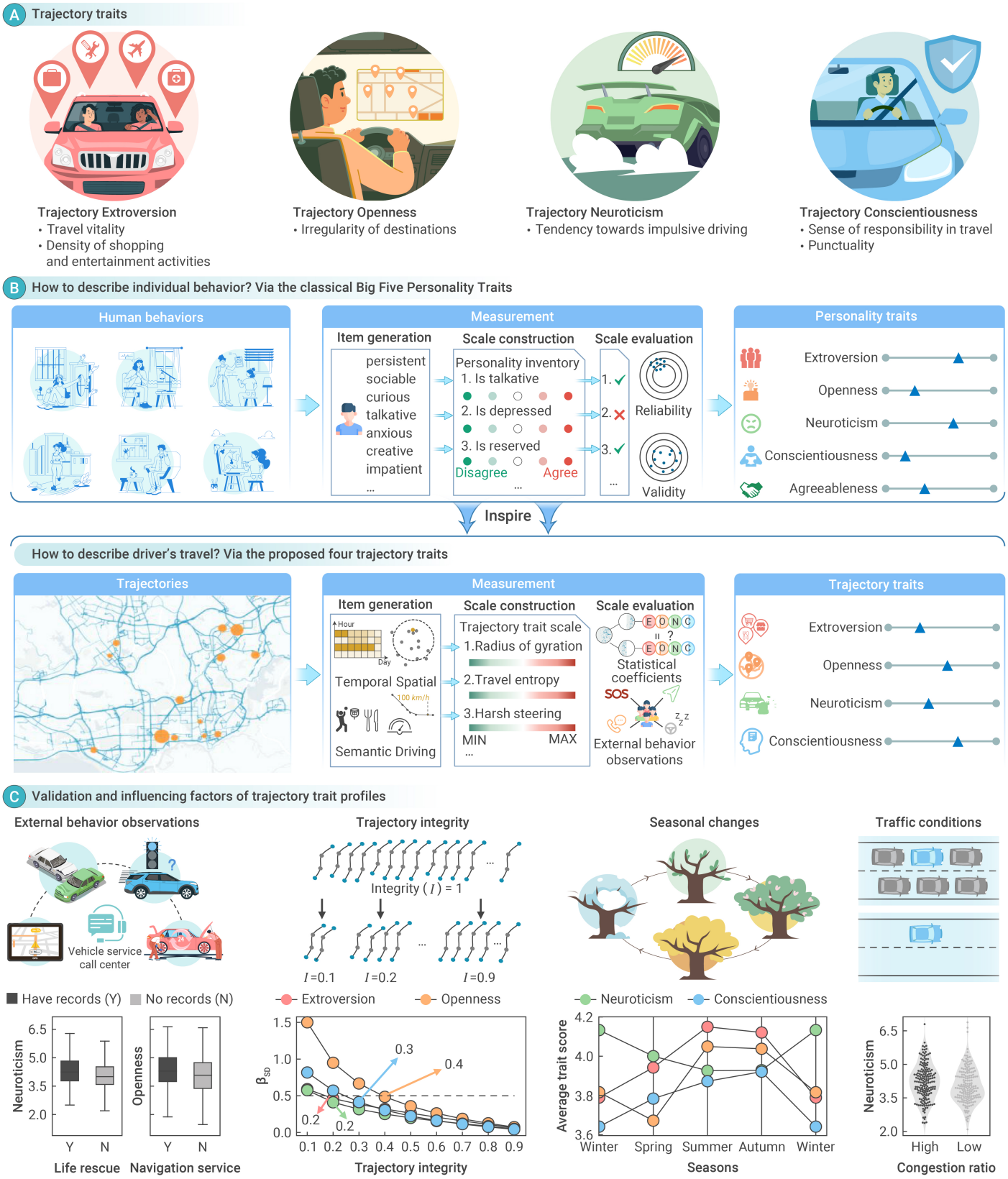



**Figure 1. Methodology and results of trajectory profiling** A. Definition of trajectory traits. B. Profiling trajectory based on the Big Five Personality Traits (BFPT) by referring to psychometrics. BFPT characterizes an individual's consistent behavior in various situations as well as differences in behavior from others in the same situation, using only five personality traits. In psychometrics, developing a scale is essential to the measurement of personality traits. Similarly, we propose trajectory traits to depict a driver's travel patterns, i.e., mobility patterns and driving behaviors, and construct the Trajectory Trait Scale (TTS) to calculate numerical trait scores for each driver. C. The associations between trajectory trait profiles and vehicle customer service records, and the influencing factors of profile stability, including trajectory integrity, seasonal changes, and traffic conditions. The box plot in the first column compares the differences in trait scores between drivers with and without certain vehicle customer service records. The line plot in the second column displays the standardized average absolute difference $\beta_{SD}$ (see Note S1) in trait scores between entire and randomly sampled trajectories at various sampling ratios ranging from 0.1 to 0.9. Each line corresponds to a trait domain. The line plot in the third column shows the changes in average trait scores with seasonal alternation. The violin plot in the last column illustrates the differences in neuroticism scores between drivers with high and low ratio of trips occurring in congested traffic.

Scale (TTS) by employing psychometrics (Figure 1B). We demonstrate the effectiveness of the proposed method by evaluating the discrimination capability of scale items and the reliability of TTS, and exploring the associations between trajectory traits and external behavior observations using visual analysis and vehicle customer service records, including life rescue, navigation service, violation query and fatigue companion, based on trajectory data of anonymized private vehicle drivers in Guangdong province, China. Moreover, we investigate the influence of trajectory integrity (i.e., sampling ratios of trajectories from the entire trajectory set of a driver) on the stability of trajectory trait profile, and the relations between profiles and spatiotemporal contexts, such as seasonal changes and traffic conditions (Figure 1C; see Figure S1 for the overall workflow). This paper seeks to address the following issues:

***Problem:*** How to model a macroscopic and coherent driver profile using trajectory data?

***Discussion:*** Does the trajectory trait profile align with the driver's performance in vehicle customer service records? Does the trajectory integrity affect the profile, and what percentage of trajectories are enough to reconstruct the trajectory profile constructed by the entire data? How does the profile vary with seasonal changes? What are the relations between trajectory traits and traffic conditions?

## MATERIALS AND METHODS

### Trajectory feature extraction for item generation

Trajectory features, which serve as the items of TTS, capture behavioral patterns from multi-dimensions. As shown in Table S1, we designed 32 trajectory features categorized into nine feature subtypes from the perspectives of time, space, activity semantics, and driving behavior. Spatial features describe the spatial extent and distribution of travel locations, temporal features reveal travel rhythm, semantic features describe travel activities, and driving behavior features suggest driving skills and risks.

To explain trajectory features more clearly, we define four trajectory-related concepts for a driver $u$, including trajectory point, trajectory, stay point, and trip. Trajectories form the spatiotemporal representation of a trip, and trajectory point is the fundamental unit of a trajectory. Stay point is the basis of extracting temporal, spatial, and semantic features, while trip contributes to the definition of driving behavior features.

***Definition 1:*** Trajectory point. A trajectory point is defined as a 6-tuple $p_i = (t, lat, lon, speed, accelaration, turningangle)$, where $t$ is the timestamp, and $lat$ and $lon$ represent latitude and longitude respectively. $speed$ is calculated as the average speed between the predecessor point $p_{i-1}$ and the successor point $p_{i+1}$. $acceleration$ is the change rate of average speed from $p_i$ to $p_{i+1}$. $turningangle$ describes the change of direction at $p_i$, and is calculated from three consecutive points $p_{i-1}, p_i$ and $p_{i+1}$ along the forward direction.

***Definition 2:*** Trajectory. A trajectory $T_j = \{p_1, p_2, \ldots, p_e\}$ is a sequence of trajectory points from the start of a vehicle to the flameout. The last point $p_e$ is the vehicle's flameout point in the trajectory.

***Definition 3:*** Stay point. The flameout points of trajectories are spatially dispersed, and may not have identical coordinates even for the same destination. As a result, they cannot accurately reflect meaningful spatial context of travel activities. Therefore, we cluster the flameout points whose stay time of $u$ is longer than the threshold, into clusters as stay points to represent active locations better. Each cluster corresponds to a stay point $sp_k$, denoted as $sp_k = (lat, lon)$, where $(lat, lon)$ is the coordinate of the geographic centroid point of the cluster. The stay points sequence $L$ represents $u$'s consecutive travel locations, defined as $L = sp_1 \rightarrow sp_k \rightarrow, \ldots, \rightarrow sp_N$. In sequence $L$, a stay point $sp_k$ can appear multiple times at different timestamps.

***Definition 4:*** Trip. A trip $tri_l = \{T_1, T_2, \ldots, T_e\}$ describes a driving route between two consecutive stay points (origin and destination) in sequence $L$, which consists of one or more trajectories.

***Temporal features.***

● Temporal entropy

Temporal entropy is defined to represent heterogeneity and uncertainty of the departure time of travel activity. Different time granularities such as daily, weekly, and even monthly should be considered to evaluate the diversity of travel rhythms. In our study, departure time is recorded as a timestamp with date and time components in the format "YYYY-MM-DD hh: mm: ss", e.g., "2019-01-01 08:29:30". The time part is for calculating the time entropy on a daily basis, i.e., time-of-day entropy; while the date part can be used to calculate date entropy on a weekly basis, i.e., day-of-week entropy. The date-time entropy is calculated by considering date and time simultaneously. To be noted, day-of-week entropy is not considered as the item in TTS, since the drivers in our dataset have similar weekly travel rhythms, and it cannot distinguish drivers effectively.

- Time-of-day entropy: $TE_d = -\sum_{i=1}^{h} p(T_i) \log(p(T_i))$, where $T_i$ represents the $i$-th timeslot in a day, and $p(T_i)$ is the frequency of travel in the $i$-th timeslot. $h$ is the number of timeslots in a day. In this study, we set $h = 24$.

- Day-of-week entropy: $TE_w = -\sum_{i=1}^{7} p(T_i) \log(p(T_i))$, where a timeslot $T_i$ represents the $i$-th day in a week, and $p(T_i)$ is the frequency of travel in the $i$-th timeslot.

- Date-time entropy: $TE_{wd} = -\sum_{i=1}^{q} p(T_i) \log(p(T_i))$. It contains more timeslots to depict a richer travel rhythm. In this study, we divide each of the 7 timeslots from Monday to Sunday into 24 equal parts, forming a total of 24×7 timeslots.

***Spatial features.***

● Radiuses of gyration

- Radius of gyration: $r_g = \sqrt{\frac{1}{n} \sum_{i=1}^{N_{sp}} n_i (sp_i - sp_{cm})^2}$, where $sp_{cm}$ is the center of all stay points, $n_i$ denotes the visiting frequency or the total time spent in $sp_i$, and $N_{sp}$ is the number of stay points. The radius of gyration indicates the characteristic distance traveled by an individual.[23] Specifically, if $n_i$ corresponds to the total time spent in $sp_i$, $r_g$ refers to *time-weighted radius of gyration*.

- Ratio of k-radius of gyration: $r_g^{(k)}/r_g$, where $r_g^{(k)} = \sqrt{\frac{1}{n_k} \sum_{i=1}^{k} n_i (sp_i - sp_{cm}^{(k)})^2}$ is the gyration radius of $k$ stay points with the most visits or the longest stay time. Profile results are robust with respect to the variation of $k$. Here, $k$ is set as 4 for a more balanced distribution of $r_g^{(k)}/r_g$, which improves its ability to differentiate among drivers (Note S2). Previous research had also observed a similar even distribution at $k = 4$, indicating a balance between individuals who frequently return to a limited number of locations (termed "$k$-returners") and those who explore a broader range of destinations (known as "$k$-explorers") .[23] This observation underscores the suitability of our parameter selection.

- Minimum number of extent-dominating locations: $K = min(k | r_g^{(k)} > \frac{r_g}{2})$. It is used to measure the spatial exploration degree of an individual. A small value of $k$ implies that $r_g$ only depends on a few frequently visited locations, i.e., frequently visited locations can convey sufficient information to characterize the travel spatial extent of an individual.

● Spatial entropy

- Random entropy: $SE_{rand} = \log_2(N_{sp})$. Assuming equal probabilities of visiting each location, it can depict the predictability of the individual $u$'s whereabouts.

- Location entropy: $SE_{loc} = -\sum_{i=1}^{N_{sp}} p(sp_i) \log(p(sp_i))$, where $p(sp_i)$ is the frequency that the stay point $sp_i$ is visited by $u$. It illustrates the uncertainty of $u$'s visit modes.

- Sequence entropy: $SE_{seq} = -\sum_L p(L) \log(p(L))$, where $L$ is a stay points sequence, and $p(L)$ is the occurrence frequency of $L$ across all trips. The sequence entropy can capture $u$'s spatiotemporal mobility pattern, because it not only depends on the locations but also relates to the order in which the locations are visited.

- OD entropy: $SE_{OD} = -\sum_{i=1}^{m} p(sp_i \rightarrow sp_{i+1}) \log(p(sp_i \rightarrow sp_{i+1}))$, where $sp_i \rightarrow sp_{i+1}$ denotes a trip with an origin $sp_i$ and destination $sp_{i+1}$, $p(sp_i \rightarrow sp_{i+1})$ is the frequency that $sp_i \rightarrow sp_{i+1}$ occurs across all trips, and $m$ is the number of unique origin-destination trips. It can be used to describe the uniformity of origin-destination trip distribution.

● Others

- Trips per month: $n_{month} = n/month$, where $n$ is the total number of trips in a given period, and $month$ is the number of months included.

- Trip length: $\bar{D} = \frac{1}{n} \sum_{i=1}^{n} dist(sp_i, sp_{i+1})$, where $sp_i$ and $sp_{i+1}$ are the origin and destination in a trip respectively, and $dist(sp_i, sp_{i+1})$ is the geographic

distance between $sp_i$ and $sp_{i+1}$, which can be Haversine distance or Great-circle distance. $n$ is the number of trips in a given period. $\bar{D}$ measures the travel vitality of an individual through the average distances of trips.

- Ratio of uninterested trips: $n_{uni}/n$, where $n_{uni}$ is the number of trips whose destinations have low visiting frequency. This feature reveals the visit rate of an individual to unfamiliar locations.

***Semantic features.***

● Home-related feature

- Ratio of stay time at home: $T_{home}/T$, where $T_{home}$ is the sum of the period when an individual stays at "home", $T$ is the sum of the period when an individual stays at all stay points. The "home" is defined as the stay point where $u$ spends the longest time at night. In this study, we set nighttime as 1 a.m. to 6 a.m.

- Distance-from-home entropy: $SE_{dh} = -\sum_{i=1}^{d} p(dh_i) log(p(dh_i))$, where $dh_i$ is the $i$-th distance from home and $p(dh_i)$ is the frequency of $dh_i$. It measures the uncertainty of travel distances from home to other stay points, and depicts whether an individual's trips are concentrated within certain distances from home. This feature can also be categorized as spatial entropy.

● POI-related feature

Points of Interest (POI) is the spatial manifestation of accessible urban infrastructures that reflect land use and urban functions.[9] Therefore, POI-related features can reveal the travel activities from the perspective of social behavior.

- Average POI importance metrics: To characterize the central tendency of an individual $u$'s travel activities for a specific POI category, we average the POI importance of that category across all stay points of the individual $u$. The POI importance of each stay point is assigned according to the semantic vector of the cell where the stay point is located in the semantic map (Figure S2A). The semantic map is a spatialized grid in which each cell has a semantic vector reflecting the importance of each POI category located in that cell. The semantic vector in the $i$-th cell, denoted as $locS_i = < m_1, m_2, \ldots, m_o >$, is computed by weighted TF-IDF algorithm (Figure S2B), where $o$ is the number of selected POI categories and $m_j = w_j \times n_j/N \times \log(C/c_j)$ represents the importance of $j$-th POI category, $n_j$ is the number of POIs in $j$-th POI category and $N$ is the totality of the POIs in that cell, $C$ and $c_j$ refer to the total number of cells and the cells containing $j$-th POI category, respectively. To distinguish cells with the same TF-IDF value but different numbers of POIs in $j$-th POI category, we use the number of POIs in $j$-th POI category within the 3 * 3 neighborhoods around that cell as the weight $w_j$.

***Driving behavior features.*** Driving behavior features are divided into overall behavior, abnormal behavior, and residential intersection behavior. Overall behavior features reflect the statistical driving characteristics of all trips, such as average speed and average acceleration; Abnormal behavior features reveal the risky driving behavior of a driver across all trips, such as harsh shifting and speeding. In particular, speeding at residential intersections, which often endangers the safety of pedestrians due to lack of supervision, can highlight the absence of drivers' sense of responsibility and moral restraint.

● Overall behavior

- Speed metrics: The speeds of all trajectory points belonging to a trip $tri_j$ form a vector $V_j = [v_1, v_2, \ldots, v_e]$. Based on the vector, the average speed $\bar{v}_j$, the standard deviation of the speed $s(V_j)$, and the maximum of the speed $max(V_j)$ can be calculated. The average speed vector for all trips of a driver is denoted as $\bar{V} = [\bar{v}_1, \bar{v}_2, \ldots, \bar{v}_n]$. The average of the vector $\bar{V}$ (named *average speed*) reflects the driver's awareness of the control of the driving speed. The standard deviation of the vector $\bar{V}$ (*standard deviation of average speed*) reveals the speed stability of a driver over different trips. The speed standard deviation vector for all trips of a driver is denoted as $s(V) = [s(V_1), s(V_2), \ldots, s(V_n)]$. The average of values in the vector $s(V)$ (*average standard deviation of speed*) and the maximum value in vector $s(V)$ (*maximum standard deviation of speed*) estimate the average and extreme standard deviations of speed over a trip, respectively. The average of the vector $max(V) = [max(V_1), max(V_2), \ldots, max(V_n)]$(*average of maximum speed*) describes the long-term awareness of controlling the extreme driving speed, where $max(V_i)$ is the maximum speed on the $i$-th trip of the driver.

- Acceleration metrics: Acceleration is calculated similarly to the speed metrics. Considering the association between acceleration and harsh shifting metrics, we only select one feature here, *maximum standard deviation of acceleration*. It reveals the extreme degree of acceleration change in a single trip of the driver.

● Abnormal behavior

- Harsh shifting metrics: Ratio of harsh shifting points in a trip $tri_j$ is defined as $R_{shift} = |\{p_i | |a(p_i)| > AT\}|/|\{p_i\}|$, where $a(p_i)$ is the acceleration of $p_i$, $|\{p_i\}|$ is the total number of trajectory points, and $|\{p_i | |a(p_i)| > AT\}|$ is the number of harsh shifting points in this trip, and $AT$ is the threshold of harsh shifting. The average $R_{shift}$ over all trips (*average ratio of harsh shifting points*) estimates the tendency of a driver to accelerate and decelerate sharply. The standard deviation $R_{shift}$ for all trips (*standard deviation ratio of harsh shifting points*) depicts the dispersion of a driver's harsh shifting in different trips.

- Harsh steering metrics: Ratio of harsh steering points in a trip $tri_j$ is defined as $R_{steering} = |\{p_i | w(p_i) > TT\}|/|\{p_i\}|$, where $w(p_i)$ is the turning angle of $p_i$, $TT$ is the threshold of harsh steering. The average and standard deviation $R_{steering}$ of all trips are calculated accordingly.

- Speeding metrics: Ratio of speeding points in a trip $tri_j$ is defined as $R_{speeding} = |\{p_i | v(p_i) > ST\}|/|\{p_i\}|$, where $v(p_i)$ is the speed of $p_i$, $ST$ is the threshold of speeding. The average $R_{speeding}$ of all trips (*average ratio of speeding points*) reflects the frequency of speeding. The average $N_{speeding}$ of all trips (*average number of speeding points*) measures the number of speeding.

● Intersection behavior

Since the code of ethics constrains conscientious drivers to behave responsibly, there should be less speeding behavior at road intersections in residential areas. Therefore, we define *intersection's average ratio of speeding points* and *intersection's average speed* to capture the drivers' awareness of speed control at residential intersections. The two features are calculated by the trajectory points within the buffer zone of the intersection in the residential area.

## Trajectory traits scale construction

The developed TTS consists of four subscales that measure trajectory extroversion, openness, neuroticism, and conscientiousness, respectively (Figure 2 and Table S2). Extroversion subscale contains items (i.e., trajectory features) that quantify travel frequency, movement range, and activity types. Openness subscale focuses on items that capture the irregularity of destinations. Neuroticism subscale involves items that reflect driving stability. Conscientiousness subscale encompasses items related to driving violations and travel temporal patterns.

***Items in scale.*** Extroversion in personality describes the density of social interaction and the vitality level of an individual from the perspective of psychological orientation.[34] Extroverts with high extroversion scores are not only sociable and gregarious, but also active, assertive, energetic, and talkative. They enjoy social interaction with people and are full of energy. In contrast, Introverts tend to reserve their social interactions and avoid large, noisy gatherings, in other words, they prefer to do things alone or with a small group of people. However, their lack of social engagement should not be interpreted as shyness or depression but rather as independence. That is, they are reserved rather than unfriendly. Due to social needs, the characteristics of trajectory extroverts are manifested as an extensive spatial extent of activities, many trips, and frequent visits to shopping or entertainment venues; introverts vice versa. Thus, we assign the items shown in Figure 2A to this domain. *Trips per month* and *ratio of stay time at home* measure the frequency of travel and willingness to stay at home respectively, while *trip length*, *radius of gyration*, and *time-weighted radius of gyration* show the driver's travel vitality level by measuring the spatial extent. Since extroverts prefer to eat and hang out with friends, geographic semantic items including *average importance of shopping POI*, *average importance of recreation POI*, and *average importance of restaurant POI* contribute to revealing driver's degree of fondness for social interactions.

Openness in personality describes a positive demand for novel experiences and a tolerance of unfamiliar situations from the perspective of thinking and acting.[34] Individuals with high openness scores are curious, creative, sensitive to beauty, and willing to accept novel ideas and unconventional

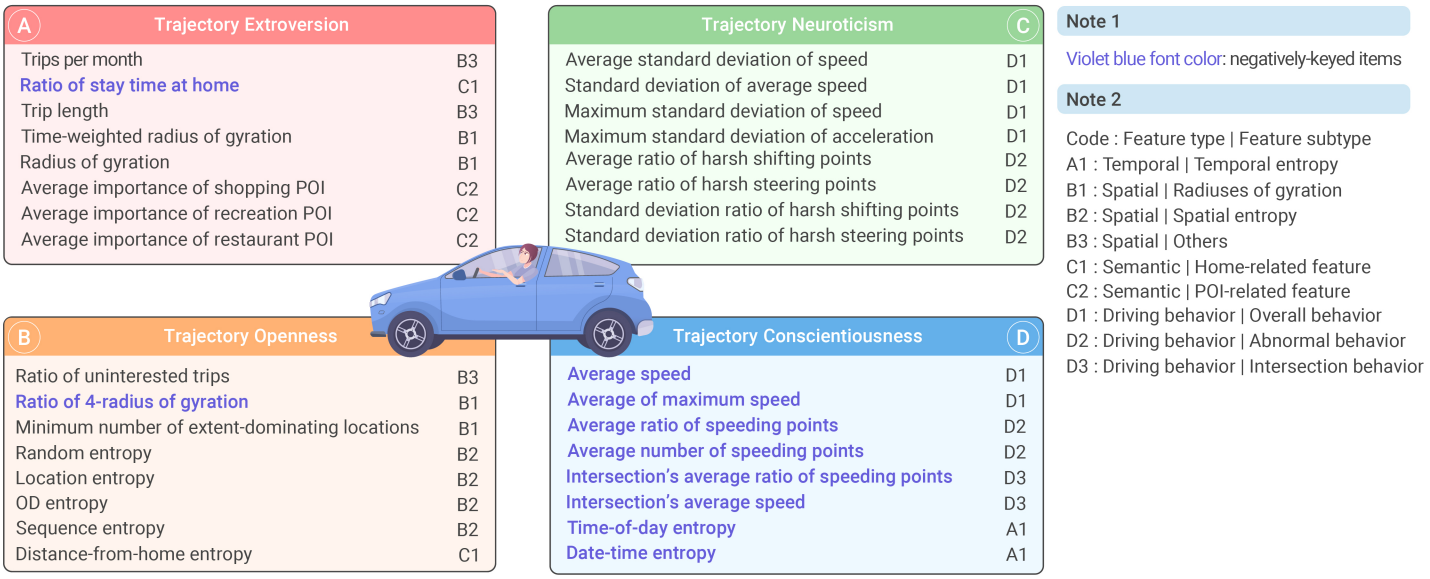



**Figure 2. Items in Trajectory Trait Scale** Each scale comprises eight items (i.e., trajectory features) extracted from nine subtypes across four dimensions: time, space, activity semantics, and driving behavior. These items include positively-keyed and negatively-keyed items (see the subsection "Scoring and interpretation" for more details).

values. In particular, they are willing to try different activities, such as exploring a new place and trying unusual foods. Conversely, people with low openness scores tend to be traditional and conservative in behavior and prefer routine work. Trajectory openness is defined as the irregularity and unpredictability of space. Therefore, we assign items to this domain as shown in Figure 2B. Four spatial features, including *random entropy*, *location entropy*, *sequence entropy*, and *OD entropy*, reflect the regularity of a driver's travel destinations, e.g., whether the driver always shops in the same mall, dines in the same restaurant, and does not like to explore new places. *Distance-from-home entropy* measures the tendency of a driver to stay within a certain distance. *Ratio of uninterested trips*, *ratio of 4-radius of gyration*, and *minimum number of extent-dominating locations* capture the proportion of frequently visited locations.

Neuroticism in personality reflects the process of an individual emotion regulation, the tendency to experience negative emotions. It is also referred to as emotional instability or stability. Individuals with high neuroticism scores are prone to experience impatience, anxiety, anger, disgust, sadness, and a variety of other negative emotions. Furthermore, they have more difficulty in controlling their impulses and coping with stress than others. On the contrary, individuals with low neuroticism scores are calm and easy to get rid of negative emotions. Previous research has shown that neuroticism is related to risky and aggressive driving and manifests in harsh acceleration, harsh braking, etc.[35] Trajectory neuroticism tends to produce impulsive driving. As shown in Figure 2C, *average ratio of harsh shifting points* and *average ratio of harsh steering points* quantify the central tendency of a driver's impulsive driving behavior, while *standard deviation ratio of harsh shifting points* and *standard deviation ratio of harsh steering points* represent the volatility and dispersion among different trips of the driver. The speed standard deviation during a trip reflects the driving stability during the trip. Hence, we design *average standard deviation of speed*, i.e., the average value in the collection of trips, to estimate the driver's driving stability in a trip, while *standard deviation of average speed* is chosen for measuring the driver's driving stability in the long-term process. Moreover, *maximum standard deviation of speed* and *maximum standard deviation of acceleration* indicate the most extreme observation of all trips.

Conscientiousness in personality evaluates the individual's organization, persistence, motivation, and action in pursuit of goals.[34] The characteristics of highly conscientious individuals can be summarized as perseverance, meticulousness, self-discipline, dedication, and reliability. Because they can organize their lives compactly, people with high conscientiousness are likely to be punctual and neat. People with high trajectory conscientiousness tend to obey traffic rules and act cautiously in traffic.[35] In addition, they will follow the travel plans in daily life. The assigned items in this domain are shown in Figure 2D. *Time-of-day entropy* and *date-time entropy* reflect whether the driver is punctual and obeys the same temporal travel pattern. Since speeding is one of the traffic violations, the six speed-relevant indicators in Figure 2D, such as *average speed* and *average ratio of speeding points*, are used to measure the driver's compliance with the rules.

***Scoring and interpretation.*** In psychometrics, the scores of the subjects' personality traits are aggregated by the item scores in each subscale. First, for each item, subjects indicate "Very Inaccurate", "Moderately Inaccurate", "Neither Accurate nor Inaccurate", "Moderately Accurate" or "Very Accurate" as a response to the item, where each response corresponds to a numerical score. For example, for the item "Feel comfortable around people"[36] measuring extroversion, the response "Very Inaccurate" is assigned a score of 1, while "Very Accurate" is assigned a score of 5. For negatively-keyed items, reverse scoring is employed to ensure the representative consistency of item responses. For example, "Am quiet around strangers",[36] the response "Very Inaccurate" is assigned a score of 5, while "Very Accurate" is assigned a score of 1. Then, the trait score for a subscale can be obtained by summing all item scores belonging to that subscale. The total score is a quick approximation for the trait being measured, meaning that each item is unit-weighted and contributes equal amount of information.[37]

In this study, since the raw trajectory features have different magnitudes, feature scaling is necessary for obtaining the trait score. Using min-max scaling to normalize trajectory features is susceptible to extreme values, and Z-score standardization does not map item scores into a bounded range. Thus, we use a logistic function $f(x)$ to normalize the features, which follows an S-shaped curve, is insensitive to outliers, and has a bounded range of (0,1).[38] For positively-keyed items, $f(x) = \frac{1}{1+e^{-k(x-x_0)}}$, while for negatively-keyed items (violet-blue font items in Figure 2), $f(x) = 1 - \frac{1}{1+e^{-k(x-x_0)}}$. We set $k = \frac{1}{IQR}$, $x_0 = X_{median}$ in this study, where IQR is the interquartile range and $X_{median}$ is the median of feature $X$. The steeper slope of the logistic function near the median widens the interval in item scores for individuals near the median of the raw trajectory features, resulting in better differentiation among individuals within the group.

When interpreting the scores for each subscale, it is important to keep in mind the following three points.[39] First, the trait score of a driver is relative to other drivers in the same sample, as the normalization parameters are derived from the sample. However, using same normalization parameters across multiple studies can facilitate result comparability. Second, scores for different trait domains are not comparable, which means that comparing the

driver's extroversion score with the openness score is meaningless. Third, we should avoid treating trait scores using categorical thinking. Although it is convenient to discuss "introverts" and "extroverts", the range of values for TTS scores is continuous. There is no single cut-off point that separates those who "have" the trait from those who "do not". Nevertheless, in Results section, classification rules with cut-off points are required to analyze the characteristics of different scorer groups. The classification rules are only valid for the intended purpose, and should not replace the dimensional interpretation of scale in terms of traits.

### Datasets and data preprocessing

Our experiments are based on the trajectory datasets D1 and D2 selected from 8,347 drivers. D1 comprises GPS trajectories of 662 drivers for eight months in 2019, selected from 1,915 private vehicle volunteers in Shenzhen, Guangdong province, China. D2 contains trajectory data from 1,389 drivers in eight months of 2022, selected from 6,432 private vehicle volunteers in Guangdong province, China. We utilized the travel duration, the number of trajectories, and the consecutive days without trajectories as selection criteria to ensure that the chosen drivers have varied travel frequencies, while each has prolonged trajectory data to support profiling (detailed in Table S3). The vehicle customer service records of the drivers in the D2 were also collected, which provides an external view to depict their travel characteristics through behavioral records.

To calculate trajectory features, we preprocess the selected trajectory datasets. 1) Use the DBSCAN algorithm[40] to cluster flameout points with stay time longer than the threshold to obtain gathered stay points and isolated stay points (noise points) (***Definition 3*** in Methods), where eps = 200m, MinPts = 3, and stay time ≥ 20 minutes. Detected noise points are defined as uninterested stay points to calculate *ratio of uninterested trips*. Profile results are robust with respect to variation of clustering parameters (Note S3). The eps is set as 200 meters, which represents the average maximum distance that drivers are willing to walk between parking spots and destinations.[41] When MinPts is set to 3, the distribution of the feature *ratio of uninterested trips* is relatively balanced, allowing for better differentiation among drivers (Note S3). 2) According to the definition of semantic features, the semantic map with cell size $500 \times 500m$ of the study area is generated using the POIs from the Amap dataset (Figure S2C). Seven categories associated with daily human activities are selected from 23 POI categories, including recreation, residence, education, hospital, shopping, restaurant, and company. We pick three elements in the semantic vector to calculate items measuring extroversion, i.e., recreation, shopping, and restaurant. 3) The road network data obtained from OpenStreetMap is used to identify trajectory points located at residential road intersections, which helps to extract residential intersection behavior in measuring conscientiousness. Based on the above preprocessing, the features shown in Figure 2 can be extracted to calculate the trait scores and build the profile.

## RESULTS

We assess TTS via item discrimination and subscale reliability, and show the validity of the generated trajectory traits through visual analysis of travel patterns and investigation of vehicle customer service records. Furthermore, we quantify the differences in trajectory trait profiles under various spatiotemporal contexts to explore the influencing factors on the stability of profiles. During the experiments, to analyze item discrimination and profile characteristics of different groups, we divide all drivers into low, medium, and high scorers in each trait domain (Figure S3).

### Trajectory trait scale validation through statistical coefficients

***Item discrimination.*** Item discrimination assesses whether an item can correctly distinguish among drivers with different trait scores.[42] It represents the degree to which drivers with high trait scores also receive high item scores. Poor discrimination implies that the item cannot measure what the subscale is intended to measure, and should be removed. The results of independent samples T-test and effect size index show that each item can discriminate between high and low scorers effectively (Note S4; Figure S4 with $P < .001$ and $d_s > 0.5$), and does not need to be removed from the subscales. Nonetheless, items perform differently in discrimination, with $d_s$ varying from 0.563 to 4.619. Specifically, all five entropy-related items in openness are strongly discriminative, i.e., *random entropy*, *location entropy*, *OD entropy*, *sequence entropy*, *distance-from-home entropy*, while the $d_s$ of *random entropy* is less than others because it does not consider the frequencies of visits to stay points but others do. In contrast, *trips per month* in extroversion, *ratio of 4-radius of gyration*, and *minimum number of extent-dominating locations* in openness, *average standard deviation of speed* in neuroticism, *time-of-day entropy*, and *date-time entropy* in conscientiousness are weakly discriminative, with $d_s$ less than 1.0. It is because their feature dimensions are different from those of other items in the corresponding trait domain, and this leads to weak correlations with the other items and small $d_s$.

***Reliability of measurement.*** Reliability refers to the consistency of a measure.[43] We consider two types of consistency for each subscale, internal consistency (across items) and data split-half reliability (across data subsets). Internal consistency indicates the degree of interrelationship or homogeneity among items in a subscale, so that items are consistent and capable of measuring the same thing.[44] Data split-half reliability, proposed in this study, evaluates the extent to which similar trait scores are obtained on independent measurements.

Internal consistency refers to the consistency of item scores. In general, all items in a subscale reflect the same underlying construct, so scores for these items should be correlated with each other. We adopt Cronbach's $\alpha$ (Note S5) to measure internal consistency, and the result suggests that items within the same subscale are consistent and the assignment of items is rational. As illustrated in Table S4, Cronbach's $\alpha$ values for extroversion (0.637, 95% confidence intervals (CI) [0.593, 0.677]) and neuroticism (0.715, 95% CI [0.681,0.747]) are acceptable, while $\alpha$ values for openness (0.872, 95% CI [0.857,0.886]) and conscientiousness (0.810, 95% CI [0.787,0.831]) are good. Extroversion and neuroticism have lower $\alpha$ than the other two trait domains, possibly because items in extroversion contain both spatial and semantic features, and items in neuroticism capture both inter-trips and intra-trip driving stability. Nonetheless, all four subscales meet the requirement for internal consistency.[45] Therefore, the meaning of the trait score is clear when using it to summarize multiple trajectory features.

If TTS is reliable, the difference in trait scores should be small when a driver is measured multiple times independently in different data partitions. The measure of data split-half reliability requires randomly dividing the trajectories of each driver into two parts, and then calculating the average absolute difference of trait scores ($\beta$ and $\beta_{SD}$ coefficient in Note S1) between the split two sub-datasets of all drivers in each trait.

Results show that trait scores are stable when measured twice independently, and the different scorer groups have different stability. As illustrated in Figure 3, most of the dots are along the diagonal, indicating minor differences between two trait scores of the same driver, and the $\beta_{SD}$ in four trait domains are all less than 0.5. Nonetheless, the stability of the four subscales varies, with $\beta_{SD}$ from low to high being conscientiousness, extroversion, openness, and neuroticism respectively. In Figure 3A, the dots in neuroticism deviate from the diagonal, while conscientiousness is relatively concentrated. Meanwhile, we find that the most stable group, whose density distribution of absolute difference gathers around zero (Figure 3B), is either high or low scorers. In addition, both high and low scorers in extroversion, openness, and conscientiousness exhibit smaller $\beta$ values than those of medium scorers. It suggests that drivers whose trait scores fall in the two tails of the distributions are more likely to maintain their characteristics on each trip. For example, the high scorers in openness ($\beta$ = 0.243) and conscientiousness ($\beta$ = 0.168) are most stable in their trait domains, indicating that they maintain similar spatiotemporal patterns and adhere to traffic rules for most trips. However, high scorers in neuroticism ($\beta$ = 0.338) are sensitive to the trajectory splitting results, because the external environment factors, such as road conditions and weather, may easily influence driving behaviors. Drivers tend to drive impulsively in unfavorable situations, which is unlikely to occur on every trip.

### Relations between trajectory traits and external behavior observations

External behavior observations can verify whether the TTS measurements conform to expectations, i.e., whether the drivers' mobility patterns and driv-

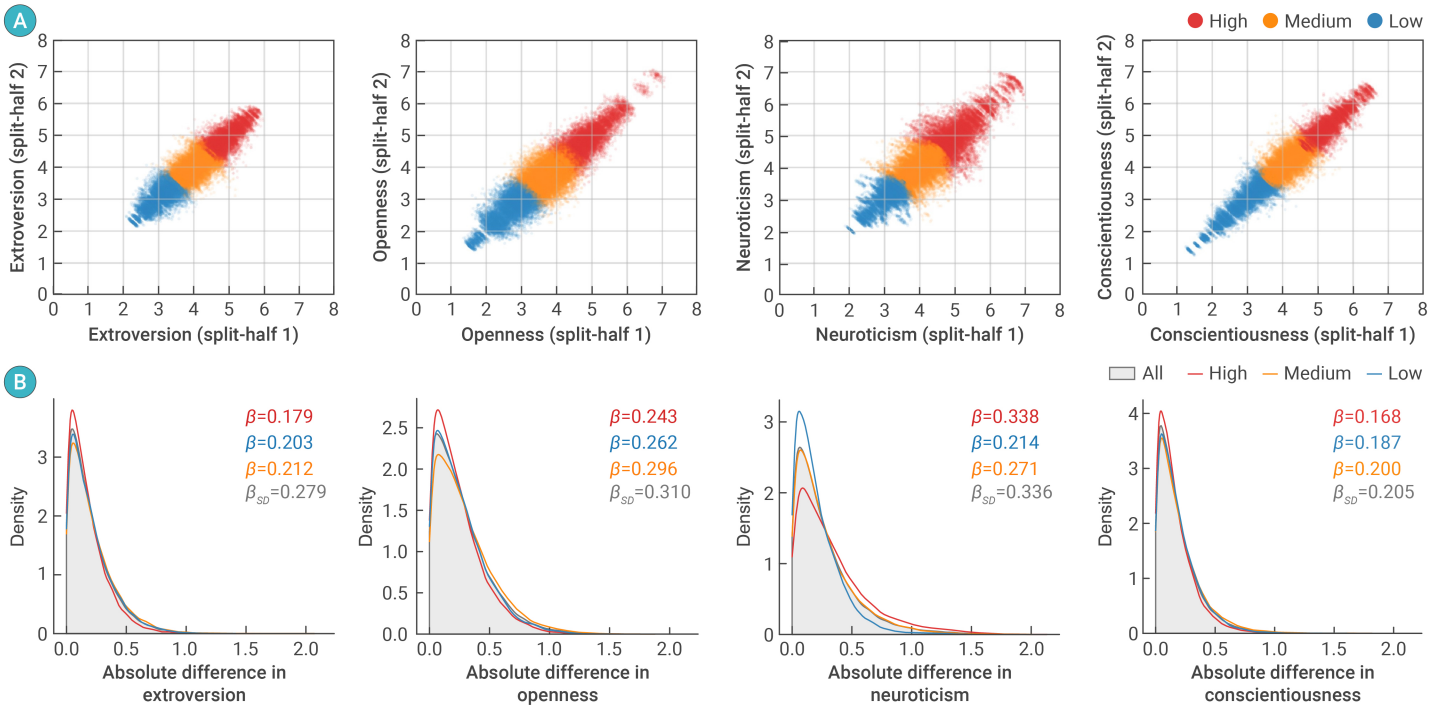



**Figure 3. Data split-half reliability of the four subscales in TTS** A. The relation between the two sets of scores for the four traits. Each dot in the scatter plot represents a pair of scores for a driver $u_i$ computed by randomly dividing $u_i$'s trajectories into two parts. B. The density distribution of the absolute difference between two sets of scores for the four traits across different scorer groups. The steeper the curve, and the closer it is to zero, the more stable the corresponding scorer group is.

ing behaviors characterized by the trait scores are consistent with the definitions and observed behaviors. We evaluate the extent to which the TTS measurements conform to the definition of trajectory trait by visual analysis. If the validity of TTS is high, drivers with similar trait scores will manifest similar general travel patterns that align with the trait definition, while drivers with distinct differences in trait scores will display different travel patterns. Furthermore, we analyze the relations between traits and vehicle customer service records. These records document the drivers' requests to the vehicle service call center during trips, which respond to the driver's travel behavior and are collected independently from trajectory data.

***Visual analysis of travel patterns.*** Compared with low scorers, high scorers in extroversion manifest a larger spatial extent in travel and a greater number of visits (Figure 4A), and they exhibit more visits to areas associated with social and leisure activities (Figure 4B). These spatial characteristics reflect those high scorers in extroversion have a higher level of vitality and social willingness, which is consistent with the definition of extroversion. In terms of openness, low scorers exhibit stay points with a large ratio of visits, e.g., 50 percent of drivers have the stay points with a ratio of visits greater than 0.46, and show a dispersed distribution of visiting ratio. In contrast, high scorers have more stay points than low scorers, and a more concentrated distribution of ratio of visits with smaller values, e.g., the ratio of visits of drivers in high scorers does not exceed 0.54 (Figures 4C-D). Meanwhile, the distribution of the relative ratio of visits for high scorers is closer to 1.0, while that of low scorers is more dispersed with two peaks at two tails, which indicates most stay points have a very small relative ratio of visits and a few stay points have dominated visiting frequency. This suggests that high scorers tend to explore more places with a relatively balanced visiting frequency, while low scorers prefer returning to frequently visited locations.

The amplitude of the acceleration fluctuations of the two highest-scored drivers in neuroticism is larger than that of the two lowest-scored drivers, and the major differences occur in the trips with high acceleration standard deviations (Figures 5A-B). This suggests that high scorers in neuroticism tend to drive impulsively on certain parts of trips, possibly influenced by the environmental factors at the time. In terms of conscientiousness, the two lowest-scored drivers have a larger number and frequency of speeding points than the two highest-scored drivers (Figures 5C-D). In the departure date-time subplots of Figure 5D, there are more blanks for the two highest-scored drivers, indicating that the departure times of the two drivers are relatively concentrated and regular within a day. These observations suggest high scorers in conscientiousness tend to travel more punctually and follow the rules. In conclusion, drivers with similar scores exhibit similar travel patterns, while apparent differences exist between the high and low scorers. These observations suggest that TTS measurements align with the definition of trajectory traits, demonstrating the validity of TTS in measuring its intended content. More details are discussed in Note S6.

***Connections with vehicle customer service records.*** We divide drivers in dataset D2 into groups based on vehicle customer service records and compare their trait scores using the Kolmogorov–Smirnov test and the Mann–Whitney U test, to investigate the consistency between trajectory traits and behavioral records (detailed in Note S7). We focus on four customer service requests: life rescue, navigation service, violation query, and fatigue companion. First, we find that drivers with records of life rescue have significantly higher neuroticism scores than those without such records ($P \leq .001$; Figure 6B & Table S5), in line with the fact that impulsive driving behavior is frequently correlated with accident occurrences and subsequent requests for life rescue services.[46] Second, drivers with records of navigation service have significantly higher openness scores compared to those without such records ($P = .003$; Figure 6C). This suggests that drivers with high openness scores, who have diverse destinations, are inclined towards utilizing navigation services, in line with the fact that drivers commonly use navigation services to reach unfamiliar destinations.[47]

Furthermore, drivers with life rescue records have significantly higher scores in conscientiousness than those without such records ($P = .001$; Figure 6D), which indicates that drivers with high conscientiousness scores tend to adhere to emergency response procedures and invoke life rescue services after a car accident. It means that drivers characterized by high conscientiousness exhibit a pronounced awareness of traffic safety. This conclusion is also supported by the observation that drivers with records of fatigue companion service requests during tired driving have lower conscientiousness scores compared to those without such records ($P = .018$; Figure 6E). Fatigue driving jeopardizes transportation safety and can reflect a driver's safety awareness and sense of responsibility.[48] Lastly, we find that drivers with records of violation query have higher conscientiousness scores compared to those without such records ($P = .007$; Figure 6F). This may be

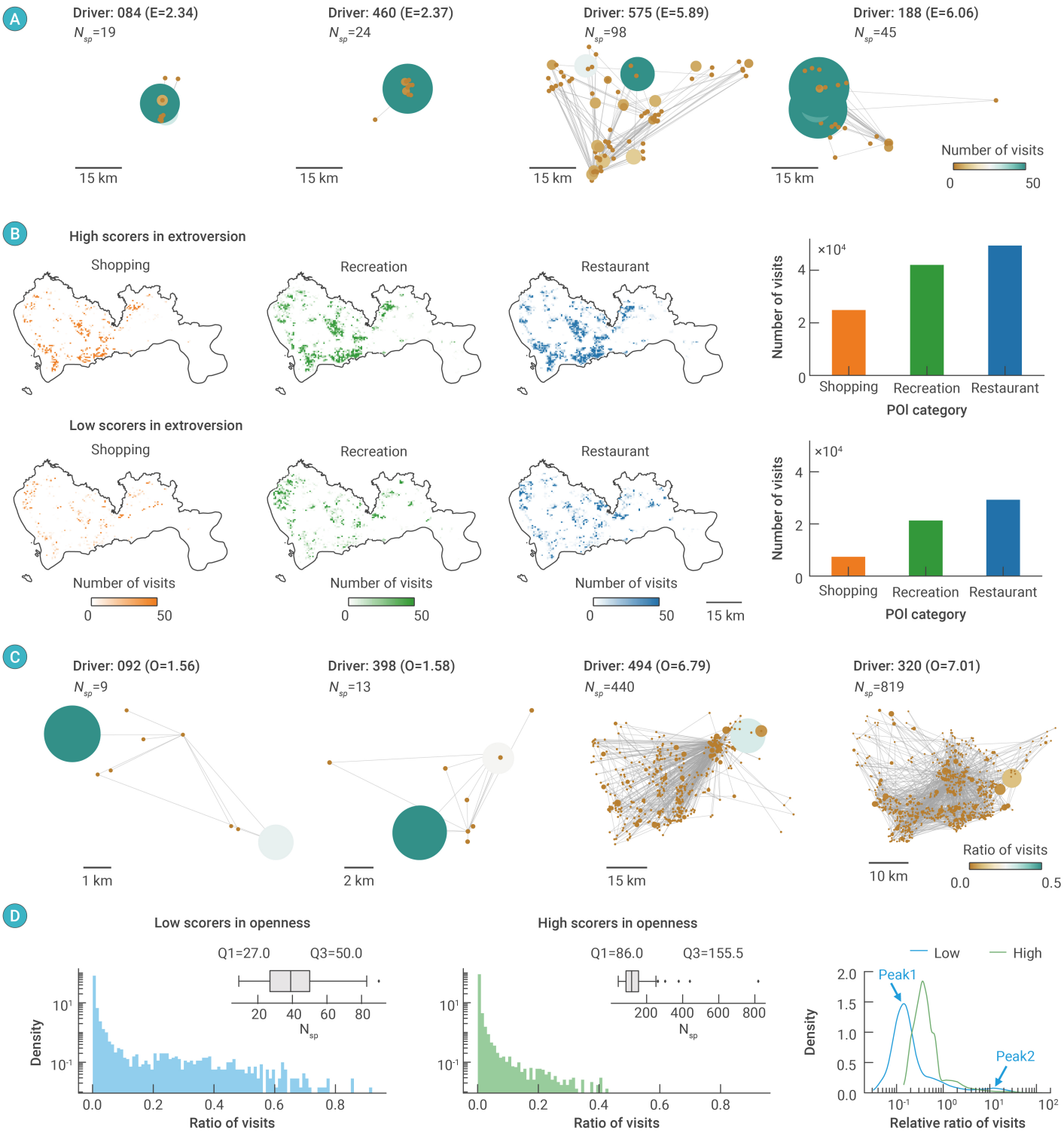



**Figure 4. Validity of extroversion and openness subscales** A. Mobility networks of four drivers, two with the lowest and two with the highest extroversion scores (denoted as E) in dataset D1. In each subplot, a node represents the stay point of the driver, and its location corresponds to the latitude and longitude of that stay point. The area of a node is proportional to the number of visits to the stay point, and the edges represent observed direct movement between two stay points. $N_{sp}$ denotes the number of stay points. B. Comparison of the spatial distribution and number of visits to stay points between high and low scorers in extroversion from the perspective of semantic activities. The colors orange, green, and blue correspond to the three categories of Points of Interest (POI) that measure extroversion, namely shopping, recreation, and restaurant, respectively. Each map, with a grid cell size $500 \times 500m$, shows the number of visits to the top 20 percent important cells (detailed in Figure S2D) for the scorer group, and the color of cell represents the number of visits. The bar chart on the right summarizes the total number of visits to the top 20 percent important cells. C. Mobility networks of four drivers, two with the lowest and two with the highest openness scores (denoted as O) in dataset D1. The nodes and edges have the same definitions as A, while the area of a node is proportional to the ratio of visits (see color bar) to the stay point. D. Comparison of the ratio of visits and $N_{sp}$ between high and low scorers in openness. The two histograms show the density distribution of the ratio of visits to stay points for drivers in the two scorer groups respectively, and boxplots summarize the distribution of $N_{sp}$, where Q1 is the 25th percentile and Q3 is the 75th percentile. The rightmost subplot compares density distributions of the relative ratio of visits $r = R/R_{equal}$ of stay points in the high and low scorers, where $R$ is the ratio of visits to the stay point, and $R_{equal} = 1/N$ is the ratio of visits when the driver visits stay points with equal probability. If the distribution of $r$ is concentrated around 1.0, drivers in the corresponding scorer group visit stay points with approximately equal probability. Otherwise, there is a large difference in the ratio of visits to stay points for drivers.

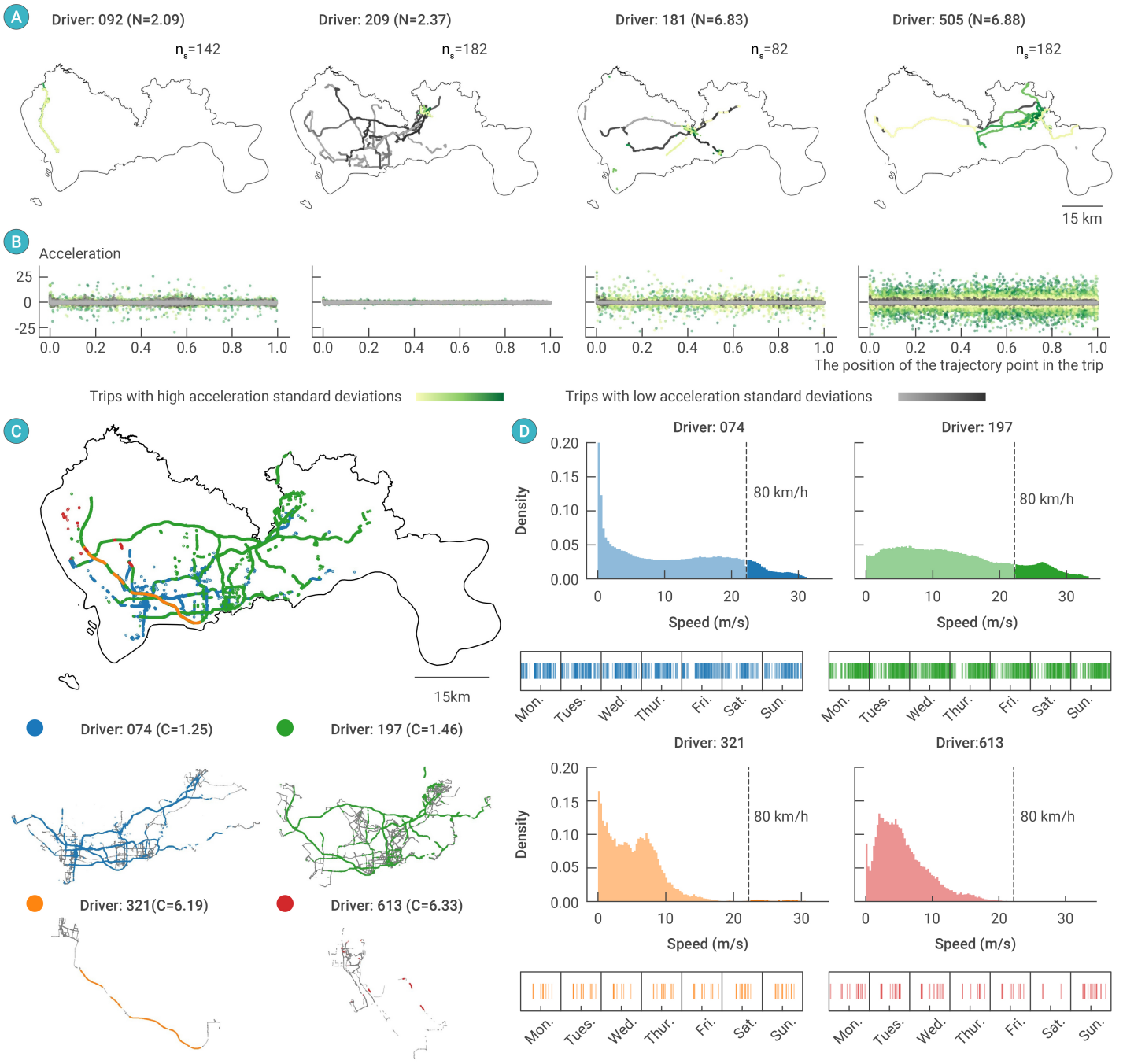



**Figure 5. Validity of neuroticism and conscientiousness subscales** A. Selected trips of four drivers, two with the lowest and two with the highest neuroticism scores (denoted as N) in dataset D1. In each subplot, the dots represent the trips with high (top 27 percent, colored dots) and low (bottom 27 percent, gray dots) acceleration standard deviations respectively. $n_s$ denotes the number of trips. B. Acceleration variation of the selected trips. To better display trips with different numbers of trajectory points, we normalize the relative position of trajectory points, dividing each position by the total number of trajectory points in the corresponding trip. C. Speeding trajectory points ($_{80km/h}$ is set as the threshold) of four drivers, two with the lowest and two with the highest conscientiousness scores (denoted as C) in dataset D1. In the four subplots at the bottom, the colored dots represent the speeding trajectory points, and the gray dots represent non-speeding points. D. Speed density distributions of all trajectory points and departure date-time modes of the four drivers. In the departure date-time subplots, each vertical line corresponds to a trip that occurred at a specific time of a day, and the darker the color, the more trips.

because drivers with high conscientiousness are more inclined towards self-monitoring and discipline, actively paying attention to whether their behavior complies with regulations. Checking for violations may be a proactive measure they take to ensure compliance with traffic rules and maintain an impeccable driving record. These observations demonstrate the consistency between trajectory traits and vehicle service records in characterizing driver behavior, and indicate that trajectory traits can reasonably depict the mobility patterns and driving behaviors of drivers. Moreover, this may reveal the connections between trajectory traits and personality traits, and contribute to potential applications such as usage-based insurance (UBI)[49] and location-based services (LBS).[50]

## Potential influencing factors on profile stability

***Trajectory integrity.*** In this experiment, we try to answer two questions: 1) Does the profile remain stable when part of a driver's trips is missing or not included? 2) Do drivers with different trait scores have different performances with varied trajectory integrity? To answer these questions, we randomly sample different ratios of trips for each driver from 0.1 to 0.9 with an interval of 0.1 to calculate trait scores, and compare them with the trait scores obtained from the entire trips using $\beta_{SD}$ coefficient (see Note S1) to

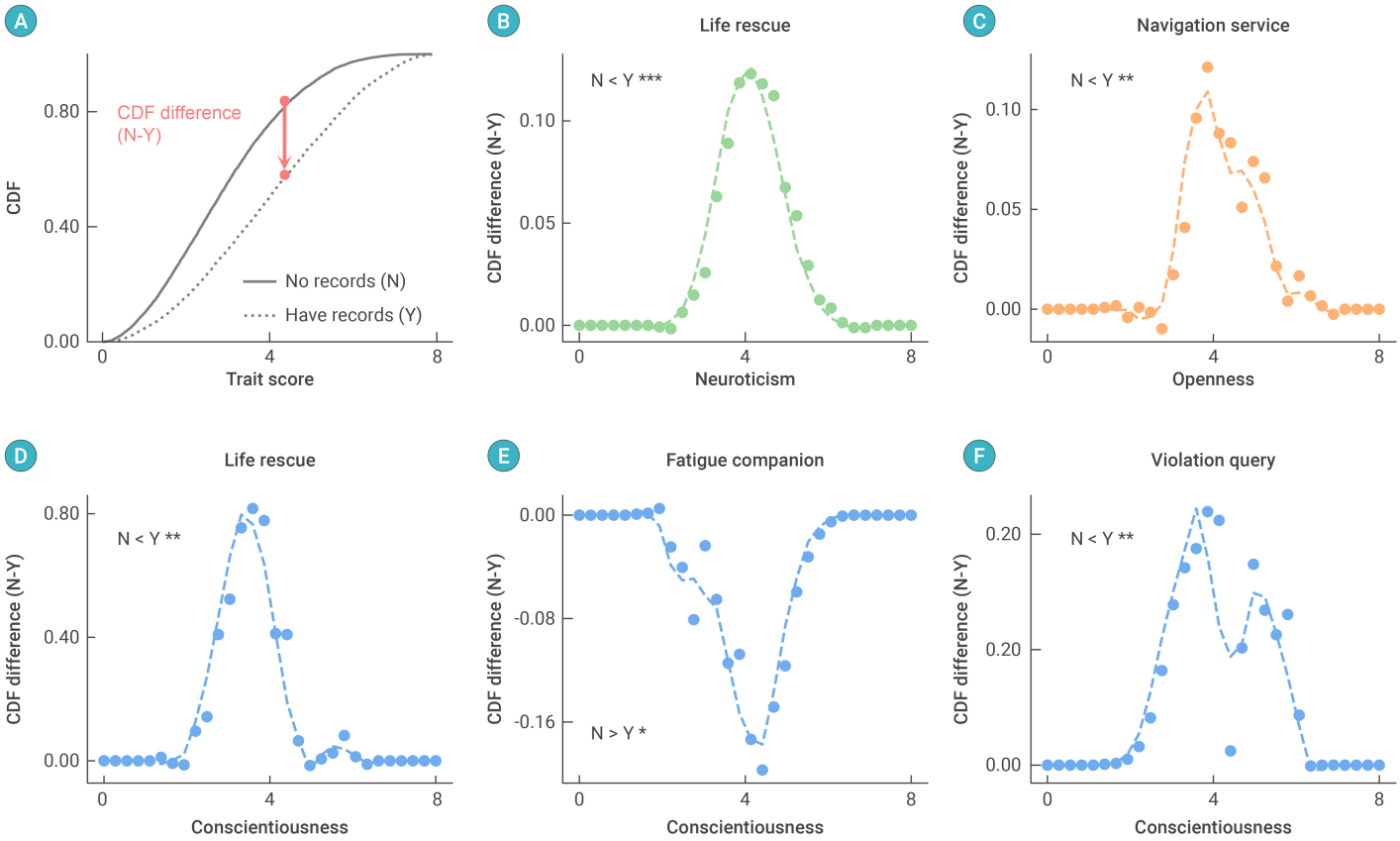



**Figure 6. Differences in trajectory traits between drivers with and without vehicle customer service records** A. A schematic of the CDF difference (N-Y), which represents the difference between the cumulative distribution function (CDF) of trait scores for drivers without vehicle customer service records (N; gray solid line) and those with such records (Y; gray dashed line). B. Difference between the CDF of the neuroticism scores for drivers with records of life rescue service requests and those without such records. N<Y*** indicates Mann–Whitney U test result $P \leq .001$, with the alternative hypothesis that trait scores from the population without service records are less than trait scores from the population with such records. The other notations for $P$ values are as follows: $P \leq .05$ (*), $P \leq .01$ (**). C. Difference in openness scores between drivers with and without navigation service records, similar to A. D, E, and F show the difference in conscientiousness scores between drivers with and without records of life rescue, fatigue companion and violation query, respectively.

analyze the impact of the trajectory integrity on trajectory trait profile.

When faced with missing data, each trait domain exhibits different degrees of stability, and the stability of different scorer groups also varies (Figure 7A). Extroversion and neuroticism are the most stable, followed by conscientiousness, while openness presents the weakest stability. The minimum sampling ratios required for a tolerance threshold $\beta_{SD} = 0.5$ are 0.2 for extroversion, 0.4 for openness, 0.2 for neuroticism, and 0.3 for conscientiousness, respectively (Table S6). That means, for neuroticism, only 20 percent of the trips are sufficient to recover the relative position of an individual's trait score in the dataset, and the same for other trait domains. The differences in stability exist because trajectory integrity has different effects on the trajectory features belonging to each trait (Figure 7B). *Trips per month* in extroversion and *random entropy* in openness are strongly affected by trajectory integrity as they are relevant to the number of trips. *OD entropy* in openness is also greatly affected as the missing data may disrupt the OD patterns. The eight features of neuroticism are all relatively stable, while the stability of conscientiousness is influenced by *date-time entropy* and *time-of-day entropy* significantly. Furthermore, trajectory integrity has a greater impact on *date-time entropy* than *time-of-day entropy*. This may be because *date-time entropy* captures fine-grained rhythms of individual travel times, which requires higher trajectory integrity. Moreover, scorer groups' performance varies also due to the impact of trajectory integrity on the features (Table S6 & Figure S5). The high scorers of extroversion and openness, and the low scorers of conscientiousness exhibit the largest average absolute differences in their respective trait domains, indicating that they require more complete data to reveal their travel patterns. For instance, compared to high scorers in conscientiousness, low scorers require higher trajectory integrity to demonstrate the uncertainty and heterogeneity of departure times in travel.

In general, missing data has a minor influence on the profiles, and these impacts can be ignored under a certain ratio of missing data. This implies that partial trips are capable of capturing the travel pattern of an individual. Meanwhile, missing data have different effects on the trajectory features. Finding the appropriate ratio of trips for different features can guide data collection and processing. For example, entropy-related features require more complete data to capture patterns, while driving behavior features, such as *average speed* and *average ratio of harsh shifting points*, can be extracted in a small ratio of trips.

***Seasonal changes.*** Through this experiment, we attempt to answer two questions similar to trajectory integrity: 1) Does the profile change with the seasons? 2) Do drivers with different trait scores have different performances under seasonal changes? To answer these two questions, we divide the trajectories by season, calculate seasonal trait scores, and then use the one sample T-test and statistics $d$ and $d_{SD}$ (see Note S8) to report the differences.

Trajectory trait profiles change slightly with the alternation of seasons (Figure 8A), indicating that the profiles for the sample group are stable. Nevertheless, different traits have different stability. Neuroticism and conscientiousness are more stable than extroversion and openness. In terms of extroversion, the average trait scores of the drivers increase from winter to spring and from spring to summer. Based on the changing trends of item scores (Figure 8B), we can infer drivers in our dataset become more active and travel more from winter to spring (as shown by *trips per month*), and stay at home less from spring to summer (as shown by *ratio of stay time at home*), contributing to the increases in extroversion trait scores. In contrast, the average trait score of the drivers in extroversion decreases significantly from autumn to winter since the two item scores show opposite trends. In terms of openness, drivers have the highest average openness score in summer, indicating that they are more willing to explore new locations during

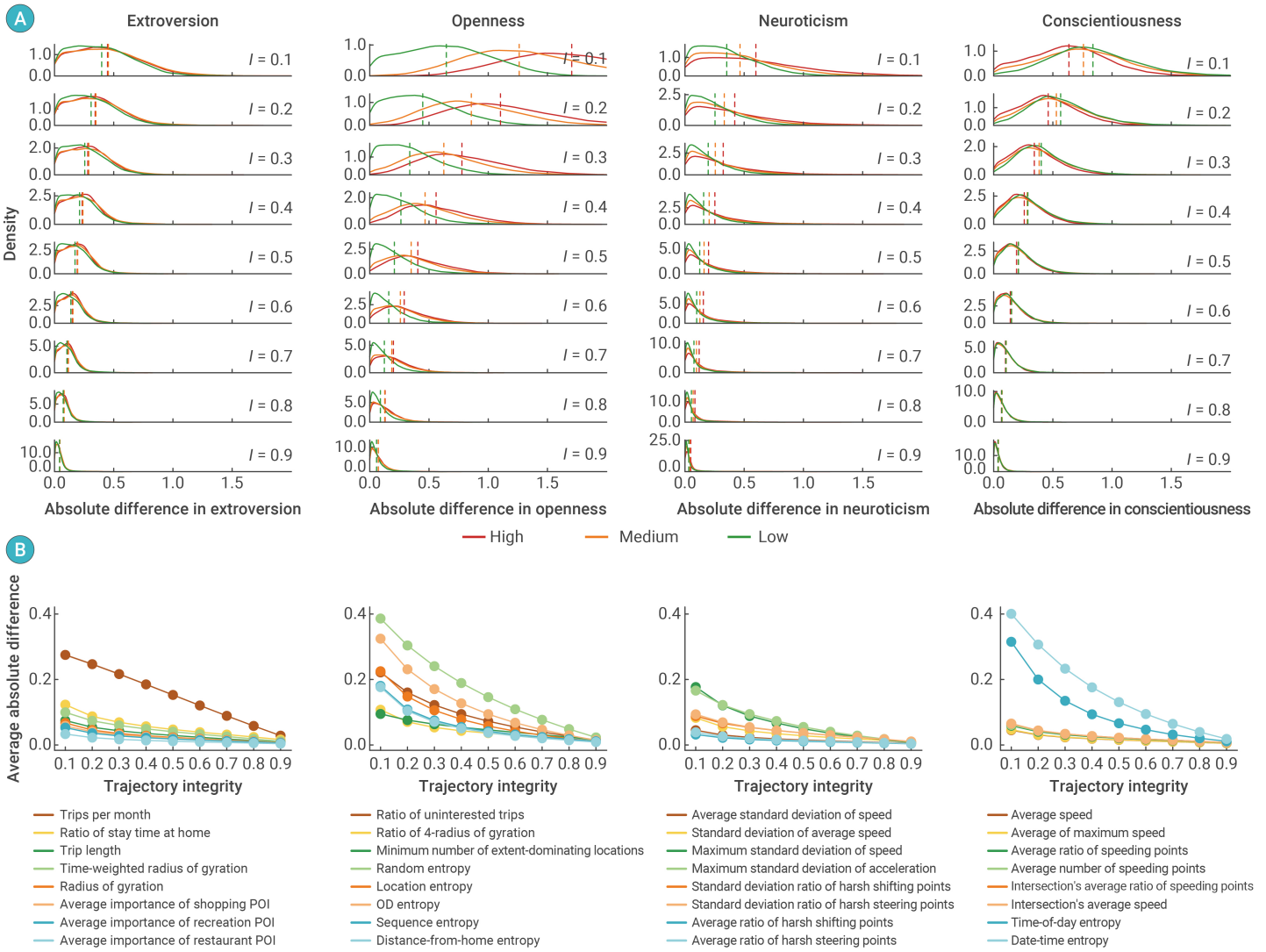



**Figure 7. The impact of trajectory integrity on trajectory trait profile** A. Distribution of the absolute difference in trait scores between the entire and randomly sampled trips for extroversion, openness, neuroticism, and conscientiousness, respectively. *I* represents the sampling ratio, i.e., trajectory integrity. B. The average absolute difference in item scores between the entire and randomly sampled trajectories for four trait domains, where each polyline corresponds to an item of the subscales.

the summer. Besides, drivers have the lowest average openness score in spring. This is because although drivers travel more in spring, they tend to go to frequently visited places, which leads to a decrease in the scores of the four spatial entropy items except for *random entropy* which is irrelevant to visiting frequency. In terms of neuroticism, the average trait score of drivers increases from autumn to winter, with an increase in item scores related to driving instability among trips (e.g., *maximum standard deviation of acceleration*, *maximum standard deviation of speed*, and *standard deviation of average speed*). One possible explanation is that road conditions are variable and complex in winter, and drivers tend to drive impulsively in unfavorable situations. For conscientiousness, drivers' overall driving speed increases, and departure times become irregular from autumn to winter, leading to a decrease in the average trait score. In addition, the two temporal entropies (*time-of-day entropy*, *date-time entropy*) are the lowest in autumn, indicating that drivers' departure times are relatively concentrated and regular.

The average scores of the different scorer groups for each trait domain have similar trends over seasons and are consistent with the entirety (Figure 8C). However, the magnitude of change in average scores for openness and neuroticism varies across different scorer groups (Table S7). From the perspective of individuals, the changes are relatively consistent within a group whose average score varies significantly with season. For example, the openness scores of drivers consistently increase from spring to summer, showing more rows in dark red. Conversely, insignificant average score change can be due to small changes of all individuals (e.g., low scorers in conscientiousness show little change in trait scores from summer to autumn) or changes in scores offsetting each other (e.g., neuroticism scores of partial drivers in the high scorers increase, while others decrease as the seasons change).

In general, seasonal changes have a limited impact on profiles, and trajectory traits remain stable among seasons within a year. Nevertheless, the directions of changes in individual trait scores have relatively high consistency in some trait domains. This indicates that these changes are relevant to seasonal environmental factors rather than occurring randomly. It would be beneficial to explore potential influencing factors on trajectory trait profiles, such as temperature, precipitation, and holidays, by including external data.

***Traffic conditions.*** We try to answer the following questions: Is there an association between the trajectory trait profile and traffic conditions? To answer the question, we compare the trait scores between drivers whose trips have high and low congestion ratios, which are defined based on the Traffic Performance Index (Note S9).

Drivers with a high congestion ratio have higher average neuroticism scores than those with a low congestion ratio (Figure S6A; independent samples T-test with $P = .042$, two-tailed). The histogram suggests that there are more drivers with low congestion ratios in the low neuroticism score range ($< 4$); in contrast, more drivers with high congestion ratios in the high score range. Besides, we compare the individual-level driving behaviors in

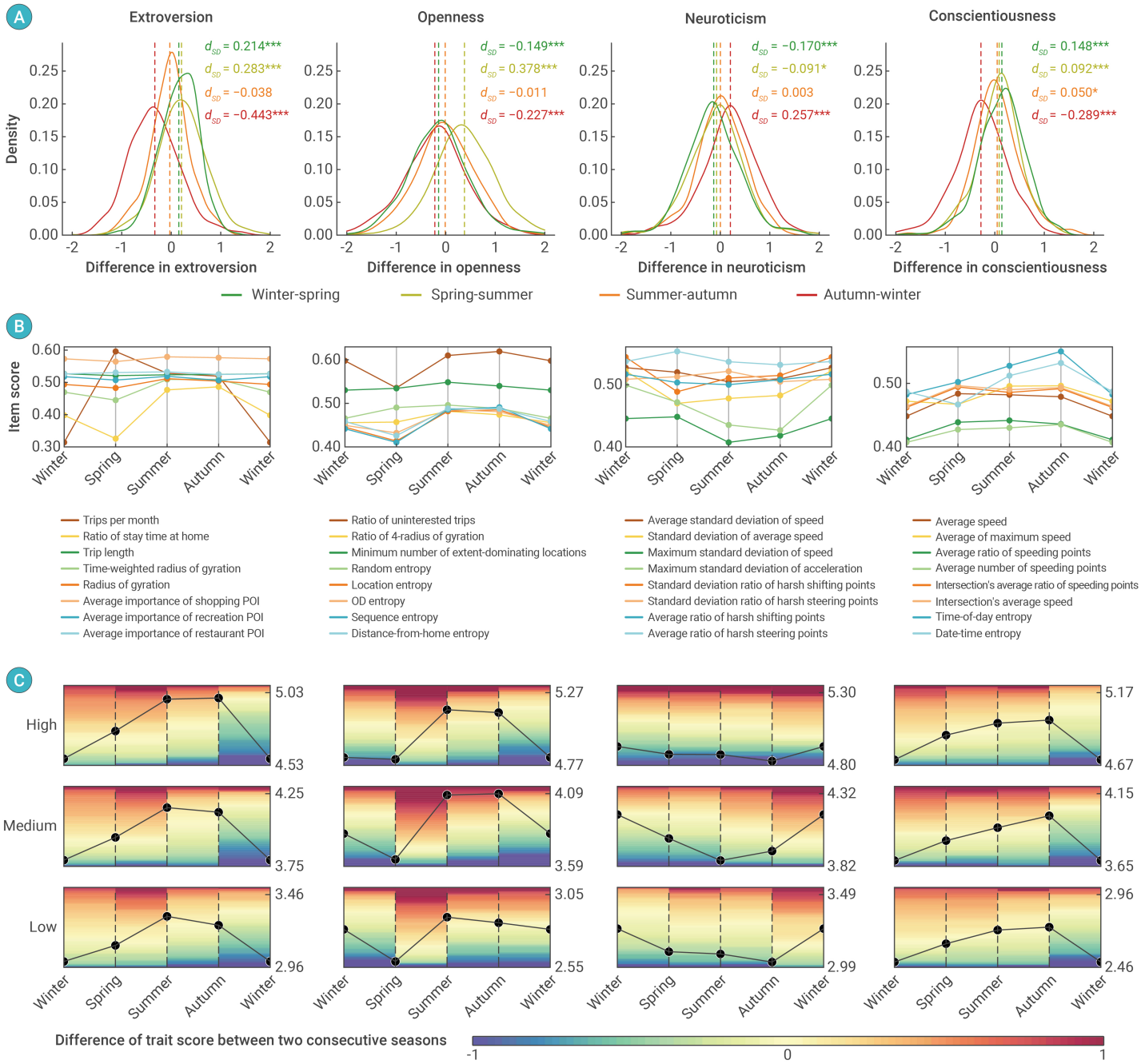



**Figure 8. The impact of seasonal changes on trajectory trait profile** A. Distribution of differences of trait scores between two consecutive seasons in four trait domains, respectively, where the four curves with different colors correspond to seasonal changes in a year. $d_{SD}$ is the statistic reporting the difference. The notation for $P$ values (two-tail) in the one sample T-test is as follows: $P \leq .05$ (*), $P \leq .01$ (**), $P \leq .001$ (***), and asterisks are not used when $P > .05$. B. Changes of item scores in each trait domain with alternating seasons, where each polyline corresponds to an item of the subscales. C. Changes of trait scores with alternating seasons for high, medium and low scorer groups, respectively. The polyline denotes the average trait score of drivers in a group. Each row in the slot of two consecutive seasons represents a driver, and color represents the change of trait scores. Red corresponds to an increase in score, while blue indicates a decrease in score, and the darker the color, the greater the change.

congested and smooth periods by taking "Driver: 125" as an example (Figure S6B). The scatter plots and amplified maps demonstrate that the fluctuation of acceleration is larger during congested period. Further, the ratio of harsh shifting points ($|\text{acceleration}| > 1.67 m/s^2$) during congested period is higher than that during smooth period.

To conclude, neuroticism is related to traffic conditions, indicating that impulsive driving behavior is not constant and is determined by not only inherent personality traits but also traffic conditions. Nonetheless, the correlation between neuroticism scores and congestion ratios is weak (Pearson correlation $r = 0.03$). This could be due to that neuroticism is influenced by various factors, including driving habits, emotions, and external distractions in addition to traffic conditions. The analysis of the relations between other trait domains and traffic conditions can be found in Table S8. It is important to recognize that trajectory trait profile is not equal to personality profile, and should be treated as a synthesis of those factors in driving.

## CONCLUSIONS

In this study, we have introduced a framework for constructing individual driver profiles using vehicle GPS trajectories, in which mobility patterns and driving behaviors of drivers are characterized by four fundamental trajectory traits: extroversion, openness, neuroticism, and conscientiousness. Trajectory traits describe and interpret drivers' travel through tendencies and preferences, offering predictive insights into their potential behavior across diverse scenarios. TTS developed in this framework serves as an effective

tool for measuring the trajectory traits, enabling the quantification of driver travel patterns and facilitating comparative analysis among drivers. We evaluated the validity and reliability of TTS and the proposed framework using long-term trajectory data from 2,051 volunteered private vehicle drivers.

Trajectory traits represent the behavior patterns of drivers under the joint influence of internal personality and external environmental factors. Experimental results reveal that trajectory traits are associated with personality traits by analyzing the connections between trajectory traits and external behavioral data of drivers. Previous research has demonstrated a positive correlation between personality traits and certain travel patterns, such as extroversion with the frequency and range of travel,[51] openness with diversity of location,[51] neuroticism with impulsive driving behavior,[35] and conscientiousness with traffic safety awareness and sense of responsibility.[52] The above correlations are also found between trajectory traits and these travel patterns. Therefore, it can be inferred that trajectory traits are influenced by personality traits to some extent, leading to drivers exhibiting personality-related travel patterns. Nonetheless, such an influence is not dominant, as trajectory traits are also shaped by environmental factors. We found that seasonal changes and traffic conditions have a small but meaningful impact on trajectory trait profiles. It suggests that trajectory traits are influenced by environmental factors, in contrast to the relatively persistent nature of personality traits. In addition, trajectory trait profiles are slightly affected by trajectory integrity. An appropriate subset of trajectories can accurately capture the travel patterns and reconstruct the profile obtained from the entire data. This implies that drivers' mobility and driving operation have inherent regularity across different trips, and such information can guide the setting of spatiotemporal granularity of trajectory recording and processing in diverse applications.

The proposed framework of trajectory trait profile is not limited to private vehicles and can be applied to a variety of trajectory sources, such as GPS location data of mobile phones, call detail records (CDRs), and taxi trajectories. Although some traits may be inapplicable to certain data sources, there is still practical value. For example, trajectory extroversion and openness measured by GPS location data of mobile phones and CDRs, can benefit destination recommendation, while taxi trajectories can be leveraged to measure trajectory neuroticism and conscientiousness for revealing driving risks and enhancing passenger safety. Furthermore, our profiling method, including trajectory features design and scale construction and evaluation, can be extended to other profile types, such as driving style profile oriented to travel safety, or profile that depicts individuals' socio-economic properties.

There are several potential avenues for further exploration. One direction is to augment the items in TTS by fusing additional data. For example, incorporating traffic violation records to improve the items measuring conscientiousness. Also, many other potential environmental and demographic factors that affect trajectory trait profile can be discussed, such as weather, road quality, driver age, income, and occupation. Trajectory traits may help identify commonalities and disparities in travel patterns across diverse regions and populations. The commonalities offer valuable perspectives on universal human mobility patterns; while investigating the underlying causality of disparities can uncover contextual factors influencing travel, such as road infrastructure, social and gender equality, and culture. Moreover, exploring the associations between trajectory traits and personality traits may provide a pathway to extrapolate personality from trajectory data.

## REFERENCES


1. Dioha M. O., Lukuyu J., Virgüez E., et al. (2022). Guiding the deployment of electric vehicles in the developing world. *Environ. Res. Lett.* **17**:071001. DOI:10.1088/1748-9326/ac765b
2. Zhang Y., Wang L., Sun T., et al. (2024). Vehicle-based monitoring and AI unravel patterns of on-road carbon and pollutant emissions. *Innov. Geosci.* **2**:100085. DOI:10.59717/j.xinn-geo.2024.100085
3. Zheng Y., Capra L., Wolfson O., et al. (2014). Urban computing: Concepts, methodologies, and applications. *ACM Trans. Intell. Syst. Technol.* **5**:1–55. DOI:10.1145/2629592
4. Nouvellet P., Bhatia S., Cori A., et al. (2021). Reduction in mobility and COVID-19 transmission. *Nat. Commun.* **12**:1–9. DOI:10.1038/s41467-021-21358-2
5. Li F., Gui Z., Zhang Z., et al. (2020). A hierarchical temporal attention-based LSTM encoder-decoder model for individual mobility prediction. *Neurocomputing* **403**:153–166. DOI:10.1016/j.neucom.2020.03.080
6. Händel P., Ohlsson J., Ohlsson M., et al. (2014). Smartphone-based measurement systems for road vehicle traffic monitoring and usage-based insurance. *IEEE Syst. J.* **8**:1238–1248. DOI:10.1109/JSYST.2013.2292721
7. Fazeen M., Gozick B., Dantu R., et al. (2012). Safe Driving Using Mobile Phones. *IEEE Trans. Intell. Transp. Syst.* **13**:1462–1468. DOI:10.1109/tits.2012.2187640
8. Júnior J. F., Carvalho E., Ferreira B. V., et al. (2017). Driver behavior profiling: An investigation with different smartphone sensors and machine learning. *PLoS One* **12**:e0174959. DOI:10.1371/journal.pone.0174959
9. Gui Z., Sun Y., Yang L., et al. (2021). LSI-LSTM: An attention-aware LSTM for real-time driving destination prediction by considering location semantics and location importance of trajectory points. *Neurocomputing* **440**:72–88. DOI:10.1016/j.neucom.2021.01.067
10. Singh H. and Kathuria A. (2021). Profiling drivers to assess safe and eco-driving behavior – A systematic review of naturalistic driving studies. *Accid. Anal. Prev.* **161**:106349. DOI:10.1016/j.aap.2021.106349
11. Chen Y., Chen X. and Gao Z. (2024). Toward equitable, transparent, and collaborative human mobility computing for smart cities. *The Innovation* **5**. DOI:10.1016/j.xinn.2024.100672
12. Ouaftouh S., Zellou A. and Idri A. (2015). User profile model: A user dimension based classification. 2015 10th International Conference on Intelligent Systems: Theories and Applications. DOI: 10.1109/SITA.2015.7358378.
13. Alonso F., Esteban C., Montoro L., et al. (2019). Conceptualization of aggressive driving behaviors through a Perception of aggressive driving scale (PAD). *Transp. Res. F: Traffic Psychol. Behav.* **60**:415–426. DOI:10.1016/j.trf.2018.10.032
14. Spano G., Caffò A. O., Lopez A., et al. (2019). Validating Driver Behavior and Attitude Measure for Older Italian Drivers and Investigating Their Link to Rare Collision Events. *Front. Psychol.* **10**. DOI:10.3389/fpsyg.2019.00368.
15. Taubman-Ben-Ari O., Mikulincer M. and Gillath O. (2004). The multidimensional driving style inventory - Scale construct and validation. *Accid. Anal. Prev.* **36**:323–332. DOI:10.1016/S0001-4575(03)00010-1
16. Banerjee T., Chowdhury A., Chakravarty T., et al. (2020). Driver authentication by quantifying driving style using GPS only. 2020 IEEE International Conference on Pervasive Computing and Communications Workshops. DOI: 10.1109/PerComWorkshops48775.2020.9156080.
17. Kashevnik A., Lashkov I. and Gurtov A. (2020). Methodology and Mobile Application for Driver Behavior Analysis and Accident Prevention. *IEEE Trans. Intell. Transp. Syst.* **21**:2427–2436. DOI:10.1109/TITS.2019.2918328
18. Fugiglando U., Santi P., Milardo S., et al. (2017). Characterizing the "Driver DNA" Through CAN Bus Data Analysis. Proceedings of the 2nd ACM International Workshop on Smart, Autonomous, and Connected Vehicular Systems and Services. DOI: 10.1145/3131944.3133939.
19. Tempelaar D., Rienties B. and Nguyen Q. (2020). Subjective data, objective data and the role of bias in predictive modelling: Lessons from a dispositional learning analytics application. *PLoS One* **15**:e0233977. DOI:10.1371/journal.pone.0233977
20. Ellison A. B., Greaves S. P. and Bliemer M. C. J. (2015). Driver behaviour profiles for road safety analysis. *Accid. Anal. Prev.* **76**:118–132. DOI:10.1016/j.aap.2015.01.009
21. Li Y., Zhao L. and Rilett L. R. (2019). Driving performances assessment based on speed variation using dedicated route truck GPS data. *IEEE Access* **7**:51002–51013. DOI:10.1109/ACCESS.2019.2909572
22. González M. C., Hidalgo C. A. and Barabási A. L. (2008). Understanding individual human mobility patterns. *Nature* **453**:779–782. DOI:10.1038/nature06958
23. Pappalardo L., Simini F., Rinzivillo S., et al. (2015). Returners and explorers dichotomy in human mobility. *Nat. Commun.* **6**:1–8. DOI:10.1038/ncomms9166
24. Schneider C. M., Belik V., Couronné T., et al. (2013). Unravelling daily human mobility motifs. *J. R. Soc. Interface.* **10**:20130246. DOI:10.1098/rsif.2013.0246
25. Iqbal M. S., Choudhury C. F., Wang P., et al. (2014). Development of origin-destination matrices using mobile phone call data. *Transp. Res. C: Emerg. Technol.* **40**:63–74. DOI:10.1016/j.trc.2014.01.002
26. Song C., Qu Z., Blumm N., et al. (2010). Limits of predictability in human mobility. *Science* **327**:1018–1021. DOI:10.1126/science.1177170
27. Wu L., Yang L., Huang Z., et al. (2019). Inferring demographics from human trajectories and geographical context. *Comput. Environ. Urban Syst.* **77**:101368. DOI:10.1016/j.compenvurbsys.2019.101368
28. Xiong Q., Liu Y., Xie P., et al. (2021). Revealing correlation patterns of individual location activity motifs between workdays and day-offs using massive mobile phone data. *Comput. Environ. Urban Syst.* **89**:101682. DOI:10.1016/j.compenvurbsys.2021.101682
29. McCann S. J. H. (2015). Big five personality and residential mobility: A state-level analysis of the USA. *Journal of Social Psychology* **155**:274–291. DOI:10.1080/00224545.2015.1007027
30. Kim D. Y. and Song H. Y. (2018). Method of predicting human mobility patterns using deep learning. *Neurocomputing* **280**:56–64. DOI:10.1016/j.neucom.2017.07.069
31. Alessandretti L., Sapiezynski P., Sekara V., et al. (2018). Evidence for a conserved quantity in human mobility. *Nat. Hum. Behav.* **2**:485–491. DOI:10.1038/s41562-018-0364-x
32. Moro E., Calacci D., Dong X., et al. (2021). Mobility patterns are associated with experienced income segregation in large US cities. *Nat. Commun.* **12**. DOI:10.1038/

s41467-021-24899-8.
33. Canzian L. and Musolesi M. (2015). Trajectories of depression: unobtrusive monitoring of depressive states by means of smartphone mobility traces analysis. Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing. DOI: 10.1145/2750858.2805845.
34. McCrae R. R. and John O. P. (1992). An Introduction to the Five - Factor Model and Its Applications. *J. Pers.* **60**:175–215. DOI:10.1111/j.1467-6494.1992.tb00970.x
35. Jovanović D., Lipovac K., Stanojević P., et al. (2011). The effects of personality traits on driving-related anger and aggressive behaviour in traffic among Serbian drivers. *Transp. Res. F: Traffic Psychol. Behav.* **14**:43–53. DOI:10.1016/j.trf.2010.09.005
36. Goldberg L. R., Johnson J. A., Eber H. W., et al. (2006). The international personality item pool and the future of public-domain personality measures. *J. Res. Pers.* **40**:84–96. DOI:10.1016/j.jrp.2005.08.007
37. McNeish D. and Wolf M. G. (2020). Thinking twice about sum scores. *Behav. Res. Methods* **52**:2287–2305. DOI:10.3758/s13428-020-01398-0
38. Rza, cedil, dkowski G., et al. (2020). A Generalized Logistic Function and its Applications. *Found. Manag.* **12**:85–92. DOI:10.2478/fman-2020-0007
39. DeYoung C. G., Quilty L. C. and Peterson J. B. (2007). Between Facets and Domains: 10 Aspects of the Big Five. *J. Pers. Soc. Psychol.* **93**:880–896. DOI:10.1037/0022-3514.93.5.880
40. Ester M., Kriegel H.-P., Sander J., et al. (1996). A density-based algorithm for discovering clusters in large spatial databases with noise. Proceedings of the Second International Conference on Knowledge Discovery and Data Mining. DOI: 10.5555/3001460.3001507.
41. van der waerden P., Timmermans H. and de Bruin-Verhoeven M. (2015). Car drivers' characteristics and the maximum walking distance between parking facility and final destination. *J. Transp. Land Use.* **10**. DOI:10.5198/jtlu.2015.568.
42. De Champlain A. F. (2010). A primer on classical test theory and item response theory for assessments in medical education. *Med. Educ.* **44**:109–117. DOI:10.1111/j.1365-2923.2009.03425.x
43. Kimberlin C. L. and Winterstein A. G. (2008). Validity and reliability of measurement instruments used in research. *Am. J. Health-Syst. Pharm.* **65**:2276–2284. DOI:10.2146/ajhp070364
44. Streiner D. L. (2003). Starting at the beginning: An introduction to coefficient alpha and internal consistency. *J. Pers. Assess.* **80**:99–103. DOI:10.1207/S15327752JPA8001_18
45. Taber K. S. (2018). The Use of Cronbach's Alpha When Developing and Reporting Research Instruments in Science Education. *Res. Sci. Educ.* **48**:1273–1296. DOI:10.1007/s11165-016-9602-2
46. Dahlen E. R., Martin R. C., Ragan K., et al. (2005). Driving anger, sensation seeking, impulsiveness, and boredom proneness in the prediction of unsafe driving. *Accid. Anal. Prev.* **37**:341–348. DOI:10.1016/j.aap.2004.10.006
47. Yared T., Patterson P. and All E. S. A. (2020). Are safety and performance affected by navigation system display size, environmental illumination, and gender when driving in both urban and rural areas. *Accid. Anal. Prev.* **142**:105585. DOI:10.1016/j.aap.2020.105585
48. Ting P. H., Hwang J. R., Doong J. L., et al. (2008). Driver fatigue and highway driving: A simulator study. *Physiol. Behav.* **94**:448–453. DOI:10.1016/j.physbeh.2008.02.015
49. Arumugam S. and Bhargavi R. (2019). A survey on driving behavior analysis in usage based insurance using big data. *J. Big Data* **6**:1–21. DOI:10.1186/s40537-019-0249-5
50. Chorley M. J., Whitaker R. M. and Allen S. M. (2015). Personality and location-based social networks. *Comput. Hum. Behav.* **46**:45–56. DOI:10.1016/j.chb.2014.12.038
51. Noë N., Whitaker R. M., Chorley M. J., et al. (2016). Birds of a feather locate together. *Foursquare checkins and personality homophily. Comput. Hum. Behav.* **58**:343–353. DOI:10.1016/j.chb.2016.01.009
52. Feng Z., Ji N., Luo Y., et al. (2021). Exploring the influencing factors of public traffic safety awareness in China. *Cogn. Technol. Work.* **23**:731–742. DOI:10.1007/s10111-020-00655-8

## FUNDING AND ACKNOWLEDGMENTS

This paper is supported by National Natural Science Foundation of China (42090010, 41971349 and 41930107), National Key R&D Program of China (2021YFE0117000, 2018YFC0809806) and Fundamental Research Funds for the Central Universities (2042024kf0005). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

## AUTHOR CONTRIBUTIONS

Z.G. and Y.L. envisioned the study. Y.X. and S.G. evolved the overarching research goals. Y.L. and Z.G. designed the method. Y.L., F.M., and L.B. prepared the datasets and conducted experiments. Y.L., Z.G., Y.X., S.G., F.M., A.Z., D.P., and F.L. performed the analysis and wrote the manuscript. H.W., Z.G., Y.X., S.G., and F.L. provided advice in analysis. H.W. and J.G. provided supervision and guidance throughout the study. All authors contributed to and approved the manuscript.

## DECLARATION OF INTERESTS

The authors declare no competing interests.

## DATA AND CODE AVAILABILITY

The raw trajectory data of private vehicles is not publicly available due to privacy considerations. However, item and trait scores for the 2,051 drivers and the sample trajectories of ten drivers are publicly available at: https://github.com/ZPGuiGroupWhu/Human-Mobility-Analysis/tree/master/Trajectory-Trait-Profile. Other datasets including POI, road network, and traffic performance index, can be obtained using their Public Search APIs, which are listed in Table S3. The code of feature extraction, trait score calculation, and TTS evaluation is publicly available at: https://github.com/ZPGuiGroupWhu/Human-Mobility-Analysis/tree/master/Trajectory-Trait-Profile.

## SUPPLEMENTAL INFORMATION

It can be found online at https://doi.org/10.59717/j.xinn-geo.2024.100114

## LEAD CONTACT WEBSITE

Zhipeng Gui: https://jszy.whu.edu.cn/ZhipengGui