Proc. of the 4th International Conference on Electrical, Communication and Computer Engineering (ICECCE)
30-31 December 2023, Dubai, UAE

Public Charging Infrastructure Optimization in Urban Neighborhood: Using DBSCAN and Map Matching with EV Trajectories

Yitian Sun
Viterbi School of Engineering
University of Southern California
Los Angeles, USA
jaspersun2001@gmail.com

***Abstract*—With the surge in electric vehicle (EV) adoption, the demand for efficient public charging infrastructure has become paramount. This paper presents a novel methodology for optimizing the deployment of public EV charging stations in urban areas which encompasses five key phases: data preprocessing, DBSCAN clustering, trajectory-road map matching, dynamic road segmentation, and optimal charging station placement through a 0/1 knapsack optimization problem. Evaluations using the EV taxi trajectory data revealed that our methodology achieved a higher coverage rate with fewer charging stations compared to the existing infrastructure within a specific urban district. The study takes place within a 4.5 square kilometers boundary box, where the traffic flow is shown to be the heaviest. This approach addresses the immediate need for efficient charging infrastructure and provides a scalable solution for other urban areas. Essential techniques such as DBSCAN clustering, road map matching, and dynamic programming are leveraged to bridge the gap between the growing demand for EV charging and the current charging infrastructure coverage. This research holds significant potential for cities globally, aiming to enhance their EV charging infrastructure in line with rising EV adoption rates.**

Keywords—EV charging station optimization, trajectory data, DBSCAN clustering, trajectory-road map matching

I. INTRODUCTION

The rapid growth of EVs in recent years has marked a new era of sustainable transportation. While countries that are the largest contributors to net carbon emissions, such as China, Japan, the U.S., and India, have all set goals to reach carbon neutrality between 2050-2070, purchasing EVs has become a trend for major purchasing powers [1]. Many reasons have propelled the transition from traditional gasoline-driven vehicles to electric alternatives, but it also presents a distinct challenge on infrastructure. As the adoption rate of EVs continues to surge, one of the primary concerns for EV owners is timely and convenient access to charging facilities [2], especially in urban neighborhoods with high population densities.

Installing private charging stations for individual EVs appear to be a straightforward solution. However, due to their low utilization rate and the substantial investment required for installation, maintenance, and electricity costs for their owners, private charging stations are an economically suboptimal choice. Constructing public charging stations is a more viable and efficient alternative. Public charging facilities can serve multiple vehicles in public parking lots, ensuring higher utilization and a better return on investment. However, establishing public charging stations introduces its own set of challenges. One of the most pressing issues is determining the optimal layout for these stations in city areas with significant land scarcity [3]. Given the constraints of budget and space, it is crucial to deploy charging stations in locations that can cater to the maximum number of EVs within the fiscal spending cap of the local government. By analyzing the flow and concentration of trajectories in different areas, it becomes possible to identify key locations where charging stations would be most beneficial. The overarching goal is to ensure maximum coverage of EV traffic flow, providing convenient charging options for the majority while optimizing the use of resources.

In this paper, we utilize DBSCAN clustering algorithm to find the optimal deployment of public charging stations especially in urban areas with the most traffic, bridging the gap between the growing demand for EV charging and the lack of charging infrastructure coverage in urban neighborhoods. Motivated by the research gap in addressing the compatibility between the charging station layout and the complex road networks in urban neighborhoods, we employ a map matching technique to ensure that the optimal layout is practical.

II. RELATED WORKS

The quest for optimizing the layout of public charging stations in urban neighborhoods has garnered significant attention from researchers over the years. The question can be framed differently under certain economic budgets and environmental considerations: some aimed to reduce the drivers' idle time searching for charging piles; some optimized conventional algorithms to enhance the utilization rate of current charging networks; others aimed to maximize the cover range of the charging station network. While there are various approaches, the end goal remains the same. As the number of EVs continues to surge, the need for a strategic and efficient charging infrastructure becomes increasingly paramount.

Researchers have developed novel models and variations of optimization algorithms to address the practical challenge.

In [4], Awasthi et al. enhanced the original genetic and particle swarm optimization algorithms. Their focus was on simulating the placement of charging stations, specifically targeting Allahabad city in India. Similarly, in [5], Li et al. took a nuanced approach by categorizing trajectory data into three distinct subcategories: normal traveling trajectory, seeking trajectory, and charging trajectory. By integrating road map gridding and integer linear programming, they achieved a notable reduction in both seeking and waiting times for charging stations. In comparison, Hu et al. employed the greedy algorithm, GSA hybrid heuristic algorithm, and GA+GA heuristic algorithm to expand Shenzhen, China's current charging station layout [6]. They posited that building fewer charging stations with a large capacity would be more efficient than developing a centralized layout. Their methodology emphasized the significance of the response time of the charging service site, a dimension often overlooked in similar studies. In [7], recognizing the underutilization of charging stations in Wuhan, China, Yang et al. wielded a three-dimensional tensor that modeled drivers' charging behavior and then a context-aware tensor collaborative decomposition method to train the model with a specific evaluation scale. The optimized result filtered out redundant charging station candidates, achieving an impressive user coverage rate of 97.63%.

Among the plethora of algorithms explored, clustering algorithms have emerged as particularly efficacious. Clustering algorithms can adroitly process vast volumes of spatial data, discerning patterns and densities within the trajectory datasets. By grouping similar data points, they can highlight areas of high EV activity, ensuring that the derived clusters truly represent common routes or stops for EVs and thus optimize the charging network's efficiency. In [8], a combination of iterative clustering techniques and modified numerical optimization methods was employed. The objective was to ascertain the most compact infrastructure that would guarantee uninterrupted electric taxi service. For instance, the shared nearest neighbor (SNN) clustering algorithm was adeptly utilized by Dong et al. [9] to craft a fast-charging network in Hainan, China. Likewise, Huang et al. [10] leveraged vehicle trajectory distribution and POI data in Chongqing to generate a heat grid map to assign charging piles based on their respective heat values. Then, they adopted the R-DBSCAN algorithm that eliminates a certain number of clusters in a region covered by radius R to avoid resource abuse. The versatility of clustering algorithms is further exemplified by research that combined Multiple Same-Type Clustering and Multiple Multi-Type Clustering Algorithms, both derivatives of the K-means clustering, to pinpoint optimal charging station locations [11]. Sánchez et al. [12] also championed the K-means clustering algorithm, noting its efficiency in computational time reduction compared to other algorithms. In summary, the research landscape on optimizing charging station layouts is vast and varied. The collective efforts of these scholars provide invaluable insights and pave the way for future innovations in this domain.

The novelty in our proposed method stems from its ability in dealing with the intricacies of urban road networks. In [8]-[11], although the researchers all put their methodologies in real-world scenarios for testing, they did not address the possibility of mismatch between the EV trajectory and the actual road network. Our approach, rooted in a meticulous analysis of real-world EV taxi GPS trajectory data, utilized trajectory-road map matching methodologies to mitigate the scenarios when the trajectory latitude and longitude data do not locate on an actual road. This ensures that all the data EV trajectory are under sufficient consideration. Besides, differently from [5], since we treat all trajectory coordinates as a general data pool, the time parameter can be neglected and therefore reduce the time complexity derived from it.

III. Proposed Method

A. *Problem Definition*

This paper employs real-world EV taxi GPS trajectory data and a strategic public charging station deployment methodology to enhance the public EV charging station coverage in an urban district, particularly focusing on its most densely populated region. To make the problem scalable, the study focuses on the charging station deployment within a 4.5 square kilometer area located at the center of the city. To evaluate the performance of our methodology, we compared the coverage of the existing 14 public charging stations to the potential coverage achieved by placing five charging stations using our framework. This research aims to ascertain the most effective placement of these charging stations to ensure maximum coverage of the EV traffic flow in the selected area.

B. *Methodology Framework*

The methodology framework shown in Figure 1 comprises five phases: 1. Data Preprocessing 2. Density- Based Clustering 3. Trajectory-Road Map Matching 4. Road Segmentation 5. Optimal Charging Station Placement.

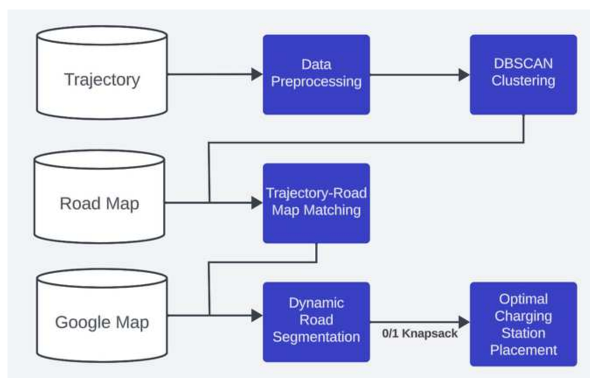

Fig. 1. Methodology Framework

Data Preprocessing. The initial phase is data preprocessing, where the first and last 20 trajectory data points from each of the 130,000 trajectories form the research's trajectory data sample pool. According to 8 interview results we conducted with five different taxi drivers who work in the downtown area, all the considerably scaled public charging stations are located at parking lots. In fact, the charging costs are dependent on the parking lots' regulations. Choosing the first and last 20 trajectory data points is based on two reasons: 1. Most public charging stations are in parking lots. 2. The beginning and end points of a trajectory typically occur either in a parking location or somewhere that offers the vehicle spaces to park, as vehicles usually start and end their journeys there. By focusing on each vehicle's start and end points, this phase effectively narrows down the data spread, filtering out scattered or erratic intermediate points and resulting in distinct clusters representing common starting and ending locations.

DBSCAN Clustering. As one of the key techniques in data mining used to discover data distribution and hidden patterns, clustering refers to dividing a large set of data points into several categories, ensuring that the data within each category is as similar as possible while the data between different categories is as distinct as possible. DBSCAN is a density-based clustering algorithm, which searches for neighboring points to gradually increase the density around an object, aiming to find areas with a high density of searched points or objects. This algorithm divides areas with sufficiently high density into clusters and can discover clusters of any shape in data with noise [13]. Multiple researchers have demonstrated the compatibility of DBSCAN to road trajectory datasets. Demonstrated in [14], GPS trajectory data exhibits significant spatiotemporal correlation due to constraints like road network structure, travel choice, and city function zoning. The paper utilizes DBSCAN to cluster the aggregated trajectory data into groups, each represented by a center point. This phase returns a set of data points that are further refined by the measures mentioned above, producing a heatmap representation of traffic density.

Trajectory-Road Map Matching. Trajectory positions recorded by GPS devices often do not match the actual road taken by the object carrying the device. This discrepancy arises from various sources of errors, such as interference in satellite signals when moving through tunnels, urban canyons or traffic jams that produce many consecutive points with minimal spatial gaps. For example, unthinkingly following the traffic flow gathered through clustering may result in placing some public charging stations in locations that are not viable, such as inside some buildings and parks, due to GPS deviations.

Embodying the impact of each individual trajectory data point, even if it is misplaced inside some buildings, we designed a map-matching model to align each trajectory point with the most proximate point on the road network. We iterated through all the cluster centers and used the distance method from the Geopandas library to compute distances between each trajectory point and all points in the road network. Existing points that are already placed on the roads are left as is. For instance, if a trajectory point is recorded inside a hospital due to GPS inaccuracies, the function will determine which point on the adjacent road is nearest to this trajectory point. This process ensures that the trajectory data is accurately aligned with the actual road network, mitigating any discrepancies introduced by GPS errors. In essence, the function refines the trajectory data by snapping each point to its most likely position on the road. The closest road point is then identified based on the shortest distance, ensuring that selected clusters are near roads, avoiding unsuitable locations like public or private buildings, which are unfit for public charging station deployment. The results returned from this phase become candidate locations for the problem.

Dynamic Road Segmentation. Traditional methods utilize fixed lengths for segments. However, a more dynamic approach could be more beneficial. For instance, highways, which have consistent traffic flow, might be represented with longer road segments, while city streets, which have more variable traffic, could have shorter segments. This dynamic segmentation can provide a more accurate representation of traffic patterns. In this paper, we implement road segmentation following the road map.

Highway road segments often have dedicated service areas, or rest stops and are abandoned since they rarely meet the requirement to build large scale public charging stations. Intuitively, traffic lights serve as natural segmentation points for roads. They often indicate intersections or junctions, which are typically areas of high traffic density.

By segmenting roads at traffic lights, the study ensures that each segment represents a distinct traffic flow pattern and simplifies the segmentation implementation. Instead of relying on complex algorithms or arbitrary distance metrics, the study uses an easily identifiable and universally understood feature of road networks. By aligning the road segmentation method with these urban planning decisions, the study ensures that charging stations are placed in areas that are not only of high traffic density but also of strategic importance to the city's overall development.

Optimal Charging Station Placement. Given W total charging stations that can be installed and n total potential charging station locations, the paper transforms the optimization problem into a 0/1 knapsack optimization problem. A dynamic programming approach then maximizes the knapsack's profit, representing the maximum traffic flow coverage by the charging station layout. This method pinpoints strategic locations that optimize traffic flow coverage while adhering to available charging station constraints and permissible. The pseudocode for optimal charging station placement is shown below:

Algorithm 1: Optimal Charging Station Placement

Input: Number of locations n, Maximum weight W, Profits array profits, Distances matrix distances, Minimum distance min_distance

Output: Maximum profit with optimized charging stations

```
Initialize dp with dimensions (n+1) x (W+1) with all values set to 0;
Initialize chosen with dimensions (n+1) x (W+1) with all values set to False;
for each location i, from 1 to n, do
    for each weight j from 1 to W, do
        no_install ← dp[i-1][j];
        k ← i - 1;
        install ← profits[i-1];
        while k >= 0 and distances[k][i-1] < min_distance do
            if k < i or not all(chosen[k][:j]) then
                install ← max(install, profits[i-1] + dp[k][j-1]);
            end if
            k ← k - 1;
        end while
```

```
dp[i][j] ← max(no_install, install);
    if install > no_install then
        chosen[i][j] ← True;
    end if
  end for
end for
return dp[n][W];
```

IV. Evaluations

The evaluation is conducted through experiments using data gleaned from the 4.5 square kilometers testing area, which assess the efficacy of our EV charging station optimization methodology. The used data for the experiment includes EV trajectory data, road map data, and latitude and longitude of the established charging stations in testing area. Detailed descriptions are presented below, respectively.

- EV Trajectory Data. The study obtains 101,600 distinctive EV taxi trajectories from an EV GPS dataset from the testing city on October 29, 2016. Each record includes its vehicle identification and all the latitude and longitude coordinates the vehicle has traveled. Since we extract the first and the last 20 trajectory data points from each trajectory, the dataset contains around 4,064,060 latitude and longitude coordinates.
- Road Map Data. The study utilizes SHP geometry data which delineates all the roads including highways within the testing area. The region is bounded respectively northwestern and southeastern by (30.654198, 104.042957) and (30.711674, 104.115727). The blue lines in Figure 2 illustrates the road network of the testing area.

Fig. 2. Road Network Data

- Google Map Data. As of October 1, 2023, there are 12 established public charging stations located in our testing area. We utilize google map data to locate the coordinates for the 12 established charging stations, displayed by the red dots in Figure 3.

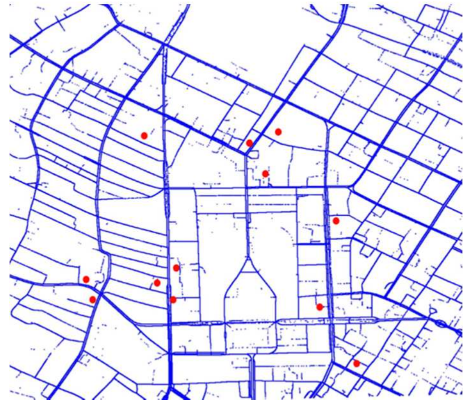

Fig. 3. Established Charging Station Locations

We extract a total of N = 4,096 EV trajectories from the dataset to implement DBSCAN clustering. Experiments showed that a plethora of data increases computational complexity and unintentionally creates excess clusters, placing the charging station in less significant locations. Conversely, a shortage of data results in overfitting, where the clusters formed may not accurately represent the actual underlying patterns but rather fit the noise and outliers, leading to misleading interpretations and predictions. To obtain the optimal cluster result, the study employs GridSearchCV, a hyperparameter tuning method, to identify the data's most suitable "eps" and "min_samples" parameters. The optimal eps parameter value is 0.001, and the optimal min_samples parameter value is 3, resulting in 256 trajectory clusters through DBSCAN. Figure 4 illustrates the trajectory data which returns 256 clusters.

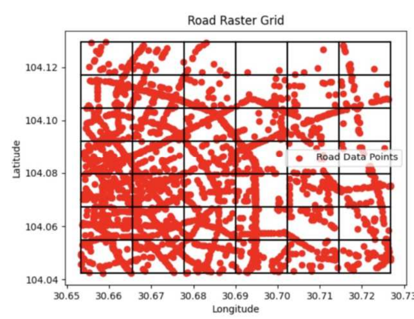

Fig. 4. Trajectory Cluster Data

We then determine the center coordinates for each cluster and conduct trajectory-road map matching technique between the 256 coordinates and the geometry data from the road network dataset. Although the cluster centers each represent the concentration of traffic flow suitable for public charging station placement, the numerical value each coordinate embodies can vary. The colored dots displayed in Figure 5 are the trajectory clusters, and each color represents their respective traffic magnitude, shown in the legend section of the chart. To accurately present the traffic flow of each cluster center, the approximate number of trajectory coordinates classified to each cluster is displayed.

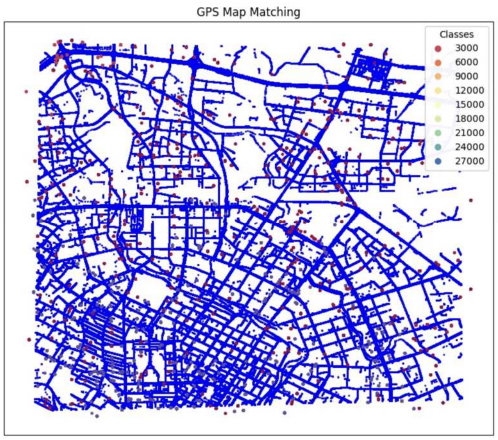

Fig. 5. Clustered Traffic Data

After implementing dynamic road segmentation and finalizing the deployment of charging stations, we conduct evaluations to assess the methodology's effectiveness. The primary metric used for evaluation is the coverage rate, measuring the percentage of EV trajectory coordinates covered by the range of charging models over the total number of trajectory coordinates from the sample pool. The current public charging infrastructure in the testing area, consisting of 12 established public charging stations, serves as the baseline for comparison. While the existing 12 charging stations achieve a 27.03% coverage rate, placing just eight public charging stations attains a 29.04% coverage rate using our methodology. This outcome indicates that with less than half the number of charging stations, our methodology exceeds the performance of the existing placement structure. Besides, the average coverage rate of individual charging facilities from both the new model and the established model are presented below. According to TABLE 1, out of the 4,064,061 total coordinates, each individual charging station of our methodology in average achieved a 1.14% increase in terms of coverage rate, spanning 46,330 more vehicle trajectory points than that of the established model.

TABLE I. AVERAGE COVERAGE RATE

Average Coverage Rate of New Model	**Average Coverage Rate of Established Model**
9.10%	7.96%

The red dots in Figure 6 represents the charging station layout devised through our methodology and the green dots denote the established charging station locations. The superior performance of our methodology in optimization and enhancing the compatibility between charging station layout and road networks in urban settings, despite using fewer charging stations, underscores its efficiency in optimizing the current charging infrastructure based on EV trajectory data. It is noteworthy that while the coverage rate achieved by our methodology is higher than that of the established ones, it is still relatively low. This can be attributed to the fact that the test dataset encompasses much bigger region than our testing area. In fact, area of the test zone accounts for only 5 percent of the total test dataset. And since we only deploy charging stations within the testing area, despite having the heaviest traffic flow, it cannot cover the traffic scattered far away from the testing boundary.

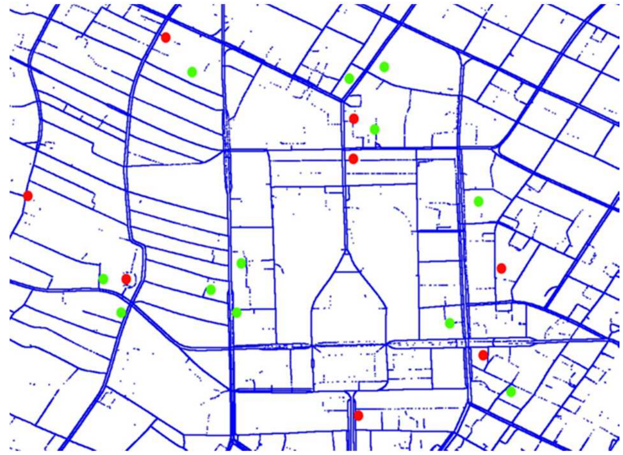

Fig. 6. Current Charging Station Locations (Red) VS. Established Charging Station Locations (Green)

V. CONCLUSION

In this study, we have presented a comprehensive methodology for optimizing the deployment of public EV charging stations, a critical infrastructure to support the burgeoning EV market. Our approach, grounded in the analysis of real-world EV taxi GPS trajectory data and employing advanced techniques like DBSCAN clustering, trajectory-road map matching, and dynamic road segmentation, has improved both the performance of individual charging stations and the overall placement model. The novelty lies in considering that the coordinates of the clusters' centers do not lie on any roads, which complicates pragmatic charging station placement. Since public charging stations are typically deployed in parking lots near roads, the map matching step ensures that all the cluster centers are located on a road segment which enhanced the feasibility of the study.

Our evaluation, conducted within a specific urban district, underscores the efficacy of our methodology. Despite utilizing fewer charging stations, we achieved a higher coverage rate compared to the existing infrastructure. This outcome not only highlights the potential for cost savings but also underscores the strategic advantage of our approach in ensuring that charging stations are deployed in areas of high EV activity, maximizing their utility and accessibility. In the future, we can expand the scale of the study generating the methodology to a larger testing area so that we can get a more comprehensive evaluation result and increase the coverage rate.

REFERENCES

[1] S. Khan, A. Adnan and N. Iqbal, "Applications of Artificial Intelligence in Transportation," 2022 International Conference on Electrical, Computer and Energy Technologies (ICECET), Prague, Czech Republic, 2022, pp. 1-6.

[2] Li, Qing, et al. "Application of Clustering Algorithms in the Location of Electric Taxi Charging Stations." Sustainability, vol. 14, no. 13, June 2022, p. 7566.

[3] Jia, Jianmin, et al. “Planning of the Charging Station for Electric Vehicles Utilizing Cellular Signaling Data.” Sustainability, vol. 11, no. 3, Jan. 2019, p. 643.

[4] Awasthi, et al. "Optimal planning of electric vehicle charging station at the distribution system using hybrid optimization algorithm." Energy 133 (2017): 70-78.

[5] Y. Li, J. Luo, C. -Y. Chow, K. -L. Chan, Y. Ding and F. Zhang, "Growing the charging station network for electric vehicles with trajectory data analytics," 2015 IEEE 31st International Conference on Data Engineering, Seoul, Korea (South), 2015, pp. 1376-1387.

[6] Hu, Dandan, et al. "Data driven optimization for electric vehicle charging station locating and sizing with charging satisfaction consideration in urban areas." IET Renewable Power Generation 16.12 (2022): 2630-2643.

[7] Yang, Yu, Yongku Zhang, and Xiangfu Meng. "A data-driven approach for optimizing the EV charging stations network." IEEE Access 8 (2020): 118572-118592.

[8] Cilio, Luca, and Oytun Babacan. "Allocation optimisation of rapid charging stations in large urban areas to support fully electric taxi fleets." Applied Energy 295 (2021): 117072.

[9] X. Dong, Y. Mu, H. Jia, J. Wu and X. Yu, "Planning of Fast EV Charging Stations on a Round Freeway," in IEEE Transactions on Sustainable Energy, vol. 7, no. 4, pp. 1452-1461, Oct. 2016.

[10] Huang, Danhui, Yuan Chen, and Xiaochen Pan. "Optimal model of locating charging stations with massive urban trajectories." IOP Conference Series: Materials Science and Engineering. Vol. 715. No. 1. IOP Publishing, 2020.

[11] Li, Qing, et al. "Application of clustering algorithms in the location of electric taxi charging stations." Sustainability 14.13 (2022): 7566.

[12] Sánchez, Danny García, et al. "A Clustering Approach for the Optimal Siting of Recharging Stations in the Electric Vehicle Routing Problem with Time Windows." Energies 15.7 (2022): 2372.

[13] Schubert, Erich, et al. "DBSCAN revisited, revisited: why and how you should (still) use DBSCAN." ACM Transactions on Database Systems (TODS) 42.3 (2017): 1-21.

[14] Liu,C.K.,etal."Uncovering the aggregation pattern of gps trajectory based on spatiotemporal clustering and 3D visualization." The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences 42 (2020): 255-260.