Research Article

TRR
JOURNAL OF THE TRANSPORTATION RESEARCH BOARD

Transportation Research Record
1–16

DOI: 10.1177/03611981221092005
journals.sagepub.com/home/trr
SAGE

Micromobility Trip Origin and Destination Inference Using General Bikeshare Feed Specification Data

Yiming Xu[1], Xiang Yan[1], Virginia P. Sisiopiku[2], Louis A. Merlin[3], Fangzhou Xing[4], and Xilei Zhao[1]

Abstract

Emerging micromobility services (e.g., e-scooters) have a great potential to enhance urban mobility but more knowledge on their usage patterns is needed. The General Bikeshare Feed Specification (GBFS) data are a possible source for examining micromobility trip patterns, but efforts are needed to infer trips from the GBFS data. Existing trip inference methods are usually based on the assumption that the vehicle identity (ID) of a micromobility option (e-scooter or e-bike) does not change, and so they cannot deal with data with vehicle IDs that change over time. This paper proposes a comprehensive package of algorithms to infer origin–destination (OD) pairs from GBFS data with static vehicle ID and unlinked trip origins and destinations from GBFS data with resetting and dynamic vehicle ID. The algorithms were implemented in Washington D.C. by analyzing one week (last week of February 2020) of GBFS data published by six vendors, and the inference accuracy of the proposed algorithms are evaluated by R-squared, mean absolute error, and sum absolute error. It is found that the R-squared measure is larger than 0.9 and the MAE measure is less than 2 when the algorithms are evaluated with a 400 m $\times$ 400 m grid. The absolute errors are relatively larger in the downtown area, and the inference error is relatively high during early morning and early nighttime. The accuracy of the trip inference algorithms is sufficiently high for most practical applications.

Keywords

data and data science, information systems and technology, urban transportation data and information systems

In recent years, shared micromobility services have experienced explosive growth in urban areas (*1*, *2*). Micromobility refers to small, single-passenger transportation modes rented for short-term use, such as e-scooters, docked bikes, and dockless bikes. Among all the micromobility options, e-scooters are growing at the fastest pace, largely owing to their flexibility, convenience, affordability, and the fun factor. They are especially attractive to travelers for short-distance trips. Micromobility also offers a potential solution to the "first-mile/last-mile" problem (the problem of public transit being unable to get passengers to the doorstep of their destinations) that has long troubled public transit. On the other hand, some are concerned that micromobility will mainly replace active travel modes such as walking. Also, some micromobility vehicles are parked inappropriately (e.g., on the sidewalks), which inhibits other productive use of the curbside. Given the potential benefits and drawbacks that micromobility brings to urban mobility, great interest exists among transportation professionals for understanding the usage patterns of micromobility services.

Survey data and web-scraping data collected by public application programming interfaces (APIs) are the main data sources for researchers to analyze and understand micromobility. Survey data covers a variety of topics

[1]Department of Civil and Coastal Engineering, University of Florida, Gainesville, FL
[2]Department of Civil, Construction, and Environmental Engineering, University of Alabama at Birmingham, Birmingham, AL
[3]Department of Urban and Regional Planning, Florida Atlantic University, Boca Raton, FL
[4]Microsoft Corporation, Redmond, WA

Corresponding Author:
Xilei Zhao, xilei.zhao@essie.ufl.edu

that cannot be gathered through API data: the characteristics of micromobility users, their perception of and willingness to use the services, and the modal shift from cars to micromobility, because it can provide the user information (e.g., age, gender, and education) that the web-scraped data does not have (*3–6*). Compared with survey data, web-scraped data usually has larger sample size and more detailed spatiotemporal information such as the latitude and longitude of trip origin and trip start time. Therefore, web-scraped data can complement the survey data to obtain a more comprehensive view of micromobility. As part of the micromobility permit requirement, cities often require micromobility providers to share data through APIs prescribed by standard formats, including the General Bikeshare Feed Specification (GBFS) and the Mobility Data Specification (MDS). GBFS was initially developed as the open data standard for bikeshare system availability in 2015, but now it is applicable for nearly all shared micromobility systems in the North America (*7*). GBFS APIs report real-time information about available vehicles, which typically includes vehicle location, vehicle type (bike or scooter), and battery level. Created by the Los Angeles Department of Transportation (LADOT) in 2018, MDS extends GBFS to require additional information from mobility providers. The additional information may include data on unavailable vehicles in the network, trip characteristics, and trip trajectories (*8*). However, the MDS has received limited adoption so far, and the MDS APIs are usually not made available to the public. Accordingly, this study focuses on GBFS.

Some recent studies have extracted trip information from the GBFS data to examine the spatiotemporal patterns of scooter usage (*2*, *9*). The trip inference method used in these studies usually assumes that GBFS APIs report *Static Vehicle IDs*, that is, the ID of a given micromobility device does not change over time. However, this assumption no longer holds for many circumstances because the GBFS data standards are updated frequently. For example, it is necessary to rotate vehicle IDs randomly after each rental in GBFS v2.0 to reduce the potential exposure of private data (*10*). To enhance rider privacy, many micromobility providers currently operate GBFS APIs that report *Resetting Vehicle IDs* (i.e., the vehicle ID will be randomly rotated once the vehicle is unlocked for a new trip) or *Dynamic Vehicle IDs* (i.e., the vehicle ID of a given scooter randomly changes every several minutes). The existing approach used for trip inference is not applicable for these new vehicle ID reporting mechanisms. In other words, new methods are needed to infer micromobility trips from the GBFS APIs that report *Resetting Vehicle IDs* and *Dynamic Vehicle IDs*.

To this end, this study develops a systematic approach to infer trip origins and destinations for all existing vehicle ID types contained in the GBFS data. Three metrics are used—R-squared (R^2), mean absolute error (MAE), and sum absolute error (SAE)—to validate the inference accuracy of proposed algorithms. All proposed algorithms have an R-squared larger than 0.9 when scooter location data is aggregated to a $400\,\text{m} \times 400\,\text{m}$ grid, which is sufficiently high for most practical purposes. The inferred e-scooter trip origins and destinations show usage patterns that are consistent with previous studies (*2*, *11*).

The contribution of this study is two-fold: (i) a comprehensive package of algorithms is proposed to infer trip origins and destinations for all existing vehicle ID types contained in the GBFS data (OD pairs for "Static Vehicle ID," and unlinked trip origins and destinations for "Resetting and Dynamic Vehicle ID"); and (2) the algorithms are implemented in Washington D.C. based on the GBFS data published by six vendors and the inference accuracy of the proposed algorithms is validated.

The remainder of the paper is structured as follows. The next section presents a literature review on previous studies on e-scooter services. The third section describes the data collection procedures and the proposed algorithms in detail. The fourth section presents an application of the trip inference algorithms in Washington D.C., followed with a validation. The final section concludes the paper by summarizing the findings, identifying limitations, and suggesting future work.

Literature Review

GBFS is the open data standard for all shared micromobility systems, including e-scooter sharing systems and docked and dockless bikesharing systems. This section summarizes existing shared micromobility studies and discusses the data these studies used.

E-Scooter Sharing Studies and Corresponding Dataset

Studies on e-scooter sharing are burgeoning. Different sources of data have been used in these studies. These data can be categorized into survey data, trip data, and real-time location data.

Since survey data provide micromobility user information such as age, gender, and education, they are usually used to analyze the characteristics of e-scooter users and for modeling people's willingness to use e-scooter services (*3–5*, *12–14*). For example, based on a survey in Vienna, Austria, Laa and Leth (*3*) found that e-scooter users were more likely to be young, male, highly educated, and local residents. Almannaa et al. (*12*) showed that gender, age, and using ride-hailing services play an important role in people's willingness to use e-scooters. Some studies also used survey data to estimate the modal shift of e-scooter users (*3*, *13*, *15*, *16*). Some studies found that shared e-scooter services mostly replaced walking and public transport trips (*3*, *13*), and that shared e-scooters were also potentially competing with ridesourcing and taxis (*16*). Survey data can also provide trip information, such as trip purposes and approximate travel time (estimated by the respondents). For example, based on survey data, Bieliński and Ważna (*4*) found that e-scooters are often used for leisure rides. However, survey data also have some limitations. Given the relatively high cost of surveys, survey data often have limited samples (less than 1,000 in most cases). In addition, the survey data cannot provide detailed spatiotemporal trip information such as latitude and longitude of trip origin and precise trip start time. Trip data and real-time location data can serve as complements to analyze the spatiotemporal e-scooter usage patterns.

Some studies analyze the trip data published on a city government's open data portal (*17*, *18*). This kind of data directly provides information (e.g., trip origin, trip destination, trip duration, and trip distance) on every trip. For example, Bai and Jiao (*17*) explored e-scooter ridership patterns and analyzed the relationship between e-scooter usage and the built environment in Austin, TX, and Minneapolis, MN, based on the shared micromobility vehicle trips data published by the City of Austin and the City of Minneapolis (*19*, *20*). However, e-scooter trip data are unavailable to the public for most cities, because these cities have not mandated trip-level data sharing for their micromobility providers. Some authors thus acquired the e-scooter trip data from the city via personal connections. For example, Liu et al. (*21*) used the data provided by the City of Indianapolis to analyze the spatiotemporal patterns of e-scooter trips.

Other studies scraped the publicly accessible GBFS or data published by the micromobility vendors and then processed the data to derive needed information (e.g., trip origin and destination, trip trajectory). The scraped data are real-time location data. For example, McKenzie (*2*) compared spatial and temporal patterns of e-scooter-share and bikeshare usage in Washington, D.C. using GBFS data. In another study, McKenzie (*11*) further explored spatial and temporal differences in usage patterns between six vendors in Washington, D.C., that is, Bird, Lime, Lyft, Skip, Spin, and Jump. The methods used in the two McKenzie studies to derive e-scooter trips from scraped GBFS data were the same: a trip was identified as the time and location when an e-scooter last appeared available in the system, to the time and location when the same e-scooter next appeared available in the set of scraped data. However, since the vehicle ID was used to identify trips, this trip inference method requires the vehicle ID of e-scooter to be consistent over time. Zou et al. (*9*) examined e-scooter travel patterns at the street-segment level based on e-scooter trip trajectories in Washington, D.C. E-scooter trip trajectories were pinpointed based on the attribute in the data that indicated whether an e-scooter was in use. However, this method requires static vehicle ID as well as information on all e-scooters in the system. Only one vendor's data met these requirements and was used in this study; this kind of data is no longer available in Washington, D.C. Baltra et al. (*22*) discussed the privacy issues of GBFS specification and used GBFS data in Los Angeles to infer trips. The logic of trip inference in Baltra et al. (*22*) was the same as McKenzie (*2*, *11*) and relied on a static scooter ID. Zhu et al. (*23*) scraped the docked e-scooter and station status data in bikesharing systems in Singapore. Based on the data, they investigated spatiotemporal heterogeneity of bikesharing and e-scooter sharing systems in two urban areas in Singapore. E-scooter trips were identified by continuously checking the vehicle ID of e-scooters in each station. If a vehicle ID disappeared, this was considered a trip origin, and if a vehicle ID appeared, this was considered a trip destination. A summary of existing studies based on inferred e-scooter trips is presented in Table 1.

Table 1. Summary of Existing Studies Based on Inferred E-Scooter Trips

Study	Data source	Inference method	Reliance on *Static Vehicle ID*
McKenzie (2)	GBFS	Disappearance and re-appearance of the vehicle in the dataset	Yes
McKenzie (*11*)	GBFS	Disappearance and re-appearance of the vehicle in the dataset	Yes
Zou et al. (9)	GBFS	Checking the attribute in the data that indicated whether an e-scooter was in use	Yes
Baltra et al. (22)	GBFS	Disappearance and re-appearance of the vehicle in the dataset	Yes
Zhu et al. (23)	Bikesharing system	Checking the vehicle ID of e-scooters in each station	Yes

Note: GBFS = General Bikeshare Feed Specification.

However, micromobility providers are moving away from *Static Vehicle ID* out of concern for privacy issues. A recent effort to preserve user privacy was to adopt *Resetting Vehicle ID*, *Dynamic Vehicle ID* , or both, in the published GBFS data, which is a growing trend since the release of GBFS v2.0 (*7*). For example, Lime published *Static Vehicle IDs* before September 24, 2019 but switched to *Dynamic Vehicle IDs* afterwards. The existing trip inference methods will not work for these kinds of data. Since vendors concentrate their vehicles at different locations, which results in distinctive spatiotemporal usage patterns across vendors (*11*), omitting trip data from vendors that publish *Dynamic Vehicle IDs* would provide an incomplete picture of e-scooter demand in the city.

In summary, the existing e-scooter trip inference methods can infer trips when providers use a *Static Vehicle ID*, but cannot deal with dockless e-scooter data with *Resetting Vehicle ID* or *Dynamic Vehicle ID*. Since the vendors are moving away from *Static Vehicle ID* to *Resetting Vehicle ID*, *Dynamic Vehicle ID*, or both, trip inference approaches (note that trip inference used in this paper refers not only to trip OD pairs, but also to separate trip origins and destinations) that can infer trips from GBFS data with different vehicle ID types are needed for a more thorough understanding of e-scooter-related travel behavior.

Bikesharing Studies and Corresponding Dataset

Bikesharing includes two schemes: docked bikesharing and dockless bikesharing. Existing studies used survey data to analyze the user characteristics, user's attitudes to bikesharing, modal shift, and usage patterns of both docked bikesharing and dockless bikesharing (*24–30*). For example, Guo et al. (*27*) explored factors affecting bikesharing usage and satisfaction of bikesharing users based on a survey conducted in Ningbo, China. The results showed that bikesharing satisfaction was associated with household income, the location of bikesharing stations, and users' perceptions. Martin and Shaheen (*24*) evaluated survey data from Washington D.C. and Minneapolis to explore who is shifting toward and away from public transit as a result of bikesharing. The results indicated that age, being male, living in lower density areas, and longer commute distances were associated with modal shift.

GBFS data have also been used by several bikesharing studies (*2*, *23*, *31–35*). In these studies, GBFS data has shown advantages in its large sample sizes and accurate spatiotemporal information compared with the survey data. For example, Qian et al. (*31*) analyzed spatial and temporal patterns of bikesharing usage patterns in San Francisco using trips inferred from GBFS data. The trip data included over 0.6 million trips, which provided a large sample for an in-depth analysis.

The GBFS data feeds for docked bikesharing and dockless bikesharing are different. Docked bikesharing trips can only start and terminate at bike stations, while the dockless shared bikes can be picked up and dropped off at any place within the area of operations. The GBFS data provide station information such as station location and number of vehicles available for docked bikesharing systems. By checking the vehicles available at each station, researchers can infer trip origins and destinations (linked OD for *Static Vehicle ID* and unlinked trip origins, and destinations for *Resetting Vehicle ID* and *Dynamic Vehicle ID*). The GBFS data on dockless bikesharing is similar to that of e-scooter sharing, so the trip inference challenges faced by dockless bikesharing are also similar to e-scooter sharing. Therefore, the methods developed in this paper can be readily applied to the dockless bikesharing system to accurately infer trip information.

Methods

This study focuses on the GBFS data collected from the e-scooter operators' public APIs and explores methods to accurately infer micromobility trips from the GBFS data with different kinds of vehicle IDs. This section describes data collection and three algorithms developed to infer trip origins and destinations using GBFS data.

Data Collection

The GBFS specification requires that vendors publish certain data feeds as JSON files, which are usually made accessible via public APIs provided by each vendor. Among these data feeds, the one that can be used to infer trip origins and destinations is *free_bike_status*, which provides location information of all currently available vehicles in the system. The attributes in *free_bike_status* are listed in Table 2. A data sample is as follows:

```
{''last_updated'':1582528501, ''ttl'':
300, ''data'':{''bikes'':[{''bike_id'':
8982, ''lat'':38.8962,''lon'':-76.9592,''
is_reserved'':0,''is_disabled'':0},
{''bike_id'':9408, ''lat'':38.8797,''
lon'':-77.0100,''is_reserved'':0,
''is_disabled'':0}]}}
```

The data feeds are updated every "time to live" (TTL) seconds. Since no specific value is required by the GBFS specification, the TTL varies among vendors. Typical TTL in GBFS includes 0 (the data should always be refreshed), 300, and 600. In this study, for the data feeds

Table 2. Attributes in *free_bike_status*

Attribute	Description	Type
last_updated	POSIX timestamp indicating the last time the data in was updated	Integer
ttl	Seconds before the data in this feed will be updated again	Integer
bike_id	Unique identifier of a vehicle	String
lat	Latitude of the vehicle location	Number
lon	Longitude of the vehicle location	Number
is_reserved	Is the vehicle currently reserved for someone else?	Integer
is_disabled	Is the vehicle currently disabled (broken)?	Integer

Note: POSIX = Portable Operating System Interface.

with a TTL below 60 s, the data is scraped every minute; otherwise, TTL seconds are used as the scraping interval.

Trip Origins and Destinations Inference Algorithms

As discussed in the previous section, the GBFS data only provide information on vehicles currently available in the system. In other words, information on vehicles in use is not available. In addition, the vendors may adopt different vehicle ID generating strategies. Currently, there are three different vehicle ID generating strategies in GBFS data:

1. *Static Vehicle ID*: Vehicle ID does not change until the vehicle is taken out of service.
2. *Resetting Vehicle ID*: Vehicle ID of corresponding vehicle randomizes after every trip, but is otherwise static. For example, when a user terminates a trip and returns the vehicle, the system will assign the vehicle a new random vehicle ID.
3. *Dynamic Vehicle ID* : Vehicle ID for all vehicles in the system randomizes every several minutes (typical time intervals are 30 min and 1 h).

The three ID generating strategies produce three different types of vehicle ID. The trip origins and destinations inference logic differs for different ID types. Therefore, for each vehicle ID type, an algorithm is developed to infer trip origins and destinations from GBFS data. The three algorithms are summarized in Table 3.

Vendors sometimes relocate e-scooters away from locations where they are not being used to high demand areas. This type of trip is called a "rebalancing trip" in this paper. To recharge low-battery e-scooters in time, some vendors (e.g., Lime) pay gig workers to recharge e-scooters at their residences. Participants are instructed to pick up e-scooters with low batteries, and drop them off at specific locations after recharging. This type of trip is named a "juicing trip" (*2*). Both rebalancing trip and juicing trip are not real trips conducted by travelers, thus it is intented to exclude such trips.

Static Vehicle ID. Since the GBFS data only provides real-time information on available e-scooters, once a user unlocks an e-scooter and starts a trip, information on this e-scooter disappears from the corresponding data feed. When the trip terminates, information on the e-scooter will reappear in the data feed. Therefore, it can be inferred that information on an e-scooter disappearing in the data feeds indicates a trip origin of this e-scooter. Similarly, information on an e-scooter reappearing in the data feeds indicates a trip destination. The inference logic is similar to previous studies (*2*, *20*, *22*). Based on this logic, an algorithm is developed to infer trip origins and destinations for e-scooters with *Static Vehicle ID*. The algorithm is presented in detail in Algorithm 1.

When e-scooter vendors publish valid data with a high update frequency, this algorithm can produce accurate inference results for origin and destination locations. Theoretically, the algorithm can identify all actual trips

Table 3. Summary of Inference Algorithms

Algorithm	Application scenario	Output	Causes of errors
Algorithm 1	Static Vehicle ID	Origin–destination pairs	GPS error, data update frequency.
Algorithm 2	Resetting Vehicle ID	Unlinked trip origins and destinations	GPS error, data update frequency, launch and elimination of vehicles.
Algorithm 3	Dynamic Vehicle ID	Unlinked trip origins and destinations	GPS error, data update frequency, launch and elimination of vehicles, movement of vehicles not in use.

Note: GPS = global positioning system.

Algorithm 1 Origin and Destination Pair Inference for Data with *Static Vehicle ID*

input all e-scooter records D_{all}, data scraping interval Δt_s, travel time threshold TT_{max}, upper bound for average speed v_{max}, lower bound for average speed v_{min}.
$V \leftarrow$ A set of unique Vehicle IDs in D_{all}
for j in $1 : length(V)$ **do**
  $D \leftarrow$ Records with vehicle ID = V_j in D_{all}
  $D_s \leftarrow$ Sort records in D by time
  $i \leftarrow 1$
  $n \leftarrow$ Number of records in D_s
  while $i<n$ **do**
    Calculate time interval Δt_i and distance d between i th and $(i+1)$ th records
    Speed $v \leftarrow d/\Delta t_i$
    if $\Delta t_s<\Delta t_i<TT_{max}$ **and** $v_{min}<v<v_{max}$ **then**
      Trip origin $Ori \leftarrow i$ th record
      Trip destination $Des \leftarrow (i+1)$ th record
    end if
    output An OD pair (Ori, Des)
    $i \leftarrow i+1$
  end while
end for

if data feeds are updated frequently enough, that is, every second. In addition, since the e-scooter vehicle ID is static, this algorithm can infer trip OD pairs. Inference error is possible if there is global positioning system (GPS) error or a low data update frequency. Launch and elimination of vehicles will not result in inference errors. The algorithm only infers linked OD pairs, thus the first record will not be identified as trip destination and the last record will not be identified as trip origin. When a e-scooter is in a rebalancing trip or juicing trip, since the e-scooter is locked, some vendors may mark this e-scooter as "available" and publish information on it to *free_bike_status*. In this case, movement of e-scooters can be observed in the data feeds but the algorithm will not categorize these trips as real trips conducted by travelers.

The analyst may seek to exclude rebalancing and juicing trips by considering additional criteria such as travel distance, trip speed, or time of day (*2*, *9*). This study uses the same criteria as in McKenzie (*2*). Trips that lasted longer than 2 h, trips with average speeds greater than 15 mph, and trips with average speeds lower than 2.2 mph are removed from this study.

Resetting Vehicle ID. When an e-scooter using *Resetting Vehicle ID* starts a trip, information on the e-scooter will disappear from the data feed. After the trip terminates, information on this e-scooter will reappear but with a different ID. The ID is randomly generated so the new ID cannot be matched with the previous one. Therefore, OD pairs cannot be inferred using this kind of data. But trip origins and destinations can be inferred separately. The e-scooter starts using a new ID when a trip terminates. When the next trip starts, this ID then disappears from the data feeds. Therefore, when an e-scooter appears in the system with an ID, this is marked as a trip destination; then at a later time point, when the e-scooter with this same ID disappears in the system, this is marked as a trip origin. In other words, for all records with a unique vehicle ID, the earliest record indicates a trip destination and the latest record indicates the origin of the next trip. Based on this logic, an algorithm to infer trip origins and destinations for e-scooters with *Resetting Vehicle ID* is developed. The algorithm is presented in Algorithm 2.

Like Algorithm 1, the inference accuracy of this algorithm is related to data update frequency and GPS error. In addition, this algorithm may overestimate trip origins and destinations. If a new e-scooter is added to the system, the first record of this e-scooter will also be incorrectly identified as a trip destination. Similarly, when a broken e-scooter is removed from the system, the last record of this e-scooter is incorrectly identified as a trip

Algorithm 2 Origin and Destination Inference for Data with *Resetting Vehicle ID*

input all e-scooter records D_{all}
$V \leftarrow$ A set of unique Vehicle IDs in D_{all}
for j in $1 : length(V)$ **do**
  $D \leftarrow$ Records with vehicle ID $= V_j$ in D_{all}
  $D_s \leftarrow$ Sort records in D by time
  A trip destination $Des \leftarrow$ the first record in D_s
  output Des
  A trip origin $Ori \leftarrow$ the last record in D_s
  output Ori
end for

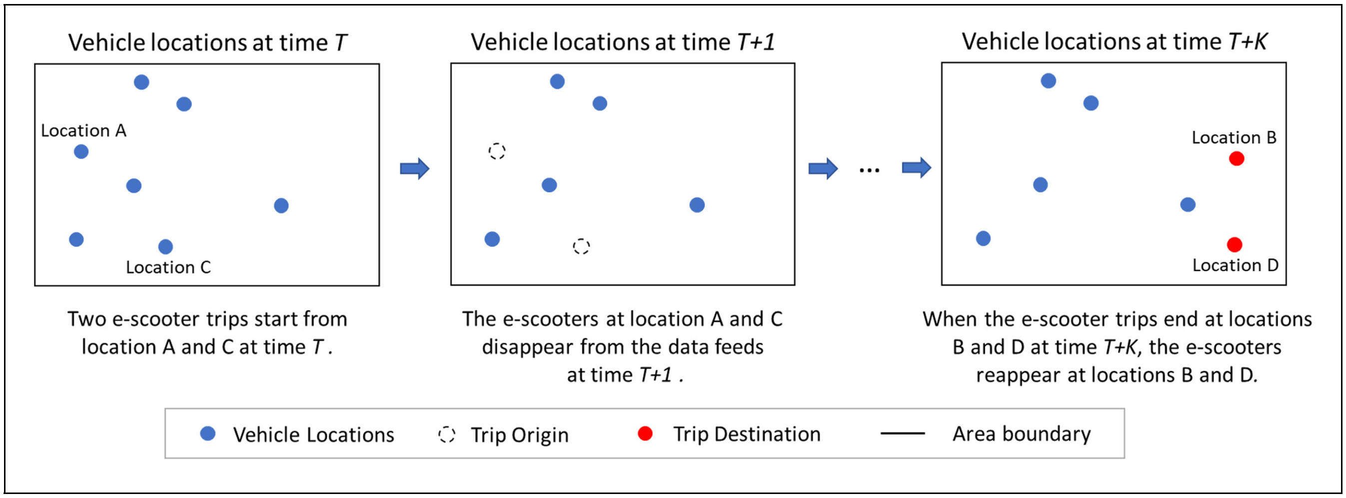

Figure 1. An example to illustrate the process of trip origin and destination identification of Algorithm 3.

origin. The sources of error are summarized in Table 3. The adverse effects of these limitations are evaluated in the section on inference accuracy that follows. Note that when an e-scooter is in a rebalancing trip, the vendors will continue publishing information about the vehicle to the data feeds. We can observe movement of e-scooters in the data feeds, but these e-scooters will not disappear from the data feed. Therefore, the algorithm will not identify trip origins and destinations of these movements.

Dynamic Vehicle ID. Scooters with *Dynamic Vehicle ID* change vehicle ID periodically, according to a certain time interval (e.g., 30 min). This ID generating strategy is a good way to protect user privacy but makes it challenging to infer trip origins and destinations. However, it is still possible to infer trip origins and destinations using the proposed method.

The first step is to create a series of zones and count the total number of scooters in each zone at regular intervals. Here A and B are used to denote specific zones. The zones should be large enough that GPS location error between time periods for the same stationary scooter does not lead to an inference of a false trip.

Consider a series of consecutive time steps $[T, T+1, ..., T+K]$, for $K>1$. Suppose two e-scooter trips start from locations A and C at T and end at locations B and D at $T+K$. When locations of all the e-scooters are compared at T and $T+1$, since the GBFS data only provide information on available e-scooters, we will find the e-scooters disappear at locations A and C. Similarly, when we compare locations of e-scooters at $T+K-1$ and $T+K$, we will find the e-scooters appear in locations B and D. However, since the vehicle ID changes over time, the trip origins and trip destinations cannot be linked in this case. This process is illustrated in Figure 1. Based on this logic, an algorithm can be developed to infer trip origins and destinations. Since the time interval of ID updates is usually larger than the data scraping interval, vehicle ID information is also used to enhance inference accuracy and reduce the computational complexity of the algorithm. For example, if a vehicle ID is found both in data feeds at T and $T+1$, we can conclude that the e-scooter with this ID made no trip during this time interval $[T, T+1]$. It is worth noting that Algorithm 3 still works without ID information. The algorithm to infer trip origins and destinations for data with *Dynamic Vehicle ID* is presented in Algorithm 3.

In the case of Algorithm 3, there are more potential sources of error. Inference error may result from GPS error, low data update frequency, and the launch or elimination of e-scooters. Besides, movements of e-scooters which are not real trips (e.g., rebalancing and juicing trips) may be recognized as trips by the algorithm, which may cause extra inference errors. The sources of error are summarized in Table 3. The inference accuracy of this algorithm is discussed below.

Note that in Algorithm 3,b is the minimum distance threshold to distinguish a trip from GPS error. In case study, $b = 100$ m.

Application and Validation of the Proposed Algorithms

Published GBFS data were collected from six e-scooter services vendors (i.e., Bird, Jump, Lime, Lyft, Skip, and Spin) in Washington D.C. from February 24, 2020 to March 1, 2020. As described in the data collection section, data feeds were scraped using API (https://ddot.dc.gov/page/dockless-api) provided by the vendors. Since

Algorithm 3 Origin and Destination Inference for Data with *Dynamic Vehicle ID*

```
input all e-scooter records D_all, buffer b
for Every time interval [T, T + 1] do
    D_T ← Records at T in D_all
    D_{T+1} ← Records at T + 1 in D_all
    if D_all has Vehicle ID information then
        V ← A set of unique Vehicle IDs in D_T
        for k in 1 : length(V) do
            if V_k in D_{T+1} then
                Remove records with vehicle ID = V_k in D_T and
                D_{T+1}
            end if
        end for
    end if

    repeat
        for Every record i in D_T do
            for Every record j in D_{T+1} do
                Calculate Euclidean distance d_ij between i and j
            end for
        end for
        d_mn ← minimum distance in d_ij
        (m, n) ← index of d_mn
        Remove record m in D_T
        Remove record n in D_{T+1}
    until d_mn > b

    Trip origin Ori ← records in D_T
    output Ori
    Trip destination Des ← records in D_{T+1}
    output Des
end for
```

the data update frequency varies among vendors, different scraping intervals were used. Specifically, the scraping interval for Bird, Jump, Skip, and Spin data is 60 s, and the scraping interval for Lime and Lyft is 300 s. For vehicle ID generating strategy, Jump, Skip and Spin use *Static Vehicle ID*. Bird uses *Resetting Vehicle ID*. Lime and Lyft use *Dynamic Vehicle ID*. After data collection, trip origins and destinations are inferred by the corresponding algorithm. The features of data are presented in Table 4.

Inference Results

Based on the data collected, 65,601 trip origins and 65,600 trip destinations are inferred in total. The temporal distribution of trip origins and destinations is presented in Figure 2. Note that each y -axis is different as the trip volume differs among vendors. Histogram colors are based on the dominant colors of the services' logos.

According to Figure 2, while the six vendors have different trip volumes, the temporal usage patterns are generally similar, which is consistent with conclusions in McKenzie (*11*). All of the six vendors have two significant trip peaks during weekdays. The morning peak hours are 8:00 a.m. to 10:00 a.m., and the afternoon peak hours are 5:00 p.m. to 7:00 p.m. But on weekends, there is only one peak, during the afternoon from 3:00 p.m. to 5:00 p.m. This difference between weekdays and weekends presumably results from commuting trips, which are more common during weekdays. Compared with the study by McKenzie (*11*), where data from December 2018 to March 2019 were used, the morning and afternoon peaks on weekdays are more prominent in this study. This result suggests that the proportion of commuting trips in e-scooter trips might be increasing over time, but more analysis is needed in the future to assess the e-scooter trip purposes. Since the travel time of most commuter e-scooter trips is less than 1 h and trips are aggregated by hour of the day, there are no significant differences between the temporal distributions of trip origins and destinations. But we can observe that the count of trip destinations is larger than the count of trip origins during the second half of the morning peak (i.e., 9:00 a.m. to 10:00 a.m.). This phenomenon may be caused by commuting trips too, most of which start from 8:00 a.m. to 9:00 a.m. and end before 10:00 a.m.

Using the data heat maps were then generated based on trip origin and destination density to explore spatial e-scooter usage patterns during peak hours. The results are presented in Figure 3. The general spatial pattern of peak-hour trips is that the trip density is significantly high in downtown areas, especially along the arterials. The typical hot spots of e-scooter usage include Dupont

Table 4. Summary of the Features of the Data

Vendor	Vehicle ID generating strategy	Data update interval (s)	Inference algorithm	Inference output
Bird	Resetting Vehicle ID	60	Algorithm 2	Unlinked trip origins and destinations (OD)
Jump	Static Vehicle ID	60	Algorithm 1	OD pairs
Lime	Dynamic Vehicle ID	300	Algorithm 3	Unlinked trip origins and destinations
Lyft	Dynamic Vehicle ID	300	Algorithm 3	Unlinked trip origins and destinations
Skip	Static Vehicle ID	60	Algorithm 1	OD pairs
Spin	Static Vehicle ID	60	Algorithm 1	OD pairs

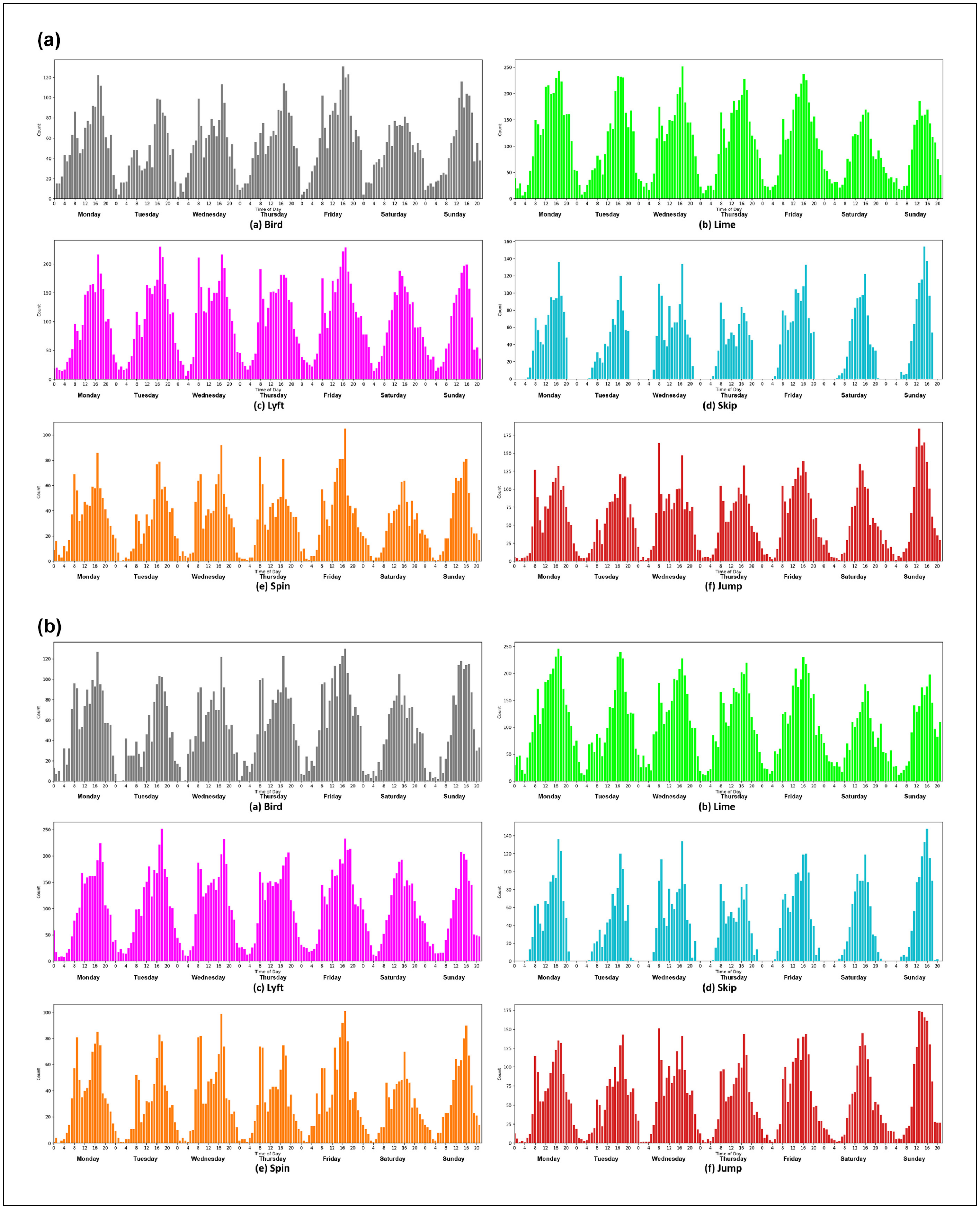

Figure 2. Temporal distributions of trip origins and destinations for different vendors in Washington, D.C.: (*a*) trip origins and (*b*) trip destinations.

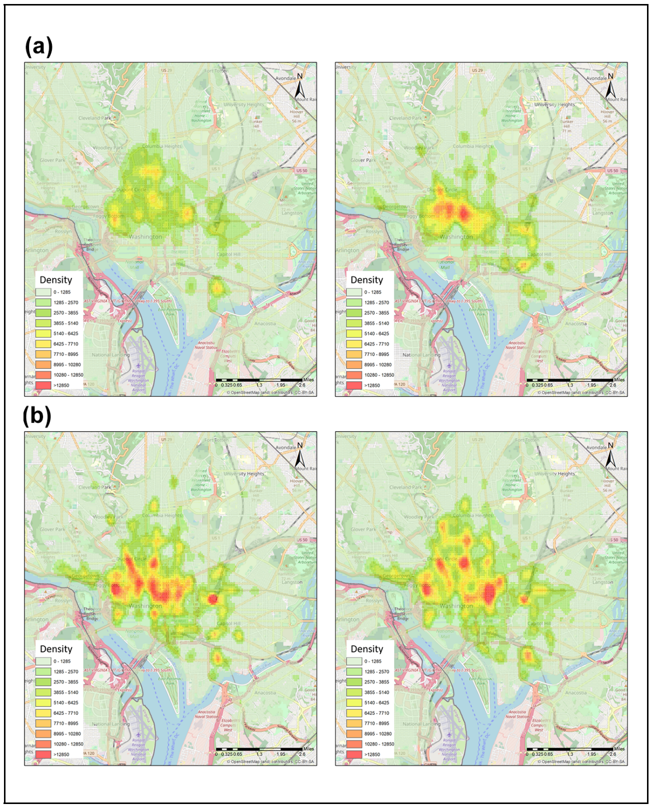

Figure 3. Spatial distributions of trip origins and destinations in Washington, D.C. (density unit: trips per day per square kilometer): (*a*) morning peak (left: trip origin; right: trip destination.) and (*b*) afternoon peak (left: trip origin; right: trip destination.).

Circle, Foggy Bottom, Mount Vernon Square, Union Station and the U.S. Department of Transportation. Interestingly, all of these areas are centered on or close to metro stations. This may indicate that the e-scooter services provide connections between transit stations and origins or destinations of commuters. If so, e-scooter services may help address the “first-mile/last-mile” problem. For morning peaks, the trip destinations are more concentrated than trip origins. This could be because most of the morning commuting trips are from home to work, and the work places are more likely to congregate in downtown areas. For afternoon peaks, there is a significant hot spot for trip destinations in the Penn Quarter, which is a popular entertainment district with bars, restaurants, and shopping. This pattern makes sense because people may use e-scooters to travel to leisure activities after work.

Evaluation of the Proposed Algorithms

GPS error and the error caused by data update frequency are irreducible. With the same data set, the two errors have the same effect on inference accuracy for the three algorithms. As discussed in the Methods section, the inference error of Algorithm 1 only results from GPS error and the error caused by data update frequency. Theoretically, Algorithm 1 is expected to have the best performance among the three algorithms. Since the vendors using Static Vehicle ID (i.e., Jump, Skip, and Spin) update the e-scooter status data every minute, the travel time error of a trip will not exceed 1 min and location errors cannot exceed the distance a scooter can travel in 1 min. Therefore, the inference results of Algorithm 1 are close to the ground truth. Therefore, the inference result of Algorithm 1 is used as *benchmark* to evaluate the performance of Algorithms 2 and 3.

The inference accuracy of proposed algorithms is evaluated by three metrics: R^2, MAE, and SAE. The research area is first divided into a regular square grid with zones of equal size. Then the trip origins and destinations are aggregated into these grids based on location and then the number of origins and destinations in each grid cell is counted. Since the actual trip count of several grids is zero, the authors choose to use absolute metrics (i.e., MAE and SAE) rather than relative metrics to evaluate the inference errors to avoid undefined values (i.e., division by zero). R^2, MAE, and SAE are calculated by:

$$SS_{tot} = \sum_i (y_i - \bar{y})^2 \tag{1}$$

$$SS_{res} = \sum_i (y_i - \hat{y}_i)^2 \tag{2}$$

$$R^2 = 1 - \frac{SS_{res}}{SS_{tot}} \tag{3}$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n} |y_i - \hat{y}_i| \tag{4}$$

$$SAE = \sum_{i=1}^{n} |y_i - \hat{y}_i| \tag{5}$$

where SS_{tot} is the total sum of squares; SS_{res} is the residual sum of squares; y_i is the count of actual origins (destinations) for grid i; $\bar{y}$ is the mean of y_i; and $\hat{y}_i$ is count of inferred origins (destinations) for grid i.

To compare the performance of the three inference algorithms, a data set with *Static Vehicle ID* is first selected. Then Algorithm 1 is used to infer trip origins and destinations as a benchmark. After that, vehicle ID is regenerated using *Resetting Vehicle ID* strategy and *Dynamic Vehicle ID* strategy. Trip origins and destinations are then inferred using the corresponding algorithms (i.e., Algorithm 2 and Algorithm 3) and the results are compared with the benchmark. Among the six vendors, Spin uses *Static Vehicle ID* and updates the data feeds at a high frequency. Therefore, data feeds

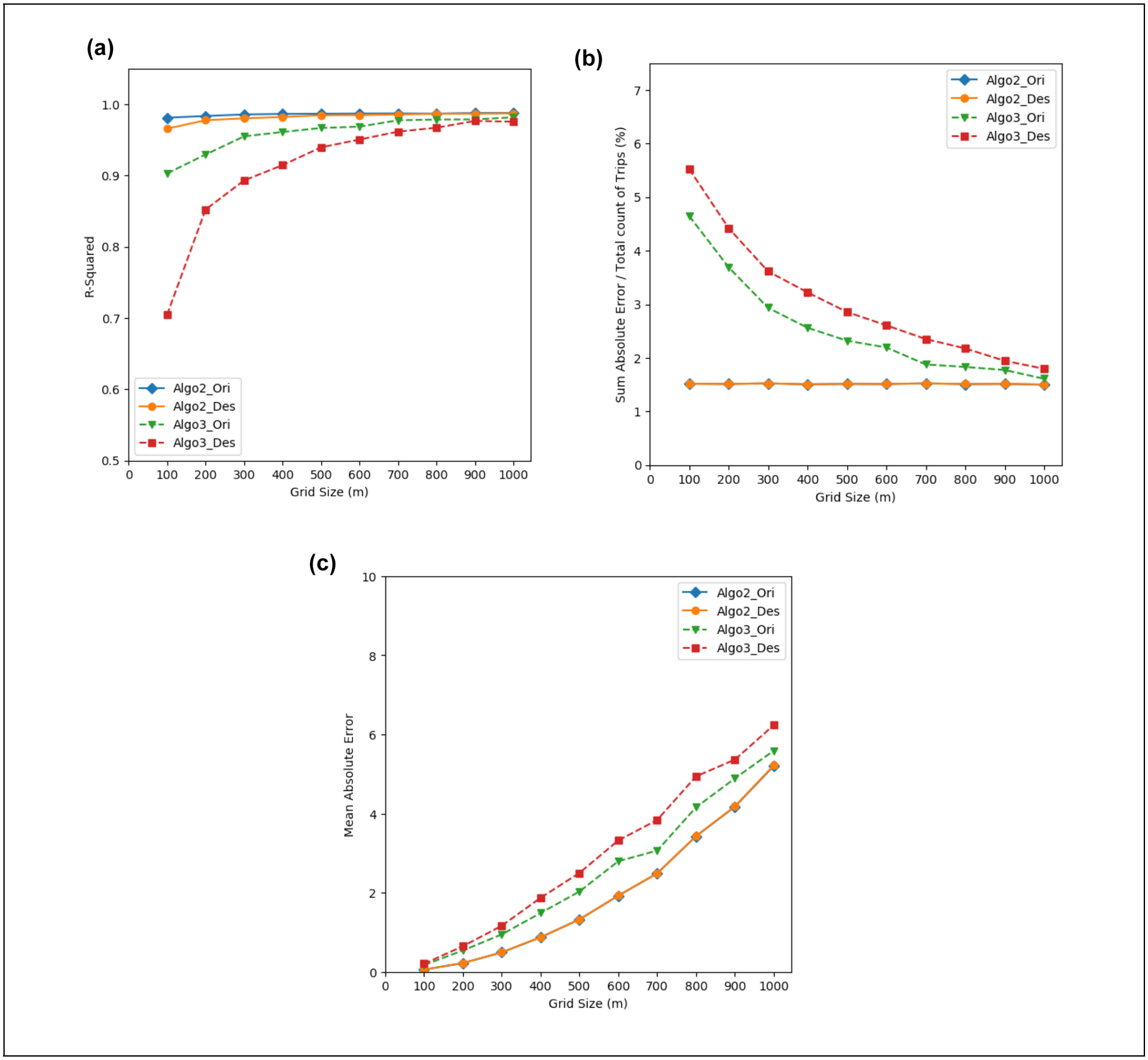

Figure 4. Inference accuracy evaluation of Algorithm 2 and Algorithm 3 for different size of grid (sensitivity analysis): (*a*) R-squared, (*b*) sum absolute error, and (*c*) mean absolute error.

published by Spin were use for algorithm evaluation. To analyze sensitivity to zone size, the fit metrics for a range of different zone sizes, ranging from 100 m square to 1,000 m square zones were calculated. The results are presented in Figure 4. Note that the y -axis of Figure 4*b* is SAE divided by total count of trips.

According to Figure 4, both Algorithm 2 and Algorithm 3 have achieved high inference accuracy. Algorithm 2 has higher inference accuracy than Algorithm 3. When the grid size is larger than 400 m $\times$ 400 m, the R^2 of every algorithm is larger than 0.9, and the MAE of every algorithm is smaller than two trips. However, we can see a significant drop in R^2 when the grid size reduces to 100 m $\times$ 100 m; this is probably because of the GPS inaccuracy. From Figure 4*b*, one can notice that as the grid size increases, the SAE of Algorithm 3 decreases while the SAE of Algorithm 2 remains flat. As the zones become larger, aggregation leads to a decrease in total error in the system. For example, a 400 m $\times$ 400 m area contains 16 100 m $\times$ 100 m areas; inference errors in each of these 16 areas may be positive or negative. The positive and negative errors will be neutralized when we calculate the error of the whole 400 m × 400 m area. However, Algorithm 2 already has an extremely high inference accuracy, so the size change of grid has little influence on the corresponding SAE. As shown in Figure 4*c*, the MAE of every algorithm increases as the grid size increases. This is because when the area of a zone becomes larger, both SAE (the numerator of MAE) and the total number of zones (the denominator of MAE) decreases, but the denominator decreases much faster than the numerator. Therefore, there is an increasing trend of MAE. In Figure 4, *b* and *c*, the MAE is always below 7 and the SAE or total count of trips is smaller than 6%.

The spatial and temporal distributions of inference error with 400 m × 400 m grid size are further examined. The spatial distributions of inference error of Algorithm 2 and Algorithm 3 are presented in Figure 5. Note that the total number of e-scooter trips taken is 8,845. We can see that the inference error (absolute value) is larger in downtown areas, where the trip density is greatest. There

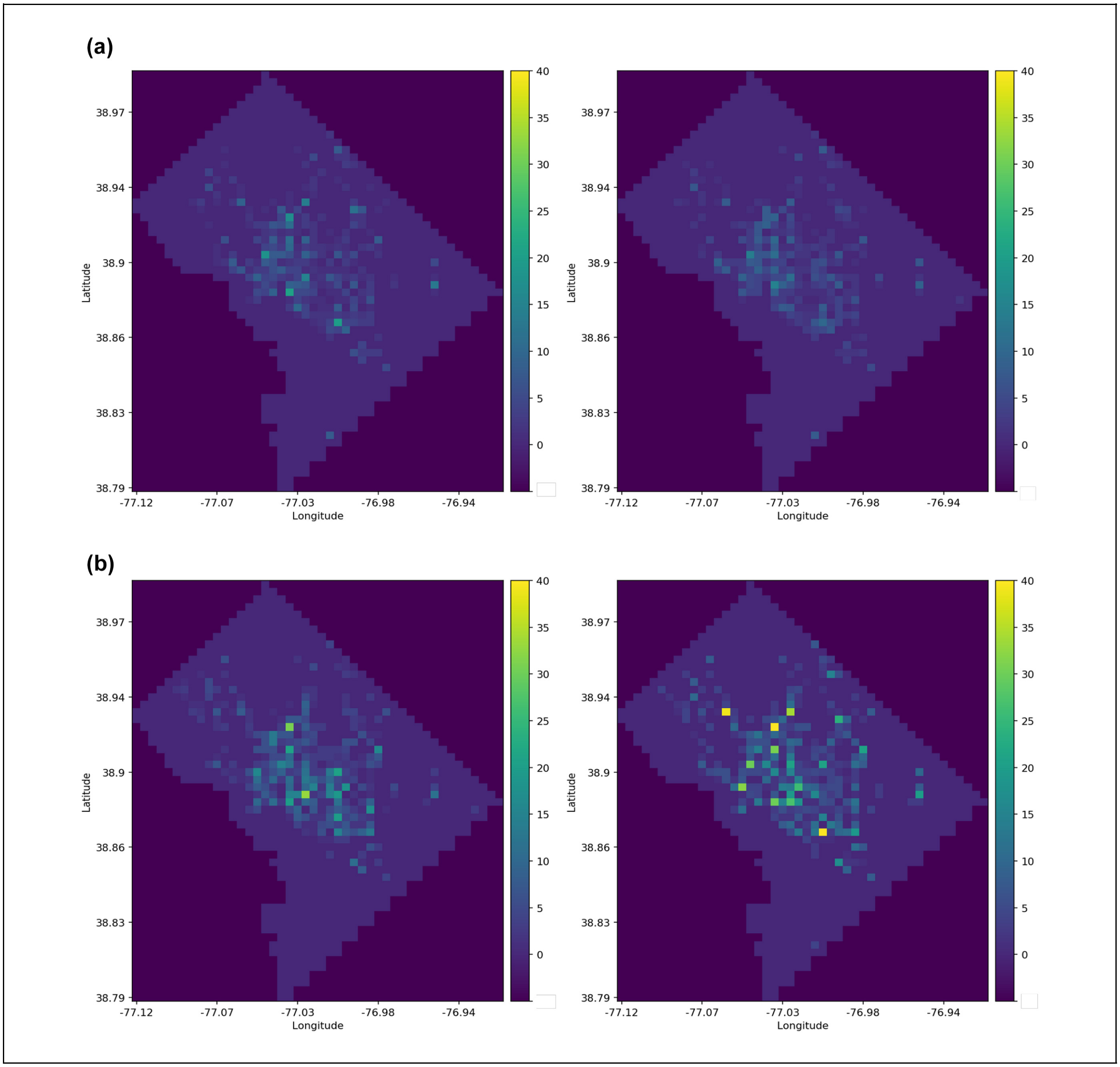

Figure 5. Spatial distributions of inference error (absolute value) of Algorithm 2 and Algorithm 3: (*a*) Algorithm 2 (left: trip origin; right: trip destination.) and (*b*) Algorithm 3 (left: trip origin; right: trip destination.).

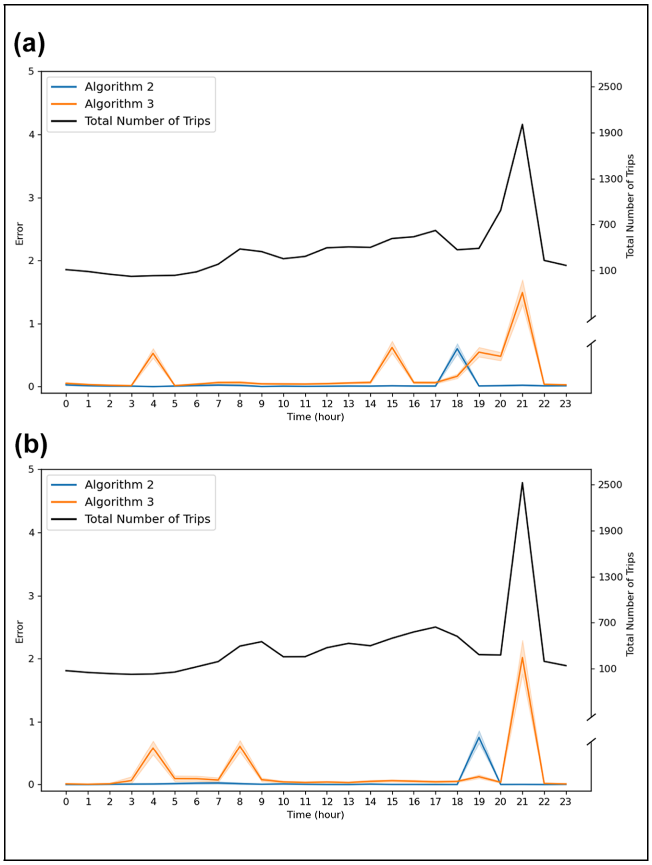

Figure 6. Temporal distributions of inference error (in the number of trips) of Algorithm 2 and Algorithm 3: (*a*) trip origins and (*b*) trip destinations.

are more movements of available e-scooters such as vehicle rebalancing in downtown areas. Most launches and eliminations of e-scooters also happen in these areas. As discussed in the Methods section, each of these activities may lead to inference errors. Therefore, the inference error is larger in the downtown area. The temporal distributions of inference error of Algorithm 2 and Algorithm 3 are presented in Figure 6. The inferred demands are grouped by time (i.e., hour). For each time interval (i.e., hour), the inference error for each cell in the grid is first calculated; then the mean and the 95% confidence interval of these errors in this time interval are calculated to generate Figure 6. According to Figure 6, the inference error is relatively high during early morning (i.e., 4:00–8:00 a.m.) and early night (i.e., 6:00–10:00 p.m.). This is because most of the vehicle rebalancing occurs during these time periods (*2*). The pick-ups and drop-offs of rebalanced vehicles result in extra inference errors for both Algorithm 2 and Algorithm 3. The inference error of Algorithm 2 is larger than that of Algorithm 3 during 6:00–7:00 p.m. After checking the data, some launches and eliminations of vehicles were observed in that time period. Since the launch and elimination of vehicles has larger adverse effect on inference accuracy of Algorithm 2 according to the algorithm mechanism, its error is larger. In addition, during 8:00–10:00 p.m., the inference error of Algorithm 3 is significantly higher than that of Algorithm 2. Relatively many movements of vehicles not in use during this time period were observed. As discussed in previous sections, these movements do not affect the inference accuracy of Algorithm 2 but cause inference errors for Algorithm 3. Therefore, the error of Algorithm 3 is higher in this time period.

Discussion and Conclusion

This study proposes a package of algorithms to infer e-scooter origins and destinations from GBFS data with various vehicle ID types. The accuracy of two proposed algorithms is examined using data from Washington D.C. Two algorithms that work with resetting or dynamic IDs are tested in comparison with an algorithm that works for static IDs, and is therefore more reliable. The inference error is measured with R^2, MAE, and SAE. With scraped GBFS data from one week, trip origins and destinations are inferred using the proposed algorithms. The inference accuracy of the proposed algorithms is evaluated by simulations based on data from Spin (which has *Static Vehicle ID*). The results show that both Algorithm 2 and Algorithm 3 have good inference accuracy with R^2 larger than 0.9 and MAE smaller than 2 when evaluated with a 400 m $\times$ 400 m grid. As the grid size increases, the R^2 of Algorithm 3 increases and the SAE decreases, while the R^2 and SAE of Algorithm 2 remain flat. The absolute errors of Algorithm 2 and Algorithm 3 are larger in the downtown areas, and the inference error is relatively high during early morning and early nighttime.

Using these proposed algorithms, a temporal analysis on scooter trip origins and destinations was conducted. These data show that e-scooter services have a morning peak and an afternoon peak on weekdays, and these peaks are more pronounced compared with the results produced by earlier studies. The findings suggest that e-scooter services may be increasingly used in commuting trips. A spatial analysis with the inferred OD data shows that many areas with high intensity e-scooter usage are close to metro stations. This suggests that travelers are using e-scooter services to connect with transit stations, possibly addressing the "first mile/last mile" problem.

Cities have to balance the need for information about transportation systems with the need for traveler privacy. The public transportation authorities require data and information to manage and guide transportation policy in a fast-changing world. For example, public authorities may seek to improve connectivity between micromobility services and the public transit system. However, a growing body of research indicates that "anonymous" mobility data, which do not specifically reference a person's name, email, or address, could still be used to re-identify specific individuals, their whereabouts, and activities (*8*, *22*). For example, information on the starts and ends of trips or the entire GPS trace of a trip can be combined with other datasets to identify specific individuals. Therefore, GPS and location data are considered as privacy-sensitive information in some U.S. states (*36*). Data standards such as MDS, which track a greater detail of location information, are becoming controversial. The requirements of MDS have been the subject of some lawsuits over data-sharing standards (e.g., Uber and ACLU sued LADOT over MDS [*37*, *38*]). The GBFS data standard, by contrast, has gradually moved away from *Static Vehicle IDs* to *Resetting Vehicle IDs* or *Dynamic Vehicle IDs* (*10*). This helps to mitigate privacy concerns. More cities are expected to adopt GBFS rather than MDS standards in the near future.

There is a trade-off between preserving user privacy and extracting useful information from the GBFS data. For GBFS with *Static Vehicle IDs*, we can directly infer OD pairs. When GBFS is used with *Resetting Vehicle IDs* or *Dynamic Vehicle IDs*, we can only get unlinked trip origins and destinations. With unlinked trip ends, it is impossible to gain information about how micromobility trips are being used as a substitute for other modes, which will partially hinder our understanding of micromobility usage patterns. Cities must carefully navigate the balance between privacy protection and data utility. If cities fail to protect sensitive data, e-scooter companies may use such failures as an excuse to claw back data-sharing agreements. On the other hand, for transportation officials to promote safe and convenient travel options for all, some mobility data (e.g., OD demand and trip trajectory) are essential to guide policymaking and mobility management. Cities must balance the goal of public data sharing with user privacy, and GBFS seems currently to offer the best balance between these competing objectives. For GBFS with a *Resetting Vehicle ID* or *Dynamic Vehicle ID*, trip origin and destination information can be inferred by the proposed algorithms; such data may be sufficient for many research and policy setting purposes.

A limitation of this study is that the inference algorithms and data cleaning process require certain assumptions to be made, such as the maximum e-scooter speed and minimum travel distance. Since the trips not made by users such as rebalancing trips are not labeled in GBFS data, we need to use some criteria such as travel distance to exclude these trips. However, these trips cannot necessarily always be eliminated. Another limitation is that the proposed algorithms for GBFS with dynamic IDs can only infer origins and destinations separately (i.e., we cannot link the origin and destination of a trip). Because of the inference error, the count of origins and destinations may not perfectly match. However, unlinked trip origins and destinations are still useful. We can still examine the spatiotemporal usage patterns using unlinked trip origins and destinations, for example, identifying hot spots of trip origins and destinations during the peak hours and exploring association between trip generation, trip attraction, and points of interest. Additionally, this information can be used to build micromobility demand forecasting models (*39*), which can facilitate effective transport planning and achieve efficient real-time management of micromobility services. Future work may consider developing optimization algorithms to infer OD pairs from the separate trip origins and destinations (*40*, *41*). Moreover, future work may also include investigating the e-scooter trip purpose by intersecting the trip inference results with the land use data. As the number of entities sharing GBFS increases, researchers may examine how the GBFS data and the algorithms developed for them such as the ones we present here have been instrumental for transportation decision-making.

Author Contributions

The authors confirm contribution to the paper as follows: study conception and design: Y. Xu, X. Yan, X. Zhao; data collection: Y. Xu, F. Xing; analysis and interpretation of results: Y. Xu, X. Yan, V. P. Sisiopiku, L. A. Merlin, X. Zhao; draft manuscript preparation: Y. Xu, X. Yan, V. P. Sisiopiku, L. A. Merlin, X. Zhao. All authors reviewed the results and approved the final version of the manuscript.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This research was supported by the U.S. Department of Transportation through the Southeastern Transportation Research, Innovation, Development and Education (STRIDE) Region 4 University Transportation Center (Grant No. 69A3551747104).

ORCID iDs

Yiming Xu https://orcid.org/0000-0002-2983-1751
Xiang Yan https://orcid.org/0000-0002-8619-0065
Virginia P. Sisiopiku https://orcid.org/0000-0003-4262-8990
Louis A. Merlin https://orcid.org/0000-0002-9267-5712
Xilei Zhao https://orcid.org/0000-0002-7903-4806

Supplemental Material

Supplemental material for this article is available online.

References

1. NACTO. *Shared Micromobility in the U.S.: 2018*. National Association of City Transportation Officials, New York, NY, 2019.
2. McKenzie, G. Spatiotemporal Comparative Analysis of Scooter-Share and Bike-Share Usage Patterns in Washington, DC. *Journal of Transport Geography*, Vol. 78, 2019, pp. 19–28.
3. Laa, B., and U. Leth. Survey of E-Scooter Users in Vienna: Who They Are and How They Ride. *Journal of Transport Geography*, Vol. 89, 2020, p. 102874.
4. Bieliński, T., and A. Ważna. Electric Scooter Sharing and Bike Sharing User Behaviour and Characteristics. *Sustainability*, Vol. 12, No. 22, 2020, p. 9640.
5. Lee, H., K. Baek, J.-H. Chung, and J. Kim. Factors Affecting Heterogeneity in Willingness to Use E-Scooter Sharing Services. *Transportation Research Part D: Transport and Environment*, Vol. 92, 2021, p. 102751.
6. Latinopoulos, C., A. Patrier, and A. Sivakumar. Planning for E-Scooter Use in Metropolitan Cities: A Case Study for Paris. *Transportation Research Part D: Transport and Environment*, Vol. 100, 2021, p. 103037.
7. North American Bikeshare Association. *General Bikeshare Feed Specification*. NABSA, 2020. https://github.com/NABSA/gbfs. Accessed July 6, 2020.
8. Clewlow, R. Finding the Right Balance Between Mobility Data-Sharing in Cities and Personal Privacy. *Medium*, May 9, 2019. https://medium.com/populus-ai/finding-the-right-balance-between-mobility-data-sharing-in-cities-and-personal-privacy-78d941d07908. Accessed July 30, 2020.
9. Zou, Z., H. Younes, S. Erdoğan, and J. Wu. Exploratory Analysis of Real-Time E-Scooter Trip Data in Washington, DC. *Transportation Research Record: Journal of the Transportation Research Board*, 2020. 2674: 285-299.
10. MobilityData. What's New in GBFS v2.0. *Medium*, 2020. https://mobilitydata.medium.com/whats-new-in-gbfs-v2-0-63eb46e6bdc4#:~:text=provider's%20native%20app.-,GBFS%20v2.,that%20particular%20vehicle%20or%20-station. Accessed November 16, 2020.
11. McKenzie, G. Urban Mobility in the Sharing Economy: A Spatiotemporal Comparison of Shared Mobility Services. *Computers, Environment and Urban Systems*, Vol. 79, 2020, p. 101418.
12. Almannaa, M. H., F. A. Alsahhaf, H. I. Ashqar, M. Elhenawy, M. Masoud, and A. Rakotonirainy. Perception Analysis of E-Scooter Riders and Non-Riders in Riyadh, Saudi Arabia: Survey Outputs. *Sustainability*, Vol. 13, No. 2, 2021, p. 863.
13. Nikiforiadis, A., E. Paschalidis, N. Stamatiadis, A. Raptopoulou, A. Kostareli, and S. Basbas. Analysis of Attitudes and Engagement of Shared E-Scooter Users. *Transportation Research Part D: Transport and Environment*, Vol. 94, 2021, p. 102790.
14. Raptopoulou, A., S. Basbas, N. Stamatiadis, and A. Nikiforiadis. A First Look at E-Scooter Users. In *Proc., Advances in Mobility-as-a-Service Systems: Conference on Sustainable Urban Mobility* (E. G. Nathanail, G. Adamos, and I. Karakikes, eds.), Skiathos Island, Greece, June 17–19, 2020, Springer, Cham, pp. 882–891.
15. Krier, C., J. Chrétien, M. Lagadic, and N. Louvet. How Do Shared Dockless E-Scooter Services Affect Mobility Practices in Paris? A Survey-Based Estimation of Modal Shift. *Transportation Research Record: Journal of the Transportation Research Board*, 2021. 2675: 291-304.
16. Guo, Y., and Y. Zhang. Understanding Factors Influencing Shared E-Scooter Usage and its Impact on Auto Mode Substitution. *Transportation Research Part D: Transport and Environment*, Vol. 99, 2021, p. 102991.
17. Bai, S., and J. Jiao. Dockless E-Scooter Usage Patterns and Urban Built Environments: A Comparison Study of Austin, TX, and Minneapolis, MN. *Travel Behaviour and Society*, Vol. 20, 2020, pp. 264–272.
18. Caspi, O., M. J. Smart, and R. B. Noland. Spatial Associations of Dockless Shared E-Scooter Usage. *Transportation Research Part D: Transport and Environment*, Vol. 86, 2020, p. 102396.
19. City of Austin. *Shared Micromobility Vehicle Trips*. City of Austin Transportation Department, TX, 2020. https://data.austintexas.gov/Transportation-and-Mobility/Shared-Micromobility-Vehicle-Trips/7d8e-dm7r. Accessed July 6, 2020.
20. City of Minneapolis. *Motorized Foot Scooter Trips 2018*. Open Data, MN, 2020. https://opendata.minneapolismn.gov/datasets/cityoflakes::motorized-foot-scooter-trips-2018/about. Accessed July 6, 2020.
21. Liu, M., S. Seeder, and H. Li. Analysis of E-Scooter Trips and Their Temporal Usage Patterns. *Institute of Transportation Engineers, ITE Journal*, Vol. 89, No. 6, 2019, pp. 44–49.
22. Baltra, G., B. Imana, W. Jiang, and A. Korolova. On the Data Fight Between Cities and Mobility Providers. *arXiv Preprint arXiv:2004.09072*, 2020.
23. Zhu, R., X. Zhang, D. Kondor, P. Santi, and C. Ratti. Understanding Spatio-Temporal Heterogeneity of Bike-Sharing and Scooter-Sharing Mobility. *Computers, Environment and Urban Systems*, Vol. 81, 2020, p. 101483.
24. Martin, E. W., and S. A. Shaheen. Evaluating Public Transit Modal Shift Dynamics in Response to Bikesharing: A Tale of Two US Cities. *Journal of Transport Geography*, Vol. 41, 2014, pp. 315–324.
25. Ma, X., Y. Yuan, N. Van Oort, and S. Hoogendoorn. Bike-Sharing Systems' Impact on Modal Shift: A Case Study in Delft, the Netherlands. *Journal of Cleaner Production*, Vol. 259, 2020, p. 120846.
26. Jia, Y., and H. Fu. Association Between Innovative Dockless Bicycle Sharing Programs and Adopting Cycling in Commuting and Non-Commuting Trips. *Transportation Research Part A: Policy and Practice*, Vol. 121, 2019, pp. 12–21.
27. Guo, Y., J. Zhou, Y. Wu, and Z. Li. Identifying the Factors Affecting Bike-Sharing Usage and Degree of Satisfaction in Ningbo, China. *PloS One*, Vol. 12, No. 9, 2017, p. e0185100.
28. Bieliński, T., A. Kwapisz, and A. Ważna. Bike-Sharing Systems in Poland. *Sustainability*, Vol. 11, No. 9, 2019, p. 2458.

29. Wang, J., and G. Lindsey. Neighborhood Socio-Demographic Characteristics and Bike Share Member Patterns of Use. *Journal of Transport Geography*, Vol. 79, 2019, p. 102475.
30. Xin, F., Y. Chen, X. Wang, and X. Chen. Cyclist Satisfaction Evaluation Model for Free-Floating Bike-Sharing System: A Case Study of Shanghai. *Transportation Research Record: Journal of the Transportation Research Board*, 2018. 2672: 21–32.
31. Qian, X., M. Jaller, and D. Niemeier. Enhancing Equitable Service Level: Which Can Address Better, Dockless or Dock-Based Bikeshare Systems? *Journal of Transport Geography*, Vol. 86, 2020, p. 102784.
32. Hirsch, J. A., J. Stratton-Rayner, M. Winters, J. Stehlin, K. Hosford, and S. J. Mooney. Roadmap for Free-Floating Bikeshare Research and Practice in North America. *Transport Reviews*, Vol. 39, No. 6, 2019, pp. 706–732.
33. Brown, A., and vs. Docked Dockless Equity: Comparing Three Micromobility Service Geographies. *Journal of Transport Geography*, Vol. 96, 2021, p. 103185.
34. Hu, S., C. Xiong, Z. Liu, and L. Zhang. Examining Spatiotemporal Changing Patterns of Bike-Sharing Usage During COVID-19 Pandemic. *Journal of Transport Geography*, Vol. 91, 2021, p. 102997.
35. Bradshaw, R., and R. Kitchin. Charting the Design and Implementation of the Smart City: The Case of Citizen-Centric Bikeshare in Hamilton, Ontario. *Urban Geography*, 2021, pp. 1–23. https://doi.org/10.1080/02723638.2021.1878439.
36. 1.5. Invasion of Privacy. In *California Law PART 1. Of Crime and Punishments: TITLE 15. Miscellaneous Crimes*, California Law, 1967.
37. Teale, C. Uber Sues LADOT Over Data-Sharing Requirements. *Smart Cities Dive*, 2020. https://www.smartcitiesdive.com/news/uber-jump-sues-los-angeles-mobility-data-sharing-requirement/574893/. Accessed July 30, 2020.
38. ACLU. Privacy Lawsuit Your Scooter GPS Data is Being Tracked. *ACLU*, 2020. https://www.aclu.org/press-releases/privacy-lawsuit-your-scooter-gps-data-being-tracked#:~:text=Today%2C%20the%20American%20Civil%20Liberties,Fourth%20Amendment%20to%20the%20U.S. Accessed July 30, 2020.
39. Merlin, L. A., X. Yan, Y. Xu, and X. Zhao. A Segment-Level Model of Shared, Electric Scooter Origins and Destinations. *Transportation Research Part D: Transport and Environment*, Vol. 92, 2021, p. 102709.
40. Ji, Y., Q. You, S. Jiang, and H. M. Zhang. Statistical Inference on Transit Route-Level Origin–Destination Flows Using Automatic Passenger Counter Data. *Journal of Advanced Transportation*, Vol. 49, No. 6, 2015, pp. 724–737.
41. Ji, Y., R. G. Mishalani, and M. R. McCord. Transit Passenger Origin–Destination Flow Estimation: Efficiently Combining Onboard Survey and Large Automatic Passenger Count Datasets. *Transportation Research Part C: Emerging Technologies*, Vol. 58, 2015, pp. 178–192.