

Adaptive Urban Congestion Modeling and Zonal Signal Control Based on Taxi Trajectory Data

Xilin Li ^{1,2}, Shucen Huo ^{1,2*}, Yongjian Huang ^{1,2}

¹ School of Management, Shanghai University, No. 99, Shangda Road, Shanghai 200444, China,

² Confucius Institute, University of Bahrain, Bahrain,

Correspondence Email: huoshucen@shu.edu.cn

Keywords: Urban Traffic Congestion, OD Heatmap, Regional Signal Control, Congestion Pressure

Abstract

With increasing urban travel demand and rising peak-hour congestion pressure, traditional static traffic signal timing methods are insufficient to support rapid response and regional coordination. This study proposes a regionalized traffic signal control strategy that integrates taxi trajectory data mining, OD heatmap clustering, and congestion index modeling. Using Yellow Taxi trip data from New York City from March to May 2025, an OD heat network is constructed to capture inter-zone travel patterns. K-Means clustering is applied to identify functionally consistent urban regions, and a “congestion pressure index” is introduced to quantify traffic load conditions. Based on this, a three-tier responsive signal timing mechanism is designed. A minute-level simulation framework is developed to evaluate the adaptability of the strategy under sudden traffic surges. Results demonstrate that the proposed strategy is highly deployable and effective in alleviating local congestion during peak periods while enhancing system stability and resilience. This study provides a data-driven and adaptive solution for intelligent urban traffic signal management.

1 Introduction

With the continuous acceleration of global urbanization, the pressure on urban transportation systems is becoming increasingly severe. Particularly during peak hours in the morning and evening, traffic congestion has become a major bottleneck affecting the operational efficiency of cities[1], the travel experience of residents, and overall environmental sustainability. Traditional traffic signal control mechanisms are often based on fixed cycles and pre-defined plans, lacking real-time adaptability to dynamic traffic flows. Consequently, they frequently result in imbalanced signal timing and localized bottlenecks when encountering sudden load surges or spatially uneven commuting demands[2]. As a result, developing more flexible and responsive signal control strategies has become a key direction in the transformation of urban traffic management.

In this context, the rise of big data and artificial intelligence has provided new technical support for intelligent

regulation of urban traffic systems. In particular, taxi trajectory data—characterized by high-frequency updates, wide spatial coverage, and continuous spatial distribution—has become a vital source for capturing urban mobility patterns and identifying peak-hour congestion dynamics. Deep mining and modeling of such data enables not only the reconstruction of OD (origin-destination) travel patterns but also facilitates regional load detection and signal strategy adaptation.

Although existing studies have made initial progress in traffic flow prediction, route optimization, and signal timing adjustment using mobility data, they still face significant challenges when addressing large-scale OD structures at the city level, coordinating multiple regions, or responding to short-term shock events. Specifically, the key research gaps lie in: (1) how to extract representative regional behavioral patterns from massive trajectory datasets for spatial traffic load delineation and evolution tracking; (2) how to construct a congestion metric with dynamic adaptability for supporting real-time signal control decisions; and (3) how to design deployable regional signal control mechanisms that enable efficient and low-cost coordination of urban traffic.

2 Literature Review

With the rapid development of intelligent transportation systems and urban computing technologies, how to optimize traffic signal control based on large-scale spatiotemporal data has become an important research focus[3]. Existing studies have approached this problem from multiple perspectives, including time-series prediction of traffic flow, intelligent signal control algorithms, and integration of real-time data[4][5]. This chapter reviews the relevant literature from three main dimensions: traffic flow prediction methods, signal control optimization strategies, and applications of urban sensing data in transportation.

Early traffic flow prediction models were mostly based on statistical methods such as Autoregressive Integrated Moving Average (ARIMA) and Kalman filtering, which offered good performance for short-term prediction under stable traffic conditions[6]. However, these models tend to struggle with complex nonlinear dynamics and fluctuating patterns during peak periods or under abnormal traffic loads[7]. In response, researchers have increasingly turned to machine learning models such as Support Vector Regression (SVR), Random Forests, and more recently, deep learning frameworks like

Long Short-Term Memory (LSTM) networks and Graph Neural Networks (GNN). These models are capable of capturing temporal dependencies and spatial correlations in traffic flow data[8][9], and have shown promising results in improving prediction accuracy.

In the field of signal control optimization, the classic Webster method remains widely used due to its simplicity and practicality. It optimizes cycle lengths and green time splits based on traffic volume and saturation flow rate[10]. However, the method assumes static and balanced flow conditions, which often do not reflect the dynamic and asymmetric nature of real-world traffic. As such, more adaptive signal control methods have emerged. For example, reinforcement learning (RL) has been introduced to optimize signal timing by treating intersections as intelligent agents that interact with their environment[11]. Multi-agent reinforcement learning (MARL) has also been proposed for coordinating signals across multiple intersections, enhancing global traffic efficiency[12]. Despite these advances, most studies remain limited to intersection-level optimization, and few address broader regional coordination challenges under variable demand.

Meanwhile, the growing availability of urban sensing data—including taxi GPS trajectories, loop detectors, mobile phone signals, and camera feeds—has enabled richer representations of traffic conditions[13]. Taxi data in particular has proven useful due to its high spatiotemporal resolution, coverage across urban areas, and strong correlation with travel demand[14]. Researchers have used taxi data to extract OD (origin-destination) matrices, detect congestion zones, and infer mobility patterns. For instance, clustering methods like K-Means and DBSCAN have been applied to identify functional regions in cities, while trajectory aggregation techniques have been used to construct dynamic heatmaps of traffic intensity[15].

However, current research still faces several challenges. First, there is often a gap between predicted traffic states and their translation into actionable signal strategies. Second, few studies have focused on signal control at the zoning or regional level, where coordinated response is necessary for mitigating network-wide congestion. Third, existing models rarely incorporate sudden traffic fluctuations (e.g., surges due to events or incidents) in the design of control mechanisms. These gaps highlight the need for integrated frameworks that couple traffic flow analysis, real-time sensing, and rule-based or data-driven control in a unified structure. Thus, This study seeks to fill these gaps by proposing a congestion-aware signal control strategy based on real-world taxi trajectory data, combining OD clustering, congestion index modeling, and dynamic strategy deployment to support intelligent urban traffic management.

3 Methodology

This study develops a congestion-responsive, zone-level signal control strategy driven by multisource urban traffic sensing data. The methodological framework includes five main components: data acquisition and preprocessing, OD

heat analysis, regional clustering, congestion index modeling, and dynamic strategy design.

At the data processing stage, we utilize the publicly available Yellow Taxi trip records released by the New York City Taxi and Limousine Commission (TLC), covering the period from March to May 2025. The dataset contains over 36 million records. For each trip, we extract key fields such as pickup time, pickup and drop-off LocationIDs, trip distance, and passenger count. Data cleaning includes removing records with missing or anomalous pickup/drop-off coordinates (e.g., LocationID equals zero), trips with a distance of less than 0.1 miles, and records outside peak hours. The analysis focuses on weekday peak periods, defined as 07:00–10:00 and 17:00–20:00. We aggregate pickup data at the minute level to generate temporal traffic heat indicators. Additionally, LocationIDs are mapped to specific zones and boroughs using the official TLC taxi zone shapefile, enabling spatial linkage and visualization.

For OD behavior analysis, we construct an OD frequency matrix based on PU-DO LocationID combinations. This matrix reflects the spatial interaction intensity between regions during peak hours. To improve spatial aggregation and interpretability, we apply clustering methods such as K-Means to partition the city into travel behavior-similar zones. The goal is to identify highly interactive zones during congestion-prone time periods, which serve as fundamental units for signal strategy implementation. The clustering results are visualized on the city map, highlighting travel intensity patterns and spatially coupled zones.

To measure the traffic burden within each cluster, we introduce a Congestion Pressure Index, defined as:

$$\text{Congestion Index} = \frac{\text{Vehicles Entering per Unit Time}}{\text{Average Travel Speed}} \quad (1)$$

Average speed is estimated using the distance and duration of each trip. This index captures the imbalance between demand and supply, dynamically reflecting the operational stress within each traffic zone. It serves as the core decision variable for signal adjustment.

Based on the congestion index, we design a responsive signal control mechanism that classifies zones into three categories: (1)Congested zones: Red light duration is extended to limit incoming traffic. (2)Free-flow zones: Green light duration is extended to enhance capacity and promote dispersion. (3)Neutral zones: Signal timing remains unchanged to maintain status quo.

This control logic supports decentralized deployment, allowing each zone to make real-time decisions based on its own pressure index. At the same time, adjacent zones can share information to achieve spatial coordination. The strategy is rule-driven and can be embedded in edge computing units within urban infrastructure, facilitating flexible and scalable implementation.

To evaluate the adaptive performance of the proposed strategy under nonstationary conditions, we construct a 15-minute time-stepped impact load simulation framework. This framework simulates sudden flow surges in specific zones and monitors the evolution of congestion levels and signal

response over time. Different rule-based strategies (e.g., red light +15s, green light +15s) are tested in combination, and their effects on traffic recovery are analyzed. Simulation results demonstrate that the system is capable of timely adjustments in high-pressure scenarios, maintaining resilience and avoiding spillover congestion.

4 Empirical Results

This chapter conducts an empirical study using New York City's peak-hour taxi trajectory data to systematically analyze OD heatmaps, regional clustering, signal strategy construction, and simulation testing.

First, through filtering and cleaning the Yellow Taxi data during peak hours on weekdays from March to May 2025, approximately 12 million valid trajectories were obtained. Using the official TLC vector map of taxi zones, spatial mapping of Location ID to geographic regions was completed. In the data visualization analysis, pickup and drop-off heatmaps were created shown in Figures 1 and 2. The results show that Midtown Manhattan, including Times Square, Midtown East, Lincoln Square East, and JFK Airport, exhibit pronounced high-frequency travel patterns, forming typical high-demand zones. Pickups are concentrated in residential and departure hubs, while drop-offs cluster in commercial centers and office areas—revealing strong OD directionality and a commuter pattern of spatial separation between home and workplace.

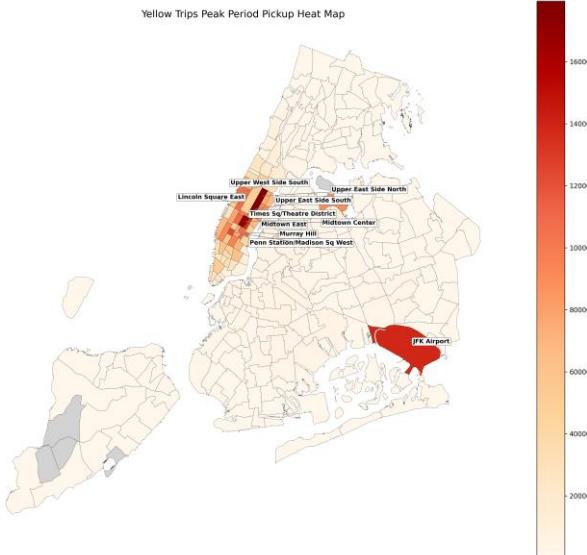


Figure 1: Yellow Trips Peak Period Pickup Heat Map.

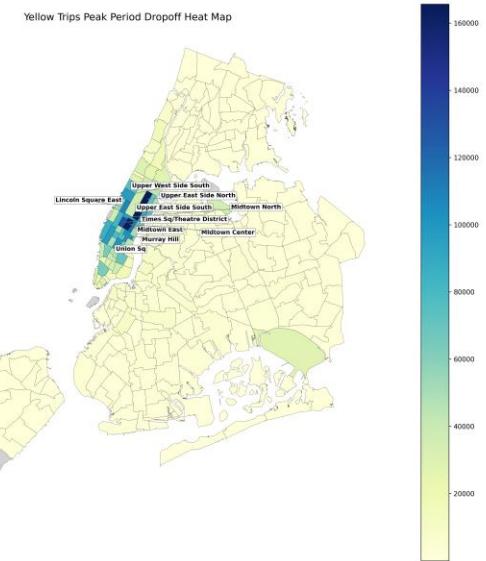


Figure 2: Yellow Trips Peak Period Dropoff Heat Map.

Furthermore, based on the PU-DO OD frequency combinations, an OD heat network was constructed, and a Sankey diagram of the top 50 OD pairs was drawn, shown in Figure 3, to visually present the traffic flow connections between key departure and destination zones. The diagram shows dense unidirectional flows from areas such as Upper East Side South and JFK Airport to destinations like Midtown Center and Times Square/Theatre District, indicating a centralized commuting pressure. These high-frequency OD links form critical pressure corridors in the traffic system and should be prioritized in signal control strategy design.

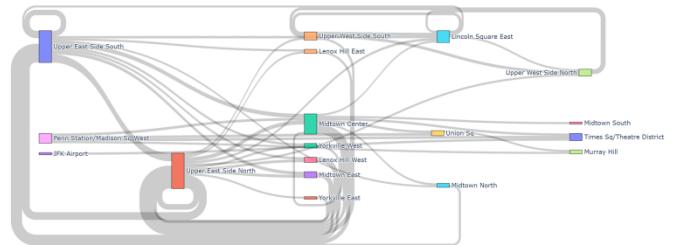


Figure 3: OD Travel Heat Sankey Diagram During Peak Hours.

To enable operable regional management and signal regulation, the K-Means algorithm was applied to OD heat metrics to spatially cluster the city into six regions with similar travel behaviors, resulting in “heat clustering zones.” As shown in Figure 4, zones within the same cluster display high spatial continuity—especially central Manhattan and central Brooklyn, which formed independent clusters—demonstrating the cohesion of travel patterns. This clustering enhances the regional coordination of signal strategies and provides a distributed foundation for large-scale traffic interventions.

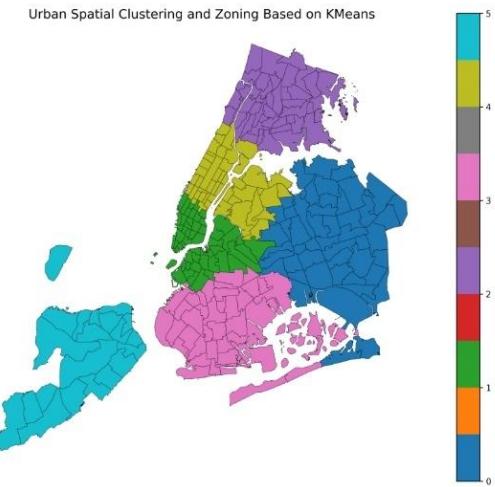


Figure 4 Urban Spatial Clustering and Zoning Based on K-Means.

Based on the clustering, a **congestion pressure index** was introduced to measure the traffic carrying status of each zone. Defined as the ratio of incoming vehicle count per unit time to average travel speed, it reflects the contradiction between instantaneous traffic load and operational efficiency. Empirical analysis found that some regions (e.g., Midtown Center, JFK Airport) had congestion pressure values exceeding 4.0 during peak hours, signaling significant bottleneck risks.

Using this congestion index, a **responsive signal control strategy** was designed. As shown in Figure 5, in highly congested areas, red light duration is extended to restrict inflows. In adjacent zones with smooth traffic, green light duration is extended to increase outflow and dispersion. For neutral zones, signal timing remains unchanged.

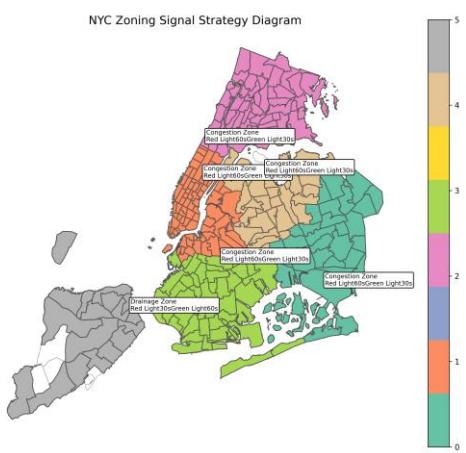


Figure 5 Urban Spatial Clustering and Zoning Based on Kmeans.

The diagram labels each zone's control type, forming a dynamic mapping of *cluster – congestion – strategy*. This not

only aids dispatchers in decision-making but also facilitates deployment in smart traffic platforms.

To test this strategy's adaptability under sudden traffic changes, a minute-level simulation framework was built. It simulates a scenario where Zone B and Zone C encounter a 250% traffic surge between 08:10 and 08:20. The system dynamically monitors congestion pressure and strategy responses, as shown in Figure 6. Results show that the strategy promptly increases red light duration during the surge and automatically reverts to “remain unchanged” after the peak, allowing traffic conditions to recover. The time-series visualization captures the coordinated evolution of strategy response and traffic load, validating the mechanism's robustness and timeliness.

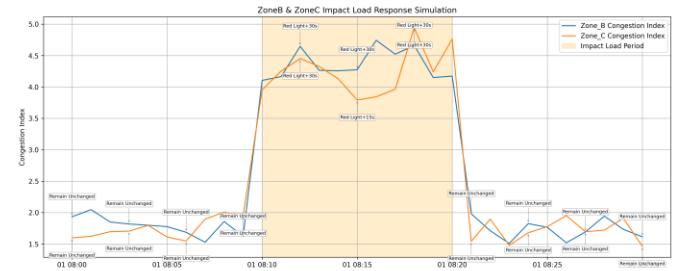


Figure 6: ZoneB & ZoneC Impact Load Response Simulation.

In conclusion, this empirical analysis demonstrates the feasibility and practicality of the proposed regional signal control model using real urban traffic data. With OD heat clustering and regional response mechanisms, the traffic signal system can more precisely and effectively adapt to dynamic traffic surges and congestion, providing strong support for next-generation intelligent traffic management.

5 Conclusion

This study addresses the problem of urban traffic congestion during peak periods by proposing a regionalized signal control method that integrates trajectory data mining with zoned control logic. By developing a “congestion pressure” index and combining spatial clustering with an adaptive strategy response mechanism, the research establishes a closed-loop system for traffic state perception, analysis, and responsive regulation. Based on peak-hour taxi trajectory data from New York City, the study conducted modeling and simulation experiments, leading to several key conclusions.

To begin with, urban travel activity displays clear patterns of spatial concentration and temporal imbalance. Certain areas experience prolonged periods of high commuting pressure during peak hours, which cannot be effectively addressed by traditional static signal timing methods. The congestion pressure index proposed in this study captures the real-time traffic load within each region and offers a quantitative basis for dynamic signal control strategies.

Secondly, the zoning approach grounded in OD heatmaps and spatial clustering demonstrates strong internal consistency in traffic behavior and strategic applicability. The responsive signal control mechanism tailored for each zone performs effectively in simulation tests, showing high adaptability and resilience. It can flexibly respond to

unexpected surges in traffic flow and help mitigate the spread of localized congestion.

Furthermore, simulation results reveal that moderately extending red-light durations in high-congestion zones, while enhancing throughput in low-load areas, contributes to achieving localized traffic balance and overall system stability. The strategy logic is both transparent and practical for deployment on edge computing infrastructures and supports future integration with intelligent transportation systems.

Finally, the overall strategy framework proposed in this study is characterized by its logical clarity, interpretability, and dynamic responsiveness. It proves the feasibility of constructing a data-driven, zone-specific signal control system using real-world urban traffic data.

In addition, the research has some limitations. The current strategy framework is rule-based and lacks a data-driven optimization or iterative improvement mechanism. Moreover, the traffic sensing model does not yet account for multimodal travel behaviors, such as public transportation or non-motorized vehicles. Future research could explore the use of reinforcement learning, adaptive modeling techniques, and multi-source data fusion to enhance the intelligence, granularity, and robustness of urban traffic control systems.

References

- [1] Hopkins, John L., and Judith McKay. "Investigating 'anywhere working' as a mechanism for alleviating traffic congestion in smart cities.", *Technological Forecasting and Social Change*, 142 (2019): 258-272.
- [2] Laanaoui, My Driss, et al. "Enhancing Urban Traffic Management Through Real-Time Anomaly Detection and Load Balancing.", *Ieee Access*(2024).
- [3] Zou, Chao, et al. "Evaluating traffic emission control policies based on large-scale and real-time data: A case study in central China.", *Science of The Total Environment*, 860 (2023): 160435.
- [4] Boukerche, Azzedine, Yanjie Tao, and Peng Sun. "Artificial intelligence-based vehicular traffic flow prediction methods for supporting intelligent transportation systems.", *Computer networks*, 182 (2020): 107484.
- [5] Zheng, Yu. "Intelligent Signal Optimization Algorithm Based on Artificial Intelligence in Intelligent Traffic Control Systems.", *Procedia Computer Science*, 247 (2024): 445-452.
- [6] Guo, Jianhua, Wei Huang, and Billy M. Williams. "Adaptive Kalman filter approach for stochastic short-term traffic flow rate prediction and uncertainty quantification.", *Transportation Research Part C: Emerging Technologies*, 43 (2014): 50-64.
- [7] Vlahogianni, Eleni I., Matthew G. Karlaftis, and John C. Golias. "Temporal evolution of short-term urban traffic flow: a nonlinear dynamics approach.", *Computer-Aided Civil and Infrastructure Engineering*, 23.7 (2008): 536-548.
- [8] Medina-Salgado, Boris, et al. "Urban traffic flow prediction techniques: A review.", *Sustainable Computing: Informatics and Systems*, 35 (2022): 100739.
- [9] Wang, Xu, et al. "Spatio-temporal analysis and prediction of cellular traffic in metropolis.", *IEEE Transactions on Mobile Computing*, 18.9 (2018): 2190-2202.
- [10] Yang, Inchul, and R. Jayakrishnan. "Real-time network-wide traffic signal optimization considering long-term green ratios based on expected route flows.", *Transportation Research Part C: Emerging Technologies*, 60 (2015): 241-257.
- [11] Wu, Zhonghao, et al. "Adaptive Traffic Signal Timing Optimization Using Deep Reinforcement Learning in Urban Networks.", *Artificial Intelligence and Machine Learning Review*, 5.4 (2024): 55-68.
- [12] Chu, Tianshu, et al. "Multi-agent deep reinforcement learning for large-scale traffic signal control.", *IEEE transactions on intelligent transportation systems*, 21.3 (2019): 1086-1095.
- [13] Mahrez, Zineb, et al. "Smart urban mobility: When mobility systems meet smart data.", *IEEE Transactions on Intelligent Transportation Systems*, 23.7 (2021): 6222-6239.
- [14] Liu, Jielun, et al. "Spatial-temporal inference of urban traffic emissions based on taxi trajectories and multi-source urban data.", *Transportation Research Part C: Emerging Technologies*, 106 (2019): 145-165.
- [15] Toshniwal, Durga, et al. "Application of clustering algorithms for spatio-temporal analysis of urban traffic data.", *Transportation Research Procedia*, 48 (2020): 1046-1059.