

TRANSPORT FINDINGS

An Algorithm for Estimating Origins and Destinations of Shared E-Scooter Trips from Public Data-Feeds

Sid Rayaprolu¹, Mohan M Venigalla¹¹ Sid and Reva Dewberry Department of Civil, Environmental, and Infrastructure Engineering, George Mason University

Keywords: shared micromobility, E-scooter, General Bikeshare Feed Specification (GBFS), First- and last-mile connectivity, Open-source data analysis, Spatiotemporal Analysis

<https://doi.org/10.32866/001c.125812>

Findings

We present an algorithm that estimates shared e-scooter trip origins and destinations using public General Bikeshare Feed Specification (GBFS) data. The model addresses challenges like GPS lag and randomized vehicle IDs. The algorithm is field-tested and validated using data for the city of Fairfax, VA. When applied to Washington, D.C., a larger city, the algorithm demonstrates scalability for larger, complex urban settings. Our methodology equips urban planners and activity center managers with a robust tool for analyzing micromobility patterns, enhancing first- and last-mile connectivity, and optimizing transportation systems—without relying on proprietary data from operators.

1. Questions

Micromobility, including shared vehicles, has evolved to be a part of the urban transportation options. Due to their relative recency and the last-mile nature, micromobility trips are not yet part of the urban travel demand modeling (TDM) processes. Understanding the impact of these trips on transportation systems is essential for urban planning (Dibaj et al. 2021; Oeschger, Carroll, and Caulfield 2020). Recent updates to the General Bikeshare Feed Specification (GBFS) have enhanced data accuracy and privacy, but these changes also limit the utility of open-source API data. Given these restrictions, how best could the GBFS data be utilized by analysts to identify micromobility trip characteristics such as origin and destination patterns between various activity centers? Could the need for data-sharing agreements be circumvented to enhance the utility of GBFS data beyond TDM needs? For example, could urban activity centers assess the utility and feasibility of dockless systems as first- and last-mile solutions using GBFS data to better serve their patrons?

To answer these questions, we propose a scalable and transferable algorithm that reliably and empirically estimates trips across different geographies, through GBFS data analysis and field-level experiments. While the estimates may not capture every trip with complete accuracy, they provide close approximations that deliver valuable insights for practitioners.

The study addresses the following research questions:

1. What are the idiosyncrasies associated with the GBFS data in deriving shared e-scooter spatiotemporal usage patterns?
2. How to reliably derive e-scooter trip origins and destinations using the GBFS data?

Past studies emphasizing vehicle appearance and disappearance or unique ID tracking, have become less effective due to dynamic randomization of vehicle IDs in recent GBFS versions (McKenzie 2019; Merlin et al. 2021; Yiming Xu et al. 2022; Zou et al. 2020). This study builds on past literature by addressing some of their limitations.

2. Methods

The GBFS data includes ‘**bike_id**’ (vehicle ID), coordinates (‘lat’, ‘lon’), and binary fields ‘**is_reserved**’ and ‘**is_disabled**’ (1 for TRUE, 0 for FALSE). Vehicles that are not disabled are deemed “Active,” while active vehicles without recent movement are deemed “Idle.” The **time-to-live (ttl)** parameter, set to 60 seconds, defines the refresh rate. We developed four field-tested scenarios ([Table 1](#)) to evaluate the accuracy of past and current methods vis-à-vis our method in detecting trip origins and destinations from the GBFS data. These scenarios reflect typical conditions in e-scooter deployment, such as GPS lag and variable update rates. Although location data often remains static for up to 10 minutes, limiting detection of short movements, we collected data at the default **ttl** interval.

We use a hexagonal tessellation used in past studies (Arias-Molinares et al. 2023; Jiao and Bai 2020; McKenzie 2019; Y. Xu et al. 2023) with 300-foot (91.44m) - apothem quadrats, which are referred to as e-scooter analysis zones (EAZs). Redacted trip data from a single operator was used to validate against the GBFS-derived trip data by our method.

[Table 1](#) highlights and [Figure 1](#) illustrates the differences between GBFS data and real-time app observations across field-tested scenarios. The app consistently captured movements in real time. GBFS updates lagged and were affected by randomized vehicle IDs, which limited detection accuracy. Prior methods relying on single attributes, like vehicle ID or net equilibrium, struggled with complex cases (McKenzie 2019; Merlin et al. 2021; Yiming Xu et al. 2022; Zou et al. 2020). Multi-zone movements and rapid exchanges caused frequent detection errors. Our method uses multiple GBFS attributes rather than just origin-destination data. This approach captures trip patterns more accurately across scenarios and scales well with evolving GBFS data.

Our method addresses gaps (see [Table 1](#)) identified in prior research (Yiming Xu et al. 2022). It consistently captures distinct origins and destinations across scenarios. We represent the GBFS data over T time intervals (i) and N zones (j) as a matrix of $T \times N$ cells, where each cell contains data on the active, idle, and reserved vehicles.

Table 1. Description of field-tested scenarios assessing e-scooter trip detection challenges and differences between app-based observations and GBFS-derived data.

Scenario	Test Description	App Observations vs. GBFS	Unique ID Method ^a	Equilibrium Method ^b	Our Method ^c
1: Movement without scanning	We moved a vehicle 15-30 ft without scanning its QR code to test update accuracy.	The app adjusted the vehicle location in real time; GBFS did not register the movement even after several update periods.	No detection, as GBFS didn't register movement.	No detection due to lack of movement update in GBFS.	No detection, as GBFS didn't register movement.
2a: Single short trip	We conducted a quick trip from A to B within a few minutes.	The app updated the location instantly, while GBFS showed a 10-minute delay with no change in "is_reserved."	Detected as 1 origin (O) and 1 destination (D), with limited clarity due to delay.	Detected as 1 origin (O) and 1 destination (D), delayed by GBFS lag.	Detected as 1 origin (O) and 1 destination (D).
2b: Multiple trips, stable count	We took two trips between A and B in the same period, keeping the net count stable.	The app tracked both trips in real-time, but GBFS only updated in the next period, missing both trips in the active period with "is_reserved" unchanged.	No detection; stable count masked trip exchanges.	No detection due to stable net count.	Detected as 2 (O) and 2 (D)
3a: Single trip across periods	We conducted a trip from A to B, starting in one period and ending in the next.	The app tracked the trip seamlessly; GBFS showed a split update with disappearance from A in the first period and appearance at B in the next, "is_reserved" TRUE (1) at A.	Partial detection; fragmented due to period split.	Detected as 1 (O) and 1 (D)	Detected as 1 (O) and 1 (D)
3b: Dual trip exchange across periods	We conducted two trips exchanged between A and B, with stable counts across periods.	The app tracked both trips in real-time, while GBFS showed stable equilibrium with no net change, "is_reserved" TRUE (1) for stationary vehicles.	Detected 4 (O) and 2 (D) due to period split.	No trips detected; stable equilibrium.	Detected 2 (O) and 2 (D) without duplication.
4: Multi-zone trips in one period	We conducted multiple trips across A, B, and C in a single period.	The app updated instantly, but GBFS showed over- or underestimation due to fluctuating zone counts, with "is_reserved" ineffective in rapid exchanges.	Frequent mismatches in multi-zone tracking.	Miscounted events due to unstable zone equilibrium.	Can detect all movements across zones.

^a GBFS feeds no longer contain unique vehicle IDs. Therefore, using this method is no longer viable.

^b Method used in past research (Yiming Xu et al. 2022)

^c Our method detected more trips accurately (4 out of 6 scenarios) than prior methods, which indicates a clear improvement over prior methods.

Trip Origins (O) and destinations (D) are calculated as follows:

$$O = A - S_{i+1,j} - R_{i+1,j} \quad (1)$$

$$D_{i,j} = A_{i,j} - S_{i,j} - R_{i,j} \quad (2)$$

where:

- $A_{i,j}$ = Active vehicles in zone j at time i
- $S_{i+1,j}$ = Idle vehicles remained in zone j at $i + 1$
- $R_{i+1,j}$ = Reserved vehicles within zone j at $i + 1$

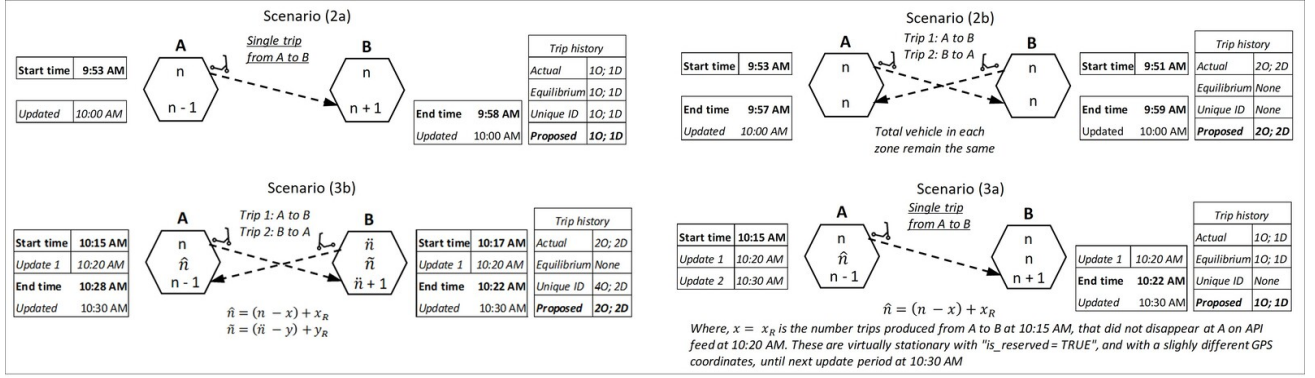


Figure 1. Clockwise from top left: (a) Scenario 2a - Single trip from A to B, (b) Scenario 2b - Two trips with same total vehicle count, (c) Scenario 3a - Single trip across update periods, (d) Scenario 3b - Multiple trips across update periods. The figure compares actual observations, equilibrium methods, unique ID-based methods, and the proposed method in detecting trip origin and destination information.

Equation (1) calculates trip origins by observing a decrease in active vehicles between times i and $i + 1$ after adjusting for vehicles that remained idle or reserved. This captures vehicle departures from the zone. Equation (2) estimates trip destinations by calculating arrivals based on changes in active vehicles, excluding idle and reserved vehicles to focus only on actual movements.

The ‘idle’ vehicle matrix tracks the number of e-scooters that remain stationary within each analysis zone during the observation period. Due to GPS inaccuracies, multiple devices may appear in the data feed with the same GPS coordinates. To differentiate them, each stationary device is assigned a unique identifier (QID) based on its GPS coordinates and timestamp.

In our method, a vehicle is classified as ‘idle’ if, over two consecutive observation periods:

1. Its reservation status $R_{i,j}$ and $R_{i+1,j}$ is FALSE, indicating it is not reserved.
2. Its QID remains consistent across these periods, an indication that it has not moved.

For each zone j , the number of idle vehicles $S_{i+1,j}$ at time $i + 1$ is the count (cardinality) of all vehicles in Q that meet these conditions:

$$S_{i+1,j} = |Q| \forall \{O_{i+1,j}\} \quad (3)$$

$$S_{i,j} = |Q| \forall \{D_{i,j}\} \quad (4)$$

where Q represents the set of vehicles that satisfy these idleness conditions in zone j .

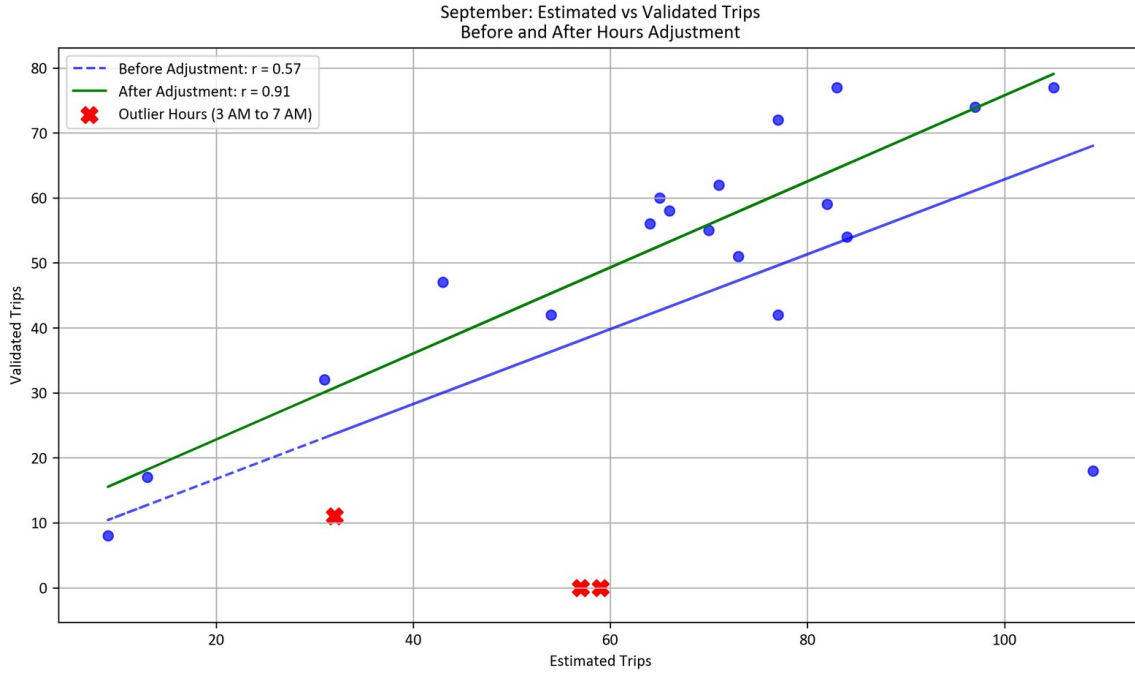


Figure 2

3. Findings

We analyzed data from September and October of 2021 to assess the accuracy of trip estimates, focusing on validating the algorithm using actual trip data from the City of Fairfax, Virginia, a small suburban jurisdiction of Washington DC. [Figure 2](#) illustrates Pearson correlation before and after residual analysis. The initial analysis yielded a correlation coefficient of 0.57 between the estimated and observed trips (blue line), indicating a moderate relationship between the two variables. To further investigate, we conducted a residual analysis and identified that hours 3 through 7 had the highest outliers, which are more than 1.5 standard deviations away from the mean. Without these outliers, the correlation improved significantly to 0.91 (green line).

[Figure 3](#) illustrates the estimated hourly trips by the algorithm vis-à-vis actual trips for a full month of data. The initial trip estimates (orange line) did not account for the time lag in GPS updates, where trips occurring in the final minutes of an hour were reported in the next hour, leading to rounding errors. Having had the advantage of obtaining actual trip O-D data, we fine-tuned the model through scaling, adjusting the estimates to account for the GPS time lag (green line), which greatly improved model performance. This refined model achieved a Mean Absolute Error (MAE) of 6.29, a Root Mean Squared Error (RMSE) of 7.85, and an R^2 (Coefficient of Determination) of 0.822, (with improvements from MAE=13.28, RMSE=16.91, and $R^2=0.822$, respectively) indicating a much better fit between the estimated and observed

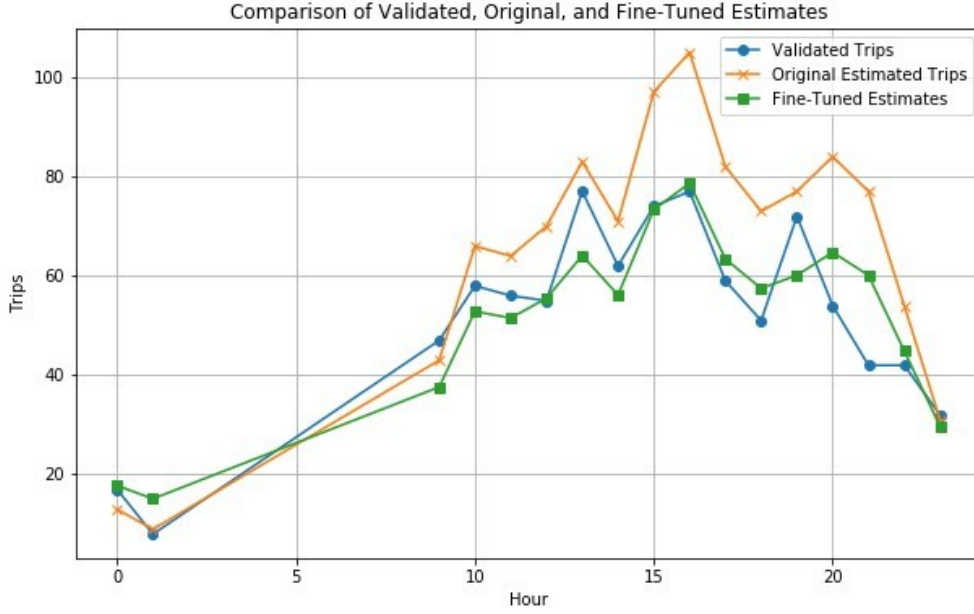


Figure 3. Comparison of validated trips, original estimates, and scaled (fine-tuned) estimates across different hours of the day. The original estimates (orange) tend to overpredict trips, whereas the scaled estimates (green) closely align with the validated trips (blue) after applying the scaling adjustment, demonstrating the improved accuracy of the model.

trips. The significance of refinement by scaling is that, when validation data is available, scaling is preferable (green). However, in the most practical scenario (where validation data is absent), the unscaled model (orange) may be used.

It should be noted that the primary emphasis of our validation is in the temporal dimension. We also attempted to validate the model spatially using Moran's I and Geary's C. Discrepancies between estimated and actual trip origins and destinations were randomly distributed (Moran's I = -0.011, p-value = 0.191) with local variation (Geary's C = 1.046, p-value = 0.007). The results of spatial validation, at best, are inconclusive.

To test the scalability of our algorithm, we applied the trip O-D estimation methodology developed using the data for the City of Fairfax to GBFS data in Washington, D.C., a much larger geography with higher trip density, diverse activity hubs, and broader geographic spread. The efficacy of this algorithm is demonstrated in [Figure 4](#), which highlights the diurnal trip activity patterns for a single operator in Washington, D.C. The figure reveals that weekday trip activity is concentrated in the core business regions of downtown D.C., particularly during the midday and afternoon hours. In contrast, weekends see higher trip activity around the National Mall (not shown in the figure), reflecting shifts in leisure-related movement patterns. The model was able to accurately capture these dynamics, estimating a total of approximately 85,000 trips for the month of May 2021 for the single operator in D.C. By scaling the model to fit the unique spatial and temporal patterns of Washington, D.C.,

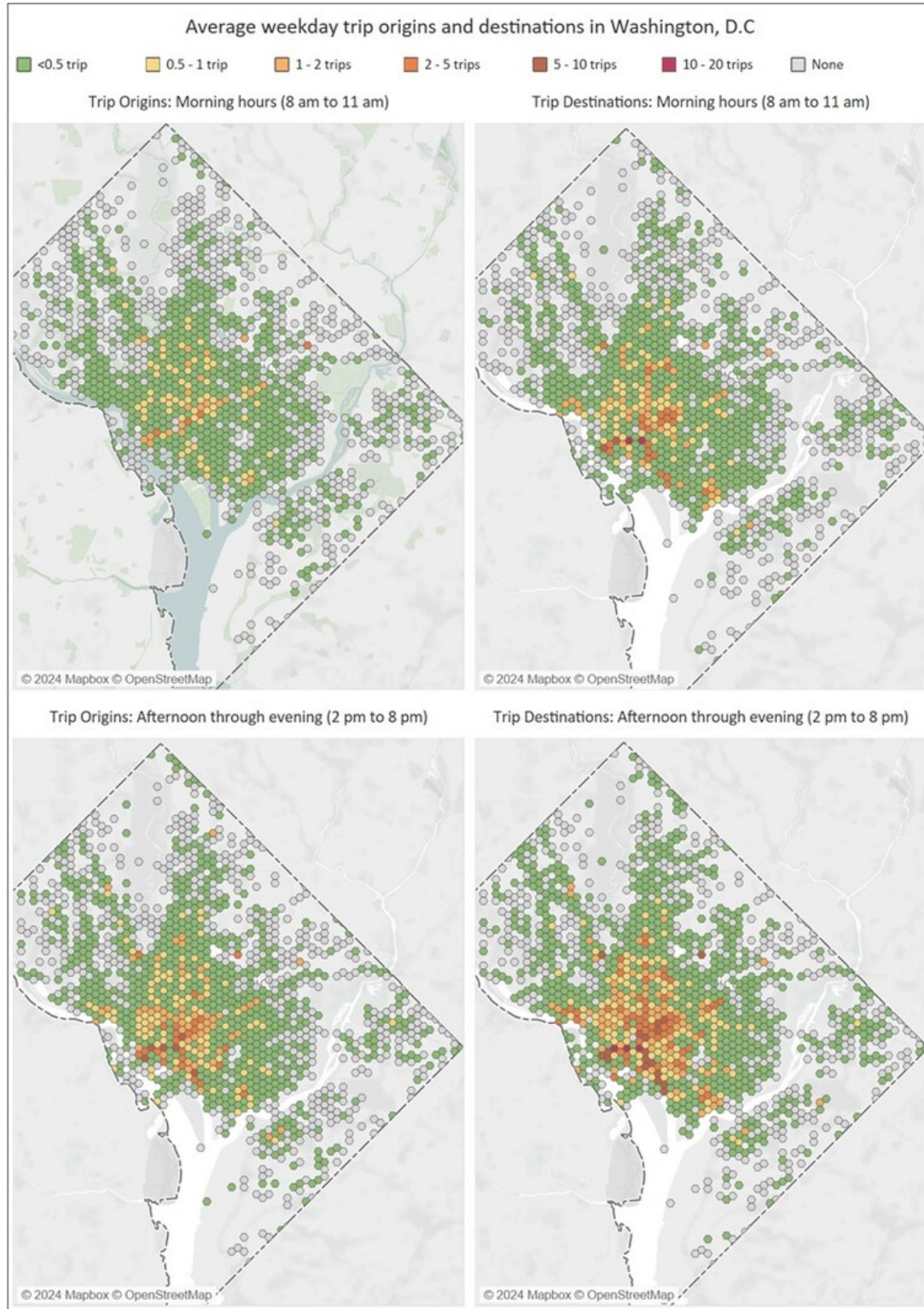


Figure 4. Average weekday trip origins and destinations in Washington, D.C., during different time periods. The top panels display trip origins (left) and destinations (right) for the morning hours (8 am to 11 am), while the bottom panels show trip origins and destinations for the afternoon through evening period (2 pm to 8 pm). Central areas exhibit higher trip activity, especially during peak commute hours, with visible variations between origins and destinations in both morning and afternoon periods.

the approach effectively illustrated how trip production fluctuates between weekdays and weekends, with peaks during commute hours on weekdays and tourist-related activity during weekends.

The algorithm and methodology provide a reliable, portable, and scalable approach for understanding shared micromobility operations across multiple U.S. cities, without requiring data sharing agreements. The validation of our model across different cities highlights its adaptability and ability to accurately capture local trip patterns. Our methodology remains applicable to newer versions of GBFS specifications that have been released after 2021, the validation year. By utilizing real-time GBFS data from open-source APIs, stakeholders can evaluate system performance throughout the year, offering critical insights for various stakeholders, particularly regarding the feasibility of first- and last-mile solutions.

Further engagement with operators and fleet managers could enhance the model's accuracy, especially during non-peak hours and for rebalancing trips. It's also important to note that regional boundaries and rebalancing cycles may introduce over- or underestimation of trips due to GPS update lags, particularly near jurisdictional geofencing boundaries. Future research should focus on addressing these boundary effects to further refine the model's accuracy in such conditions.

Submitted: October 04, 2024 AEDT, Accepted: November 12, 2024 AEDT



This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CCBY-SA-4.0). View this license's legal deed at <https://creativecommons.org/licenses/by-sa/4.0> and legal code at <https://creativecommons.org/licenses/by-sa/4.0/legalcode> for more information.

REFERENCES

- Arias-Molinares, Daniela, Juan Carlos García-Palomares, Gustavo Romanillos, and Javier Gutiérrez. 2023. "Uncovering Spatiotemporal Micromobility Patterns through the Lens of Space-Time Cubes and GIS Tools." *Journal of Geographical Systems* 25 (3): 403–27. <https://doi.org/10.1007/s10109-023-00418-9>.
- Dibaj, Samira, Aryan Hosseinzadeh, Miloš N. Mladenović, and Robert Kluger. 2021. "Where Have Shared E-Scooters Taken Us So Far? A Review of Mobility Patterns, Usage Frequency, and Personas." *Sustainability* 13 (21). <https://doi.org/10.3390/su132111792>.
- Jiao, Junfeng, and Shunhua Bai. 2020. "Understanding the Shared E-Scooter Travels in Austin, TX." *ISPRS International Journal of Geo-Information* 9 (2). <https://doi.org/10.3390/ijgi9020135>.
- McKenzie, Grant. 2019. "Spatiotemporal Comparative Analysis of Scooter-Share and Bike-Share Usage Patterns in Washington, D.C." *Journal of Transport Geography* 78 (June):19–28. <https://doi.org/10.1016/j.jtrangeo.2019.05.007>.
- Merlin, Louis A., Xiang Yan, Yiming Xu, and Xilei Zhao. 2021. "A Segment-Level Model of Shared, Electric Scooter Origins and Destinations." *Transportation Research Part D: Transport and Environment* 92 (March):102709. <https://doi.org/10.1016/j.trd.2021.102709>.
- Oeschger, Giulia, Páraic Carroll, and Brian Caulfield. 2020. "Micromobility and Public Transport Integration: The Current State of Knowledge." *Transportation Research Part D: Transport and Environment* 89 (December):102628. <https://doi.org/10.1016/j.trd.2020.102628>.
- Xu, Y., X. Zhao, X. Zhang, and M. Paliwal. 2023. "Real-Time Forecasting of Dockless Scooter-Sharing Demand: A Spatio-Temporal Multi-Graph Transformer Approach." *IEEE Transactions on Intelligent Transportation Systems* 24 (8): 8507–18. <https://doi.org/10.1109/TITS.2023.3239309>.
- Xu, Yiming, Xiang Yan, Virginia P. Sisiopiku, Louis A. Merlin, Fangzhou Xing, and Xilei Zhao. 2022. "Micromobility Trip Origin and Destination Inference Using General Bikeshare Feed Specification Data." *Transportation Research Record* 2676 (11): 223–38. <https://doi.org/10.1177/03611981221092005>.
- Zou, Zhenpeng, Hannah Younes, Sevgi Erdoğan, and Jiahui Wu. 2020. "Exploratory Analysis of Real-Time E-Scooter Trip Data in Washington, D.C." *Transportation Research Record*, May. <https://doi.org/10.1177/0361198120919760>.