Australasian Transport Research Forum 2025 Proceedings
18-21 November, Auckland, New Zealand
Publication website: https://australasiantransportresearchforum.org.au/

Detecting bus stop congestion and bus bunching using citywide bus GPS data: A Brisbane case study

Junyan Liu[1], Dan He[2], Zetong Zhang[2], Jiwon Kim[2]

[1]Northern Arizona University

[2] The University of Queensland

Email for correspondence: d.he@uq.edu.au

Abstract

Public bus systems play a vital role in urban mobility, especially in large metropolitan areas such as Brisbane. Ensuring the reliability and efficiency of these systems is essential for reducing passenger wait times, improving service satisfaction, and supporting sustainable transport goals. This study focuses on detecting two common operational issues, bus stop congestion and bus bunching, using citywide real-time GPS data from Brisbane's entire bus fleet. Bus stop congestion is identified when multiple buses from different routes arrive at the same stop within a short time window, while bus bunching refers to successive arrivals of same-route buses with abnormally short headways. Using one day of historical GPS data, we perform stop-by-stop analysis to detect these events based on simple threshold-based rules. The outcome highlights spatial and temporal hotspots of inefficiency in the network, offering valuable insights for improving real-time operations, scheduling strategies, and long-term transit planning.

1 Introduction

Public bus transportation serves as a vital component of urban mobility systems, offering an affordable, accessible, and sustainable travel option for millions of residents. As urban populations grow and traffic conditions become increasingly dynamic, ensuring the reliability and efficiency of bus operations has become a pressing concern for city planners and transit authorities. Service irregularities such as bus stop congestion (Gibson, Baeza and Willumsen, 1989) and bus bunching (Ganguly and Maitra, 2024) can significantly degrade transit performance, leading to increased passenger wait times, unpredictable arrivals, and overall dissatisfaction with public transport. Detecting and addressing these operational issues is essential for improving service quality, optimising network design, and encouraging public transport usage.

In this study, we leverage real-time bus GPS trajectory data in GTFS format provided by Translink (Queensland Government, 2020), which records the continuous movement of all buses operating across Brisbane. Although the data does not directly capture stop-level events, it includes high-frequency location updates with timestamps, allowing us to infer arrival patterns at stops through spatial and temporal analysis. Using one day of historical GPS data, we aim to investigate two specific operational issues: (1) bus stop congestion, defined as situations where multiple buses—regardless of route—arrive in close proximity to a stop within a short time window, and (2) bus bunching, where buses from the same route arrive near a stop in rapid succession with significantly shorter-than-expected headways. These issues are examined using a stop-by-stop, rule-based detection approach, where bus arrivals are inferred from trajectory data. This method offers simplicity, interpretability, and scalability across the city's extensive bus network while remaining adaptable for real-time monitoring applications.

Numerous prior studies (Mignone, Malerba and Ceci, 2022; Wrona, Grzenda and Luckner, 2022) have addressed public transport anomalies using a range of techniques, from time-

series analysis and rule-based detection to deep learning (Gangwani and Gangwani, 2021) and spatio-temporal clustering. While these advanced models can offer high accuracy, they often depend on complex architectures, extensive labelled data, or domain-specific configurations, making them less accessible for large-scale deployment or real-time operational use. Additionally, many existing approaches (Kong *et al.*, 2024) focus on network-level or trajectory-based analysis, which may overlook localized issues such as stop-specific congestion or irregular headways. As an initial step toward fully leveraging the potential of citywide bus GPS data, this study adopts a straightforward, rule-based approach that analyses each stop independently using clearly defined temporal thresholds. By focusing on practical detection of high-density bus arrivals and same-route headway deviations, we aim to establish a transparent and scalable baseline. At the conclusion of this paper, we provide an insightful analysis of the findings and discuss potential directions for future research, including the integration of spatio-temporal models, machine learning techniques, and real-time detection systems to enhance operational monitoring and decision-making.

The outcome of this study includes the identification of key congestion hotspots, frequent bunching locations and times, and visualisations such as heatmaps that provide actionable insights into Brisbane's bus network performance. These findings not only contribute to understanding service reliability issues from a data-driven perspective but also provide a foundation for developing real-time detection systems, integrating predictive models, and informing future strategies for scheduling adjustments and dynamic fleet control. This work represents an important step toward operationalizing large-scale transit data for practical transport planning and real-time management applications.

2 Related work

A wide range of studies have been conducted to detect and address such service anomalies (Mignone, Malerba and Ceci, 2022). Rule-based approaches are commonly used due to their simplicity and interpretability. These typically rely on threshold definitions for headway deviation (for bunching) (Moreira-Matias *et al.*, 2012) or arrival counts within time windows (for stop congestion). While effective in certain settings, they often require manual tuning and may not adapt well to dynamic traffic or schedule variations.

2.1 Bus stop congestion detection

Traffic congestion detection and prediction have been extensively studied in the literature (Jilani *et al.*, 2023), with approaches ranging from traditional statistical models to advanced deep learning techniques. In contrast, bus stop congestion detection—a more localized and operationally specific problem—has received comparatively less attention. Existing work in this area has primarily focused on detecting high-density vehicle arrivals at bus stops, often leveraging Automatic Vehicle Location (AVL) data or smart card transaction records to infer congestion patterns. One common approach involves analysing dwell times, where prolonged bus stoppage may signal congestion due to increased passenger boarding, alighting, or blocked bay access (Rashidi, Ataeian and Ranjitkar, 2023). In parallel, spatio-temporal clustering methods have been used to identify persistent congestion hotspots at high-traffic stops or transfer hubs, particularly during peak hours (Li *et al.*, 2019). While these methods provide valuable insights, they are often applied to a limited number of locations or routes and may require extensive tuning or domain-specific customization. Our study differs by applying a scalable, stop-by-stop rule-based method to detect bus stop congestion across the entire city network, offering a practical framework for large-scale monitoring and spatial-temporal analysis of operational pressure points.

2.2 Bus bunching detection

Bus bunching—where two or more buses on the same route arrive at a stop in close succession—has been a long-standing challenge in transit operations. Numerous studies have explored holding control strategies to mitigate bunching, including stochastic control

policies and headway-based holding methods that aim to regularize bus intervals dynamically (Hickman, 2001). These approaches often rely on estimated or real-time headways (Daganzo, 2009) to determine whether a vehicle should be held at a stop to prevent clustering with the following bus. In recent years, the availability of high-resolution AVL (Automatic Vehicle Location) data and other automatic sensing systems has enabled the development of more sophisticated solutions. These include real-time control algorithms, proactive prediction models, and data-driven evaluations of operational policies aimed at reducing bunching (Rezazada, Nassir and Tanin, 2023). Studies have employed predictive modelling techniques (Ma *et al.*, 2021) to forecast future headways or delays and applied control decisions accordingly.

3 Preliminaries

3.1 Data description

The real-time transit data used in this study were obtained from the Translink Real-Time GTFS feed (Queensland Government, 2020), provided by the Queensland Department of Transport and Main Roads via the Queensland Government's Open Data Portal (Queensland Government, 2020). This dataset offers dynamic Automatic Vehicle Location (AVL) information for all public buses operating in South East Queensland, formatted according to the GTFS-realtime specification. The real-time data contains high-frequency positional records of buses throughout the day, with the selected attributes summarized in Table 1. In addition to the dynamic data, a static GTFS dataset is used as a reference to describe the underlying transit infrastructure, including route definitions, stop locations, and trip sequences. The selected attributes from the static dataset are presented in Table 2.

Table 1: Data attribute description of vehicle position data

Attribute	Description
Vehicle ID	A unique identifier assigned to each bus
Trip ID	An identifier of a trip for a specific route
Route ID	The bus route number or code
Timestamp	The time at which the GPS coordinate was recorded
Latitude and Longitude	The real-time geographic position of the bus
Direction	The inbound or outbound direction of the bus

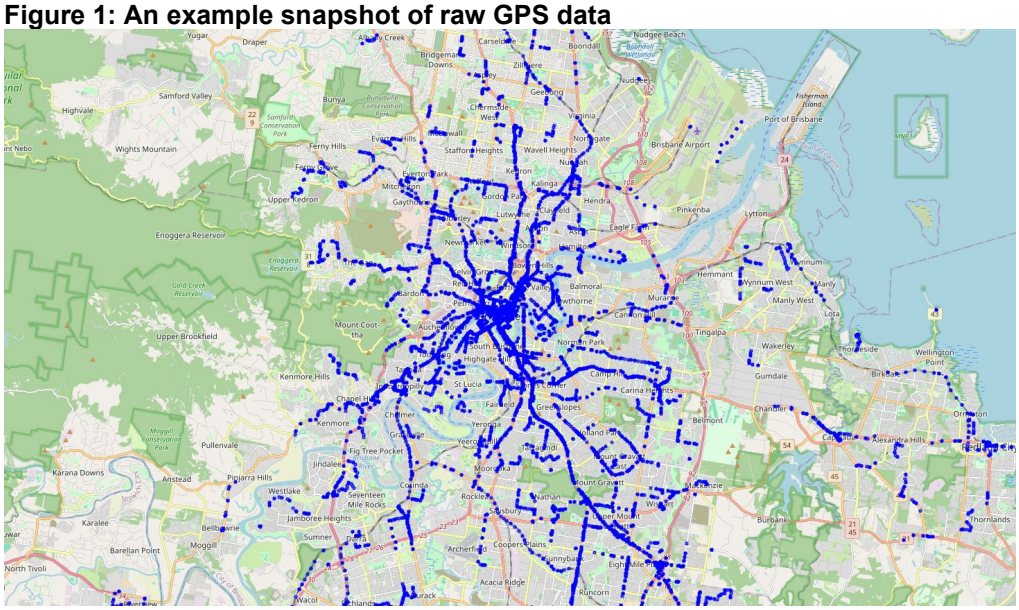
Table 2: Data attribute description of schedule data

Attribute	Description
Route ID	The bus route number or code
Stop Sequence	The sequence number of bus stop
Stop ID	The identifier assigned to each bus stop
Stop Name	The name of the bus stop
Latitude and Longitude	The location of the bus stop

The real-time GPS data does not include explicit arrival or departure events at stops. Therefore, stop-level arrival events must be inferred by spatially matching GPS points to nearby stop locations defined in the static GTFS dataset. A spatial buffer (e.g., 50 meters) around each stop is used to detect when a bus is sufficiently close to be considered as having

"arrived." Temporal sequences of these inferred stop visits are then used to detect operational anomalies. The dataset is representative of a complete operational cycle, covering both peak and off-peak hours, and enables a high-resolution reconstruction of vehicle trajectories across all bus routes in Brisbane. Figure 1 provides a visual example of the raw GPS data used in this study. The figure displays a portion of the bus location points recorded within a 3-minute time window, illustrating the spatial density and coverage of Translink's real-time AVL feed. Each blue dot corresponds to a single location record from a bus, reflecting the high-frequency nature of the data stream.

Figure 1: An example snapshot of raw GPS data



3.2 Problem statement

This study focuses on identifying two common operational issues in large-scale urban bus networks: bus stop congestion and bus bunching. Both phenomena negatively impact the reliability and efficiency of public transport services and are of critical importance to transit operators and urban planners.

1. **Bus Stop Congestion** refers to situations where multiple buses—regardless of their route—arrive at the same stop within a short time window. Such simultaneous arrivals may cause queuing at the stop, delays in boarding or alighting, and extended dwell times, especially at high traffic stops or during peak hours.
2. **Bus Bunching** occurs when buses assigned to the same route arrive at a stop with significantly shorter-than-expected headways, indicating the loss of regular service spacing. Bunching reduces service reliability, as it leads to uneven passenger loads, long gaps between buses, and passenger dissatisfaction.

The goal of this work is to use one day of citywide bus movement data to examine the frequency, spatial distribution, and temporal patterns of these two types of anomalies across Brisbane's bus network. By analysing these issues at scale, we aim to generate insights into systemic inefficiencies and inform future strategies for improving service reliability.

4 Methodology

4.1 Data preprocessing

The raw GPS data provided by Translink captures high-frequency location updates for all buses operating across Brisbane. Each record includes a vehicle identifier, route identifier, timestamp, latitude, and longitude, and direction information. However, this data is collected continuously along the vehicle trajectory and does not explicitly indicate stop-level arrival or

departure events. To enable the detection of bus stop congestion and bus bunching, it is necessary to preprocess the raw data and infer structured stop-level arrival events by integrating both the real-time trajectory data and static GTFS information.

A challenge in the raw GPS dataset is the presence of a large number of duplicate records. These typically arise when a bus is stationary or moving slowly, while the sampling rate of GPS is frequent. Therefore, the first preprocessing step involves deduplication, where, for each bus on a given day, consecutive records with the same timestamp and location are filtered out. Once duplicates are removed, the GPS records are sorted chronologically for each vehicle, ensuring the data accurately reflects the temporal sequence of bus movements throughout the day. This ordered data is then used to reconstruct the trajectory of each bus. To identify stop-level interactions, we utilize the static GTFS feed, which provides the geographic locations of all bus stops as well as the sequence of stops for each route.

For each GPS point, we determine whether it falls within a circular buffer zone around a stop location, typically defined with a radius of 50 meters. If a bus enters this buffer and remains within it for a sequence of points over a short time period, it is assumed to have visited the stop. To avoid falsely detecting multiple arrivals at the same stop due to lingering dwell time or minor location jitter, we impose a nearest distance rule: we keep only the data point where it is nearest to the bus stop as stop arrival event. Each inferred stop arrival event is then recorded with the following information: vehicle ID, route ID, stop ID, and the timestamp of the first point in the stop buffer. This results in a sequence of discrete arrival events for each stop throughout the day. Because arrival times are estimated from GPS, and not recorded directly at the vehicle-door level, these events are approximate but sufficiently accurate for service pattern analysis.

To prepare for bunching detection, the arrival events at each stop are further grouped by route ID and stop ID, and then sorted by timestamp. The headway time between consecutive arrivals of same-route buses at the same stop is then computed as the difference between their arrival timestamps. This preprocessing pipeline transforms raw, noisy trajectory data into a structured format that supports the subsequent detection of congestion (based on the volume and timing of arrivals at each stop) and bunching (based on irregularities in headways between same-route buses). The resulting data can be used directly in rule-based detection algorithms or serve as input to more advanced models in future work.

4.2 Bus stop congestion detection

The goal of this task is to identify periods when multiple buses arrive at the same stop within a short time window, potentially causing queuing, extended dwell times, and delays. This form of localized congestion often occurs at major transfer points or during peak hours, especially when stop infrastructure is limited (e.g., a single bus bay shared by many routes).

To detect such events, we use the inferred stop-level arrival events generated during preprocessing. For each stop, we aggregate all arrival events across all routes and group them by time. We define a congestion event at a stop if the number of arrivals within a specified time window exceeds a predefined threshold.

Formally, for a given stop s, let $A_s = \{t_1, t_2, \ldots, t_n\}$ be the sequence of inferred arrival timestamps over the analysis period. For each arrival time t_i, we count the number of arrivals that fall within a forward time window of length ΔT(e.g., 30 seconds):

$$C_s(t_i) = \{\, t_j \in A_s \mid t_j \in [t_i, t_i + \Delta T] \,\}$$

If $|C_s(t_i)| \geq N$, where N is the congestion threshold (e.g., 3 buses), a congestion event is recorded at stop s during the window starting at t_i.

This process is repeated for every stop in the network across the full day. By tuning the values of ΔT and N, we can adjust the sensitivity of the detection. In this study, we use fixed parameters to provide a uniform comparison across all stops. The output of this process is a list of congestion events, each represented by the stop ID, time window, number of arrivals,

and involved bus IDs. These results can be aggregated to identify hotspots (stops with frequent congestion) and peak periods (hours with recurring multi-route convergence).

4.3 Bus bunching detection

Bus bunching refers to a situation where two or more buses assigned to the same route arrive at a stop with shorter-than-expected headway intervals. This disrupts the even spacing of service and is often caused by delays propagating along a route, leading to clustered arrivals and large service gaps behind the bunched vehicles.

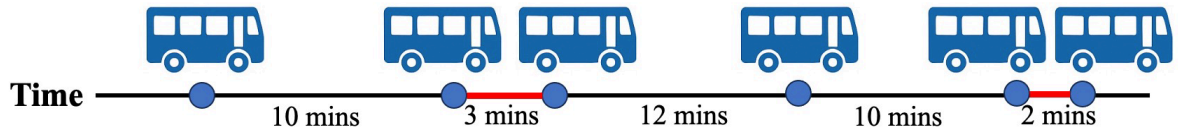
To detect bunching, we first group the inferred stop-level arrival events by route ID and stop ID, and then sort the arrivals by timestamp. For each pair of consecutive arrivals $t_i$and t_{i+1} from the same route at the same stop, we compute the headway time:

$$H = t_{i+1} - t_i$$

We define a bunching event if this headway H is significantly shorter than the scheduled or typical headway for that route and time of day. In the absence of detailed scheduled headways, we apply a fixed headway threshold θ (e.g., 3 minutes) and flag any pair where $H < \theta$.

Figure 2 provides a visual illustration of the bus bunching concept. In the timeline, each bus icon represents a vehicle serving the same route, and the time intervals between arrivals are annotated below. Ideally, headways between buses should remain consistent, such as the first two intervals of 10 and 12 minutes. However, when the time gap between consecutive buses drops significantly, such as the 3-minute and 2-minute intervals highlighted in red, it indicates a bunching event. These uneven headways typically occur when a leading bus experiences delays (e.g., longer dwell time or traffic), allowing the following bus to catch up. This results in clustered arrivals and uneven service distribution, which undermines schedule reliability and reduces service quality for passengers.

Figure 2: Example of bus bunching



To improve robustness, we exclude consecutive arrivals from the same vehicle, which may occur due to short-turning or looping, and only consider pairs of distinct vehicle IDs. The final output includes the stop ID, route ID, timestamp of each bunching event, headway value, and vehicle IDs of the involved buses. These events can be further analysed to identify routes with frequent bunching, temporal patterns, and correlation with congestion events.

5 Results analysis and visualisation

After processing one full day of real-time GPS data from all buses operating in Brisbane, we applied the rule-based detection methods described in Section 4 to identify instances of bus stop congestion and bus bunching. The congestion detection algorithm yielded a total of 3624 congestion events, representing cases where three or more buses arrived at the same stop within a 30-second time window. These events were distributed across 352 unique bus stops, with varying degrees of frequency and severity.

For bus bunching, we identified 23441 bunching events, defined as occurrences where two or more buses from the same route arrived at a stop with a headway below a fixed threshold of three minutes. These events spanned 347 different routes and 5622 stops, highlighting spatial and operational patterns across the network.

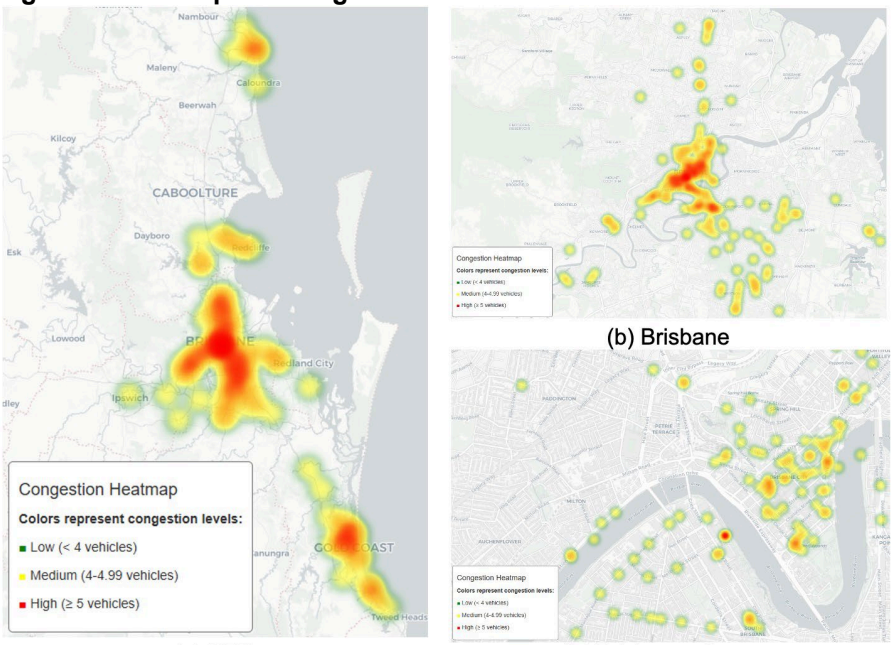
The results provide a foundation for analysing where and when service anomalies occur, which routes or stops are most affected, and how these disruptions vary throughout the day. In the following sections, we present detailed analyses and visualisations of both congestion

and bunching events, exploring their spatial distribution, temporal trends, and network-level implications.

5.1 Bus stop congestion analysis

This section presents an analysis of the congestion events identified using the 30-second window-based detection method described in Section 4.2. A congestion event is defined as the arrival of three or more buses (regardless of route) at the same stop within a 30-second interval. These events are critical indicators of local capacity stress and potential delays due to queuing or stop access conflicts. Over the course of the observation day, the average number of buses involved in each event was 1.6, with the most severe instance involving 10 buses arriving simultaneously at a major interchange location.

Figure 3: Heatmaps of congestion events



(a) SEQ

(b) Brisbane

(c) Brisbane city area

5.1.1 Spatial distribution of congestion events

To visualise the spatial intensity of bus stop congestion, we generated heatmaps by aggregating the number of congestion events recorded at each stop over the course of a full day. Each congestion event was detected based on the arrival of three or more buses at the same stop within a 30-second window. The resulting event counts were mapped geographically using a kernel density estimation method, where colour intensity represents the level of congestion: green for low, yellow for medium, and red for high concentrations. As shown in Figure 3, panel (a) presents a regional view of South East Queensland (SEQ). The most congested areas are concentrated in major city centres, including Brisbane, the Sunshine Coast, and the Gold Coast, reflecting regions with high transit demand and dense route intersections. Panel (b) provides a closer view of the Brisbane metropolitan area, where congestion hotspots appear along major corridors, highlighting the alignment of pressure points with high-frequency routes. Panel (c) zooms further into the Brisbane city centre, revealing that congestion is most intense at major intersections and interchange hubs, such as King George Square, Cultural Centre, and Roma Street, where multiple routes converge. This spatial visualisation confirms that congestion is not uniformly distributed, but rather localized around key operational bottlenecks and transfer-heavy areas.

5.1.2 Temporal distribution of congestion events

The hourly distribution of congestion events is shown in Figure 4, which combines a histogram with a smoothed density curve to better illustrate underlying trends. The results exhibit a clear bimodal pattern, with two prominent peaks occurring during the typical weekday morning and

afternoon rush hours. The highest volume of congestion is recorded around 7:00 AM, with a sharp spike exceeding 550 events, reflecting intense inbound commuting demand. A secondary peak occurs at approximately 4:00 PM, coinciding with the afternoon outbound travel period. Between these peaks, event frequency declines but remains non-negligible, suggesting moderate congestion levels persist throughout the day. From late evening into early morning (8:00 PM to 5:00 AM), the number of events drops significantly, consistent with reduced service frequency during off-peak hours. This temporal trend reinforces the notion that bus stop congestion is strongly linked to commuter demand cycles, and that operational adjustments—such as staggered scheduling, increased vehicle dispatch frequency, or infrastructure upgrades—should be prioritized during peak hours to alleviate pressure at high-demand stops.

Figure 4: Temporal distribution of congestion events

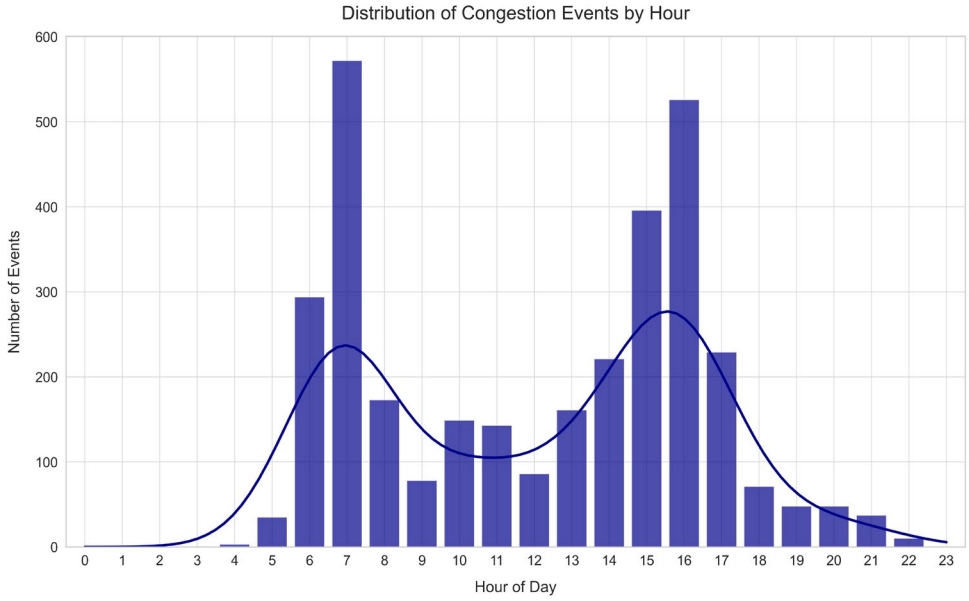
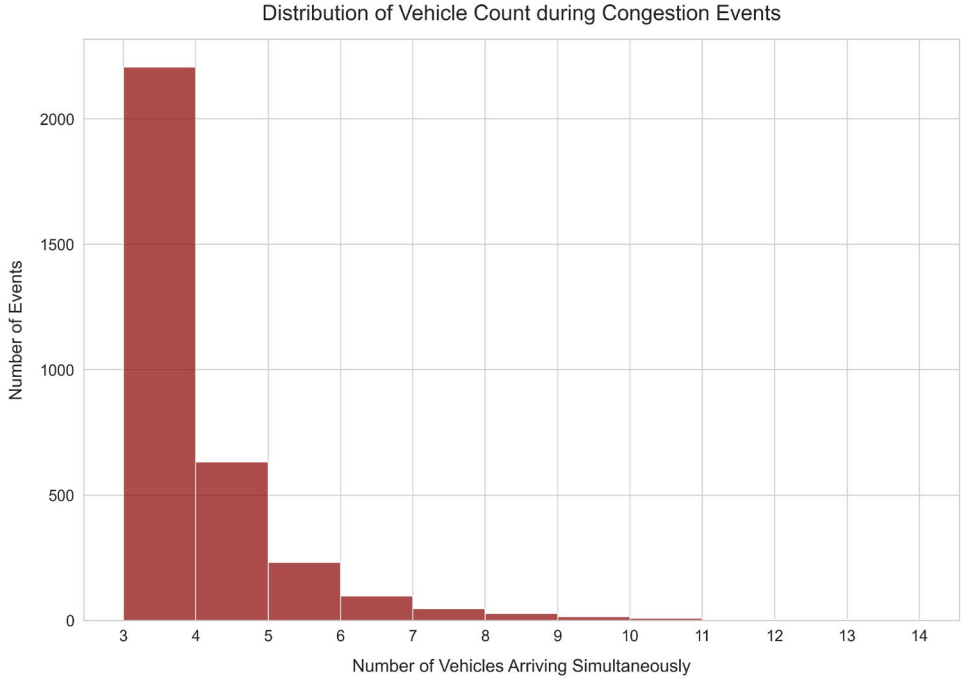


Figure 5: Histogram of number of buses per event



5.1.3 Severity of congestion events

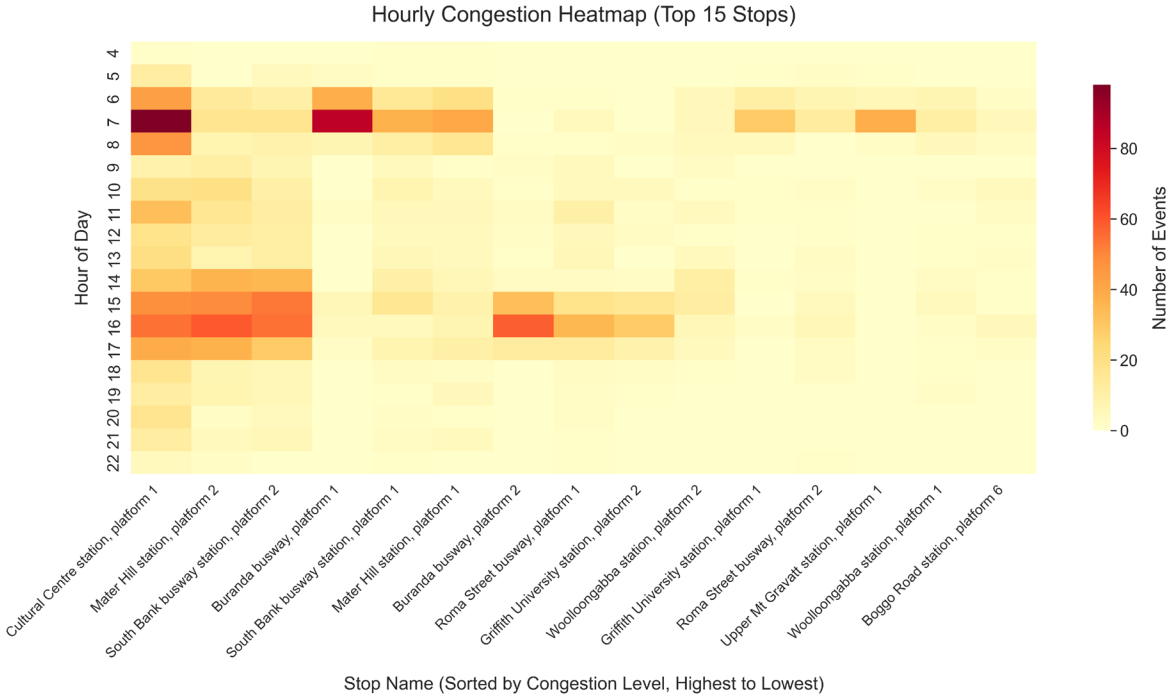
To assess the severity of bus stop congestion, we analysed how many vehicles arrived simultaneously at the same stop during each detected event. Figure 5 illustrates the

distribution of vehicle counts per congestion event. The vast majority of congestion cases involved the minimum threshold of three buses, accounting for over 2,000 events, indicating that mild congestion is the most common form encountered. A smaller number of events involved four to six buses, with a rapid decline in frequency as the number of simultaneously arriving vehicles increased. Notably, only a few extreme cases involved seven or more buses, and events with ten or more buses were exceptionally rare. This long-tail distribution suggests that while most congestion is manageable, a subset of high-severity events may place significant strain on bus bay capacity and disrupt overall schedule adherence. These findings support the need for tiered operational responses, where routine congestion can be managed with minor schedule smoothing, but severe events may require targeted interventions such as dynamic dispatching, holding strategies, or infrastructure redesign.

5.1.4 Time-Station heatmap of congestion frequency

To further explore the interplay between stop-level congestion and time of day, we generated a heatmap of the top 15 most congested stops based on total event count. As shown in Figure 6, the heatmap displays the number of congestion events by stop (sorted by overall congestion level) and hour of the day. The colour intensity corresponds to the event frequency, with darker shades indicating higher congestion levels. The results reveal strong temporal clustering of congestion at several major stops. For example, Cultural Centre Station, Platform 1 and South Bank Busway Station, Platform 2 exhibit severe congestion during the morning peak (6:00–8:00 AM), while Burnside Busway Station and Mater Hill Station show elevated congestion during both morning and afternoon peaks. Some stops, such as Griffith University Station and Roma Street Busway, show more dispersed activity with congestion spread across midday and afternoon hours. This stop-hour matrix highlights the heterogeneity in congestion timing across the network, suggesting that different stops are affected by varying commuter patterns and route overlaps. Such fine-grained temporal insights are valuable for designing targeted strategies—for instance, adjusting stop-level dispatching during specific hours or coordinating arrival times across intersecting routes.

Figure 6: Temporal distribution of top 15 congested stations



5.1.5 Route-Level participation in congestion events

To investigate which routes, contribute most frequently to congestion events, we analysed the total number of times each bus route was involved in a detected stop-level congestion scenario. Figure 7 presents the top 15 most frequently involved routes, ranked by congestion event count. Route 111 dominates the chart with over 700 event participations, significantly

higher than the next most involved routes such as 160 and 222, each with over 400 events. Many of the top-ranked routes serve major corridors and pass through high-traffic zones such as the Brisbane CBD and major interchange stations. The prevalence of these routes in congestion events indicates their high operational frequency and route overlap with other services. Understanding which routes are most involved in congestion can inform targeted scheduling strategies, including headway adjustments, dedicated bus lanes, or stop design improvements. These insights also suggest the potential benefit of coordinated control strategies between overlapping routes to alleviate crowding at critical bottlenecks.

Figure 7: Top 15 routes involved in the highest number of congestion events

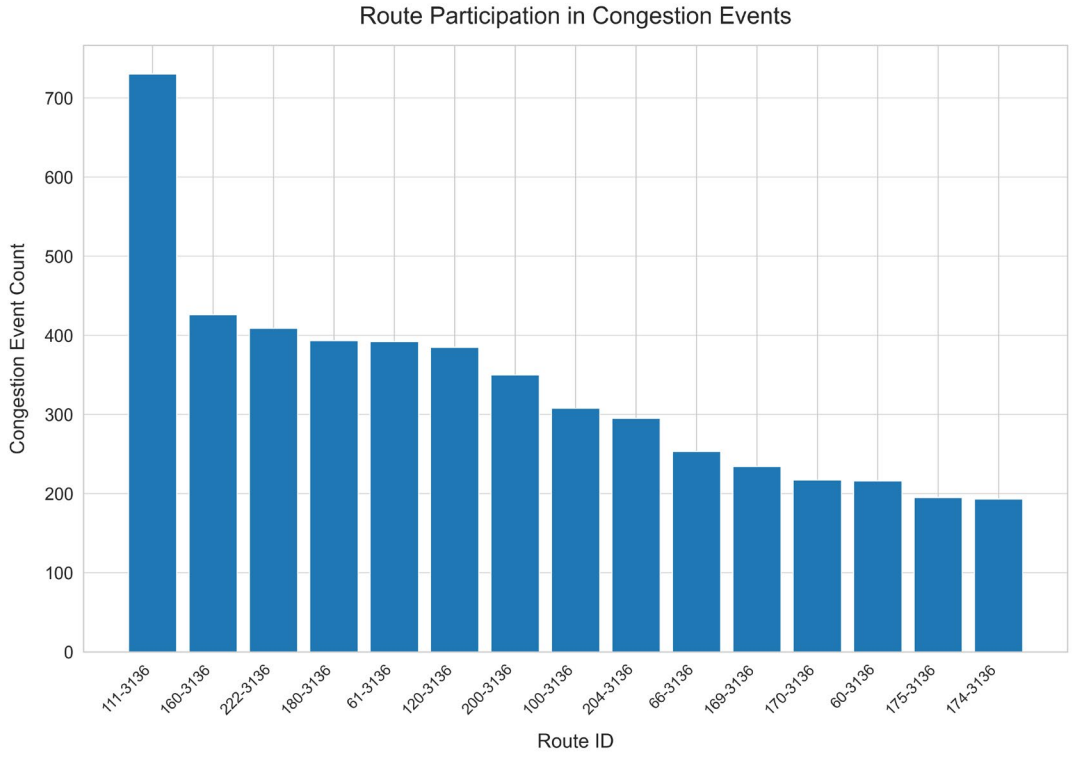
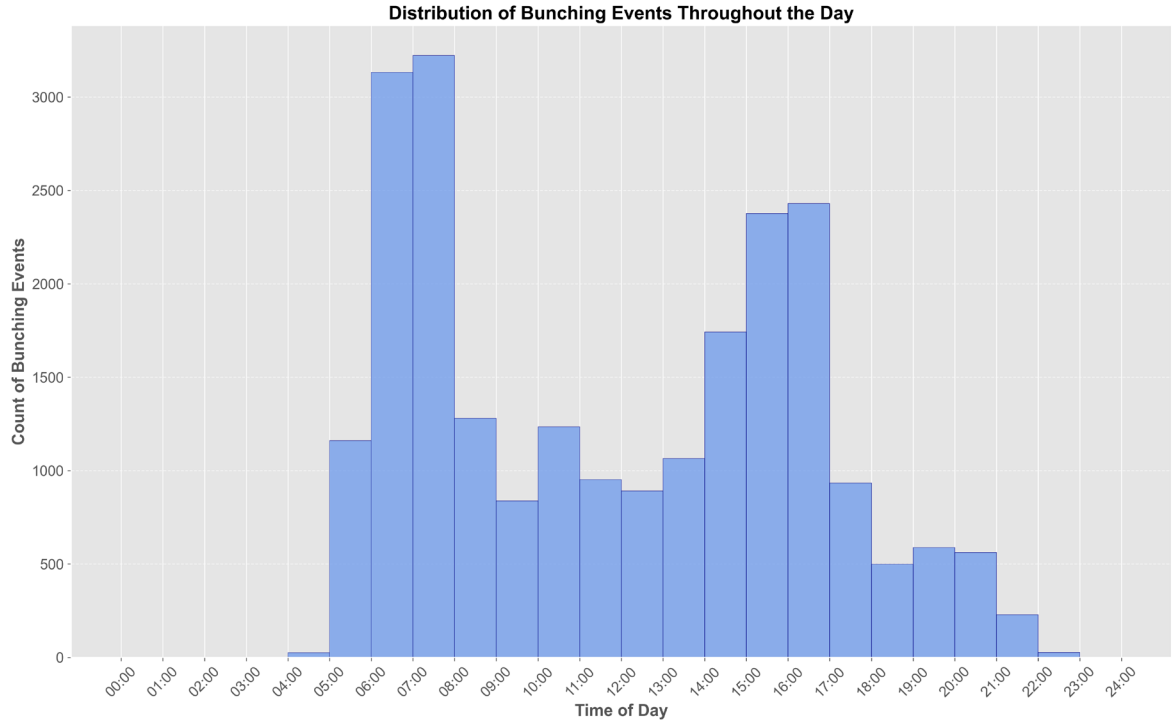


Figure 8: Temporal distribution of bus bunching events



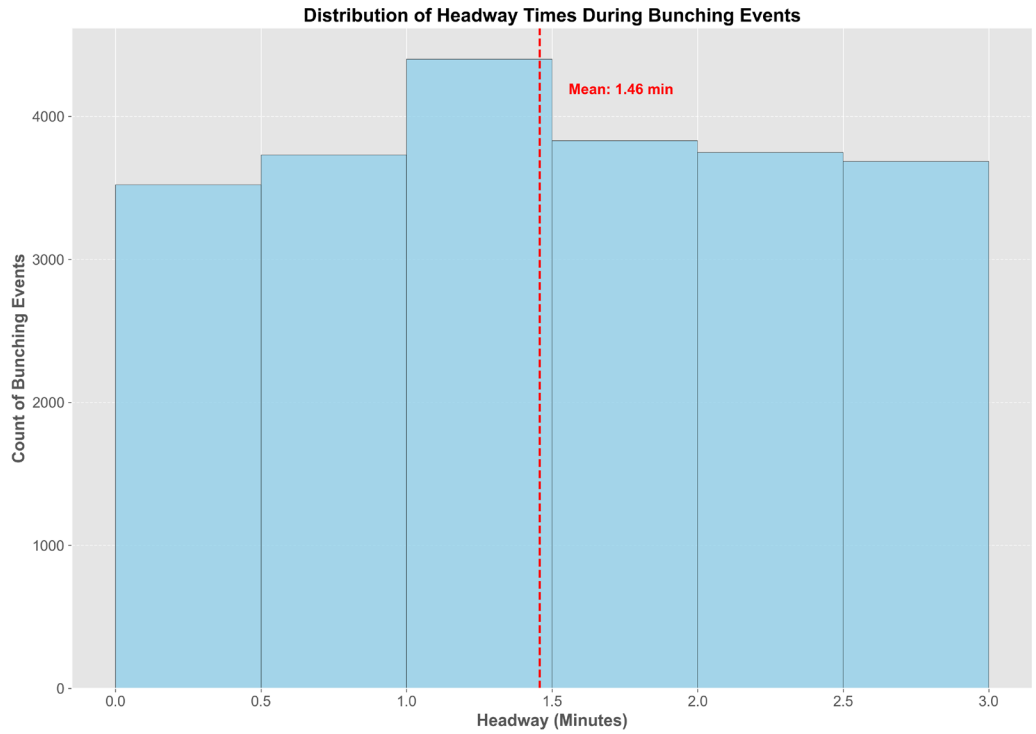
5.2 Bus bunching analysis

In this section, we examine the occurrence of bus bunching—an operational issue where buses from the same route arrive at a stop in close succession, reducing service regularity and passenger satisfaction. Given that scheduled headways in the Brisbane network typically range between 10 to 15 minutes, we define a bunching event as any instance where the headway between two consecutive buses on the same route at the same stop is less than 3 minutes. This threshold captures significant deviations from expected spacing and highlights cases where the following bus is likely catching up due to delays affecting the lead bus. This definition enables a systematic, rule-based detection of bunching patterns across the network and lays the foundation for identifying affected routes, times of day, and high-risk stations.

5.2.1 Temporal distribution of bus bunching events

Figure 8 presents the temporal distribution of bus bunching events detected across the day. A clear bimodal pattern emerges, with sharp peaks during the morning (6:00–8:00 AM) and afternoon (3:00–5:00 PM) rush hours, aligning with the periods of highest passenger demand and vehicle frequency. The most pronounced spike occurs between 7:00 and 8:00 AM, where the number of bunching incidents exceeds 3,000, suggesting intense operational stress in the early morning commute period. Between these peaks, bunching events persist at a moderate level, reflecting variability in service reliability even during off-peak hours. In the evening, events taper off but remain non-negligible until around 9:00 PM, likely due to schedule recovery limitations or late-day demand surges. This distribution highlights the susceptibility of the network to service regularity breakdowns during high-frequency periods, reinforcing the need for dynamic headway management and real-time intervention strategies during peak demand windows.

Figure 9: Histogram of headway per event



5.2.2 Headway characteristics of bunching events

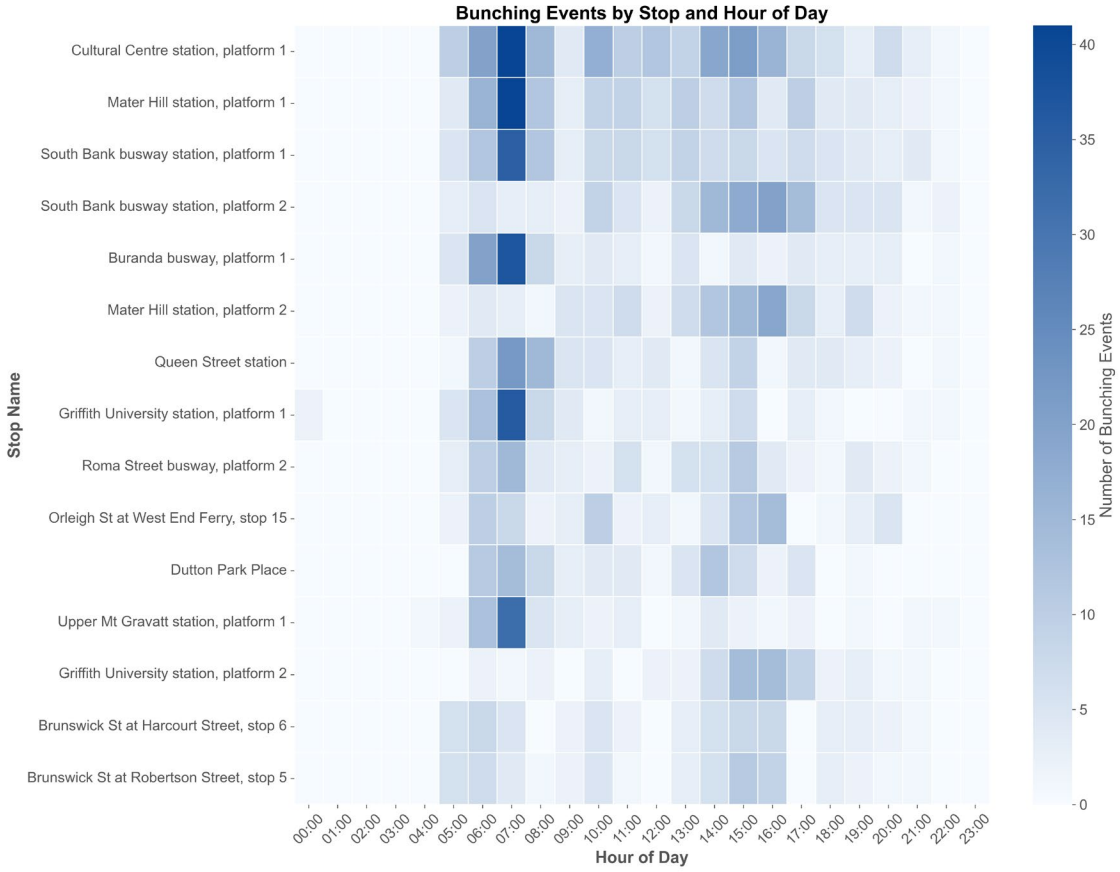
To further characterize the severity of detected bunching incidents, we analysed the actual headway durations—i.e., the time intervals between consecutive same-route buses arriving at the same stop—for all identified bunching events. Figure 9 illustrates the distribution of these headway values. The distribution is fairly uniform across the sub-3-minute range, with a mean headway of approximately 1.46 minutes, as marked by the red dashed line. This indicates that in most bunching events, the following bus arrives within 1 to 2 minutes of the preceding one,

which is significantly shorter than the scheduled spacing of 10–15 minutes. While some headways are close to zero—reflecting near-simultaneous arrivals—a large volume of events clusters around the 1 to 2-minute range, underscoring the operational inefficiencies and uneven passenger loads these events may introduce. The consistency of low headway intervals across such a large number of events reinforces the need for real-time spacing strategies, such as adaptive holding or headway-based control, to maintain service regularity and reduce cascading delays.

5.2.3 Time-Station heatmap of bus bunching frequency

Figure 10 illustrates the distribution of bus bunching events across the top 15 affected stops, broken down by hour of day. The heatmap reveals that Cultural Centre Station (platform 1), Mater Hill Station, and South Bank Busway Station are among the most frequently affected locations, showing persistent bunching activity throughout both morning and afternoon peak periods. Most stops experience heightened bunching between 6:00 and 9:00 AM and again between 3:00 and 5:00 PM, aligning with typical commuter flows. Notably, some stops such as Griffith University Station and Roma Street Busway also exhibit midday bunching, suggesting broader schedule disruption or high-frequency service routes during off-peak periods. These patterns underscore the spatial concentration of reliability issues at major transit hubs and high-demand corridors. This highlights the importance of targeted mitigation strategies—such as stop-level bus holding, vehicle spacing control, or dispatch adjustments—at these chronically impacted locations.

Figure 10: Temporal distribution of top 15 affected stations



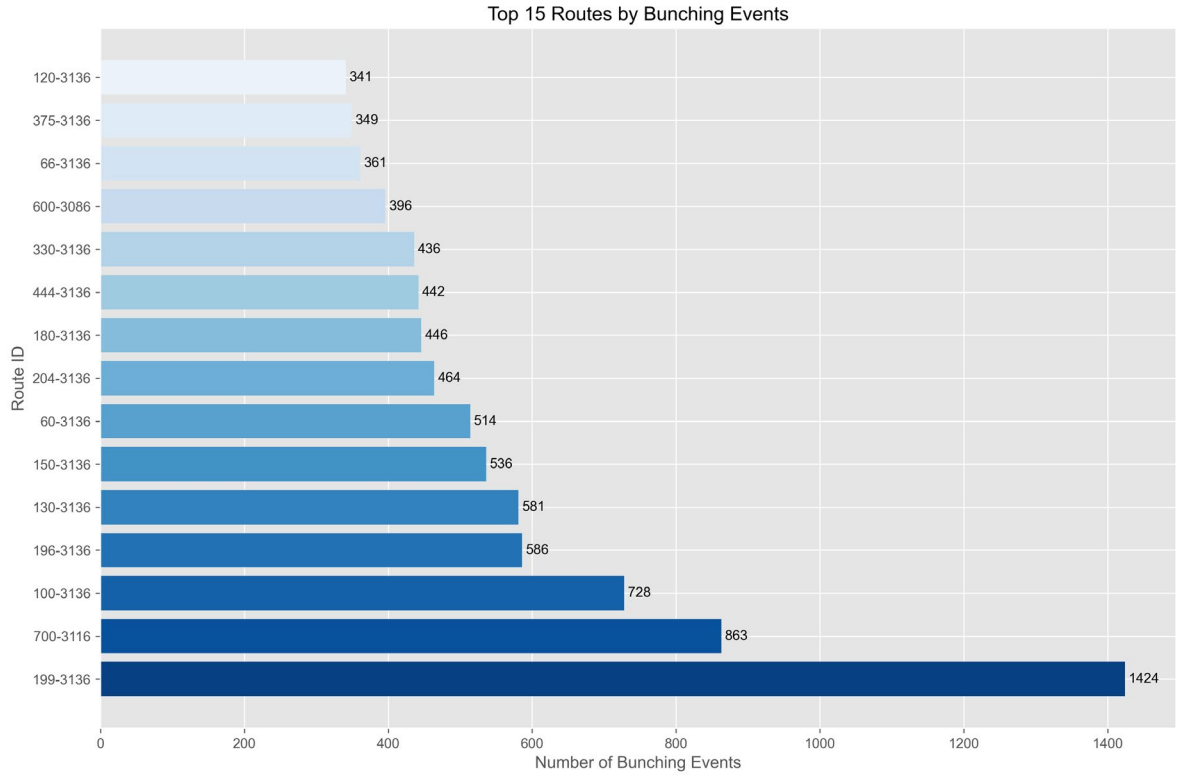
5.2.4 Route-Level participation in congestion events

Figure 11 presents the top 15 bus routes ranked by the number of bunching events observed throughout the study day. Route 199 exhibited the highest number of bunching occurrences (over 1,400), followed by routes 700 and 100. These routes not only show high event counts but also feature a substantial number of daily trips and stops, as detailed in the accompanying table. For example, route 700 had 114 trips and 53 stops, suggesting that high service

frequency and route length may contribute to the likelihood of bunching. Conversely, some routes with fewer trips, such as route 204-3136 (22 trips), still appear on the list, implying that operational factors like traffic conditions or scheduling may also influence bunching. This route-level analysis highlights candidates for targeted investigation or schedule optimisation to mitigate reliability issues.

Further investigation of Route 199, which recorded the highest number of bunching events, reveals contributing operational characteristics. As illustrated in Figure 12, Route 199 spans a densely served corridor with a high concentration of stops distributed closely along the inner Brisbane area, including South Bank, West End, and New Farm. The short inter-stop distances may lead to frequent stopping and acceleration/deceleration cycles, increasing the likelihood of delay propagation and ultimately bunching. This spatial configuration, combined with relatively high trip frequency (72 trips per day), makes Route 199 particularly susceptible to bunching issues and highlights the importance of considering route geometry and stop density in operational diagnostics. These findings suggest that trip frequency planning should be carefully calibrated based not only on demand but also on the spatial layout of stops; excessive frequency on closely spaced routes may exacerbate inefficiencies. Incorporating inter-stop distances and total stop counts into scheduling decisions could help prevent over-saturation and ensure more reliable public transport service.

Figure 11: Top 15 routes involved in the highest number of bus bunching events



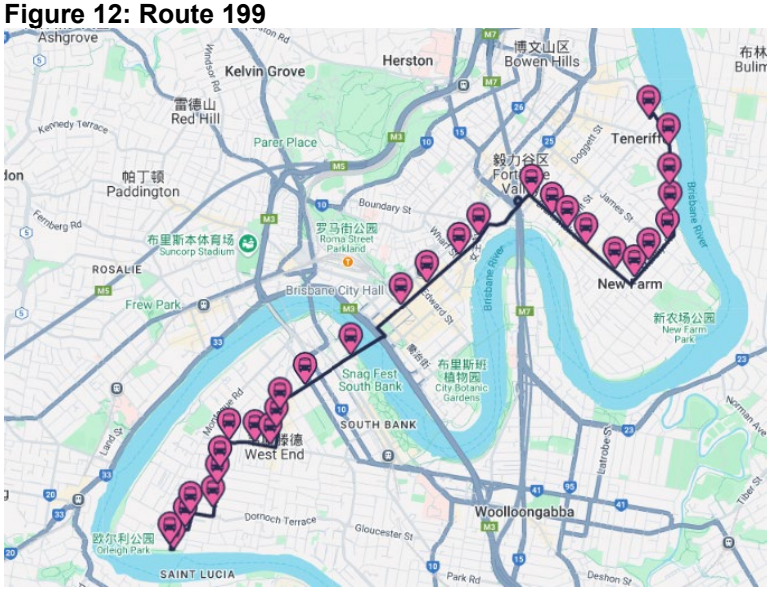
6 Discussion

6.1 Summary of key findings

This study conducted an exploratory analysis of bus stop congestion and bus bunching using high-frequency AVL data from Brisbane's Translink GTFS-real-time feed. Our analysis was designed to detect two types of events: congestion events, where multiple buses arrive at a stop within a short temporal window, and bunching events, where consecutive buses on the same route arrive with unusually short headways. Using a 30-second threshold for congestion and a 3-minute threshold for bunching, we were able to detect and visualise these events across time, space, stop, and route dimensions. The spatial heatmaps revealed that

congestion events are particularly dense in key urban centres including Brisbane CBD, the Sunshine Coast, and the Gold Coast. At the stop level, high-frequency congestion occurred consistently at key busway stations such as Cultural Centre, South Bank, and Mater Hill. Temporally, congestion and bunching followed the expected daily commuting rhythm, peaking during morning (7:00–9:00) and afternoon (16:00–18:00) rush hours. Furthermore, we found that some specific routes—such as Route 199—suffered from frequent bunching. Our visual inspection of this route's stop layout confirmed a high stop density with many closely spaced stops, which may contribute to bus delay propagation and vehicle clustering.

Figure 12: Route 199



These findings have meaningful implications for transit planning and operations in Brisbane. The identified hotspots can guide planners in prioritising targeted interventions, such as adding additional layover space at highly congested stops, staggering bus departure times for overlapping routes, or revising stop spacing on routes prone to bunching. For example, dynamic headway control or bus holding strategies could be piloted on high-frequency routes like 199 to mitigate cascading delays. Moreover, the temporal patterns of congestion support time-of-day scheduling adjustments, such as increasing off-peak spacing or deploying additional resources during peak congestion windows. The spatial-temporal outputs from this study can also serve as a decision-support tool for transit agencies evaluating the effectiveness of operational changes over time.

Beyond operational planning, the study also offers insights for policy-making and economic evaluation. Persistent bus bunching and stop congestion can lead to missed connections, increased operating costs due to schedule disruptions, and reduced passenger satisfaction—all of which undermine the efficiency of public transport investment. By quantifying the spatial and temporal extent of these inefficiencies, our results can support cost-benefit analyses for infrastructure upgrades, priority lane implementation, or real-time control technologies. Integrating such data-driven evidence into transport policy can help align resource allocation with the most critical areas of network pressure, ultimately contributing to more reliable and economically sustainable transit systems.

6.2 Limitations

While the results offer important insights, several limitations should be addressed in future work. Data quality is a notable concern. The real-time GPS data suffer from redundancy, missing values, and timestamp noise, which required considerable preprocessing. Although we removed duplicate records and applied rule-based filters, some subtle inconsistencies may still remain, potentially affecting the accuracy of event detection. Threshold-based rule simplicity represents another limitation. Our use of fixed thresholds (30 seconds for congestion, 3 minutes for bunching) provides interpretability but lacks flexibility. These values may not

capture variability across routes with different service frequencies, vehicle types, or operational characteristics. A one-size-fits-all threshold may lead to under- or over-identification of events in some cases. Scope and scale are also restricted. Our current analysis is based on one day of data. While this allows for a clear case study, it limits the generalizability of conclusions. More robust insights would require analysis across multiple days, weeks, or months, incorporating variability due to weather, special events, and service changes. Technical depth of the analysis is preliminary. While our rule-based approach is simple and scalable, it may not capture complex, latent patterns in space-time mobility data. More advanced modelling approaches could enhance detection accuracy and provide predictive capabilities.

6.2 Future work

There are several directions in which this work can be extended to further enhance its value and applicability. First, expanding the analysis to a multi-day or longitudinal framework would allow for a more comprehensive understanding of recurring congestion and bunching patterns. By incorporating data from multiple days or weeks, researchers could examine variability across weekdays, weekends, and different seasons, which would also facilitate more robust statistical analyses and trend evaluations.

Second, integrating the real-time AVL data with static scheduled data from GTFS could provide deeper insights into the causes of congestion and bunching. This would enable the comparison of actual bus arrivals against scheduled times, making it possible to quantify delays, assess schedule adherence, and detect anomalies that deviate from planned operations. Such comparisons can further inform the classification of events as systemic or exceptional.

Third, the current approach uses a fixed threshold (e.g., a 3-minute headway to define bunching), which may not capture the nuanced differences across routes and times of day. Future work could explore dynamic or adaptive thresholding strategies that account for route-specific characteristics, expected dwell times, or empirical distributions of headways. This refinement would improve the accuracy and generalizability of event detection.

Finally, an important extension of this work lies in predictive modelling. Leveraging both historical and real-time data, machine learning models could be developed to forecast the likelihood of congestion or bunching events before they occur. These models may include time-series forecasting techniques, decision-tree-based models, or deep learning architectures tailored for spatio-temporal prediction. A predictive framework would allow for proactive mitigation strategies and ultimately support smarter transit operations.

References

Daganzo, C.F. (2009) ‘A headway-based approach to eliminate bus bunching: Systematic analysis and comparisons’, *Transportation Research Part B: Methodological*, 43(10), pp. 913–921.

Ganguly, S. and Maitra, B. (2024) ‘Bus Bunching at Bus Stops and Its Spatiotemporal Patterns along Urban Bus Routes’, *Journal of Urban Planning and Development*, 150(3), p. 04024017.

Gangwani, D. and Gangwani, P. (2021) ‘Applications of machine learning and artificial intelligence in intelligent transportation system: A review’, *Applications of Artificial Intelligence and Machine Learning: Select Proceedings of ICAAAIML 2020*, pp. 203–216.

Gibson, J., Baeza, I. and Willumsen, L. (1989) ‘Bus-stops, congestion and congested bus-stops’, *Traffic engineering & control*, 30(6).

Hickman, M.D. (2001) ‘An analytic stochastic model for the transit vehicle holding problem’, *Transportation Science*, 35(3), pp. 215–237.

Jilani, U. *et al.* (2023) ‘A systematic review on urban road traffic congestion’, *Wireless Personal Communications*, pp. 1–29.

Kong, X. *et al.* (2024) ‘Mobile trajectory anomaly detection: Taxonomy, methodology, challenges, and directions’, *IEEE Internet of Things Journal* [Preprint].

Li, Z. *et al.* (2019) 'Relieve the congestion by shuttle bus in rush hours using aggregation clustering algorithm on group travel pattern', *Concurrency and Computation: Practice and Experience*, 31(9), p. e4847.
Ma, Q. *et al.* (2021) 'Robust optimal predictive control for real-time bus regulation strategy with passenger demand uncertainties in urban rapid transit', *Transportation Research Part C: Emerging Technologies*, 127, p. 103086.
Mignone, P., Malerba, D. and Ceci, M. (2022) 'Anomaly detection for public transport and air pollution analysis', in *2022 IEEE International Conference on Big Data (Big Data)*. IEEE, pp. 2867–2874.
Moreira-Matias, L. *et al.* (2012) 'Bus bunching detection by mining sequences of headway deviations', in *Advances in Data Mining. Applications and Theoretical Aspects: 12th Industrial Conference, ICDM 2012, Berlin, Germany, July 13-20, 2012. Proceedings 12*. Springer, pp. 77–91.
Queensland Government (2020) *Translink Real-Time GTFS Data Feed, Queensland Open Data Portal. [Online]*. corporateName=The State of Queensland; jurisdiction=Queensland. Available at: https://www.data.qld.gov.au/dataset/translink-real-time-data/resource/d92375b5-a291-49cc-aae8-4a7180d7984f (Accessed: 13 May 2025).
Rashidi, S., Ataeian, S. and Ranjitkar, P. (2023) 'Estimating bus dwell time: A review of the literature', *Transport Reviews*, 43(1), pp. 32–61.
Rezazada, M., Nassir, N. and Tanin, E. (2023) 'What influences passenger's arrival rate at stops in Melbourne: Wasea-Lstm, a novel deep learning model leveraging multi-source data fusion', in *Australasian Transport Research Forum (ATRF), 44th, 2023, Perth, Western Australia, Australia*.
Wrona, P., Grzenda, M. and Luckner, M. (2022) 'Streaming detection of significant delay changes in public transport systems', in *International Conference on Computational Science*. Springer, pp. 486–499.