Physica A 651 (2024) 129864

Contents lists available at ScienceDirect

Physica A: Statistical Mechanics and its Applications

journal homepage: www.elsevier.com/locate/physa

Oversaturated intersections: A real-world assessment of polynomial fluid queue models

Alisa Doll [a], Mohammad Abbasi [a,*], Ming Zhao [b], Xuesong (Simon) Zhou [a]

[a] School of Sustainable Engineering and the Built Environment, Arizona State University, Tempe, AZ 85281, United States
[b] School of Computing and Augmented Intelligence, Arizona State University, Tempe, AZ 85281, United States

ARTICLE INFO

Keywords:
Queuing theory
Signalized intersections
Oversaturation
Polynomial Fluid Queue
Model Order Reduction
Time-dependent arrival rates
Queue lengths
Congestion patterns

ABSTRACT

This paper investigates the challenges of quantifying stochasticity and dynamics in traffic states, crucial to efficient transportation systems, particularly during oversaturated conditions at signalized intersections. We build upon Newell's polynomial fluid queuing-theoretic ordinary differential equation (ODE) model as an example of model order reduction, with a focus on time-dependent stochastic arrival rates and departure rates. These assumptions facilitate dynamic arrival rate approximation using polynomial functions, which, when integrated with queue lengths, yield total delay per episode. Validation of Newell's polynomial fluid-queue based dynamic patterns is sought by analyzing cumulative cycles during peak periods, segmented based on residual queue values. A comprehensive real-world experiment employs emerging high resolution video detector cameras to capture various queue lengths, revealing recurring congestion patterns. Real-time arrival and departure count during peak periods further allow for the creation of cumulative curves, enhancing our understanding of congestion dynamics at signalized intersections under partially oversaturated conditions. This foundational understanding is essential for applying deep reinforcement learning in such environments, where the goal is to observe environment states to maximize expected long-term reward.

1. Introduction

1.1. Overview

Stakeholders in the transportation sector are exploring a range of strategies to tackle the persistent challenge of escalating traffic volumes and its resulting congestion problems by investing more in public transit and promoting alternative modes of transportation, repairing and increasing the capacity of existing roadways, and implementing intelligent transportation systems.

The evaluation of signalized intersection performance is widely acknowledged to depend significantly on the utilization of queue estimation and traffic delay analysis methodologies. Nevertheless, delay remains a challenging parameter to ascertain accurately. For instance, Dion et al. [1] noted that analytical delay models, such as deterministic queuing, shock wave, steady-state stochastic, and time-dependent stochastic delay models, yield comparable results for signalized intersections experiencing low traffic demand, whereas notable differences increase as the traffic demand approaches saturation. Many researchers in this domain have primarily

* Correspodning author.
E-mail addresses: abonz@asu.edu (A. Doll), mohammad_abbasi@asu.edu (M. Abbasi), mingzhao@asu.edu (M. Zhao), xzhou74@asu.edu (X.(S. Zhou).

https://doi.org/10.1016/j.physa.2024.129864
Received 25 November 2023; Received in revised form 14 May 2024; Accepted 30 May 2024
Available online 18 July 2024

focused on the expectation value of queues, while giving limited attention to the uncertainty surrounding queue evolution dynamics, including queue propagation and dissipation, and its consequent implications. Furthermore, it is worth noting that a multitude of queuing models have been established, predicated on the assumption of stationary Poisson arrival processes [2]. In such instances, it is customary to presume that the expected queue lengths at underutilized traffic signals attain finite values during steady-state conditions. However, real-world queue behavior is considerably more intricate, particularly when the average number of arrivals approaches and exceeds the signal service rate. Under such circumstances, queue dynamics become highly sensitive to even minor changes in traffic conditions, rendering them considerably unpredictable.

Efficient transportation systems require a profound understanding of the randomness and unpredictability (stochasticity) and temporal fluctuations (dynamics) in traffic states, especially when dealing with episodes of oversaturated conditions at signalized intersections. However, there is a significant lack of reliable metrics to quantify these phenomena, particularly in the context of oversaturation. As such, the development of robust, practical techniques to measure and understand the intricate impacts of these elements on traffic flow dynamics and variations is critical. Moreover, it is crucial to appreciate the relationship between local disturbances—such as short-term inflows from upstream intersections or platoons—and the broader oversaturated state. In this regard, the roles of local and global variability are key: local variability refers to equilibrium fluctuations within a given area, while global variability encompasses the entire state-space of the system. Addressing the challenges associated with measuring stochasticity and dynamics in traffic states, this paper emphasizes the importance of accounting for both local and global variability in the effective evaluation of transportation systems operating under episodes of oversaturated conditions.

1.2. Literature review

The process of queue formulation and dissipation at signalized intersections, along with its impact on traffic delay, can be effectively studied using queuing theory and summarized as an M/D/n/FIFO system where vehicles arrive randomly (M) and leave at a constant rate (D) in multiple lanes (n), following the first in-first out (FIFO) rule [2]. Various delay estimation models based on queuing theory have been proposed, including those by Webster [3], Akcelik [4], Zhou et al. [5] and the Highway Capacity Manual (HCM) [6], Additionally, studies focusing on traffic flow dynamics and stochastic transitions also offer valuable insights [7,8].

Within available literature, methods for estimating queue lengths can be categorically divided into two principal classifications: input-output models [1,9-13] and shockwave models [14-19]. Input-output models derive queue length estimates by leveraging the cumulative traffic input-output relationship observed at a given link, yielding straightforward conceptual approximations. Nonetheless, they may exhibit constraints when attempting to account for the intricate spatial queuing intricacies characteristic of real-world traffic scenarios. Conversely, shockwave models, founded on the principles of traffic shockwave theory, have garnered notable attention for their capacity to elucidate the genesis and dissipation of queues [14,16]. They offer a more dynamic approach to queue length estimation by considering the propagation of traffic shockwaves, providing valuable insights into queue behavior in diverse traffic scenarios.

Recent studies explore the impact of intelligent transport systems (ITS) and vehicle-to-everything (V2X) technology on traffic management strategies [20-24]. Tang et al. [23] proposed a speed guidance strategy for multiple signalized intersections based on car-following models, aiming to reduce travel time, fuel consumption, emissions, and enhance traffic safety and operational efficiency. Similarly, a speed guidance model was also developed that specifically studies driving behaviors at signalized intersections to effectively address queue evolution process and ultimately reduce delay [23]. Thus, introducing ITS has enabled the exploration of major breakdowns induced by traffic jams at intersections in coupled-cycle urban-scale networks. More specifically, Nagatani [20] investigates the dynamic transition induced by route choice in two-route traffic networks with onramps. The study highlights the pivotal role of real-time traffic information in influencing route choice and the impact of bottlenecks and signals on traffic flow dynamics, which underscores the significance of information dissemination and queue evolution process for traffic behavior and management in urban settings [20-22].

Additionally, Zhu et al. [24] investigate the braking process of vehicles at signalized intersections based on car-following theory, emphasizing the significance of understanding microscopic behavior of traffic flow in traffic jams and bottlenecks under signal influence. Meanwhile, Jiang and Wu [25] explore the impact of traffic signals on vehicular flow using a stopped time-dependent randomization cellular automata model, shedding light on traffic flow control methodologies.

Despite the dynamic nature of traffic during significant breakdowns, fluid dynamics have demonstrated their ability to offer a continuous representation of the queue system within traffic breakdown scenarios. In 1965, Newell [26] introduced a more comprehensive algebraic framework for estimating intersection delays, taking into account the cumulative arrivals and departures over a specified time interval. Nevertheless, several of these early models leaned heavily on the assumption of steady-state conditions, thereby constraining their utility in situations where traffic flow surpassed its capacity [27]. Additionally, this class of models have difficulties in capturing the transition between under-saturation and oversaturation conditions [28].

During the 1970s, studies introduced the coordinate transformation method to address oversaturation at signalized intersections. This technique involved repositioning the steady-state curve in such a manner that it asymptotically approached the deterministic oversaturation delay as outlined by May and Keller [29]. Kimber and Hollis [30] augmented this model by incorporating an initial queue length parameter, thereby further enhancing its predictive accuracy.

Given the frequent non-stationary nature of traffic flow, particularly during peak periods, researchers advanced heuristic formulations and time-dependent models to account for traffic conditions near capacity [31-35]. These formulations introduced complexities such as time-varying overflow queue lengths and addressed the temporal variability of traffic arrivals [36-39].

Within congested freeway networks, demand momentarily surpasses the available supply in both space and time, resulting in

oversaturated traffic conditions. Vickrey's bottleneck model describes the complexities of traffic dynamics during peak periods, with subsequent research studies centering on network equilibrium and the optimal toll problems. However, empirical observations have demonstrated complex nonlinear traffic flow patterns, which challenge the simplistic step function assumptions inherent in these models. Newell's fluid-based queue models [40-43] have been extended to consider time-dependent arrival rates. However, the dependence on linear and quadratic functions has imposed limitations on their ability to comprehensively capture the nonlinear dynamics inherent in traffic flow.

When extending from one-dimensional queue models to two-dimensional traffic problems, various spatial and temporal discretization schemes can be applied. In recent studies, Zhao et al. [44,45] have explored simultaneous longitudinal and lateral vehicle control within intersections, providing valuable insights in predicting traffic operations for assessing roadway designs' efficiency and safety. The studies introduce an integrated microscopic traffic flow model that considers human-driven vehicle maneuvers under interactions, optimizing costs to minimize undesirable situations. Fundamentally, as Newell's work indicates, fluid queue models are extendable to Cell Transmission Models (CTM) and Cellular Automata (CA(m)), a concept further developed by Daganzo [46].

Furthermore, recent studies on the multi-resolution network detailed by Hadi et al. [47] and Lu and Zhou [48], facilitate the integration of external data into road driving conditions through a CA model, enabling sophisticated maneuver and platoon simulations. The CA(M) model, introduced by Daganzo [46], enhances traffic modeling by incorporating lane changes and routing within a comprehensive graph, augmented by data fusion from sensors and cameras for a detailed, real-time traffic analysis.

To address time-dependent queues, one strategy involves segmenting the period into shorter intervals and applying steady-state queuing formulas to each. This approach, however, has its challenges. It may not adequately capture oversaturation during peak times and can overlook the residual queues from earlier intervals. Computer simulation-based methods, like Monte Carlo simulations, are frequently employed to capture these time-dependent queue dynamics, particularly in oversaturated conditions. Yet, achieving a representative average system performance typically requires numerous simulations runs, each using different random number seeds. This necessity complicates the assimilation of these simulation methods into mathematical programming frameworks that prioritize tractable analytical functions. Continued research and refinement of queueing models will help us better understand and manage traffic congestion in various real-world scenarios.

In this study, we propose an extension of Newell's classic fluid-based queue model using polynomial-function-approximated arrival rates for episodes of oversaturated traffic systems at signalized intersections. This extension aims to enhance the estimation of queue profiles, including time-dependent queue lengths, delays, average delays, and travel times for system performance evaluation.

1.3. Traffic signal performance measures

Performance monitoring for arterial traffic control and management systems is gaining prominence in the United States, emphasizing the critical need for routine traffic data collection and system monitoring. This focus aims to assess performance, identify operational issues, and optimize signal control to enhance system service, enabling transportation agencies to improve traffic control and service levels effectively [49]. The Second Edition Signal Timing Manual (STM2) serves as a comprehensive guidebook designed to assist transportation agencies in efficiently managing and optimizing traffic signal systems [50]. It offers a systematic approach to signal timing that aligns with modern traffic management practices and technology.

While the operational objectives, mitigation strategies, and performance measures outlined in the STM2 process are effective in managing traffic under unsaturated conditions, they face challenges in oversaturated traffic systems due to their complexity and differing operational objectives focused on optimizing system throughput and mitigating queue length and effects [51]. When oversaturated conditions occur, performance evaluation typically revolve around three periods: the loading period, the oversaturated period, and the recovery period [50].

This heightened focus on signal optimization underscores the critical need for effective methodologies to assess and enhance traffic signal performance [52]. One pioneering solution leading this effort is the implementation of Automated Traffic Signal Performance Measures (ATSPMs). These advanced systems are fundamentally reshaping traditional signal timing and operational practices by seamlessly integrating high-resolution data-logging capabilities into existing traffic signal infrastructure. They also leverage cutting-edge data analysis techniques, providing a cost-effective means of accessing essential information for proactive signal performance management [53-55].

1.4. Modeling opportunities from emerging data sources

The emergence of ATSPMs underscores the growing demand for improved data collection and processing, recognizing the pivotal role of emerging data sources in enhancing traffic simulation and calibration when implementing Performance Measures. Transportation agencies now have a multitude of options for data collection and processing, encompassing devices, storage solutions, and software, offered by both vendors and open-source platforms [56].

This shift towards advanced data sources is integrated with established technologies like loop detectors, radar, LiDAR sensors, GPS, and video cameras, collectively providing valuable insights and opportunities to enhance traffic simulation calibration [57]. This diverse array of sources encompasses a wide spectrum of traffic data sets, including probe vehicles, High-Resolution Vehicle Trajectories, Connected and Autonomous Vehicles (CAVs), and Crowdsourcing. To further enhance traffic mobility and safety, transportation agencies require precise traffic data, and as advanced data collection technologies become increasingly available, professionals in the field must select the most suitable technology for their agency's needs [58,59]. As discussed in Table 1, these advancements are gradually supplanting traditional data collection methods, such as road tube, inductive loop, and piezo systems, as

manufacturers adapt to the evolving demands of traffic data collection [60].

The assortment of surveillance techniques detailed in Table 1 has improved our understanding of traffic dynamics in over-congested systems. Techniques such as Automatic Vehicle Identification and Mobile GPS location sensors not only facilitate real-time monitoring but also assist in predicting and managing congestion by providing granular data on traffic flow and vehicle behaviors. The trajectory data from video image processing, as implemented in this study, enables us to dissect complex traffic patterns and understand the stochastic nature of queues in real-time, a crucial element for accurate traffic simulation and the development of effective traffic management strategies. Continuous technological advancements in this domain present exciting opportunities to further enrich traffic modeling, necessitating ongoing assessment to leverage the full potential of these emerging data sources.

Continuous evaluation of these innovations is essential to explore potentially more efficient alternatives [58,59]. The application of these new technologies for collecting a wide array of traffic data has led to the development of publicly available datasets. These datasets have significantly enriched the research community's resources, providing ample access to a variety of traffic data that can be employed in modeling dynamic transportation systems. The integration of emerging data sources, such as ATSPMs and other advanced technologies, presents both opportunities and challenges in the field of traffic signal management and simulation. These sources offer valuable insights into traffic behavior and real-time conditions but require rigorous data processing and calibration to ensure accuracy and usability.

In our study, oversaturation is characterized by the intermittent and random inability of vehicle queues at an intersection to dissipate within several signal cycles, attributable to variable traffic inflow. In contrast, complete spill-over is a more severe and persistent condition where the queues exceed the roadway's available capacity, consequently obstructing adjacent intersections or lanes and disrupting the cohesive flow of traffic. Addressing the latter necessitates the implementation of more comprehensive and multifaceted traffic management strategies beyond mere adjustments to signal timings. We acknowledge the limitations inherent in field data measurements, especially when residual traffic backs up to the upstream intersection, making accurate measurement of arrivals challenging. In such scenarios, the traffic arrivals are constrained by the actual throughputs at the upstream intersection, underscoring the need for considering these limitations in our study.

1.5. Research objectives and potential contributions

This paper's research objectives center on the elucidation and quantification of stochasticity and dynamics in traffic states at signalized intersections, particularly under oversaturated conditions. Through this lens, we aim to study queue dynamics at signalized intersections, which occur when the arrival rates on a road segment surpass the departure rates. Our research specifically targets the role of traffic signals during such oversaturated conditions, and how factors like effective red and green times contribute to the formation and dissipation of vehicle queues.

Our study expands upon Newell's research [43] on the spatial queuing-theoretic ordinary differential equation model, by further considering oversaturated signalized intersections with time-dependent arrival rates. Accordingly, our study aims to address several key questions:

1. How do traffic signals affect the formation and dissipation of vehicle queues (specifically residual queue) when the demand reaches and exceeds capacity during peak periods?
2. How can a polynomial fluid queuing-theoretic partial differential equation model be expanded to consider oversaturated signalized intersections with time-dependent arrival rates?

Table 1
Observations from various surveillance techniques.(1).

Surveillance Type	Key Observations on Partially Over-congested System	Methodology
Point Detectors	- Captures Real-time Arrival and Departure Counts from individual detection points - Provides point speed information aiding in congestion level and flow rate determination	Utilizes inductive loop technology to measure vehicle presence and count at specific points
Automatic Vehicle Identification (AVI)	- Delivers specific flow details of tagged vehicles, providing insights into potentially congested corridors - Facilitates travel time data collection, aiding in delay time assessment and verification of stochastic nature of traffic flows	Employs RFID technology for tracking individual vehicles, capturing detailed travel patterns
Mobile GPS location sensors	- Offers semi-continuous path trajectories, yielding dynamic insights into individual vehicle movement patterns during congestion - Enables data collection on residual queue values as GPS-equipped vehicles traverse the intersection	Uses GPS data from mobile devices and vehicles to track location and movement in real-time
Trajectory data from video image processing	- Ensures continuous monitoring across all vehicle lanes for a detailed understanding of queue lengths and their fluctuations - Provides updates on stochastic arrival and departure rates by tracking vehicle movements	High-resolution video detection cameras capture traffic flow, processed manually for detailed analysis

(1) The surveillance technique deployed in this study utilized high resolution video detection cameras. Recorded footage was analyzed in slow motion to manually collect the detailed traffic data including real-time arrival and departure counts, vehicle delay times, stochastic arrival and departure flow rates, and residual queue values.

3. Can polynomial functions effectively approximate dynamic arrival rates considering the stochastic nature of vehicle arrivals at signalized intersections and what factors influence the accuracy of this approximation?

The Kalman Filter (KF) technique (see Table 2) plays a crucial role in enhancing the accuracy of traffic flow predictions, especially when dealing with noisy data and fluctuating traffic conditions [61,62]. The KF, known for its versatility and applicability in modeling systems with multiple inputs and outputs, offers a unified approach for predicting various traffic processes that can be represented in a state space [61-64]. This technique harnesses both historical and real-time data, allowing it to adapt to non-stationary traffic conditions, from stable to highly volatile traffic circulation [61-64]. By calibrating the KF equations with real-time data, accounting for factors like fluctuating arrival rates and signal settings, it becomes a powerful tool for short-term traffic prediction [61,64]. The KF's state-space model, introduced by Kalman, facilitates the modeling of traffic states, and its parameters can be fine-tuned to capture variations in flow patterns, thus improving prediction accuracy. Utilizing KF in conjunction with partial differential equations and polynomial approximations for arrival flows and queue lengths provides a comprehensive framework to account for noise and refine traffic predictions, making it a promising tool for optimizing traffic management and control [61,62,64,65-67].

The use of Newell's model [43] can be viewed as the application of the principles of Model Order Reduction (MOR), a methodology aimed at reducing the computational complexity of mathematical models used in numerical simulations. To validate Newell's dynamic traffic patterns, a meticulous assessment of cumulative cycles during peak times is conducted, with partitioning into distinct oversaturation episodes guided by residual queue values. To capture a comprehensive picture of queue lengths throughout peak periods, the study implements a wide-ranging real-world experiment, leveraging emergent technologies such as video cameras and other data sources. From a conceptual Kalman filtering perspective shown in Table 2, this approach aims to reveal the dynamics and stochasticity of queue build-up and dispersal, shedding light on the behavior of traffic queues at signalized intersections.

This paper is structured as follows. Section 2 presents the development of a theoretical framework of the queue evolution process using fluid queue approximation models at oversaturated intersections. Section 3 discusses the data collection process, followed by the implementation and discussion of the results of the proposed model in Section 4. Finally, Section 5 concludes the paper.

2. Model formulation

2.1. Embracing fluid queue models: theoretical framework and pragmatic measurement approaches for stochasticity and dynamics

2.1.1. Notation

The parameters used in this paper are listed.(Table 3)

2.1.2. Basic concepts and definitions

When we consider the virtual arrival curves, we consider the theoretical concept of virtual arrival rates, referring to the time when vehicles reach the virtual end of a road section. This approach allows us to consistently analyze virtual arrival and departure curves. For data collection, timestamps are recorded when vehicles reach the physical end of queues. One can employ the method proposed by Lawson et al. [68]. This method adjusts the actual arrival times at the queue to virtual arrival times using the free-flow speed and the remaining distance to the stop bar. Additionally, it involves mapping the actual waiting time to the virtual waiting time at the queue's start. Thise vehicles at the intersection adhere to a first-in, first-out order, this adjustment of timestamps is unlikely significantly alter the arrival curves.

Table 2
Mapping of the conceptual Kalman Filtering framework for traffic control to research keywords.

Process	Formula/Equation	Variables/Parameters	Research Linkage
State Update Process	$X_t = F_t X_{t-1} + G_t U_t + w_t$ Describes the temporal dynamics of traffic using state, control, and noise models.	State Transition Model F_t: Captures the change from the previous state X_{t-1}, such as polynomial process model. Control Model G_t: Incorporates the effect of various control strategies U_t such as signal control plans. State Variables X_t: Includes traffic flow states, such as arrival rates and departure rates and performance metrics (like travel time and queue lengths). State Process Noise w_t: Represents uncertainty in the model, distributed as $N(0, Q_t)$.	Traffic Signals & Dynamics: Relating to the effects of traffic signals on traffic dynamics. Oversaturated Intersections: Considering intersections beyond their capacity. Dynamic Arrival Rates: Time-dependent variation in traffic inflow.
Measurement update Process	$Y_t = H_t X_t + v_t$ Relates the actual measurements with the state of the system.	Measurement Model H_t: Maps the state space X_t to the measurement space Y_t. Measurement Noise v_t: Represents uncertainty in measurements, distributed as $N(0, R_t)$.	Polynomial Model Mapping: Establishing a connection between polynomial fluid queuing-theoretic models and the state/measurement models. Measurement Types & Observability: Different sensor types contributing to the observation matrix H.

Table 3
Symbols and definitions used in this study.

Symbols	Definitions
$A(t)$	cumulative arrival (vehicles) rate
A_c	cumulative arrivals (vehicles)
$A_s(t)$	cumulative service arrival (vehicles) rate
C	total cycle length, $C = r_e + g_e$
D	congested demand during the entire peak period
$D(t)$	cumulative departures (vehicles) rate
D_c	cumulative departures (vehicles)
g_e	effective green time duration (s)
$g(m)$	the conversion factor of the polynomial arrival queue
m	Peak Queue Fraction (PQF), The queue peaks at a fraction m of the total congestion duration, $m = (t_2 - t_0)/(t_3 - t_0)$
m_o	observed PQF
m_t	approximated PQF
$NFII$	net flow intensity index or normalized highest net flow volume, $NFII = \left[\frac{\lambda(t_1) - \mu}{\gamma}\right]$
P	congestion period (cycles), $P = t_3 - t_0$
P_o	observed congestion period (cycles)
P_t	approximated congestion period (cycles)
ρ	traffic intensity ratio, $\rho = \lambda_{\max}/\mu$
$Q(t)$	queue length rate at time t
$Q_s(t)$	queue length at time t associated with the traffic signal delay
$Q_{od}(t)$	queue length at time t associated with the overflow delay
$Q_o(t)$	observed queue length at time t
$Q_t(t)$	approximated queue length at time t
$Q_{\max}$	maximum residual queue length for episode
$Q_r(t)$	residual queue length at cycle termination
$\Delta(Q)$	residual queue length percentage approximation error, $\Delta(Q) = Q_o(t) - Q_t(t)$
QRII	Residual Queue Intensity Index, (total residual queue lengths for the episode/ maximum residual queue length for the episode), $QRII = \sum Q_r/Q_{\max}$
r_e	effective red time duration (s)
R^2	coefficient of determination
s	saturation flow rate (veh/cycle), $s = \lambda/\mu$
t	time
t_f	total travel time at free flow speed
t_t	total travel time
t'	the time the vehicle departs the link or is serviced by the signal
t_0	appoximated time at which the congestion period begins
t_1	approximated time at which maximum arrival rate occurs
t_2	approximated time at which the maximum residual queue occurs
t_3	approximated end time of congestion period
T_0	observed time at which the congestion period begins
T_1	observed time at which maximum arrival rate occurs
T_2	observed time at which the maximum residual queue occurs
T_3	observed end time of congestion period
$\widetilde{t}$	third root associated with cubic approximation function
TS1	time stamp 1
TS2	time stamp 2
TS3	time stamp 3
$w(t)$	delay experienced by a vehicle entering the link at time t
w_{od}	delay due to overflow experienced by a vehicle entering the link at time t
w_s	delay due to traffic signal experienced by a vehicle entering the link at time t
W	total delay experienced during a congestion period
$\overline{w}$	average delay experienced by a vehicle during a congestion period
$\lambda(t)$	arrival rate function at time t (veh/cycle)
$\lambda_o(t)$	observed arrival rate function at time t (veh/cycle)
$\lambda_t(t)$	approximated arrival rate function at time t (veh/cycle)
$\Delta(\lambda_t - \lambda_o)$	arrival flow rate percentage approximation error, $\Delta(\lambda) = \lambda_o(t) - \lambda_t(t)$
$\lambda_{\max}$	maximum arrival rate for an episode (veh/cycle)
γ	arrival curvature gradient used in polynomial form
μ	discharge rate, assumed to be a constant value (veh/cycle)
μ_{avg}	average discharge rate for an episode (veh/cycle)

Undersaturated Conditions: An intersection is classified as operating in undersaturated conditions when the incoming traffic flow does not exceed the capacity or service rate of the signal cycle. However, Fig. 1, initially interpreted as depicting undersaturation, actually represents a critical state where the intersection operates at maximum capacity just before becoming saturated. In this state, the queue dissipates exactly at the end of the green phase, signifying that any increase in traffic inflow would lead to oversaturation. Hence, Fig. 1 illustrates the threshold between undersaturation and the onset of saturation, emphasizing the delicate balance maintained at signalized intersections [69]. It is also important to note that Fig. 1 represents the average expected queue length and not the actual queue length. This modeling choice is based on prior studies that have found using expected values to be more representative for

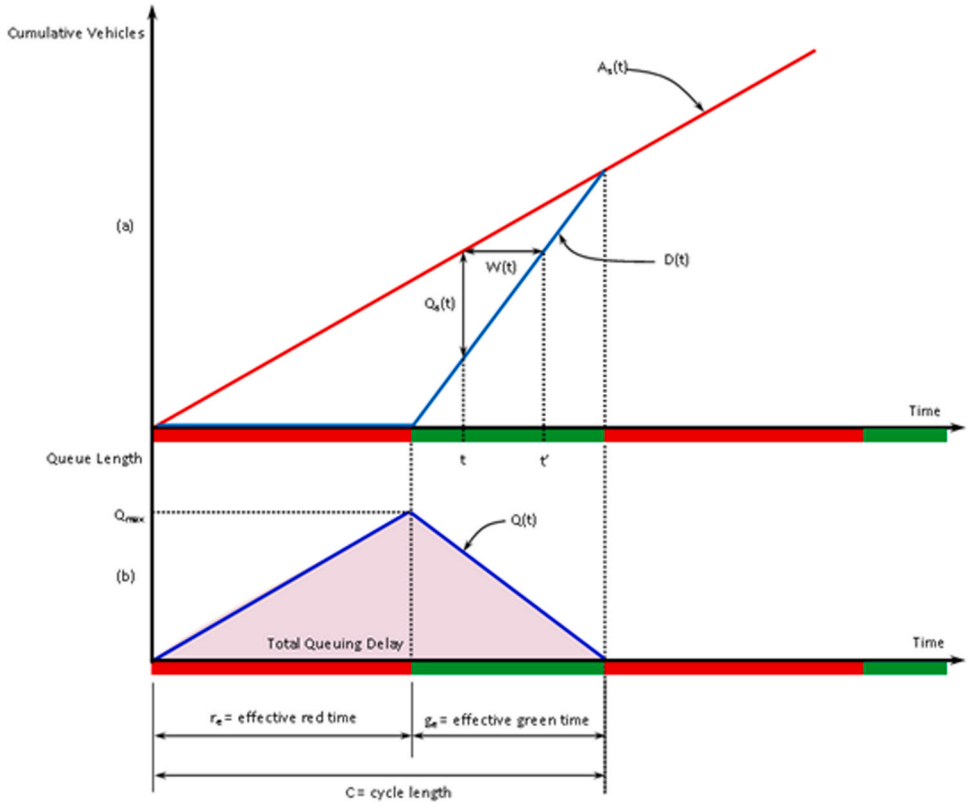


Fig. 1. Cumulative arrival and departures for undersaturated signalized intersection with linear arrival rates. (Adapted from Rouphail et al) [70]

capturing the discrete nature of vehicle arrivals at intersections. Such consideration of expected values is essential in accounting for the stochastic variability of traffic flow, providing a more realistic portrayal of queue dynamics, especially in approaching saturation [69].

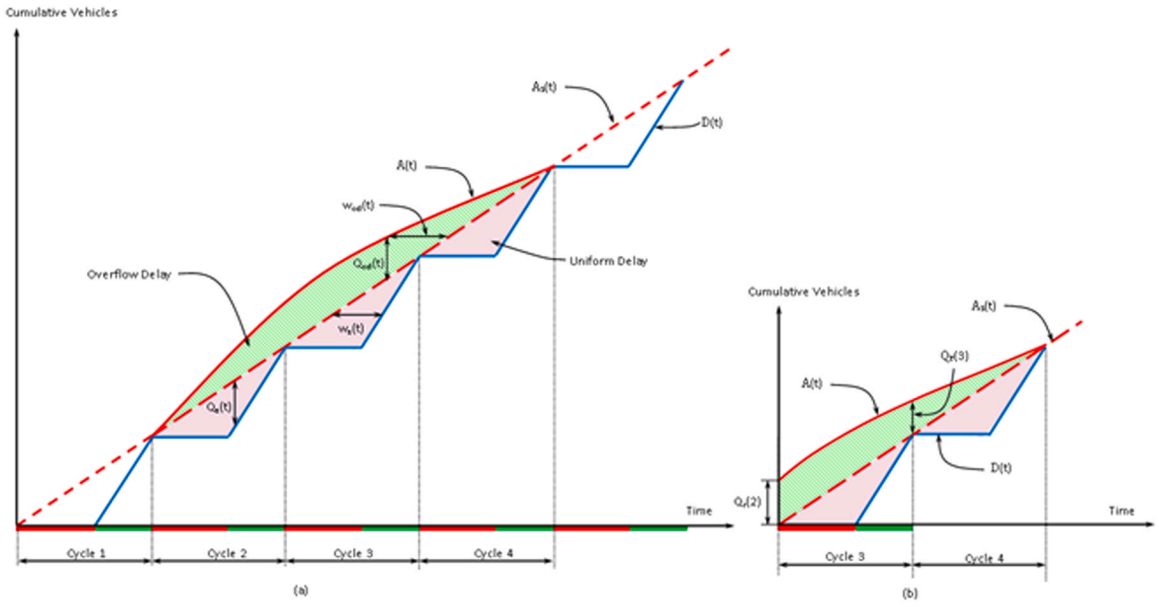


Fig. 2. Cumulative arrival and departures for oversaturated signalized intersection with non-linear polynomial arrival rates. (Adpated from Dion et al.) [1]

In a broader sense, the red linear line in Fig. 1**(a)** signifies the service cumulative arrival rate, assuming a uniform linear rate denoted as $A_s(t)$. The service cumulative arrival rate represents the total number of vehicles the traffic signal can serve without leaving a residual queue at the cycle's termination. The cumulative departure rate, $D(t)$, is divided into two segments based on the effective red and green times. During the effective red time, vehicles can't exit (or be serviced by the traffic signal), resulting in a departure rate of zero. Conversely, the effective green time allows vehicles to be serviced uniformly at the saturation rate.

Throughout the effective red time, vehicles continue to enter the link without departing, leading to a linearly increasing queue extending from the stop bar. This dynamic queue length formed due to the delay of the traffic signal, denoted as $Q_s(t)$, can be mathematically expressed as:

$$Q_s(t) = A_s(t) - D(t) \tag{1}$$

As a result, vehicles entering the link experience delay, with the traffic signal service delay encountered by a vehicle entering at time t and exiting at time t′ represented as:

$$w_s(t) = t' - t \tag{2}$$

The total queuing delay during a cycle can be found by calculating the area of the shaded pink triangle in the queuing profile (refer to Fig. 1**(b)**).

In undersaturated conditions, at the cycle's termination the queue length build-up will have completely dissipated.

In Fig. 1, we illustrate the dynamic behavior of queues, assuming constant cumulative arrivals and service times under conditions void of congestion. However, it's essential to recognize that actual vehicle arrivals and departures in real-world scenarios typically exhibit irregular intervals. In our study, we delve further into this stochasticity by considering a nonlinear cumulative arrival rate during peak periods, as visualized in Fig. 2.

During congested periods, the cumulative arrival rate surpasses the signal's capacity (cumulative service arrival rate), the traffic conditions become oversaturated, leading to overloading and the formation of overflow queues, as highlighted in Fig. 2**(a)**. The service cumulative arrival rate, $A_s(t)$, is depicted as the red linear dashed line in Fig. 2**(a)**, while the peak period cumulative arrival rate, denoted as $A(t)$, follows a red nonlinear curve. At the end of Cycle 1 we see that $A(t)$ continues to escalate, exceeding the service rate ($A_s(t)$), resulting in a vertical separation between the two functions, forming an overflow queue termed $Q_{od}(t)$. Consequently, the time-dependent total queue length, $Q(t)$, becomes:

$$Q(t) = Q_{od}(t) + Q_s(t) \tag{3}$$

Similarly, the total waiting time of a vehicle arriving at time t, denoted $w(t)$, can be expressed as

$$w(t) = w_{od}(t) + w_s(t) \tag{4}$$

The total control delay due to the traffic signal is the area of the pink triangle and can be expressed as

$$w_s = \frac{1}{2}C^2\left(1 - \frac{\mu}{s}\right)\mu \tag{5}$$

As the episode progresses into cycle 2, if $A(t)$ remains greater than $A_s(t)$, the queue will persist, requiring service during Cycle 3. This remaining queue at the cycle's termination is termed a residual queue, Q_r.

Residual queues can also develop when an undersaturated period follows an oversaturated period, assuming a non-zero initial queue at the cycle's start. At lower arrival rates, cycles often start and end with a zero residual queue. However, with increasing saturation levels, the likelihood of cycles initiating or concluding with a residual queue rises. An illustration of a saturated Cycle 4 succeeding an oversaturated Cycle 3 is presented in Fig. 2**(b)**. Cycle 3 concludes with a residual queue $Q_r(3)$, necessitating Cycle 4 to commence with a residual queue. Throughout this cycle, $A(t)$ diminishes, allowing the overall queue length to be reduced until it has completely dissipated at the end of the cycle.

Residual queues play a critical role in determining the temporal behavior of queues and inherent delays in the oversaturated system. Deterministic models that rely on fluid theory have demonstrated their applicability in addressing highly oversaturated traffic scenarios [70]. However, these models have a limitation in accounting for the temporal effects that arise when traffic signals operate near their capacity. In real-world settings, traffic signals are often optimized to operate as closely as possible to saturation during peak hours, making it essential to model these effects.

The following section seeks to critically analyze fluid queue models, particularly in oversaturated traffic conditions. We will explore the impact of the cumulative arrival rate exceeding the capacity of the signal (service rate) on the queue dynamics and delay incurred. We focus on the conventional signalized intersection model's applicability and propose an enhancement through a non-linear polynomial function of the arrival flow rate, λ(t). We ground our model in real-world observations and leverage reduced order modeling to capture the complexity of traffic conditions efficiently.

3. Theoretical framework

The fluid approximation method replaces individual discrete vehicles with a continuous fluid representation. This approach differs in the way that they consider stochasticity. This method considers a deterministic continuum, treating the arrival of vehicles as flowing fluid with a rate λ(t) into a reservoir (representing the traffic system). The service process is approximated by a deterministic outflow (discharge rate) from the reservoir. The level of fluid in the reservoir serves as an approximation for the number of vehicles in the

system. The fluid approximation represents one of the first approaches for the analysis of time-dependent queueing systems. It is described in Newell's book [43] as an engineering approach for the performance evaluation of systems for which temporary overload, rather than randomness, is the primary reason for the existence of queues [38].

In Fig. 3**(a)**, the stochastic cumulative arrivals are depicted by a red non-linear curve, $A(t)$, while the accompanying blue zig-zag line represents the intermittent flow of cumulative departures, $D(t)$, both as functions as time. During the effective red time, $D(t)$ is equal to zero, conversely during the effective green time, cumulative departures align with the saturation rate.

The notion of a temporary congestion period (P), or an oversaturation episode, comes into play in situations where the traffic signal experiences brief periods of oversaturation within a larger congested timeframe, typically peak hours. Such temporary oversaturation arises when the rate of arrival flow momentarily surpasses the departure flow rate, causing short-lived instances of congestion. This congestion period interval begins at time t_0, marked by the cumulative arrivals overtaking cumulative departures. It concludes at time t_3, signaling the return to undersaturated conditions, where capacity is no longer exceeded.

Fig. 3**(a)** showcases the cumulative vehicle arrivals and departures during a congestion period, establishing a link to the polynomial fluid queue model.

The congestion period is represented by the expression:

$$P = t_3 - t_o \tag{6}$$

The queue length at time t can be expressed as:

$$Q(t) = A(t) - D(t) \tag{7}$$

The vehicle waiting time at time t, $w(t)$, can calculated as the ratio of the queue length and the constant slope μ, as

$$w(t) = \frac{Q(t)}{\mu} \tag{8}$$

The total waiting time in the queue, W, is the ratio of total queue length, Q, over the congested period, P, as

$$W = \frac{Q}{P} \tag{9}$$

The arrival flow rate for a congestion period can be approximated by a quadratic polynomial function. Fig. 3(b) illustrates the approximated polynomial arrival flow function (red), λ(t), and the assumed constant departure flow rate (blue) μ. The congestion period begins at time t_o when the number of incoming vehicles is equal to the number of vehicles that are serviced (μ), reaches a maximum at time t_1, then gradually decreases until the period concludes at time t_3.

The time dependent function, $Q(t)$, for the queue formation is seen in Fig. 3(c). The queue length at time t_0 is zero. As time continues, the number of vehicles remaining after each cycle begins to grow until a maximum queue length is reached at time t_2, then as the number of arrivals decrease, so does the queue length until there are no vehicles remaining in the queue after the last cycle has terminated at time t_3. Mathematically, the total queue length at time t can be expressed as (area under the curve, shaded in gray)

$$Q(t) = \int_{t_0}^{t} [\lambda(t) - \mu] dt \tag{10}$$

Derivation of the system state dynamics using the quadratic arrival rate function.

In Newell's classic model, two key assumptions are made: (1) the virtual arrival flow rate $\lambda(t)$ can be effectively approximated using a polynomial function, often quadratic in nature, and (2) the departure flow rate μ remains constant. By employing these assumptions, we can derive a series of functions in different orders of magnitude for the arrival flow rate, the queue length, and delay. Furthermore, this allows for the analytical examination of the queueing system in terms of both time-dependent and averaged system metrics.

As shown in Fig. 3**(b)**, the quadratic function $\lambda(t)$ evolves throughout the congestion period. It exhibits a gradual increase, reaching its peak arrival flow at time t_1, followed by a gradual decrease until the period concludes at time t_3, signaling a return to an undersaturated state. Newell's approach rested on the assumption that the stochastic nature of the arrival flow rate could be most effectively approximated through the quadratic Taylor expansion [40]. Refer to **Appendix A** for the full derivation. Therefore, the quadratic function for approximating the arrival flow rate λ(t) centered about time t_1 can be shown as:

$$\lambda(t) = \lambda(t_1) + \lambda'(t_1) \bullet (t - t_1) + \frac{1}{2}\lambda''(t_1) \bullet (t - t_1)^2 \tag{11}$$

Given $\lambda'(t_1) = 0$, let $\gamma = -\frac{1}{2}\lambda''(t_1)$, then Eq. (10) can be transformed to

$$\lambda(t) = \lambda(t_1) - \gamma \bullet (t - t_1)^2 \tag{12}$$

Since γ describes the curvature or shape of the time-dependent inflow arrival rates, we term it as the inflow curvature parameter.

By assuming a constant departure flow rate within a single period, the departure flow rate μ (where $\mu = \lambda(t_0) = \lambda(t_2)$), as shown in Fig. 3**(b)**, can be estimated by the two equations:

$$\mu = \lambda(t_1) - \gamma \bullet (t_0 - t_1)^2 = \lambda(t_1) - \gamma \bullet (t_2 - t_1)^2 \tag{13}$$

Then the two real roots, t_0 and t_2, can be obtained as follows:

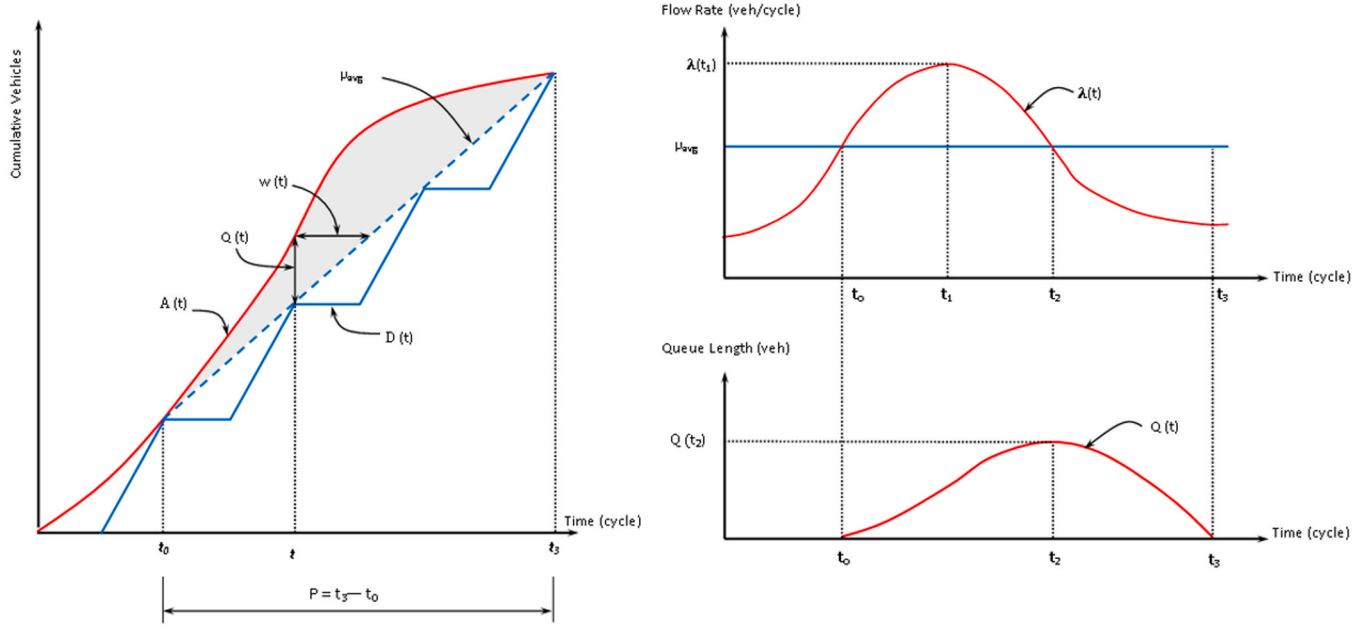


Fig. 3. Model illustration based on the fluid approximation model and spatial queue model. (Adapted from Cheng et al.) [71]

$$t_0 = t_1 - \left[\frac{\lambda(t_1)-\mu}{\gamma}\right]^{\frac{1}{2}}, \quad t_2 = t_1 + \left[\frac{\lambda(t_1)-\mu}{\gamma}\right]^{\frac{1}{2}} \tag{14}$$

From Fig. 3**(b)**, Therefore Eq. (12) can be expressed as:

$$\lambda(t) - \mu = \gamma(t - t_0)(t_2 - t) \tag{15}$$

The residual queue length at any time, $Q(t)$ can be expressed by:

$$Q(t) = \int_{t_0}^{t} [\gamma(t - t_0)(t_2 - t)]dt \tag{16}$$

Through substitution from the equations above, the equation for the queue length, $Q(t)$ becomes:

$$Q(t) = \frac{\gamma}{3}(t - t_0)^2(t_3 - t) \tag{17}$$

The maximum queue length occurs at time, t_2 and can be expressed as

$$Q(t_2) = \frac{\gamma}{6}(t_2 - t_0)^3 \tag{18}$$

The delay per vehicle at time, t can be expressed as:

$$w(t) = \frac{Q(t)}{\mu} = \frac{\gamma}{3\mu}(t - t_0)^2(t_3 - t) \tag{19}$$

The total delay for the period from t_0 to t_3:

$$W = \int_{t_0}^{t_3} Q(t)dt = \frac{9}{4}\left(\frac{\lambda(t_1)-\mu}{\gamma}\right)^2 \tag{20}$$

The average delay $\overline{w}$ for the period from t_0 to t_3:

$$\overline{w} = \frac{W}{D} = \frac{\gamma}{36\mu}\left[\frac{D}{\mu}\right]^3 \tag{21}$$

Total travel time, t_t:

$$t_t = t_f + W \tag{22}$$

Through substitution the total travel time becomes:

$$t_t = t_f + \left[1 + \frac{\gamma}{36\mu t_f}\left[\frac{D}{\mu}\right]^3\right] \tag{23}$$

It is noteworthy to highlight the similarity in functional form to the Bureau of Public Roads (BPR) Equation:

$$t_t = t_f\left[1 + \alpha\left[\frac{V}{C}\right]^{\beta}\right] \tag{24}$$

$\alpha = \frac{\gamma}{36\mu t_f}$, $V = D$, $C = \mu$, β=3

However, for the connection between the polynomial arrival queue model and the volume-delay function with the inflow demand-to-capacity ratio on freeway facilities, we refer to the work of Zhou et al. [5].

Hypothesis A. According to Eq. (14), the congestion duration, $P = t_3 - t_0$, should then be proportional to the square root of the Net Flow Volume Index (NFII). The NFII is the normalized highest net flow volume, calculated as the difference between the highest arrival flow rate and the constant discharge rate. Mathematically, that is $\left[\frac{\lambda(t_1)-\mu}{\gamma}\right]^{\frac{1}{2}}$. This relationship is analogous to the inverse of the curvature's second-order gradient γ. Through substitution of Eq. (14) into Eq. (17a) (Appendix A)

$$t_0 = t_1 - \left[\frac{\lambda(t_1)-\mu}{\gamma}\right]^{\frac{1}{2}}, \quad t_2 = t_1 + \left[\frac{\lambda(t_1)-\mu}{\gamma}\right]^{\frac{1}{2}} \tag{14'}$$

$$t_3 = \frac{2}{3}(t_2 - t_0) + t_0 \tag{17a}$$

We can obtain t_3-t_0

$$t_3 - t_0 = \frac{4}{3}\left[\frac{\lambda(t_1) - \mu}{\gamma}\right]^{1/2} \tag{25}$$

Eq. (25) is substituted into Eq. (6) ($P = t_3 - t_0$), P can now be expressed as:

$$P = \frac{3}{4}\left[\frac{\lambda(t_1) - \mu}{\gamma}\right]^{1/2} \tag{26}$$

Therefore,

$$P \sim \left[\frac{\lambda(t_1) - \mu}{\gamma}\right]^{1/2} \tag{27}$$

Hypothesis B. In accordance with Eq. (18), the longest queue length, $Q_{\max}$, should be proportional to the NFII, raised to the power of 3/2, expressed as $\left[\frac{\lambda(t_1)-\mu}{\gamma}\right]^{\frac{3}{2}}$. This is akin to the inverse of the curvature's second-order gradient γ. To find the relationship between $Q(t_2)$ or $Q_{\max}$ and NFII, we will solve for $Q_{\max}$ through substitution of Eq. (14) into Eq. (18).

$$t_0 = t_1 - \left[\frac{\lambda(t_1) - \mu}{\gamma}\right]^{\frac{1}{2}}, \quad t_2 = t_1 + \left[\frac{\lambda(t_1) - \mu}{\gamma}\right]^{\frac{1}{2}} \tag{14}$$

$$Q(t_2) = \frac{\gamma}{6}(t_2 - t_0)^3 \tag{18}$$

$Q_{\max}$ can now be expressed as:

$$Q(t_2) = Q_{\max} = \frac{\gamma}{3}\left[\frac{\lambda(t_1) - \mu}{\gamma}\right]^{\frac{3}{2}} \tag{28}$$

Therefore,

$$Q_{\max} \sim \left[\frac{\lambda(t_1) - \mu}{\gamma}\right]^{\frac{3}{2}} \tag{29}$$

Hypothesis C. If $Q_{\max}$ is proportional to the net flow intensity index, raised to the power of 3/2 and P is proportional to the NFII raised to the power of ½, then $Q_{\max}$ is a cubic function of P. By substituting Eq. (26), $P = \frac{3}{4}\left[\frac{\lambda(t_1)-\mu}{\gamma}\right]^{1/2}$, into Eq. (28), $Q_{\max} = \frac{\gamma}{3}\left[\frac{\lambda(t_1)-\mu}{\gamma}\right]^{\frac{3}{2}}$, we can obtain to proportionality between P and $Q_{\max}$,

$$Q_{\max} \sim P^3 \tag{30}$$

Hypothesis D. In accordance with Eq. (20), the total delay per period should be proportional to the NFII raised to the power of 2, expressed as $\left[\frac{\lambda(t_1)-\mu}{\gamma}\right]^2$, then W is the 4th power function of P. This is akin to the inverse of the curvature's second-order gradient γ. From Eq. (20), W can be expressed as,

$$W = \frac{9}{4}\left(\frac{\lambda(t_1) - \mu}{\gamma}\right)^2 \tag{20}$$

By substituting Eq. (26), $P = \frac{3}{4}\left[\frac{\lambda(t_1)-\mu}{\gamma}\right]^{1/2}$, into Eq. (20), the proportionality between total delay and the congestion period is,

$$W \sim P^4 \tag{31}$$

Remarks: If this set of hypotheses is not validated, it may indicate that the current model of the arrival process doesn't sufficiently align with the equations associated with quadratic arrival flow rates. As a result, users should approach these equations with caution if they're based on quadratic arrival rates. It may also be prudent to consider alternative higher-order models to match the observed behavior more closely.

Two primary indices for evaluating the signal timing performance are the congestion duration and the longest queue lengths. (1) Congestion duration P not only impacts the immediate vehicles in the queue but can also have cascading effects on subsequent traffic flow and even adjacent intersections or routes. (2) Longest queue lengths give insights into the peak congestion levels. A longer queue can lead to increased waiting times for vehicles, potential blockages of adjacent lanes or intersections, and contribute to increased emissions and fuel consumption.

Given the significance of these two metrics, the hypotheses based on analytical derivations mentioned earlier still have important implications:

1. Active Demand Management: Traffic systems can modulate demand by comprehending inflow rates. This can entail extending demand across broader time frames, recommending alternative paths, or setting dynamic road pricing based on congestion.
2. Supply Management: By fine-tuning discharge rates to cater to incoming traffic, congestion risks decrease. Implementing measures like adjusting traffic signals or dynamically altering lane directions aids this.
3. Flow Shape Management: Predicting flow patterns, such as sudden vehicle influxes from events or work shifts, allows traffic managers to pre-emptively counteract potential congestion.

Derivation of the system state dynamics using an approximated cubic arrival flow rate function.

Newell's work on approximating stochastic arrival flows at signalized intersections was expanded upon with Cheng's et al. [71] further assumption that higher order polynomials would be a better fit for capturing their dynamic nature. They proposed a similar system of dynamic equations to capture the traffic elements during an oversaturated period with a higher order cubic polynomial function. A third root is introduced with the application of a cubic function. This additional root does not occur within the congestion period, it occurs at time $\tilde{t}$ (either before or after the oversaturation episode). To address this, we will introduce the Peak Queue Fraction, $m = \frac{t_2 - t_0}{t_3 - t_0}$, the ratio of the time required to the reach maximum queue length over the total congestion period.

The quadratic arrival rate function (as previously derived) can be expanded to have three roots and become a cubic arrival rate function:

$$\lambda(t) - \mu = \gamma(t - t_0)(t_2 - t)\left(t - \tilde{t}\right) \tag{32}$$

The queue length at any time can be approximated by:

$$Q(t) = \int_{t_0}^{t} [\gamma(t - t_0)(t_2 - t)(t - \bar{t})]dt \tag{33}$$

A summarized version is provided below (refer to Appendix B for the full derivation). The time-dependent queue length, $Q(t)$, maximum queue length, $Q(t_2)$, time dependent delay, $w(t)$, total delay, $W(t_3)$, conversion factor, $g(m)$, and average waiting time, $\overline{w}$, can be expressed using Eqs. (34)-(41).

$$Q(t) = \gamma(t - t_0)^2\left[\frac{1}{4}(t - t_0)^2 - \frac{1}{3}(t - t_0)(t_3 - t_0)\left[m + \left[\frac{3 - 4m}{4 - 6m}\right]\right] + \frac{1}{2}(t_3 - t_0)^2 m\left[\frac{3 - 4m}{4 - 6m}\right]\right] \tag{34}$$

$$Q(t_2) = \gamma(t_3 - t_0)^4 m^3\left[\frac{(m - 1)^2}{8 - 12m)}\right] \tag{35}$$

$$W(t) = \frac{Q(t)}{\mu} = \frac{\gamma(t - t_0)^2}{\mu}\left[\frac{1}{4}(t - t_0)^2 - \frac{1}{3}(t - t_0)(t_3 - t_0)\left[m + \left[\frac{3 - 4m}{4 - 6m}\right]\right] + \frac{1}{2}(t_3 - t_0)^2 m\left[\frac{3 - 4m}{4 - 6m}\right]\right] \tag{36}$$

$$W(t_3) = \int_{t_0}^{t_3} Q(t)dt = \gamma(t_3 - t_0)^5\left[\frac{1}{20} - \frac{1}{12}\left[m + \left[\frac{3 - 4m}{4 - 6m}\right]\right] + \frac{1}{6}m\left[\frac{3 - 4m}{4 - 6m}\right]\right] \tag{37}$$

$$g(m) = \left[\frac{1}{20} - \frac{1}{12}\left[m + \left[\frac{3 - 4m}{4 - 6m}\right]\right] + \frac{1}{6}m\left[\frac{3 - 4m}{4 - 6m}\right]\right] \tag{38}$$

$$W(t_3) = \gamma g(m)(t_3 - t_0)^5. \tag{39}$$

$$\overline{w} = \frac{W}{D} = \frac{\gamma g(m)}{\mu}(t_3 - t_0)^5 = \frac{\gamma g(m)}{\mu}\left(\frac{D}{\mu}\right)^4 \tag{40}$$

$$t_t = t_f + \overline{w} = t_f\left[1 + \frac{\gamma g(m)}{\mu t_f}\left[\frac{D}{\mu}\right]^4\right] \tag{41}$$

Hypothesis E: For cubic polynomial functions with one additional root, the congestion duration, $P = t_3 - t_0$, should be proportional to the cubic root of the net flow intensity index, NFII. Mathematically, that is. $\left[\frac{\lambda(t_1) - \mu}{\gamma}\right]^{\frac{1}{3}}$. This relationship is analogous to the inverse of the curvature's second-order gradient γ. By substituting Eq. (5b) (Appendix B) into Eq. (2b) (Appendix B) and further into Eq. (6), $P = t_3 - t_0$,

$$Q_r(t_3) = 0 = \gamma(t_3 - t_0)^2\left[\frac{1}{4}(t_3 - t_0)^2 + \frac{1}{3}(t_3 - t_0)(2t_0 - \bar{t} - t_2) + \frac{1}{2}(t_0 - t_2)(t_0 - \bar{t})\right] \tag{2b}$$

$$(\bar{t}-t_0)=\frac{3(t_3-t_0)^2-4(t_2-t_0)(t_3-t_0)}{4(t_3-t_0)-6(t_2-t_0)} \tag{5b}$$

We arrive at the proportional expression,

$$P\sim\left[\frac{\lambda(t_1)-\mu}{\gamma}\right]^{1/3} \tag{42}$$

Hypothesis F: According to Eq. (35), the longest queue length, $Q_{\max}$, should be proportional to the net flow intensity index, NFII, raised to the power of 4/3, expressed as $\left[\frac{\lambda(t_1)-\mu}{\gamma}\right]^{\frac{4}{3}}$. This is akin to the inverse of the curvature's second-order gradient γ. To find the relationship between $Q(t_2)$ or $Q_{\max}$ and NFII, we will solve for $Q_{\max}$ through substitution of Eq. (6), $P=t_3-t_0$, into Eq. (35)

$$Q_{\max}=Q(t_2)=\gamma(t_3-t_0)^4m^3\left[\frac{(m-1)^2}{8-12m)}\right] \tag{35}$$

We can now express $Q_{\max}$ as,

$$Q_{\max}=\gamma(P)^4m^3\left[\frac{(m-1)^2}{8-12m)}\right] \tag{43}$$

From Hypothesis E, we know, $P\sim\left[\frac{\lambda(t_1)-\mu}{\gamma}\right]^{1/3}$, which can be substituted into Eq. (43) to find the proportional relation between $Q_{\max}$ and NFII,

$$Q_{\max}\sim\left[\frac{\lambda(t_1)-\mu}{\gamma}\right]^{4/3} \tag{44}$$

Hypothesis G: If $Q_{\max}$ is proportional to the net flow intensity index, NFII, raised to the power of 4/3 and P is proportional to the NFII raised to the 4th power, then $Q_{\max}$ is a 4th power function of P. From Eqs. (42) and (44), the proportionality can be expressed as

$$Q_{\max}\sim P^4 \tag{45}$$

Hypothesis H: The total delay per period should be proportional to the net flow intensity index (NFII), raised to the power of 5/3, expressed as $\left[\frac{\lambda(t_1)-\mu}{\gamma}\right]^{\frac{5}{3}}$, then W is the 5th power function of P. This is akin to the inverse of the curvature's second-order gradient γ. The total delay is expressed as,

$$W(t_3)=\gamma g(m)(t_3-t_0)^5 \tag{39}$$

And the congestion period, $P=t_3-t_0$, we can obtain,

$$W(t_3)=\gamma g(m)(P)^5 \tag{46}$$

Using the proportionality between the P and NFII from Hypothesis E, $P\sim\left[\frac{\lambda(t_1)-\mu}{\gamma}\right]^{1/3}$, the delay proportional to the congestion period can be express as,

$$W\sim P^{5/3} \tag{47}$$

Remarks:

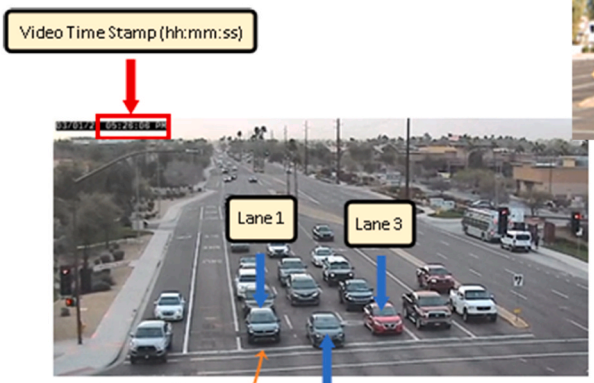
1. One can also define traffic queue intensity index, QRII, as $QRII=\sum Q_r/Q_{max}$ which is proportional to the relationship $\frac{\left[\frac{\lambda(t_1)-\mu}{\gamma}\right]^{\frac{5}{3}}}{\left[\frac{\lambda(t_1)-\mu}{\gamma}\right]^{\frac{4}{3}}}=\left[\frac{\lambda(t_1)-\mu}{\gamma}\right]^{\frac{1}{3}}$ and the congestion period is proportional to the NFII raised to the 1/3 power, therefore the traffic queue intensity index should be proportional to the congestion period.
2. From Hypothesis E, we can see the congestion period, P, should be linked to normalized net flow rates, if the curvature's second order gradient, γ, remains constant, but γ is does not remain constant and continues to change throughout the congestion period, then additional tests can be conducted as γ fluctuates under varying lengths of P's.

(a) Aerial view of limits of observed link on Chandler Blvd between Carriage Way and Dobson Rd

(b) Aerial view of three eastbound through travel lanes on Chandler Blvd

(d) FLIR TRANSFISENSE2 Dual Camera

(c) Video Camera Mounted on Luminaire Mast Arm



(e) FLIR Camera view of Eastbound Traffic

Fig. 4. Aerial perspective: Study area and operational equipment.

4. Data collection

Today, a plethora of richer data sources exists to capture real-time traffic dynamics. Traditional traffic signal systems relied on loop sensors that offered only limited insights into passing vehicles. In contrast, modern technology, such as high-definition video detection cameras, not only provide intricate portrayals of traffic scenarios but also illuminate the patterns of vehicle arrivals at intersections. This treasure trove of mobility data significantly enhances our comprehension of traffic dynamics, making substantial contributions to the advancement of traffic control systems.

For this study, the traffic data collection took place in Chandler, Arizona, specifically along a signalized stretch of Chandler Boulevard extending from Carriage Way to Dobson Road. Positioned in a densely populated sector of Chandler, this intersection is encircled by major activity hubs, including significant employment corridors, bustling commercial centers, a hospital, and a freeway interchange. According to the Chandler Transportation Master Plan [72], the study link resides within the top ten highest volume segments in the city, registering over 41,700 vehicles daily, along with the Chandler/Dobson intersection also ranking among the city's top ten highest volume intersections.

To accumulate real-time traffic insights, data collection spanned three consecutive weekdays during the PM Peak hours of 15:30–18:00, focusing solely on eastbound traffic. An aerial overview of the observed link is presented in Fig. 4**(a)**, commencing at point A (Carriage Way intersection) and concluding at point B (Dobson Road intersection). Fig. 4**(b)** provides a closer inspection of the three eastbound through lanes, the focal point for collecting vehicle traffic data.

Recording the eastbound through traffic necessitated the use of a FLIR TRAFISENSE2 Dual Camera, affixed to the luminaire mast arm of the adjacent traffic signal (as depicted in Fig. 4). Data collection was cycle-based across the observation period, employing high-resolution video recordings scrutinized in slow motion. The ensuing data was meticulously gathered for each lane: arrivals and departures per cycle, timestamps of vehicle arrivals and departures, cycle lengths, queue lengths, and residual queue lengths.

4.1. Collection process

Two distinct data collection processes were employed to capture real-time vehicle arrival and departure timestamps. The initial process, denoted as Data Collection Process A (depicted in Fig. 5), was applied to cars that accessed the link at any point during the cycle (whether in the red or green phase) and managed to exit the link within the same cycle, specifically during the green phase. Subsequently, Data Collection Process B (illustrated in Fig. 6) was implemented alongside Data Collection Process A. This was activated when a cycle concluded with a residual queue of any magnitude. The collection of timestamps encompassed both the vehicle's entry and exit for all three eastbound through lanes throughout the designated three-day observation period.

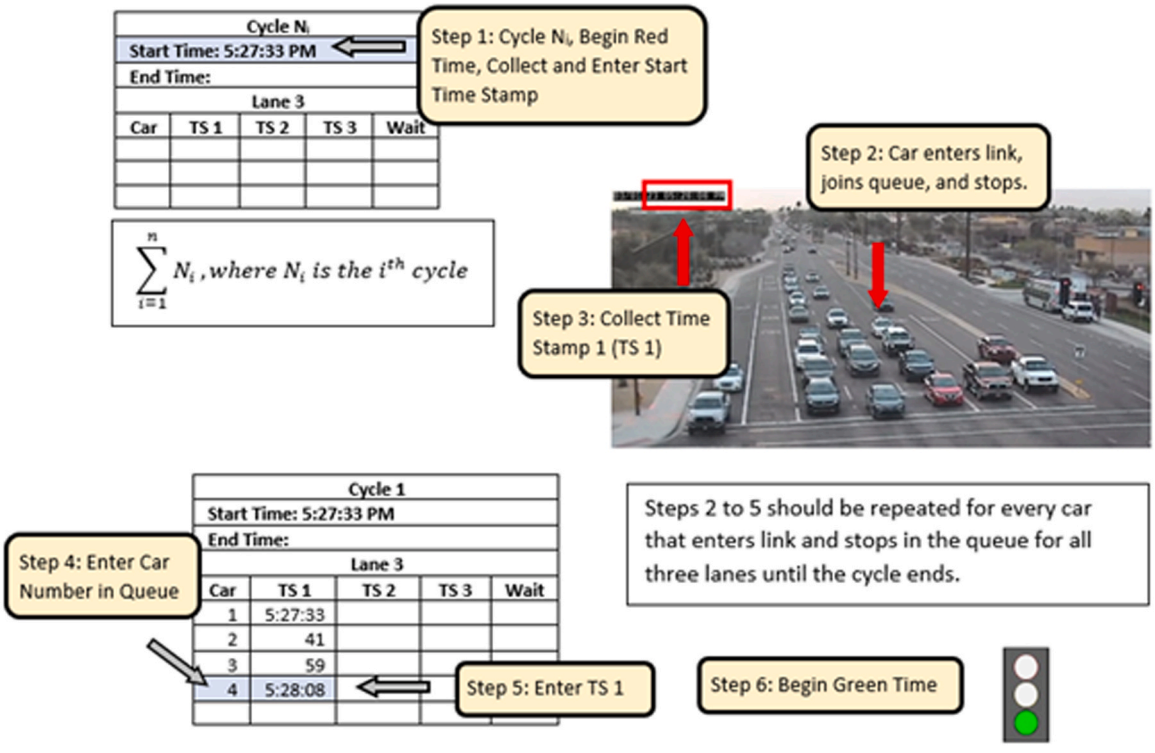
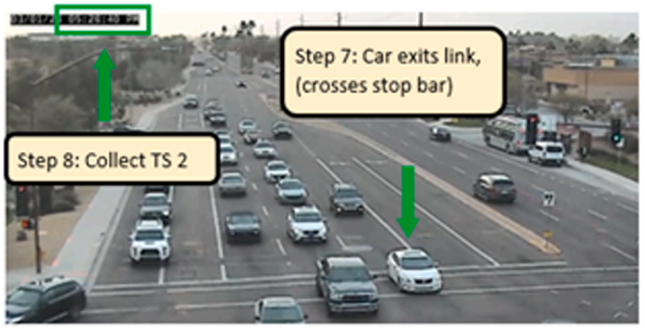


Fig. 5a. Data collection process A: Cycles with no residual queue (Step 1 to Step 6).

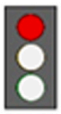


Cycle 1				
Begin: 5:27:33 PM				
End:				
Lane 3				
Car	TS 1	TS 2	TS 3	Wait
1	5:27:33	5:28:35		
2	41	37		
3	59	39		
4	5:28:08	5:28:40		

Step 9: Enter TS 2

Steps 7 to 9 should be repeated for every car that exits the link during the green time for all three lanes.

Step 10: End Green Time



Step 11: Cycle N_i Ends, Collect and Enter Time Stamp

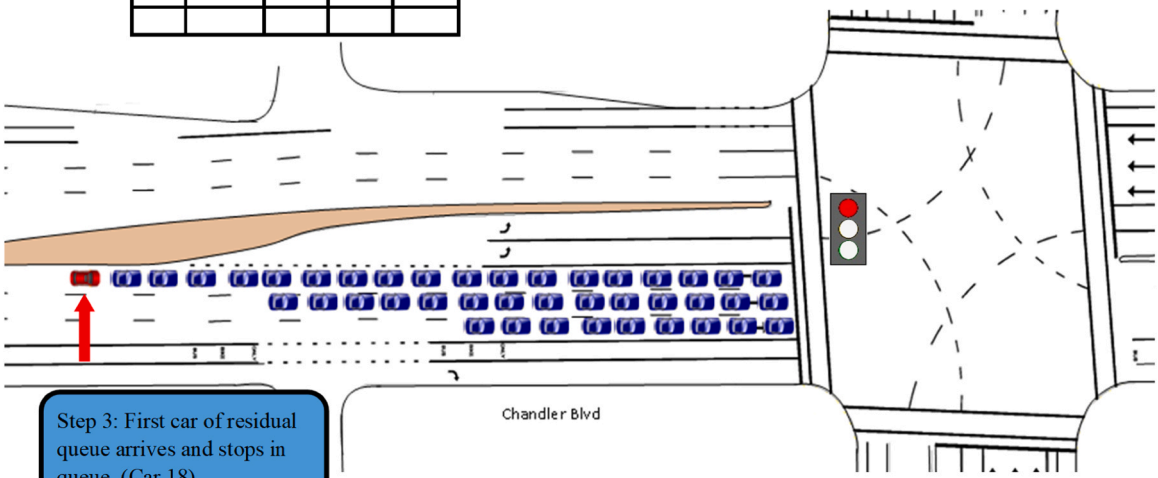
Cycle 1				
Begin: 5:27:33 PM				
End: 5:29:12 PM				
Lane 3				
Car	TS 1	TS 2	TS 3	Wait
1	5:27:33	5:28:35		
2	41	37		
3	59	39		
4	5:28:08	5:28:40		

Fig. 5b. Data collection process A: Cycles with no residual queue (Step 7 to Step 11).

Step 1: Cycle N_{i+1} Start time is equal to Cycle N_i End Time. Enter Start Time

Cycle N_{i+1}				
Start Time: 5:29:12 PM				
End Time:				
Lane 3				
Car	TS 1	TS 2	TS 3	Wait

Step 2: Follow Data Collection Process A: Step 2 to Step 5 for all cars and that enter and exit the link within the same cycle. Once Steps are Complete, Proceed to Step 3 Below



Step 3: First car of residual queue arrives and stops in queue. (Car 18)

Fig. 6a. Data collection process B: Cycles with a residual queue (Step 1 to Step 3).

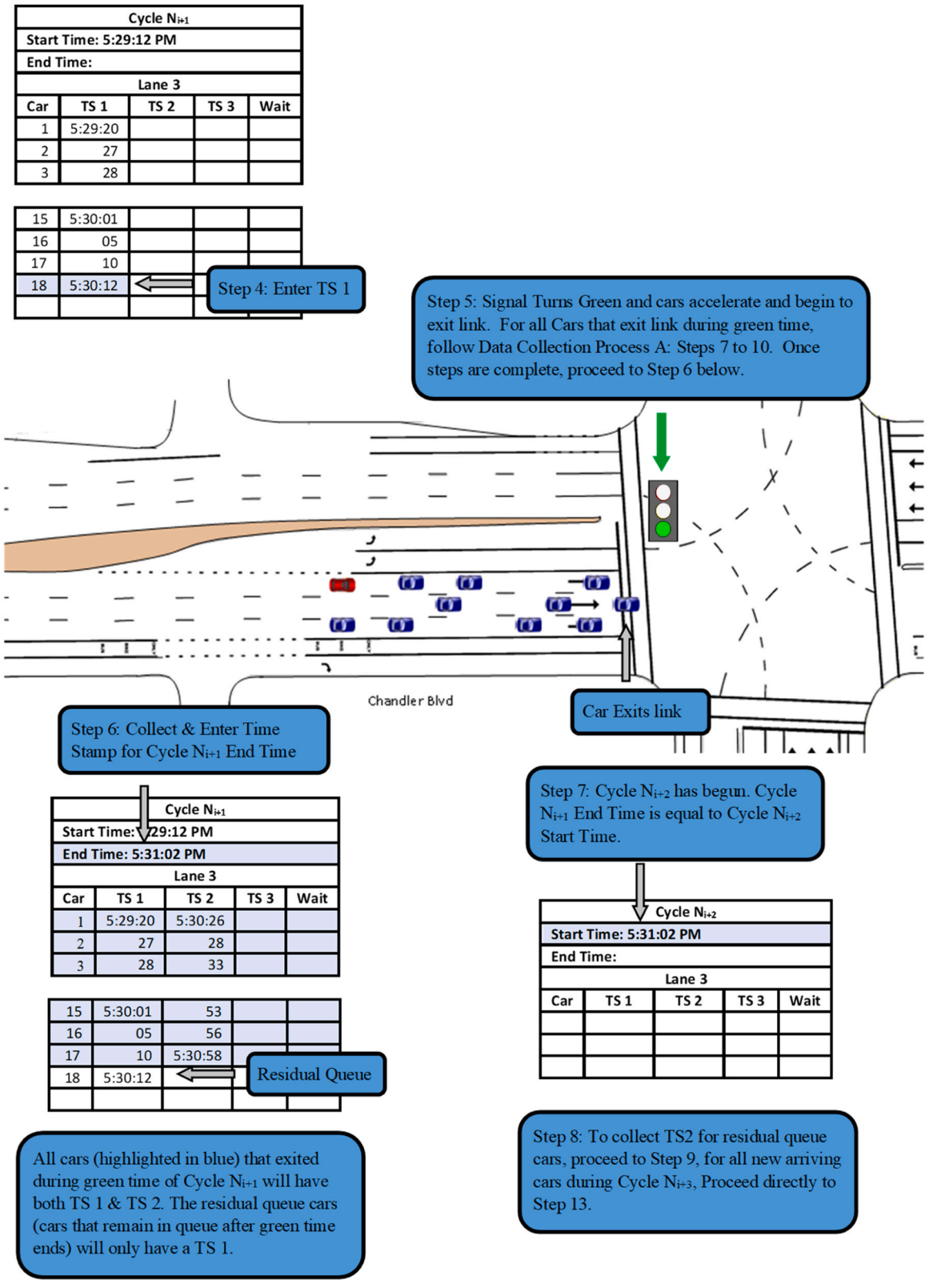
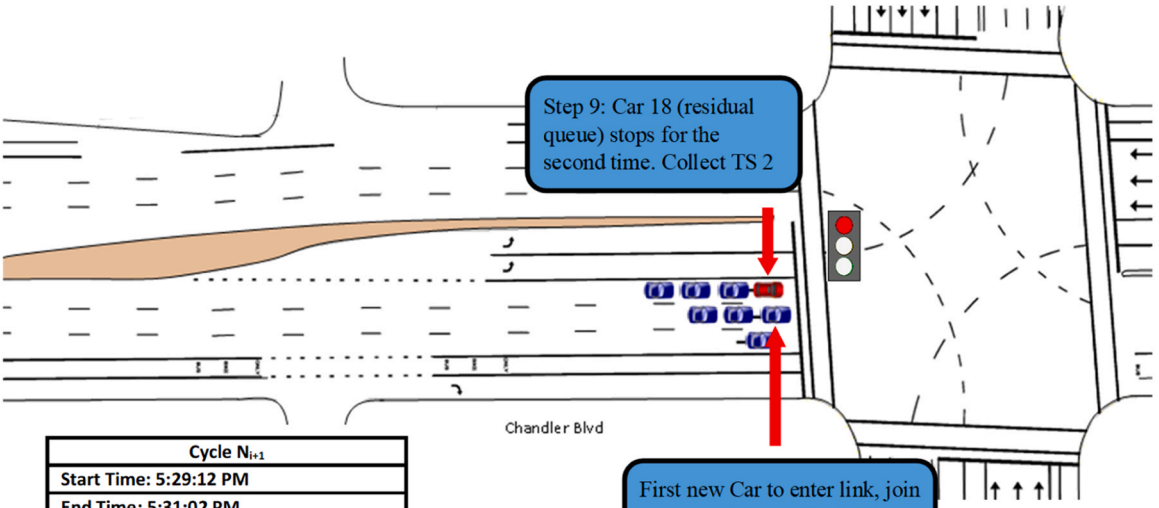


Fig. 6b. Data collection process B: Cycles with a residual queue (Step 4 to Step 8).



Cycle N_{i+1}				
Start Time: 5:29:12 PM				
End Time: 5:31:02 PM				
Lane 3				
Car	TS 1	TS 2	TS 3	Wait
1	5:29:20	5:30:26		
2	27	28		
3	28	33		
15	5:30:01	53		
16	05	56		
17	10	5:30:58		
18	5:30:12	5:31:06		

Step 10: Enter TS 2 for Car 18 in Cycle N_{i+1} Time Table.

Step 11: Repeat Step 9 and Step 10 for all residual queue cars from Cycle N_{i+1}.

Step 12: For new all cars that enter link, join queue, and stop during Cycle N_{i+2}, follow Data Collection Process A: Step 2 to Step 5. All times shall be entered into the TS 1 Column in the Table for Cycle N_{i+2} (see below). Once the above Steps are Complete, Proceed to Step 13 Below

Cycle N_{i+2}				
Start Time: 5:31:02 PM				
End Time:				
Lane 3				
Car	TS 1	TS 2	TS 3	Wait
1	5:31:10			
2	11			
3	12			
4	18			
5	22			
6	32			
7	5:32:01			

First Car to Arrive in Queue for Cycle N_{i+2}

Last Car to Arrive in Queue for Cycle N_{i+2}

Fig. 6c. Data collection process B: Cycles with a residual queue (Step 9 to Step 12).

4.2. Data aggregation

To facilitate analysis, the vehicle arrival and departure time stamps must be transformed into vehicle wait times, cumulative counts of arrivals and departures, total queue lengths, and residual queue lengths. For vehicles with only two time stamps, the total wait time or delay experienced by a vehicle due to the traffic signal or queuing was determined by calculating the difference between Time Stamp

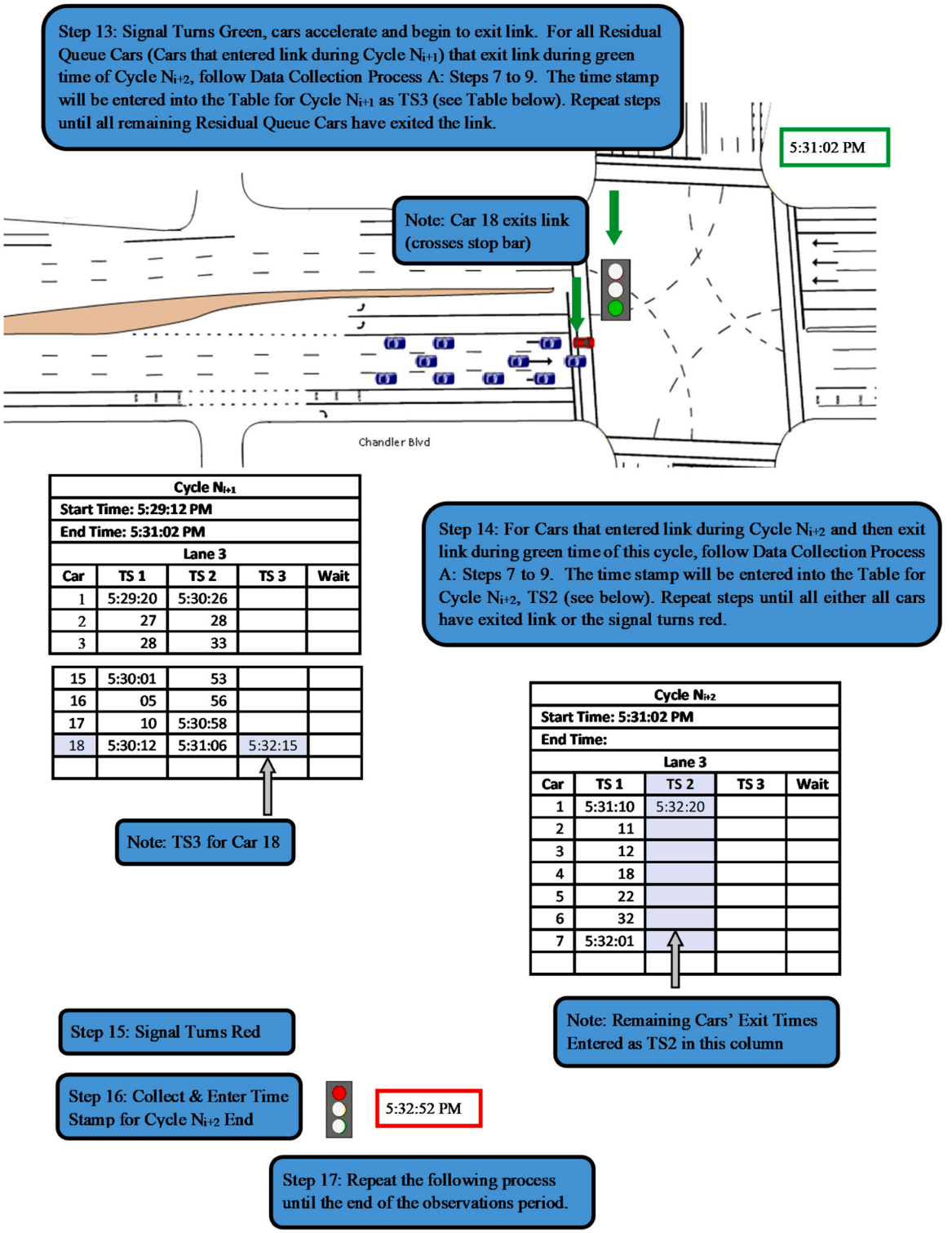


Fig. 6d. Data collection process B: Cycles with a residual queue (Step 13 to Step 17).

1 (TS 1) and Time Stamp 2 (TS 2). This resulting time duration, expressed in seconds, was then recorded in the fourth column of the table (refer to Fig. 7).

For vehicles in residual queues, which possess three time stamps, the total wait time (delay) incurred by a vehicle was calculated by finding the difference between Time Stamp 1 (TS 1) and Time Stamp 3 (TS 3). This resultant time span was also entered into the fourth

column of the table (see Fig. 7). This comprehensive process was conducted across all lanes within each cycle throughout the three-day observation period. Table 4 below is an illustrative instance of a consolidated compilation depicting the total wait time for vehicles during Day 1, specifically within Cycle 2.

It is important to note, although the observation period for the three days was from 15:30–18:00, data was not collected for the entire period. Data collection commenced once the intersection began to operate in an oversaturated condition. For Day 1, the time in which the data collection began was approximately 16:30–17:30, Day 2 approximately from 16:15–17:30, and Day 3 approximately from 16:00–17:30.

The aggregate count for the arriving and departing vehicles per cycle, the total residual queue length and cumulative arrivals and departures for each day, were integrated for each day. Table 5 illustrates these metrics for the 34 cycles of observation during Day 1 (similar tables were generated for the subsequent two days).

5. Discussion and results

5.1. Statistical analysis

Table 6 presents a sample of data collected over the three consecutive days, organized into episodes, each representing a complete queue evolution process including queue propagation and dissipation which span over multiple traffic signal cycles during peak periods. The cumulative cycles for each day's peak periods were carefully examined and divided into oversaturation episodes.

Day 1's peak period comprised of 34 cycles, which were further segmented into 5 oversaturation episodes. For Day 2, there were 43 cycles during the peak period, resulting in 6 oversaturation episodes. Similarly, Day 3 had 42 cycles, leading to 6 oversaturation episodes as well. The number of cycles per oversaturation episode varied, ranging from as little as 4 cycles to as high as 17 cycles before the residual queue eventually dissipated.

5.2. Oversaturation episode identification

To identify distinct oversaturation episodes within each day's peak period, a scatterplot (see Fig. 8) was generated using real-time residual queue lengths for each cycle (similar to the data presented in Table 5 for Day 1). The data points were connected by straight lines to facilitate visual comparison of individual cycle queue lengths over the peak period. Remarkably, the queue values exhibited a pattern of buildup to a maximum value, followed by dissipation either completely or to a minimal value (less than 4). This cyclic behavior repeated continuously throughout the oversaturation period. The clear and repetitive nature of this phenomenon made the task of dividing the peak period into episodes straightforward.

5.3. Episode breakdown

The traffic signal governing this intersection adheres to a fixed cycle length (C) of 110 seconds, with individual phase durations of red time (r_e) and green time (g_e) amounting to 72 seconds and 38 seconds, respectively. Notably, there are no occurrences of gap-outs during the observed peak period. The cycle's initiation is marked at t=0 cycles (0 seconds), culminating at t=1 cycle (110 seconds). For the sake of analytical clarity, time is herein expressed in cycles, unless explicitly indicated otherwise.

For a comprehensive dissection of the dataset, Fig. 9 elucidates the breakdown of episodes used for data analysis, particularly

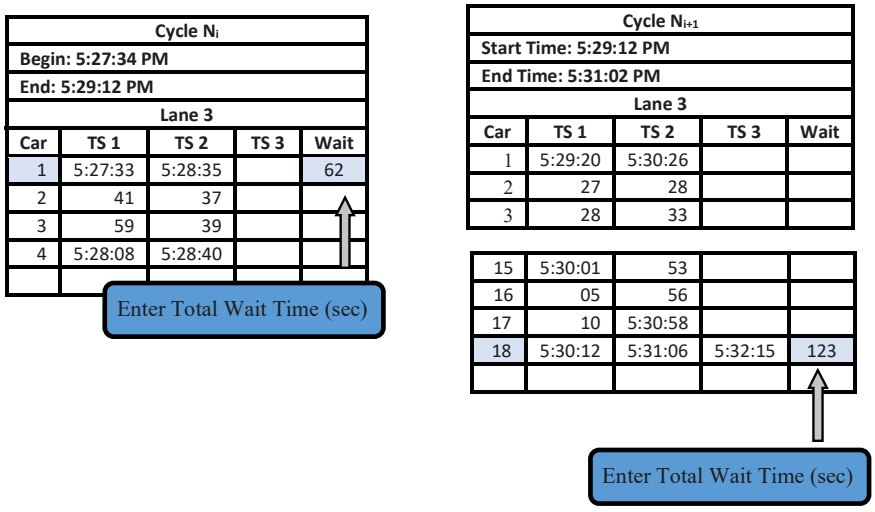


Fig. 7. Vehicle wait times.

Table 4
Example vehicle wait times for each lane during Cycle 2 of Day 1.

Day 1, Cycle 2		Day 1, Cycle 2		Day 1, Cycle 2	
Lane 1		Lane 2		Lane 3	
Car	Wait (s)	Car	Wait (s)	Car	Wait (s)
1	68	1	45	1	61
2	45	2	44	2	62
3	31	3	18	3	46
4	23	4	16	4	39
5	24	5	16	5	18
6	21	6	17	6	18
7	25	7	18	7	19
8	27	8	20	8	19
9	29	9	20	9	19
10	29	10	22	10	21
11	30	11	12	11	21
12	31	12	13	12	22
13	27	13	15	13	22
14	100	14	15	14	21
15	100	15	16	15	102
16	92	16	10	16	104
		17	6	17	105
		18	84	18	106
		19	87	19	107
		20	86		
		21	90		
		22	91		
		23	93		

focusing on the flow rate (measured in vehicles per cycle) over sequential cycles. This designated episode spans eight cycles, commencing at t=0 cycles and concluding at t=8 cycles. In specific terms, Cycle 1 commences at t=0 cycles, concluding at t=1 cycle, thus yielding a cycle length of 1. This logical progression continues with Cycle 2, commencing at t=1 cycle and concluding at t=2 cycles, maintaining a cycle length of 1 cycle.

The red dots placed throughout the graph denote the arrival flow rate, λ(t), quantified per cycle. This rate is determined at each cycle's culmination and mirrors the total count of vehicles arriving during the cycle's duration, encompassing both the red and green phases. For instance, Point A signifies the arrival rate, $\lambda(t)$=56 vehicles/cycle, at t=5 cycles (or equivalently, the cumulative vehicles that arrived during cycle 4) within the observed traffic conditions.

Now that the raw data has been transformed, we are able to start analyzing the real-time traffic conditions. The scatter plot depicting the maximum residual queue lengths against the number of cycles to dissipate the residual queue revealed no discernible correlation. Interestingly, episodes with the same maximum queue length displayed varying cycle lengths, indicating the dissipation process is influenced by factors beyond the maximum queue value. For instance, three episodes exhibited a maximum queue length of 12 vehicles, but their respective episode lengths were 5, 8, and 9 cycles. Additionally, an episode with a 13-vehicle queue had a shorter duration of only 6 cycles, highlighting the complexity of the dissipation dynamics.

Similarly, an attempt to correlate the observed *PQF* (m_o) from Table 7 with the corresponding maximum queue lengths didn't yield any clear relationship. Noteworthy is the instance where one episode demonstrated a *PQF* of 0.31 coupled with a 21-vehicle maximum queue extending over 17 cycles. In contrast, another episode exhibited a slightly higher PQF of 0.33, accompanied by a more substantial maximum residual queue of 29 vehicles spanning only 7 cycles. This observation accentuates the multifaceted interplay of factors governing dissipation dynamics, transcending a solitary reliance on the maximum queue length or the Peak Queue Fraction.

However, strong correlations were found between the traffic intensity coefficient, ρ, with both the *PQF* and maximum queue length, as well as the residual queue intensity index, QRII, with the congestion period. It is recommended to reevaluate the relationships among the variables listed in Table 7 by segregating the oversaturation episodes based on the order of the polynomial arrival rate function. This approach considers the possibility that these relationships might be influenced by additional factors associated with the order of the approximated polynomial arrival and queue length functions.

For a clearer perspective, Fig. 10 illustrates the Residual Queue Intensity Index (QRII)in relation to the total congestion period per episode. This graphical representation unveils a discernible pattern where higher residual queue ratios correspond to shorter episodes. In essence, episodes characterized by a majority of residual queue vehicles within the maximum queue cycle tend to exhibit briefer periods of oversaturation. This deduction implies that the extent of the maximum residual queue significantly influences the overarching duration of the saturation episode.

To unravel the intricate dynamics at play, further comprehensive investigations are warranted, aimed at comprehending the underlying determinants governing the temporal behavior of traffic queues during peak traffic periods.

Table 5
Example observed aggregated data for day 1 (34 cycles), including cumulative arrivals and departures, maximum residual queue, and arrival and departure flow rates per cycle.

Day 1					
Cycle	**Arrivals**	**Departures**	**Q_r**	**A_c**	**D_c**
1	46	46	0	46	46
2	46	43	3	92	89
3	56	49	10	148	138
4	43	43	10	191	181
5	43	44	9	234	225
6	41	38	12	275	263
7	39	45	6	314	308
8	43	49	0	357	357
9	60	45	15	417	402
10	56	42	29	473	444
11	43	50	22	516	494
12	38	50	10	554	544
13	36	41	5	590	585
14	36	41	0	626	626
15	46	44	2	672	670
16	46	45	3	718	715
17	45	42	6	763	757
18	45	45	6	808	802
19	36	36	6	844	838
20	49	40	15	893	878
21	45	47	13	938	925
22	52	48	17	990	973
23	47	45	19	1037	1018
24	42	50	11	1079	1068
25	46	45	12	1125	1113
26	37	48	1	1162	1161
27	60	52	9	1222	1213
28	35	42	2	1257	1255
29	46	43	5	1303	1298
30	46	46	5	1349	1344
31	44	41	8	1393	1385
32	50	48	10	1443	1433
33	47	47	10	1490	1480
34	34	44	0	1524	1524

Table 6
Summary of data collected for each oversaturation Episode over the three day observation period.

Episode	**λ_c**	**μ_c**	**Max Q_r**	**Max λ**	**# Cycles**						
1.1	357	357	12	56	8						
1.2	312	318	29	60	7						
1.3	572	576	19	52	13	**Episode**	**λ_c**	**μ_c**	**Max Q_r**	**Max λ**	**# Cycles**
1.4	132	142	9	60	3	3.1	188	184	15	65	4
1.5	302	311	10	50	7	3.2	280	285	13	54	6
						3.3	766	771	21	60	17
Episode	**λ_c**	**μ_c**	**Max Q_r**	**Max λ**	**# Cycles**	3.4	218	226	15	55	5
2.1	265	265	21	59	6	3.5	295	307	15	51	7
2.2	258	264	9	50	7	3.6	362	366	14	58	8
2.3	406	407	12	56	9						
2.4	217	225	12	60	5						
2.5	558	562	17	60	12						
2.6	424	432	18	55	9						

5.4. Analysis of cumulative arrival and departure rates

For each instance of oversaturation, we collected and graphed real-time cumulative vehicle arrival and departure data to gain insights into the cyclic nature and degree of saturation. Depicted in Fig. 11 is Episode 1.1's cumulative arrival rate, $A(t)$ and the cumulative departure rate, $D(t)$ per time (cycle).

For the analysis and polynomial approximation of arrival flow rates, we treat the departure rate as a constant. Consequently, the cumulative departure rate should be represented as a uniform linear rate rather than a piece-wise function (depicted by the solid blue zigzag line). To achieve this conversion, we explored two methods: the first involved using Excel's trend line tool, resulting in a linear rate with a slope of 45.50, while the second method employed the average cumulative departure rate of 44.63 as the slope of the linear

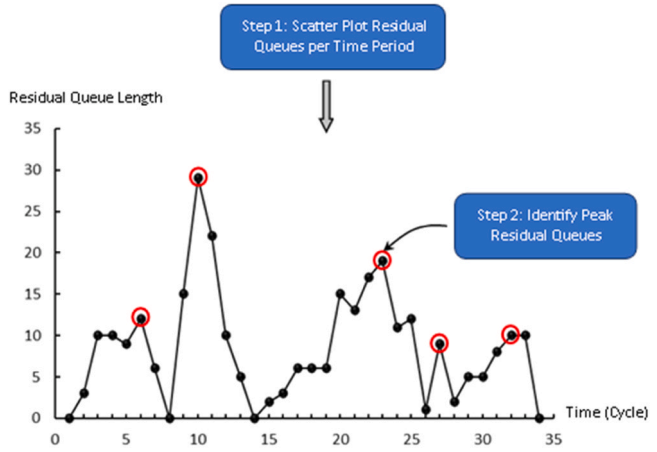


Fig. 8a. Procedure for breakdown of episodes for Day 1 (residual queue lengths per cycle).

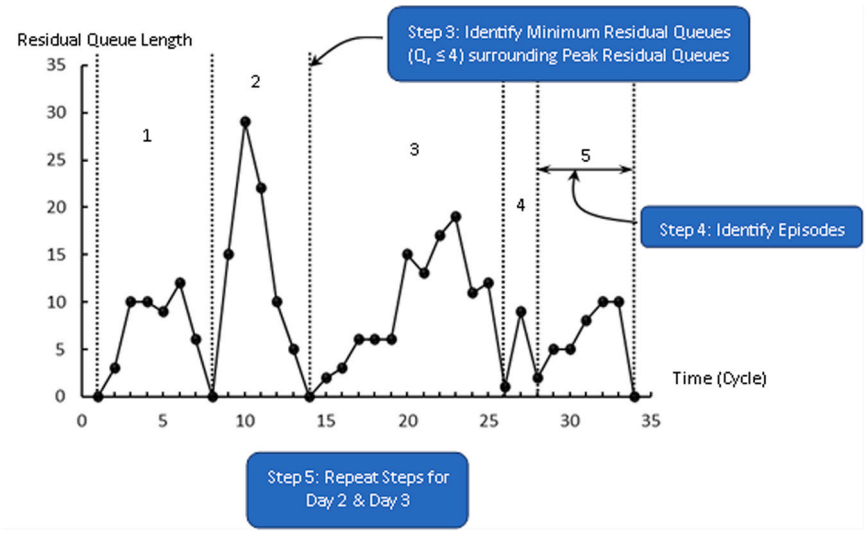


Fig. 8b. Procedure for breakdown of episodes for Day 1 (identification of episodes).

rate. The average departure rate (shown as the blue dashed line in Fig. 11 aligned better with observed values and was selected for subsequent flow rate analysis.

For the episode, the calculated average cumulative arrival rate closely matches the average cumulative departure rate, both at approximately 44.60 vehicles per cycle. This suggests a saturation rate of 1.0 for the intersection, however, further observation (see Table 8) reveals the saturation level begins to exceed 1.0, after the first cycle, and forcing the intersection to operate in an over-saturated state. This leads to the formation of a residual queue at the end of each subsequent cycle, which is represented by the vertical gap between the cumulative arrival and departure curves in Fig. 11. Importantly, this oversaturation is temporary and persists only until the conclusion of cycle 8, returning the saturation rate to 1.0.

5.5. *Approximation of arrival rates with polynomial curves*

An additional set of scatter plots was created, each corresponding to an oversaturation episode, visually representing arrival flow rates per cycle alongside the average departure flow rate specific to the respective episode. To determine the most suitable approach, careful consideration was given to the inherent characteristics and stochastic tendencies within the observed data. Depending on these factors, a deliberate choice was made between quadratic and cubic functions to achieve the best fit for the data. The systematic process for approximating real-time observed arrival flow rates using polynomial functions is clearly outlined in Fig. 12. (The flow rate for the following process is in terms of vehicles per cycle.)

Following a similar methodology, we created a scatterplot and the corresponding trend function for the residual queue length, like the approach employed for the arrival flow rate. Building upon the processes established in our previous derivation, it becomes

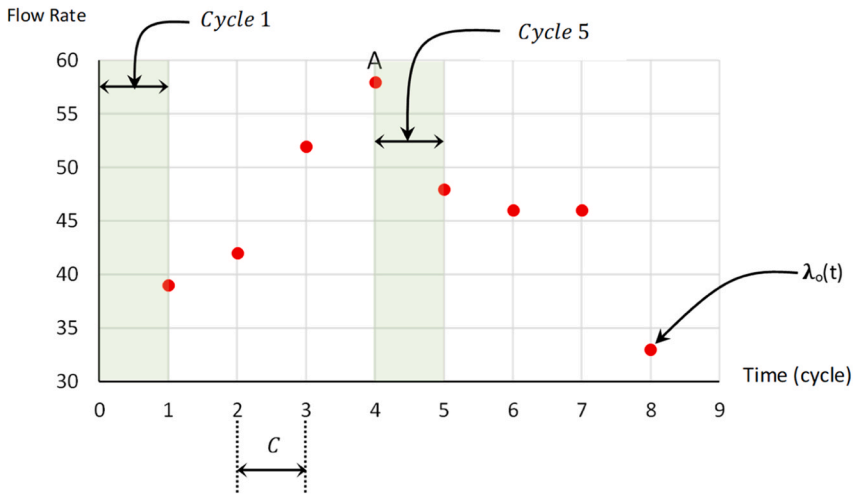


Fig. 9. Example illustration of cycle lengths and arrival rate observations for an episode.

Table 7
Summary of collected and aggregated data for all episodes.

Episode	Max Q_r	Max λ	# Cycles	PQF m_o	Total Q_r	QRII	ρ	P
1.1	12	56	8	0.71	50	4.17	1.25	7
1.2	29	60	7	0.33	81	2.79	1.36	6
1.3	19	52	13	0.75	111	5.84	1.17	12
1.4	9	60	3	0.50	12	1.33	1.27	2
1.5	10	50	7	0.67	40	4.00	1.13	6
2.1	21	59	6	0.60	58	2.76	1.34	5
2.2	9	50	7	0.50	30	3.33	1.33	6
2.3	12	56	9	0.50	67	5.58	1.24	8
2.4	12	60	5	0.25	27	2.25	1.33	4
2.5	17	60	12	0.55	116	6.82	1.28	11
2.6	18	55	9	0.38	89	4.94	1.15	8
3.1	15	65	4	0.33	24	1.60	1.41	3
3.2	13	54	6	0.60	41	3.15	1.14	5
3.3	21	60	17	0.31	145	6.90	1.32	16
3.4	15	55	5	0.75	35	2.33	1.22	4
3.5	15	51	7	0.50	49	3.27	1.16	6
3.6	14	58	8	0.43	49	3.50	1.27	7

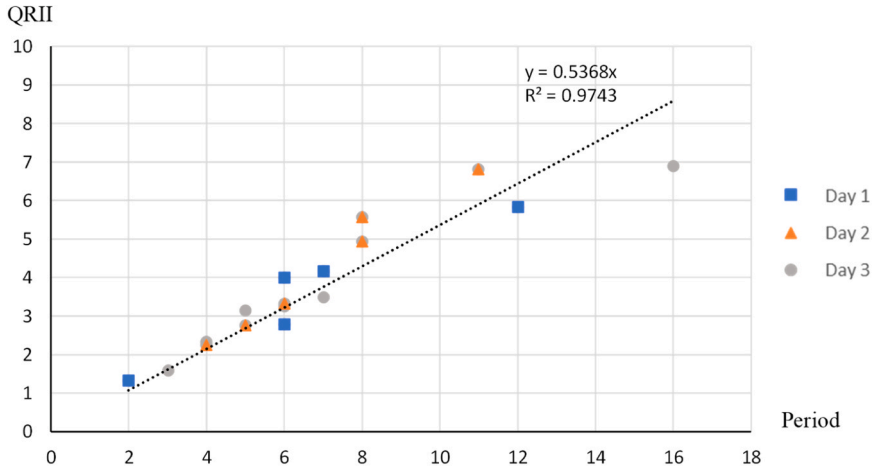


Fig. 10. Congestion period with associated residual queue intensity index per episode.

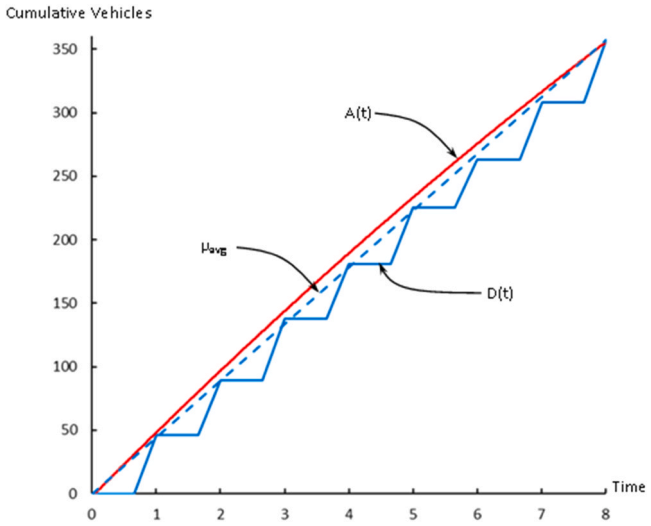


Fig. 11. Episode 1.1 cumulative arrival and departure curves.

Table 8
Episode 1.1 cumulative arrival and departures, maximum residual queue lengths, and volume to capacity ratio for each cycle.

Cycle	A_c	D_c	Q_r	λ/μ
1	46	46	0	1.00
2	92	89	3	1.03
3	148	138	10	1.07
4	191	181	10	1.06
5	234	225	9	1.04
6	275	263	12	1.05
7	314	308	6	1.02
8	357	357	0	1.00

apparent that queue length can be determined by integration or by finding the antiderivative of the arrival flow rate function. To clarify, in Episode 1.1, where the arrival flow rate function takes on a cubic form, the resulting approximated queue length function exhibits a fourth-order characteristic (See Fig. 13).

During Step 3 A of the arrival flow rate approximation process, a piecewise function was employed to fit the observed data, and the average R-squared value for this series of functions was calculated. In Step 4 of the queue length approximation process, a quadratic polynomial function was selected, and its associated R-squared value was also collected. The R-squared values associated with the polynomial functions selected in Step 3 C or Step 3D and in Step 4 were compared to those given to the piece wise function and associated queue length function. For all but three of the episodes compared, the R-squared values from the set of functions associated with the quadratic or cubic arrival flow functions was higher than with the set of functions associated with the piece wise. Two of the three, had lower or nearly equal R-squared values for the arrival flow functions, but the R-squared values for the queue length functions for the polynomial arrival flow functions were higher than those associated with the piece wise function, therefore the quadratic or cubic arrival flow function was selected and its associated cubic or 4th order queue length function. The last episode, with only three cycles, presented identical R-squared values for both types of functions. However, the polynomial arrival flow function was selected as it performed better in predicting the onset of congestion. Further studies with a larger dataset of episodes featuring short congestion periods are recommended to determine the 'best' fit more conclusively.

5.6. *Episodes featuring time-dependent stochastic arrival rate functions ($\gamma > 0$)*

In oversaturation episodes featuring a cubic function characterized by a positive curvature parameter, the anticipated pattern of arrival flow rates (veh/cycle) aligns with the graph presented below in Fig. 14. The confinement of the arrival rate function's Peak Queue Fraction, m_t, is expected to coincide with the stipulated boundary: $m_t \in \left[\frac{1}{2}, \frac{2}{3}\right)$

Table 9 presents a compilation of the twelve oversaturation episodes approximated by cubic functions with positive curvature and their corresponding associated observed and approximated traffic data. The percentage approximation error for arrival rates at the peak demand using the continuous-time cubic model, is denoted as $\Delta(\lambda)$, Results from Table 9 show an average absolute percentage deviation of approximately 8.0 % in terms of $\Delta(\lambda)$, indicating a satisfactory quality of approximation for the proposed model.

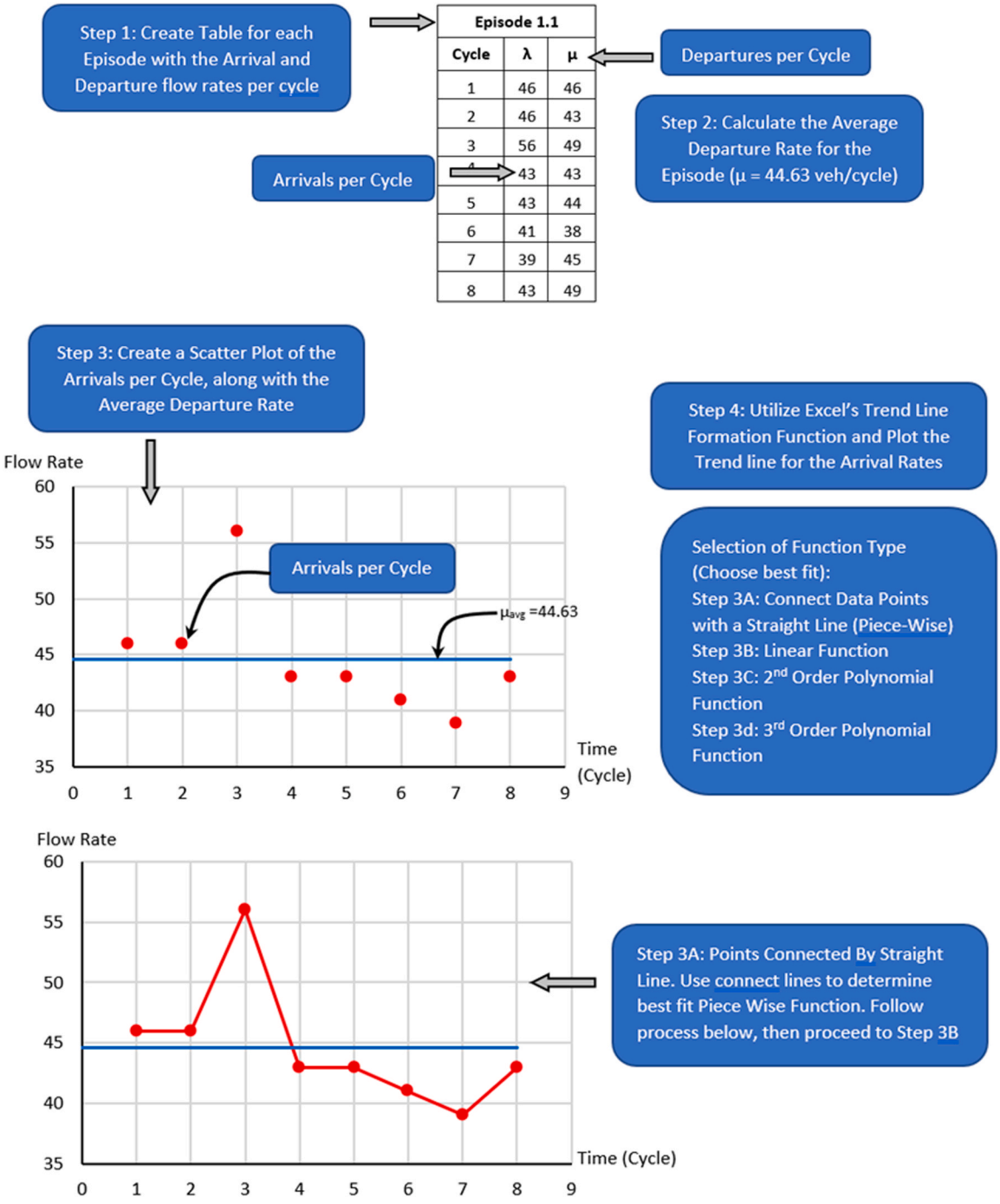


Fig. 12a. Procedure for approximating arrival flow rate functions (Steps 1 to Step 3A).

Table 10 summarizes the R-squared values associated with aligning the approximated arrival rate functions with observed traffic conditions. About half of the twelve episodes achieved acceptable R-squared values, while only two episodes (2.1 and 3.6) deviated from prescribed PQF boundary conditions.

Episodes (1.1, 2.2, 2.3, 2.5, 3.2, and 3.3) with lower R-squared values showed a diminished correlation with the cubic's parabolic shape before and after t_1. Episode 3.2, with the highest R-squared value (0.78) of the six episodes, maintained escalating arrival rates to the maximum without losing momentum, while others displayed declining rates before a surge at t_1. The magnitude of the surge correlated with lower R-squared values.

Most episodes with lower R-squared values faced challenges predicting maximum residual queues, deviating within 1.19 cycles on

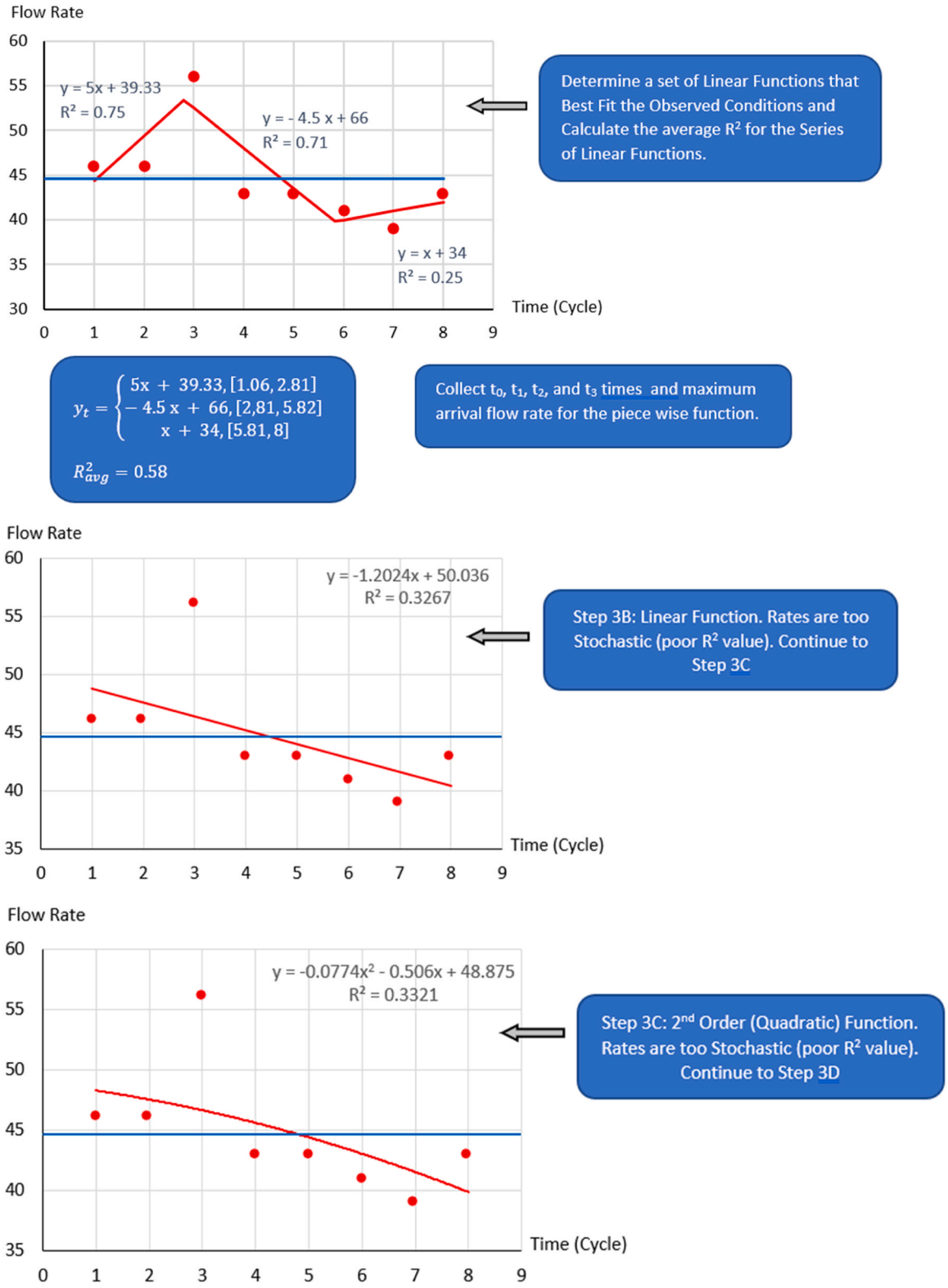


Fig. 12b. Procedure for approximating arrival flow rate functions (Step 3 A (cont.) to Step 3 C).

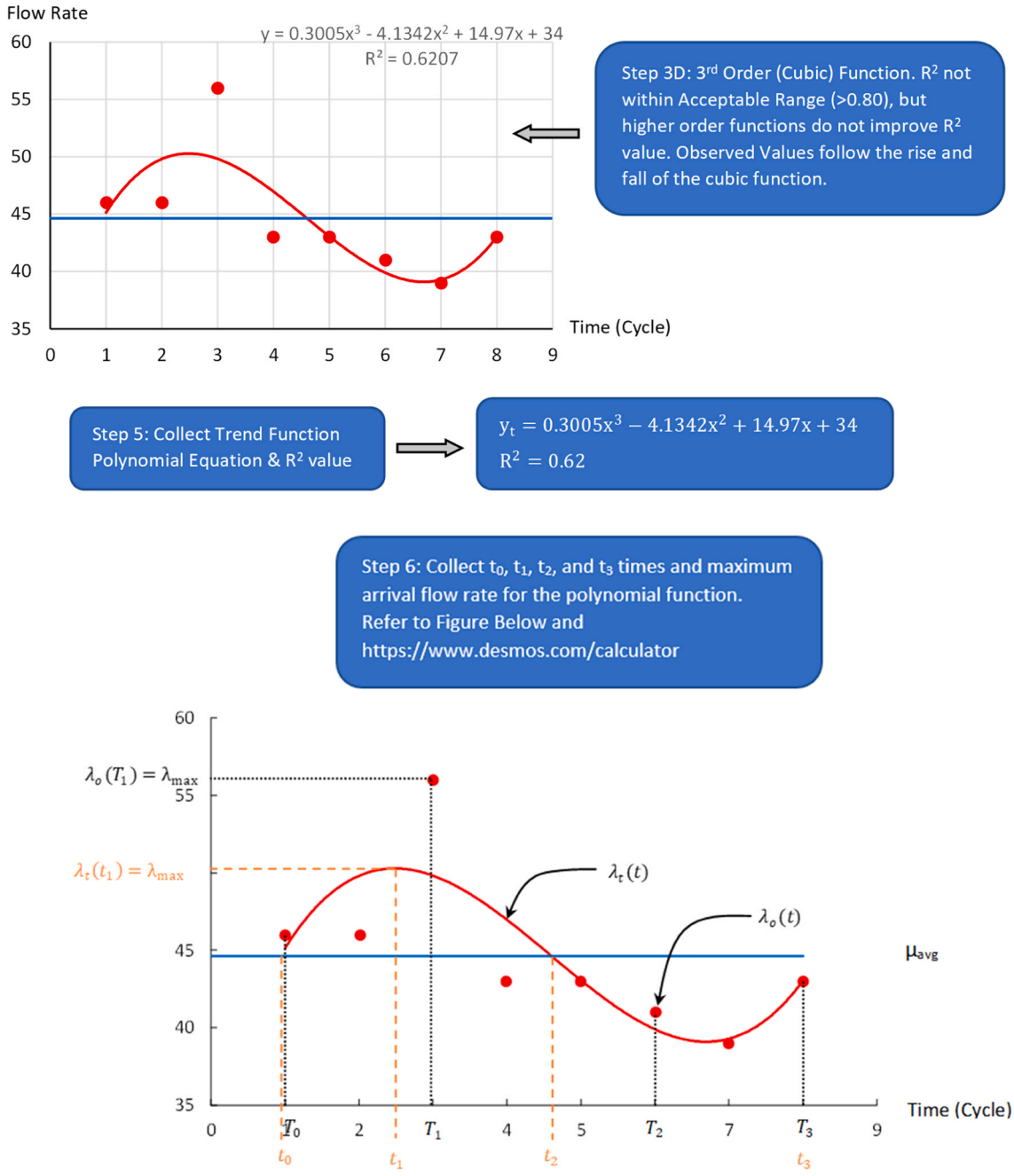


Fig. 12c. Procedure for approximating arrival flow rate functions (Step 3D to Step 6).

average. Highly stochastic episodes (2.3, 2.5, and 3.3) struggled more, with average deviations of 2.19 cycles, particularly Episode 3.3.

Predicted λ_{max} values averaged 7.6 % lower than observed values. Surprisingly, no clear correlation emerged between R-squared values and the discrepancy in maximum arrival flow rates. Episodes with low R-squared values predicted maximum arrivals within 0.59 cycles on average, similar to those with acceptable R-squared values.

Episodes 2.3, 2.5, and 3.3 had very low R-squared values, showcasing stochastic arrivals flows (as show in Fig. 15) characterized by oscillations between surges and drops, challenging cubic curve alignment. Signal coordination dynamics likely influenced these patterns.

Conversely, episodes with strong correlations had non-stochastic arrivals but faced challenges predicting the occurrence of the maximum residual queues, with deviations averaging 1.18 cycles late and in some episodes 4 cycles later. Despite these variations, these cubic curvature approximations appear to be generally compatible with the roadway conditions.

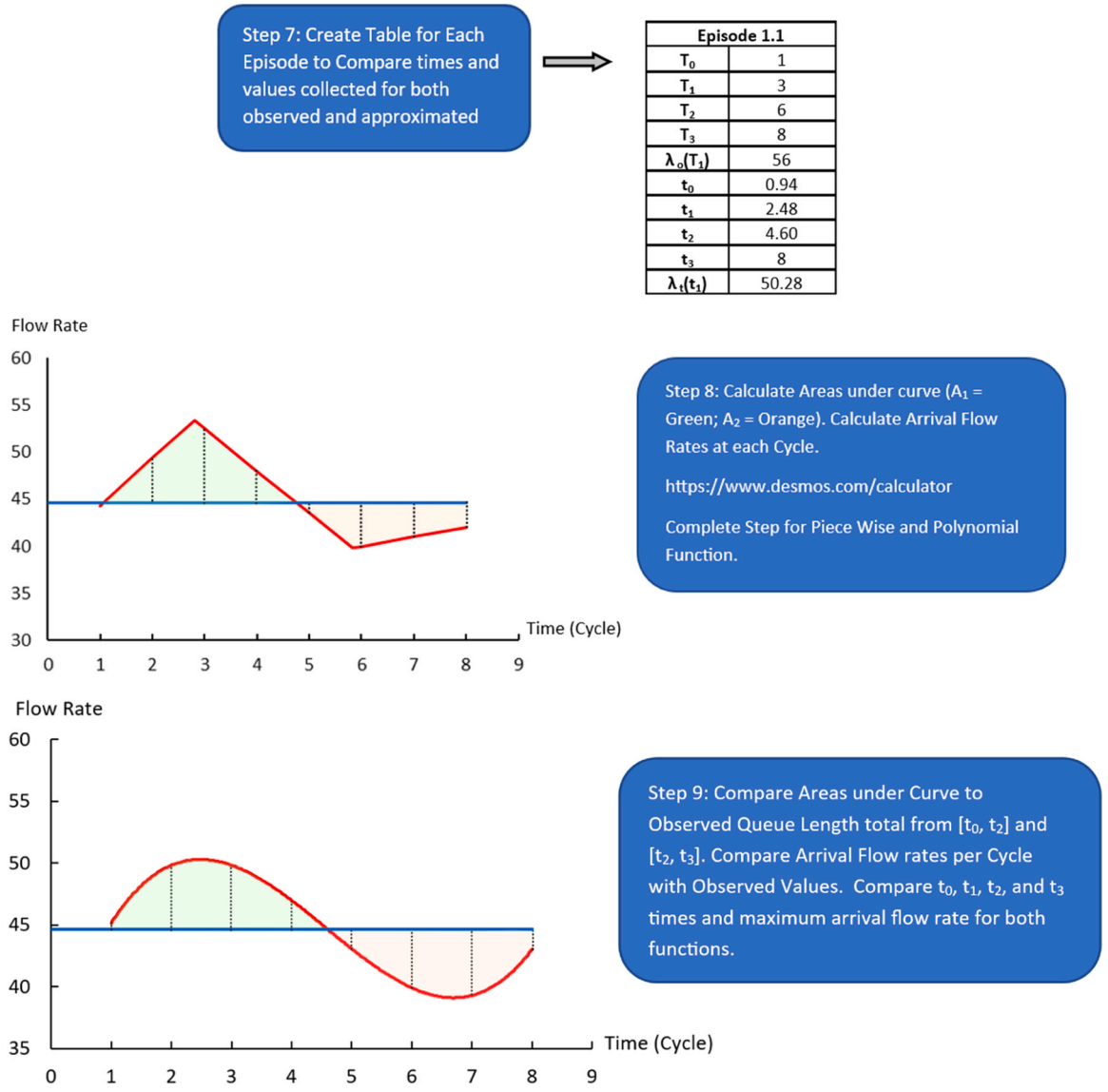


Fig. 12d. Procedure for approximating arrival flow rate functions (Step 7 to Step 9).

Only two episodes had approximated PQF (m_t) values that deviated from boundary conditions. The t_3 value occurred before the local minimum and third root, resulting in most arrivals preceding maximum residual queue formation, contrary to the characteristic of the PQF, but this discrepancy didn't strongly correlate with R-squared values.

5.6.1. Approximated residual queue length function of higher order

As previously discussed, episodes characterized by cubic polynomial approximations for arrival flow rates are associated with fourth-order polynomial functions used to approximate residual queue lengths, as illustrated in Fig. 16. Table 11 presents data on observed and predicted values, maximum residual queue times, Δ queue length, and R-squared values.

Among these functions, eight episodes exhibit strong correlations (R-squared $\geq$ 0.91), while four have lower R-squared values (0.63–0.78).

Three of the lower performing episodes, 2.3, 2.5, and 3.3, resulted from erratic, stochastic arrivals causing queue length variations. Departure rates also surges, 14.1 % above average, following the maximum queue length, further complicating predictions. Similar trends appear in maximum queue lengths, with predictions averaging 26.7 % below observed values. Episodes 2.3 and 2.5 differ by 15.4 %, while Episodes 3.3 and 3.6 show 38.1 % disparities. While the first three episodes contend with dynamic departure rates, Episode 3.6 lacks such dynamism, beginning with a departure rate notably lower (15 % below the episode's average) – likely contributing to the function's prediction of a much later t_2 (almost three cycles late).

Episodes with acceptable R-squared values average 7.0 % lower maximum queue lengths. Dynamic arrivals and departures, rapid

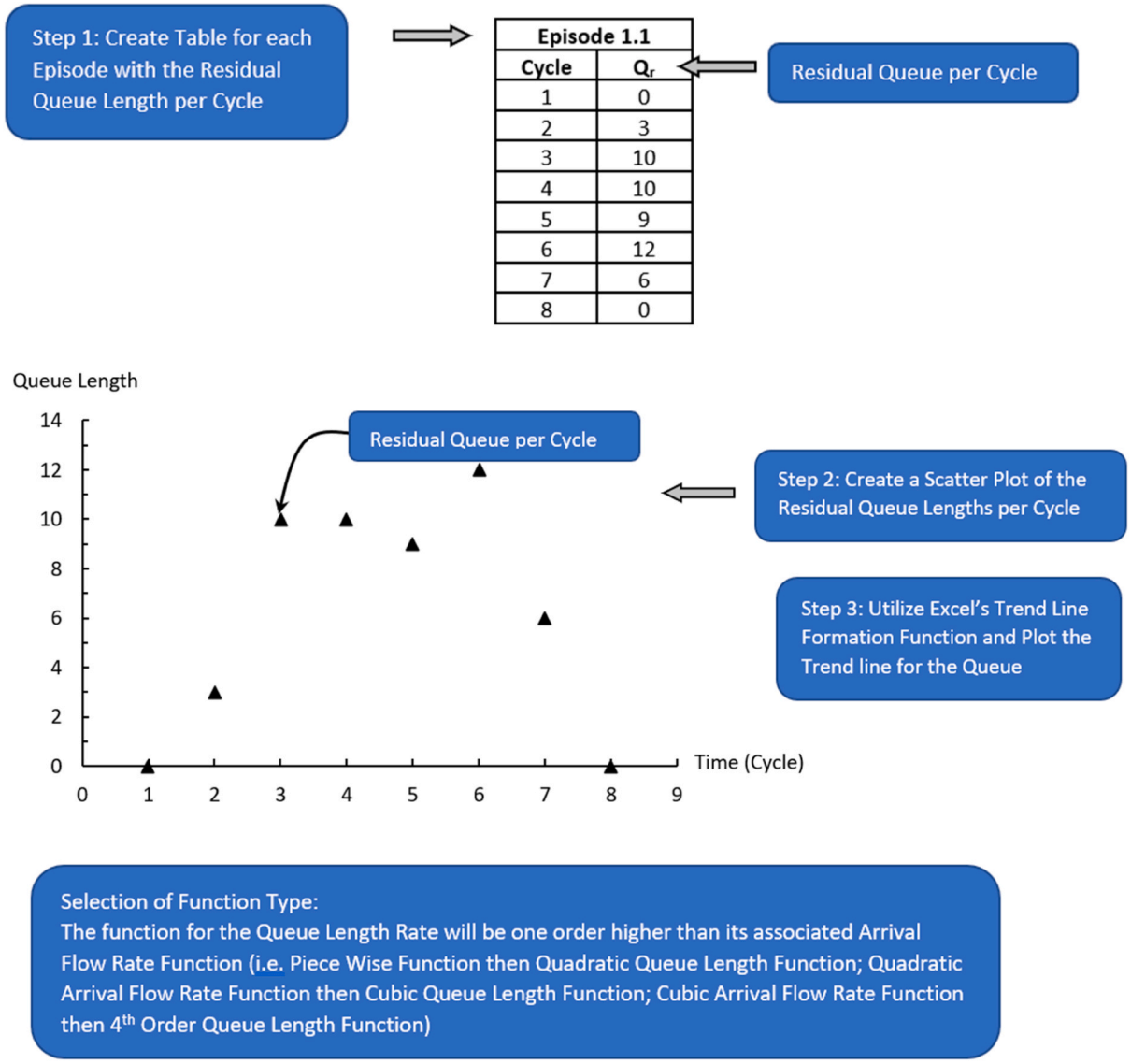


Fig. 13a. Procedure for approximating queue length functions (Step 1 to Step 3).

or gradual queue buildup, and t_1 and t_2 disparities contribute to this complexity. While some episodes exhibit rapid queue buildup due to dynamic arrivals, others experience a more gradual accumulation influenced by dynamic departures. Additionally, disparities in t_1 and t_2 alignment further underscore the complexity of the observed phenomena. Despite the slight variations from observed conditions, these episodes' fourth-order functions align both visually and computationally with observed data, indicating compatibility with traffic dynamics during oversaturation episodes.

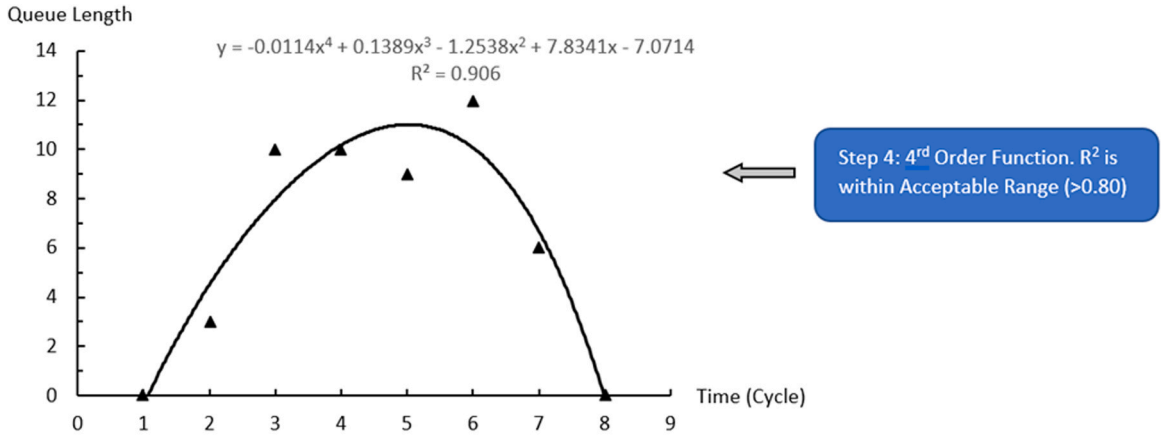
5.6.2. Total delay

The total delay for the episode is made up of two types of delay (see previous discussion) the control (uniform delay) and the overflow(queueing) delay. From our derivation for cubic functions, we know the queue delay can be expressed as (Eq. (28))

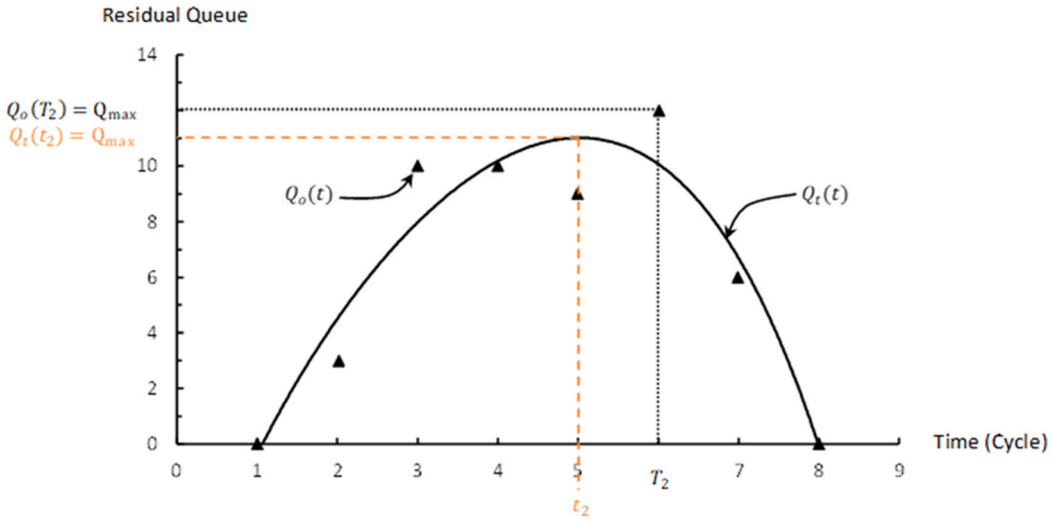
$$W(t) = \frac{Q(t)}{\mu} = \frac{\gamma(t-t_0)^2}{\mu}\left[\frac{1}{4}(t-t_0)^2 - \frac{1}{3}(t-t_0)(t_3-t_0)\left[m+\left[\frac{3-4m}{4-6m}\right]\right] + \frac{1}{2}(t_3-t_0)^2 m\left[\frac{3-4m}{4-6m}\right]\right]. \tag{28}$$

The approximated total queueing delay for each episode can be calculated by taking the area underneath the $Q(t)$ curve for the congestion period. Table 12 illustrates the comparison between the observed and approximated total queueing delays for the episodes who have 4th order queue length functions and cubic arrival flow rate functions with positive curvature.

When comparing the observed total queuing delay with the calculated total queueing delay, the episodes that show a larger discrepancy are not associated with the performance of the queue lengths function with the observed conditions. The average delta in



Step 6: Collect t_0, t_2, and t_3 times and maximum Queue Length for Episode for the polynomial function. Refer to Figure Below and https://www.desmos.com/calculator



Step 7: Create Table for Each Episode to Compare times and values collected for both observed and approximated

Episode 1.1	
T_0	0
T_2	5
T_3	8
$Q_o(T_2)$	12
t_0	1.06
t_2	5.01
t_3	7.98
$Q_t(t_2)$	10.99

Fig. 13b. Procedure for approximating queue length functions (Step 4 to Step 7).

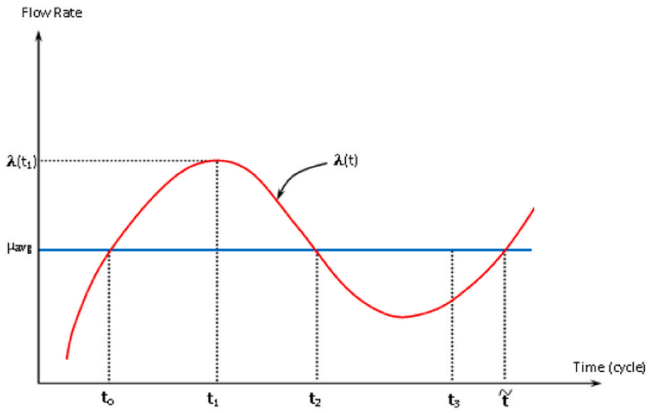


Fig. 14. Graph of approximated cubic arrival rate function (positive curvature).

Table 9
Episodes with time-dependent stochastic cubic arrival rate functions (positive curvature), where Δ(λ) denotes the percentage approximation error.

Episode	T_0	t_0	T_1	t_1	T_2	t_2	T_3	t_3	P_o	P_t	$\tilde{t}$	$\lambda_o(T_1)$	$\lambda_t(t_1)$	Δ(λ)
1.1	1	0.94	3	2.48	6	5.15	8	8	7	7.06	8.22	56	50.28	11 %
1.2	1	1.04	2	2.38	3	4.2	7	7	6	5.96	7.48	62	57.46	8 %
2.1	1	1.19	4	2.83	4	4.67	6	6.00	5	4.81	18.68	59	56.99	3 %
2.2	1	1.69	4	2.88	4	4.57	7	6.89	6	5.2	6.89	50	46.18	8 %
2.3	1	2.08	3	3.77	5	6.19	9	9	8	6.92	9.32	56	53.82	4 %
2.4	1	1.12	2	1.95	2	3.05	5	5	4	3.88	5.14	60	56.15	7 %
2.5	1	2.58	5	4.78	7	7.9	12	12	11	9.42	12.37	60	56.54	6 %
3.1	1	1.08	2	1.74	2	2.69	4	4	3	2.92	4.00	65	67.24	−3 %
3.2	1	1.21	3	2.42	4	4.03	6	6	5	4.79	7.25	54	53.68	1 %
3.3	1	1.43	5	5.33	6	10.47	17	17	16	15.57	21.43	60	50.90	16 %
3.5	1	1.77	2.00	3.07	4.00	4.93	7	7	6	5.23	8.13	51	53.45	−5 %
3.6	1	2.12	4	4.17	4	6.52	8	8	7	5.88	20.73	58	53.42	8 %

Table 10
R-squared values with observed & approximated peak queue fraction factors for cubic arrival rate functions (positive curvature).

Episode	R^2	PQF (m_t)
1.1	0.62	0.60
1.2	0.92	0.53
2.1	0.85	0.72
2.2	0.75	0.55
2.3	0.34	0.59
2.4	0.81	0.50
2.5	0.20	0.56
3.1	1.00	0.55
3.2	0.78	0.59
3.3	0.37	0.58
3.5	0.90	0.60
3.6	0.86	0.75

total delay is only 4 % for the twelve episodes and the average delta for the average delay per vehicle per episode is only 0.05 cycles or 5.5 seconds.

5.7. Episodes with time-dependent stochastic cubic arrival rate functions (negative curvature)

For saturation episodes featuring a cubic function characterized by a negative curvature parameter, the anticipated pattern of arrival rates aligns with the graph presented below in Fig. 17. The confinement of the arrival rate function's PQF factor, $_t$, is expected to coincide with the stipulated boundary: $m_t \in \left(\frac{2}{3}, \frac{3}{4}\right]$

This analysis delves into relationships between the function's, R-squared values, third root behavior, scale parameter, adherence to boundaries, and arrival pattern variations in Tables 13 and 14 for the three episodes.

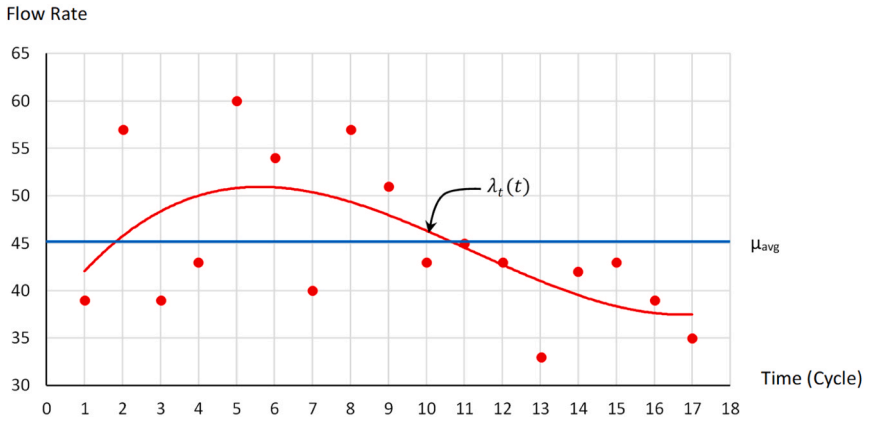


Fig. 15. Example of highly stochastic arrival flow during a saturation episode with low R^2 value for cubic function.

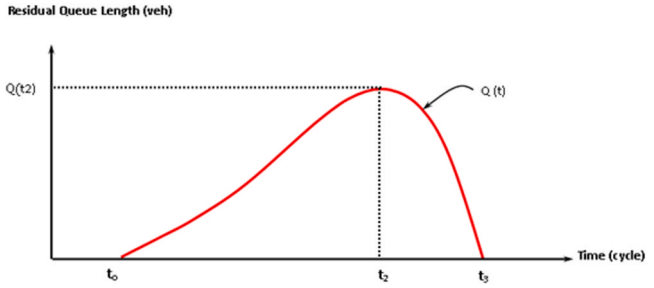


Fig. 16. Example of 4th order approximation of residual queue values.

Table 11
R-squared values with observed & approximated queue lengths for episodes.

Episode	T_2	t_2	$Q_o(T_2)$	$Q_t(t_2)$	ΔQ	R^2
1.1	6	5.01	12	10.99	9 %	0.91
1.2	3	3.20	29	25.52	13 %	0.95
2.1	4	3.35	21	19.04	10 %	0.91
2.2	4	4.17	9	8.64	4 %	0.93
2.3	5	5.33	12	10.66	12 %	0.71
2.4	2	1.73	12	12.68	−6 %	1.00
2.5	7	6.39	17	14.07	19 %	0.78
3.1	2	1.88	15	15.18	−1 %	1.00
3.2	4	3.30	13	11.82	10 %	0.92
3.3	6	6.34	21	13.88	41 %	0.65
3.5	4	4.06	15	14.56	3 %	0.95
3.6	4	4.59	14	9.80	35 %	0.63

Episode 1.3's lower R-squared value results from significant random arrival flow rate fluctuations cycle-to-cycle, resulting in the delayed start of four cycles and a three-cycle shorter period. The episode has an anomaly where the occurrence of the lowest arrival rate in the middle of the episode results in a delay of the maximum queue length by two cycles.

Episode 1.5's diminished acceptability is linked to inconsistent initial arrivals and a sudden 13.6 % surge in peak arrivals. Episode 3.4 shows a similar pattern, with declining arrivals before t_1 followed by a 17 % surge. Episodes 1.5 and 3.4 both have aligning t_2 and t_1 in the same cycle, which the function cannot duplicate and therefore, places the maximum queue length in the cycle after the maximum arrival flow rate. This pattern introduces demand dynamics: a drop (avg. 4.5 %) followed by a surge (avg. 14.5 %), aligning with R-squared values.

All three episodes accurately predict peak arrival flow rates in the same cycle as observed. No clear pattern connects predictive accuracy and approximated peak arrival rates. Cubic approximation conformity to observed conditions seems to be inherently connected to the maximum arrival rate of the episode.

Analysis of the three episodes reveals consistent PQF deviations (0.04) from boundary conditions, $_t \in (2/3, 3/4]$, along with the third root ($\tilde{t}$) occurring on average about 7.10 cycles before the start of the episodes. Episode 1.5, R-squared 0.76, has a third root

Table 12
Total queueing delay (cycles) for 4th order queue length functions.

Episode	W_o	$W_{o(avg)}$	W_t	$W_{t(avg)}$	ΔW
1.1	41.57	0.84	51.25	1.03	21 %
1.2	79.48	0.94	85.52	0.97	7 %
2.1	53.34	0.92	59.91	1.03	12 %
2.2	27.5	0.92	28.48	0.95	4 %
2.3	65.64	0.98	66.48	0.99	1 %
2.4	25.05	0.93	28.86	1.07	14 %
2.5	113.67	0.99	117.79	1.02	4 %
3.1	24.18	1.01	23.98	1.00	−1 %
3.2	40.73	0.99	39.88	0.97	−2 %
3.3	143.39	0.98	131.98	0.91	−8 %
3.5	46.9	0.96	46.39	0.95	−1 %
3.6	48.15	0.98	47.26	0.96	−2 %

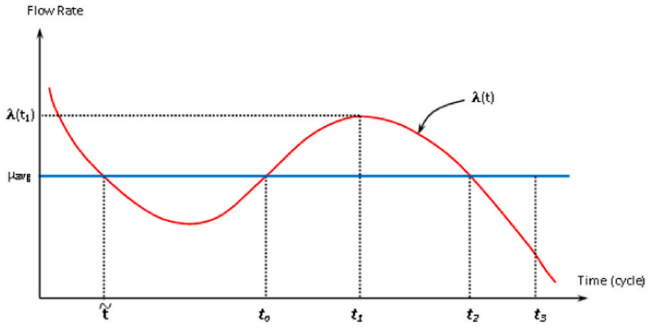


Fig. 17. Graph of approximated cubic arrival rate function (negative curvature).

Table 13
Episodes with time-dependent stochastic cubic arrival rate functions (negative curvature).

Episode	T_0	t_0	T_1	t_1	T_2	t_2	T_3	t_3	P_o	P_t	$\lambda_o(T_1)$	$\lambda_t(t_1)$	$\Delta(\lambda_t-\lambda_o)$
1.3	1	4.52	9	8.31	10	11.24	13	13	12	8.48	52	47.04	10 %
1.5	1	2.54	5	4.42	5	5.99	7	7	6	4.46	50	49.08	2 %
3.4	1	1.74	4	3.17	4	4.33	5	5	4	3.26	55	54.62	1 %

Table 14
R-squared values with observed & approximated peak queue fraction for cubic arrival rate functions (negative curvature).

Episode	R^2	PQF (m_t)
1.3	0.27	0.79
1.5	0.76	0.77
3.4	0.80	0.79

closest to the boundary at 7.56 cycles before. The last two episodes each have a PQF of 0.79. Episode 1.3, R-squared 0.27, has the furthest third root at 9 cycles before. Episode 3.4, R-squared 0.80, aligns closest to the third root at 4.73 cycles. The functions alignment with observed conditions (R-squared values) highlights the relationship the with third root proximity. Unusual data patterns, like t_2 preceding t_1 in Episode 1.3, coinciding t_1 and t_2 in Episodes 1.5 and 3.4, also may prevent boundary condition adherence.

Table 15
R-squared values with observed & approximated queue lengths for cubic arrival rate functions (negative curvature).

Episode	T_2	t_2	$Q_o(T_2)$	$Q_t(t_2)$	R^2
1.3	10	9.75	19	16.48	0.90
1.5	5	5.51	10	10.78	0.99
3.4	4	4.03	15	15.01	1.00

5.7.1. Approximated residual queue length function of higher order

The analysis of the observed residual queue data was executed through the application of a 4th order polynomial function, yielding commendable R-squared values of 0.90 or higher across all three episodes under consideration, indicating a strong alignment between the modeled predictions and the actual data (refer to Table 13). However, a noteworthy deviation from this pattern is evident in Episode 1.3, which displayed a slightly lower R-squared value compared to the other two episodes. (Table 15)

Examining the scatter plot and trend curve for Episode 1.3 reveals unique queue length patterns. This episode experiences extended periods of stable queue lengths, followed by a significant surge (averaging 145 %) likely due to random arrivals and departures flow rates. Consequently, the polynomial function for Episode 1.3 deviates notably from expectations, with a 14.1 % difference in the predicted maximum queue length, delayed by 2 cycles. In contrast, the other two episodes show strong alignment between approximated functions and observed values. Episode 1.5, while predicting a slightly delayed maximum queue occurrence, maintains overall compatibility, highlighting the reliability of the majority of the analyzed episodes.

5.7.2. Total delay

The approximated total queueing delay for the three episodes is shown in Table 16. The episodes that display a higher R-squared value associated with their 4th order queue length functions struggle more to approximate the total queue delay. The episodes have an average delta in total delay of 9 % and an average of 0.06 cycles (6.6 seconds) higher for the average delay per vehicle per episode.

5.8. Episodes with time-dependent stochastic arrival rate functions (quadratic curvature)

In saturation episodes characterized by a quadratic function, the expected pattern of arrival flow rates should align with the graph presented in Fig. 18. The confinement of the arrival rate function's Peak Queue Fraction, m_t, is expected to coincide with the stipulated boundary: $m_t = \frac{2}{3}$

Tables 17 and 18 display the episodes whose real-time traffic in oversaturated conditions were approximated by quadratic functions. The limited number of functions restricts our ability to make a conclusive analysis on their accuracy in approximating traffic with stochastic arrival patterns under oversaturation.

Episode 1.4 demonstrates a perfect fit with an R-squared value of 1.0, aligning perfectly with observed data, except for a slightly delayed t_2 due to the unique alignment of the maximum arrival flow rate and queue length in the same cycle, which is challenging for the function to replicate.

Conversely, Episode 2.6 presents a less favorable scenario with a lower R-squared value of 0.68. Episode 2.6's lower performance can be attributed to duplicate instances of maximum arrival flow rates within the episode, causing deviations in the overall period, t_1, t_2, and the timing of the maximum residual queue. Further examination of Episode 2.6 reveals an accurate prediction of the maximum arrival flow rate but a significant overestimation of the maximum queue length occurrence, placing it 4 cycles later than observed.

Episode 1.4, despite a slightly delayed t_2, maintains alignment with observed conditions for arrivals until the final two cycles, where an unexpected 7 % surge in arrivals occurs. Both episodes share an equivalent number of cycles before and after t_0 and t_2, contributing to a higher Peak Queue Fraction factor than anticipated.

5.8.1. Approximated residual queue length cubic functions

This next analysis explores the use of cubic functions to approximate queue length based on the integral of time-dependent arrival functions, focusing on two episodes (1.4 and 2.6) characterized by quadratic arrival rate functions and their alignment with observed data. Table 19 displays the arrival time, maximum residual queue length, and the corresponding R-squared value for the cubic function.

Cubic functions were fitted to the observed data, resulting in an R-squared value of 0.74 for Episode 2.6 and a perfect R-squared value of 1.0 for Episode 1.4. Episode 1.4's function closely aligns with all observed values, while Episode 2.6's cubic function matches most values except for a slight delay in predicting Q_{max} by one cycle and a 14 % discrepancy in the maximum queue length. This discrepancy may be due to the substantial residual queue observed in the second-to-last cycle, affecting the timing of Q_{max} prediction. Both episodes exhibit consistent departure rates modeled by parabolic functions, indicating uniform vehicle departure from the queue.

5.8.2. Total delay

The approximated total queueing delay for both episodes is shown in Table 20. Both episodes struggles to approximate the total queueing delay, as before, there does not seem be a correlation between this finding and the R-squared values associated with the arrival flow and queue length functions. The episodes have an average delta in total delay of 11 % and an average of 0.11 cycles (12.1 seconds) higher for the average delay per vehicle per episode.

Table 16
total queueing delay (cycles) for 4th order queue length functions.

Episode	W_o	$W_{o(avg)}$	W_t	$W_{t(avg)}$	ΔW
1.3	105.97	0.96	109.31	0.98	3 %
1.5	36.1	0.9	40.8	1.02	12 %
3.4	35.31	1.01	40	1.14	12 %

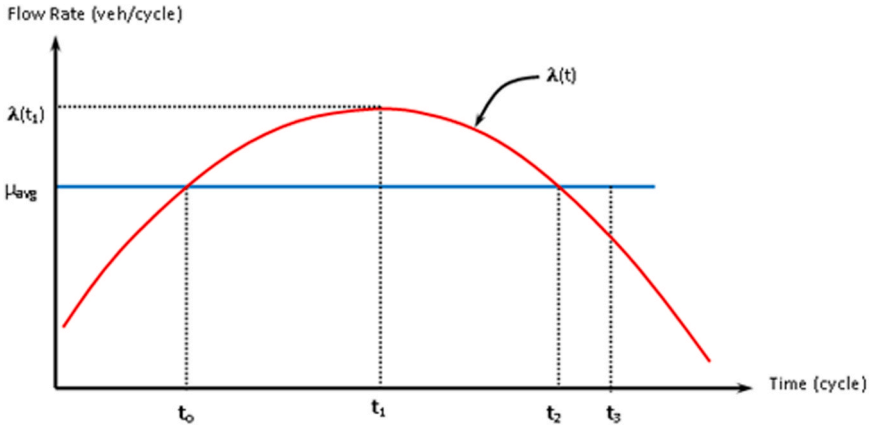


Fig. 18. Graph of approximated quadratic arrival rate function.

Table 17
episodes with time-dependent stochastic quadratic arrival rate functions.

Episode	T_0	t_0	T_1	t_1	T_2	t_2	T_3	t_3	P_o	P_t	$\lambda_o(T_1)$	$\lambda_t(t_1)$	$\Delta(\lambda_t-\lambda_t)$
1.4	1	1.25	2	1.98	2	2.71	3	3	2	1.75	60	60.01	0 %
2.6	1	2.42	4	4.86	4	7.31	9	9	8	6.58	56	55.57	1 %

Table 18
R-squared values with observed & approximated peak queue fraction factors for quadratic functions.

Episode	R^2	m_o	m_t
1.4	1	0.50	0.83
2.6	0.68	0.38	0.74

Table 19
R-squared values with observed & approximated queue lengths for quadratic functions.

Episode	T_2	t_2	$Q_o(t_2)$	$Q_t(t_2)$	R^2
1.4	2	2.03	9	9.01	1.00
2.6	4	5.06	18	15.65	0.74

Table 20
Total queueing delay (cycles) for cubic queue length functions.

Episode	W_o	$W_{o(avg)}$	W_t	$W_{t(avg)}$	ΔW
1.4	9.76	0.81	13	1.08	28 %
2.6	94.75	1.06	89.38	1	−6 %

6. Hypothesis analysis

6.1. Quadratic arrival flow rates and cubic queue length functions

Hypothesis A:

$$P \sim \left[\frac{\lambda(t_1)-\mu}{\gamma}\right]^{1/2}$$

We investigated proportional relationships, as hypothesized, between the congestion period (P) and the net flow intensity index (NFII) using three variable sets: NFII [$\lambda_o(T_1)$ and μ_{avg}] and P_o; NFII [$\lambda_o(T_1)$ and $\mu_o(T_1)$] and P_o; NFII [$\lambda_t(t_1)$ and μ_{avg}] and P_t.

The three variable sets yielded R-squared values of 0.61, 0.67, and 0.61 in sequence. To understand the differences in R-squared values, it's crucial to examine the accuracy of the estimated arrival flow rate and queue length functions compared to observed traffic conditions. The R-squared values for the two episodes' arrival flow rates were 1.0 and 0.68, and for queue length functions, they were

1.0 and 0.74. Episode 2.6's lower value, coupled with a 14 % lower calculated maximum queue length compared to observed conditions, may explain the suboptimal performance in the analysis.

To address potential sources of error associated with the lower R-squared value, eliminating underperforming approximating functions from the experimental setup is an option. However, this approach is infeasible in our context because it would leave only one function for comparison, making it impossible to conclusively prove or disprove our hypothesis.

Hypothesis B:

$$Q_{\max} \sim \left[\frac{\lambda(t_1) - \mu}{\gamma}\right]^{\frac{3}{2}}$$

A similar approach of applying three variable sets was applied to evaluate Hypothesis B, for the relationship between the maximum queue length, Q_{max}, and the net flow intensity index (NFII): NFII [$\lambda_o(T_1)$ and μ_{avg}] and $Q_r(T_2)$; NFII [$\lambda_o(T_1)$, and $\mu_o(T_1)$] and $Q_r(T_2)$; NFII [$\lambda_t(t_1)$ and μ_{avg}] and $Q_r(t_2)$.

The three variable sets yielded R-squared values of 0.59, 0.74, and 0.61, respectively. It is imperative to underscore that Episode 2.6's discrepancy between the observed and approximated results, particularly the low maximum residual queue length, may provide key insights into the subpar performance of the quadratic curvature functions. This, in conjunction with the limited function diversity, contributes to the uncertainty in our ability to draw conclusive inferences regarding Hypothesis B.

Hypothesis C:

$$Q_{\max} \sim P^3$$

The three variable sets used for evaluation of the relationship between the maximum queue length, Q_{max}, and congestion period, P, are: $Q_r(T_2)$ and P_o; $Q_r(t_2)$ and P_{Qt}; $Q_r(t_2)$ and $P_{\lambda t}$.

In this analysis, proportionality between the maximum queue length and the congestion period, exhibited R-squared valued of 0.81, 0.77 and 0.75, respectively. As in the previous cases, the lack of alignment between the two functions and the observed arrival flow and queue length in one of the episodes may have contributed to the disproportional relationship between P and Q_{max}. To establish a more conclusive validation of Hypothesis C, it is highly recommended to conduct a more extensive and comprehensive study, incorporating additional episodes.

Hypothesis D:

$$W \sim P^4$$

For this analysis, three variable sets used for evaluation of the relationship between the total wait time and the congestion period, P, are: w_o and P_o; w_t and $P_{\lambda t}$; and w_t and P_{Qt}.

Remarkably, this revealed a significant proportionality between the two variables, yielding extremely high R-squared values of 0.99, 0.98, and 0.98, respectively. While the limited number of episodes available for analysis prevents us from definitively proving or disproving the hypothesis, these results indicate a promising starting point.

6.2. Cubic arrival flow rates and cubic queue length functions

Hypothesis E: $P \sim \left[\frac{\lambda(t_1)-\mu}{\gamma}\right]^{1/3}$

We investigated proportional relationships, as hypothesized, between the congestion period (P) and the net flow intensity index (NFII) using three variable sets: NFII [$\lambda_o(t_1)$ and μ_{avg}] and P_o; NFII [$\lambda_o(t_1)$ and $\mu_o(t_1)$] and P_o; NFII [$\lambda_t(t_1)$ and μ_{avg}] and P_t.

Cubic Functions with Negative Curvature Gradient (3 episodes): This subset of episodes resulted in R-squared values of 0.81, 0.84, and 0.73. We closely examined the R-squared values associated with arrival flow functions, which were 0.26, 0.76, and 0.80. To mitigate potential errors, we excluded this underperforming function, resulting in a significantly improved R-squared value of 0.93. This elimination underscored the importance of accurately approximating real-time conditions and yielded a strong correlation in support of the hypothesis.

Cubic Functions with Positive Curvature Gradient (12 episodes): A similar analysis yielded R-squared values of 0.81, 0.80, and 0.74, respectively. To enhance the robustness of the analysis, we excluded the three functions with the lowest R-squared values for the arrival flow rate functions. This step resulted in an improved R-squared value of 0.89 for the approximated conditions, reinforcing the strong correlation with the hypothesis. Our analysis supports Hypothesis E, demonstrating alignment between the data and the hypothesis, particularly after removing the lowest-performing functions.

Hypothesis F: $Q_{\max} \sim \left[\frac{\lambda(t_1)-\mu}{\gamma}\right]^{4/3}$

A similar approach of applying three variable sets was applied to evaluate Hypothesis B, for the relationship between the maximum queue length, Q_{max}, and the net flow intensity index (NFII): NFII [$\lambda_o(T_1)$ and μ_{avg}] and $Q_r(T_2)$; NFII [$\lambda_o(T_1)$, and $\mu_o(T_1)$] and $Q_r(T_2)$; NFII [$\lambda_t(t_1)$ and μ_{avg}] and $Q_r(t_2)$.

Cubic Functions with Negative Curvature Gradient (3 episodes): Analysis of these cubic functions resulted in high R-squared values of 0.94 and 0.97 for observed conditions, a significantly lower R-squared value of 0.69 for the approximated conditions. Eliminating the lowest-performing function improved the overall fit with an enhanced R-squared value of 0.94, affirming strong alignment with

Hypothesis B in the context of negative curvature cubic functions.

Cubic Functions with Positive Curvature Gradient (12 episodes): Similarly, the analysis extended to observed conditions for these twelve cubic functions produced slightly lower R-squared values of 0.89 and 0.83 and generated a subpar R-squared value of 0.71 for approximated conditions. We removed the four lowest-performing functions for both the arrival flow rate and queue length functions. This approach yielded an acceptable R-squared value of 0.81, indicating a satisfactory alignment with Hypothesis B.

However, it's important to note a critical observation: there was a significant difference in the approximated queue lengths, especially for positive curvature functions, despite their high R-squared values. Specifically, the four functions with the lowest R-squared values for their queue predictions forecasted queue lengths 12–41 % lower than observed, averaging a 27 % deviation. In contrast, the remaining functions had an average discrepancy of 7 % below observed queue lengths. When compared to functions with negative curvature, the latter exhibited a much smaller average deviation of 4 %, with two functions nearly matching observed values. This deviation likely contributed to the satisfactory R-squared value of only 0.81 for the positive curvature episodes.

Hypothesis G: $Q_{max} \sim P^4$

The three variable sets used for evaluation of the relationship between the maximum queue length, Q_{max}, and congestion period, P, are: $Q_r(T_2)$ and P_o; $Q_r(t_2)$ and P_{Qt}; $Q_r(t_2)$ and $P_{\lambda t}$.

Cubic Functions with Negative Curvature Gradient (3 episodes): These three episodes revealed relatively low R-squared values for both observed and approximated conditions, with values of 0.57, 0.50, and 0.49, respectively. Attempting to elevate the robustness of the analysis, we eliminated the lowest-performing queue length function, resulting in slightly higher R-squared values of 0.61 and 0.53, indicating a moderate degree of correlation between Q_{max} and P within the context of negative curvature cubic functions.

Cubic Functions with Positive Curvature Gradient (12 episodes): Conversely, these cubic functions exhibited markedly lower R-squared values of 0.22, 0.12, and 0.18. After eliminating the four under-performing queue length functions in both approximated scenarios, we achieved only one modestly improved R-squared value of 0.51 and one decreased value of 0.15, respectively. To further improve proportionality, we eliminated five underperforming arrival flow functions, but this approach resulted only in a significant increase in the R-squared value for the graph that utilized the arrival flow function to predict the period (0.81), while the R-squared value dropped further to 0.14 for the last graph.

It is crucial to acknowledge that the queue length functions encountered challenges in precisely aligning with the observed episodes' Q_{max} values. On average, these predicted values were roughly 27 % lower than the observed values, indicative of potential sources of error within the model or underlying data. Additionally, our real-time data collection encompassed various congestion period lengths, contributing to variability in the results.

In summary, Hypothesis C garners moderate support, as evidenced by R-squared values that indicate a potential correlation between Q_{max} and P.

Hypothesis H: $W \sim P^{5/3}$

For this analysis, three variable sets used for evaluation of the relationship between the total wait time and the congestion period, P, are: w_o and P_o; w_t and $P_{\lambda t}$; and w_t and P_{Qt}.

Cubic Functions with Negative Curvature Gradient (3 episodes): Analysis of these three episodes revealed acceptable R-squared values for both observed and approximated conditions, attaining 0.80, 0.81, and 0.84, respectively. An attempt to increase the R-squared values by eliminating the lowest-performing episode, did the exact opposite and we observed an unexpected drop in the R-squared values.

Cubic Functions with Positive Curvature Gradient (12 episodes): Conversely, these cubic functions exhibited significantly lower R-squared values of 0.47, 0.43, and 0.37. After the elimination of the four under-performing queue length functions, we achieved a significant increase in the R-squared value from 0.43 to 0.89. With elimination of the five underperforming arrival flow functions the resultant R-squared value increased to 0.84 from 0.37.

In conclusion, higher-order polynomial functions, such as those raised to the fourth or fifth power, are more sensitive to data deviations, which can result in significant variations in model outputs, thus only showing moderate support for the above hypothesis.

7. Conclusion

This study explores the dynamics and stochastic nature of traffic flows at signalized intersections under partially oversaturated conditions, focusing primarily on three critical aspects. It identifies diverse stochastic factors that influence signal performance, specifically emphasizing random vehicle arrivals, while acknowledging the need for future research to incorporate factors like accidents, lane closures, construction activities, and poorly coordinated signals. Secondly, it explores temporal fluctuations in signal behavior. Lastly, it tackles the challenges in quantifying stochasticity and dynamics under oversaturated conditions, calling for the consideration of both local and global traffic state variabilities. The research underscores the need to develop robust, universally applicable traffic models that can efficiently manage the entire corridor. Such progress would ultimately pave the way for more efficient traffic management strategies and real-world applicable models.

Future research in the field of traffic management and signalized intersections can leverage the findings of this study to advance our understanding of complex traffic phenomena. To achieve this, researchers can explore advanced modeling techniques to capture the intricate dynamics of traffic flows at intersections, incorporating methodologies from Li et al. [73] and Li and Zhou [74] for simultaneous route guidance and traffic signal optimization, as well as connected-and-automated-vehicle (CAV) scheduling.

This research also underscores the pivotal role of foundational understanding gleaned from polynomial queue models and volume delay functions [75] for applying Deep Reinforcement Learning (DRL) in traffic environments [76,77], aiming to maximize expected long-term rewards. DRL, recognized for its diverse applications, leverages principles from these models to define states and actions in

traffic systems, focusing on the positions, velocities, and accelerations of CAVs and surrounding vehicles. A crucial component of DRL is reward shaping, aligning with the system's objectives and incorporating several performance indices proposed in our study, such as analyzing cumulative cycles and queue lengths during peak periods. Comprehensive real-world experiments in our study, utilizing emerging video detectors and real-time data, reveal congestion patterns and enhance our understanding of congestion dynamics at signalized intersections under partially oversaturated conditions. In essence, the interplay between polynomial queue models and DRL provides valuable insights and methodologies to optimize traffic flow and alleviate congestion in challenging intersection environments.

In concluding the present study, it is important to acknowledge one of its limitations, namely, the absence of consideration for short lanes. Spillback from a short turn pocket and blockage of its entrance can reduce signal capacity. In this scenario, one movement obstructs the other, resulting in inefficient utilization of green time. To address this limitation, future research endeavors will explore the impact of short lane effects by employing effective methodologies, such as the approach demonstrated in ↱[78], where a computational method was presented to integrate the impact of short left turn pockets on sustained service rates in mesoscopic models. Mesoscopic models, which handle large networks while retaining individual vehicle identity, offer efficiency but often overlook queuing effects. The proposed model introduces a gating mechanism to address this, resulting in favorable comparisons with micro-simulation models like VISSIM.

CRediT authorship contribution statement

Alisa Doll: Writing – review & editing, Writing – original draft, Visualization, Validation, Methodology, Formal analysis, Data curation, Conceptualization. **Mohammad Abbasi:** Writing – review & editing, Writing – original draft, Methodology, Investigation, Formal analysis, Conceptualization. **Ming Zhao:** Writing – review & editing, Writing – original draft, Resources, Funding acquisition, Conceptualization. **Xuesong Zhou:** Writing – review & editing, Supervision, Resources, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Conceptualization.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data Availability

Data will be made available on request.

Acknowledgement

The authors would like to express their sincere gratitude to the National Science Foundation (NSF) for the support provided through Grant No. 2126291. This research has been facilitated by the generous funding under the CC* Integration-Large: (BLUE) Software-Defined CyberInfrastructure to enable data-driven smart campus applications.

Author contributions

The authors confirm contribution to the paper as follows: study conception and design: Doll. A, Zhao. M, Zhou. X; data collection: Doll. A; analysis and interpretation of results: Doll. A, Zhou. X; draft manuscript preparation: Doll. A, Abbasi. M, Zhao. M, Zhou. X. All authors reviewed the results and approved the final version of the manuscript.

Appendix A. Quadratic arrival flow-based fluid queue performance derivation

In this Appendix, we provide self-contained derivation which are adapted from the work by Newell [40]. Newell assumes the arrival flow $\lambda(t)$ and developed a quadratic function to approximate the arrival flow at any given time with a maximum arrival at b with the application of the Taylor Series Approximation for quadratic functions:

$$\lambda(t) = \lambda(b) + \lambda'(b)(t-b) + \int_b^t \lambda'(t)(t-b)dt$$

After integrating the above function, it can now be expressed as:

$$\lambda(t) = \lambda(b) + \lambda'(b)(t-b) + \frac{1}{2}\lambda''(b)(t-b)^2$$

The quadratic function for approximating the arrival flow rates is centered about t_1 or the maximum critical point of the curve, therefore $b = t_1$.

Given $b = t_1$, the equation above is transformed to:

$$\lambda(t) = \lambda(t_1) + \lambda'(t_1)(t - t_1) + \frac{1}{2}\lambda''(t_1)(t - t_1)^2 \tag{1a}$$

From the Fig. 3(b), the local maximum or critical point for the function occurs at t_1 or $\lambda(t_1)$. The derivative of a function (slope of the tangent line is zero) therefore, $\lambda'(t_1) = 0$. The Eqn 1a now simplifies to:

$$\lambda(t) = \lambda(t_1) + \frac{1}{2}\lambda''(t_1)(t - t_1)^2 \tag{2a}$$

A function that is concave down at its maximum critical point, implies the second derivative is less than zero, therefore, $\lambda''(t_1) < 0$. Let γ be the inflow curvature parameter since it describes the shape or curvature of the time-dependent inflow arrival function, where $\gamma = -\frac{1}{2}\lambda''(t_1)$. Eqn 2a can now be expressed in the general form:

$$\lambda(t) = \lambda(t_1) - \gamma(t - t_1)^2 \tag{3a}$$

We have made the assumption that the time dependent departure rate is constant and will be equal to the departure rate, μ, for the episode. From Fig. 3(a), the departure rate, μ, is equal to the arrival inflow rate at time t_0 and time t_2, where $\mu = \lambda(t_0) = \lambda(t_2)$, and can be expressed as:

$$\mu = \lambda(t_1) - \gamma(t_0 - t_1)^2 = \lambda(t_1) - \gamma(t_2 - t_1)^2 \tag{4a}$$

The two real roots are as follows:

$$\mu = \lambda(t_1) - \gamma(t_0 - t_1)^2 \tag{5a}$$

$$\mu = \lambda(t_1) - \gamma(t_2 - t_1)^2 \tag{6a}$$

Solving for t_1 and t_2:

$$t_0 = t_1 - \left(\frac{\lambda(t_1) - \mu}{\gamma}\right)^{\frac{1}{2}} \tag{7a}$$

$$t_2 = t_1 + \left(\frac{\lambda(t_1) - \mu}{\gamma}\right)^{\frac{1}{2}} \tag{8a}$$

Through further substitution Eqn 3a can be written in the form:

$\lambda(t) - \mu = \gamma(t - t_0)(t_2 - t)$

From the Fig. 3**(b)** the queue length Q(t) at any time can be approximated by finding the vertical difference between the cumulative arrival rate , $A(t)$ and the cumulative departure rate, $D(t)$. The time dependent function for the queue length, Q(t) can be expressed:

$$Q(t) = A(t) - D(t) \tag{9a}$$

The queue length at any time is $Q(t) = A(t) - D(t)$ can also be calculated by finding the area between the arrival curve $\lambda(t)$ and the departure curve $\mu(t)$:

$$Q(t) = \int_{t_0}^{t} [\lambda(t) - \mu]dt = \int_{t_0}^{t} [\gamma(t - t_0)(t_2 - t)]dt \tag{10a}$$

With the application of substitution, the integral can be solved:

Let $v = t - t_0$ and $t = v + t_0$

$$Q(t) = \int_0^{t-t_0} [\gamma v(t_2 - (v + t_0)]dt$$

$$Q(t) = \gamma(t - t_0)^2\left[\frac{(t_2 - t_0)}{2} - \frac{(t - t_0)}{3}\right]$$

We know that the maximum queue length occurs at time, t_2 and can be expressed as:

$$Q(t_2) = \gamma(t_2 - t_0)^2\left[\frac{(t_2 - t_0)}{2} - \frac{(t_2 - t_0)}{3}\right] \tag{11a}$$

$$Q(t_2) = \frac{\gamma}{6}(t_2 - t_0)^3 \tag{12a}$$

And through substitution from Eqn 7a, maximum queue length can also be expressed as:

$$Q(t_2) = \frac{4\gamma}{3}\left(\frac{\lambda(t_1)-\mu}{\gamma}\right)^{\frac{2}{3}} \tag{13a}$$

Solving for the shape parameter:

$$\gamma = \frac{Q(t_2)}{(t_2-t_0)^3} \tag{14a}$$

The queue dissipates at time t_3, therefore $Q(t_3) = 0$, and t_3 can be expressed as:

$$Q(t_3) = \gamma(t_3-t_0)^2\left[\frac{(t_2-t_0)}{2} - \frac{(t_3-t_0)}{3}\right] = 0 \tag{15a}$$

$$\frac{(t_2-t_0)}{2} = \frac{(t_3-t_0)}{3}$$

$$t_3 = \frac{2}{3}(t_2-t_0)+t_0 \tag{16a}$$

Substituting

$$Q(t) = \gamma\left[\frac{(t_2-t_0)}{2} - \frac{(t-t_0)}{3}\right]$$

$$Q(t) = \frac{\gamma}{3}(t-t_0)^2(t_3-t) \tag{17a}$$

The delay per vehicle at time, t can be expressed as:

$$w(t) = \frac{Q(t)}{\mu} = \frac{\gamma}{3\mu}(t-t_0)^2(t_3-t) \tag{18a}$$

The total delay for the period from t_0 to t_3:

$$W = \int_{t_0}^{t_3} Q(t)dt = \int_{t_0}^{t_3}\left[\frac{\gamma}{3}(t-t_0)^2(t_3-t)\right]dt = \frac{\gamma}{36}(t_3-t_0)^4 = \frac{9}{4}\left(\frac{\lambda(t_1)-\mu}{\gamma}\right)^2 \tag{19a}$$

The average delay $\overline{w}$ for the period from t_0 to t_3:

$$\overline{w} = \frac{W}{D} = \frac{\gamma}{36}\frac{(t_3-t_0)^4}{D} \tag{20a}$$

The period,

$$P = t_3 - t_0 = \frac{D}{\mu} \tag{21a}$$

$$\overline{w} = \frac{\gamma}{36\mu}\left[\frac{D}{\mu}\right]^3 \tag{22a}$$

The average total travel time is equal to the travel time at free flow speed and the average delay:

$$t_t = t_f + \overline{w} \tag{23a}$$

$$t_t = t_f\left[1 + \frac{\gamma}{36\mu t_f}\left[\frac{D}{\mu}\right]^3\right] \tag{24a}$$

Appendix B. Cubic arrival flow-based fluid queue performance derivation

In this Appendix, we provide self-contained derivations adapted from the work of Cheng's et al. [49] on approximating stochastic arrival flows based on a cubic assumption.

$$t = u + t_0, \text{ and substitute into } Q(t) = \gamma\int_{t_0}^{t}[\gamma(t-t_0)(t-t_2)(t-\bar{t}\,]dt$$

Let $u = t - t_0$ and *therefore*

$$Q(t) = \gamma\int_0^{t-t_0}[u(u+t_0-t_2)(u+t_0-\bar{t})\,]du$$

$$Q(t) = \gamma(t-t_0)^2\left[\frac{1}{4}(t-t_0)^2 + \frac{1}{3}(t-t_0)(2t_0 - \bar{t} - t_2) + \frac{1}{2}(t_0 - t_2)(t_0 - \bar{t})\right] \tag{1b}$$

The queue dissipates at time t_3, therefore $Q(t_3) = 0$, and then Eqn 1b can be expressed as:

$$Q(t_3) = 0 = \gamma(t_3-t_0)^2\left[\frac{1}{4}(t_3-t_0)^2 + \frac{1}{3}(t_3-t_0)(2t_0 - \bar{t} - t_2) + \frac{1}{2}(t_0 - t_2)(t_0 - \bar{t})\right] \tag{2b}$$

Then, we can obtain the relationship between *t*0, *t*2, *t*3 and *t* as:

$$(\bar{t} - t_0) = \frac{3(t_3-t_0)^2 - 4(t_2-t_0)(t_3-t_0)}{4(t_3-t_0) - 6(t_2-t_0)} \tag{3b}$$

$$\bar{t} = \frac{3(t_3-t_0)^2 - 4(t_2-t_0)(t_3-t_0)}{4(t_3-t_0) - 6(t_2-t_0)} + t_0 \tag{4b}$$

$$(t_0 - \bar{t}) = -\left[\frac{3(t_3-t_0)^2 - 4(t_2-t_0)(t_3-t_0)}{4(t_3-t_0) - 6(t_2-t_0)}\right] \tag{5b}$$

Denote the Peak Queue Fraction factor m by the ratio between the time duration from the start of congestion to the time with maximum queue length and the congestion duration,

$$m = \frac{t_2 - t_0}{t_3 - t_0} \tag{6b}$$

then we can obtain the time-dependent queue length function after substituting Eq. (4b) and Eqn 5b into Eq. (1b):

$$Q(t) = \gamma(t-t_0)^2\left[\frac{1}{4}(t-t_0)^2 + \frac{1}{3}(t-t_0)(2t_0 - t_2 - \left[\frac{3(t_3-t_0)^2 - 4(t_2-t_0)(t_3-t_0)}{4(t_3-t_0) - 6(t_2-t_0)} + t_0\right] + \frac{1}{2}(t_0 - t_2)(-1)\left[\frac{3(t_3-t_0)^2 - 4(t_2-t_0)(t_3-t_0)}{4(t_3-t_0) - 6(t_2-t_0)}\right]\right] \tag{7b}$$

Next, substitute, $m = \frac{t_2-t_0}{t_3-t_0}$, into Eq. (7b),

$$Q(t) = \gamma(t-t_0)^2\left[\frac{1}{4}(t-t_0)^2 - \frac{1}{3}(t-t_0)(t_3-t_0)\left[m + \left[\frac{3-4m}{4-6m}\right]\right] + \frac{1}{2}(t_3-t_0)^2 m\left[\frac{3-4m}{4-6m}\right]\right] \tag{8b}$$

Maximum queue length occurs at $Q(t_2)$

$$Q(t_2) = \gamma(t_2-t_0)^2\left[\frac{1}{4}(t_2-t_0)^2 - \frac{1}{3}(t_2-t_0)(t_3-t_0)\left[m + \left[\frac{3-4m}{4-6m}\right]\right] + \frac{1}{2}(t_3-t_0)^2 m\left[\frac{3-4m}{4-6m}\right]\right] \tag{9b}$$

$$Q(t_2) = \gamma(t_3-t_0)^2 m^2\left[\frac{1}{4}m^2(t_3-t_0)^2 - \frac{1}{3}(t_3-t_0)m(t_3-t_0)\left[m + \left[\frac{3-4m}{4-6m}\right]\right] + \frac{1}{2}(t_3-t_0)^2 m\left[\frac{3-4m}{4-6m}\right]\right]$$

$$Q(t_2) = \gamma(t_3-t_0)^4 m^3\left[\frac{(m-1)^2}{(8-12m)}\right] \tag{10b}$$

Time dependent delay:

$$W(t) = \frac{Q(t)}{\mu} = \frac{\gamma(t-t_0)^2}{\mu}\left[\frac{1}{4}(t-t_0)^2 - \frac{1}{3}(t-t_0)(t_3-t_0)\left[m + \left[\frac{3-4m}{4-6m}\right]\right] + \frac{1}{2}(t_3-t_0)^2 m\left[\frac{3-4m}{4-6m}\right]\right] \tag{11b}$$

The total delay from t_0 to t_3:

$$W(t_3) = \int_{t_0}^{t_3} Q(t)dt$$

$$= \int_{t_0}^{t_3}\left[\gamma(t-t_0)^2\left[\frac{1}{4}(t-t_0)^2 - \frac{1}{3}(t-t_0)(t_3-t_0)\left[m + \left[\frac{3-4m}{4-6m}\right]\right] + \frac{1}{2}(t_3-t_0)^2 m\left[\frac{3-4m}{4-6m}\right]\right]\right]dt \tag{12b}$$

Let $u = t - t_0$ and $t = u + t_0$

$$= \gamma\int_0^{t_3-t_0}\left[u^2\left[\frac{1}{4}u^2 - \frac{1}{3}u(t_3-t_0)\left[m + \left[\frac{3-4m}{4-6m}\right]\right] + \frac{1}{2}(t_3-t_0)^2 m\left[\frac{3-4m}{4-6m}\right]\right]\right]du$$

$$= \gamma(t_3 - t_0)^5 \left[\frac{1}{20} - \frac{1}{12}\left[m + \left[\frac{3-4m}{4-6m}\right]\right] + \frac{1}{6}m\left[\frac{3-4m}{4-6m}\right]\right] \tag{13b}$$

$$g(m) = \left[\frac{1}{20} - \frac{1}{12}\left[m + \left[\frac{3-4m}{4-6m}\right]\right] + \frac{1}{6}m\left[\frac{3-4m}{4-6m}\right]\right] \tag{14b}$$

$$W(t_3) = \gamma g(m)(t_3 - t_0)^5 \tag{15b}$$

The average waiting time, $\overline{w}$

$$\overline{w} = \frac{W}{D} = \frac{\gamma g(m)}{\mu}(t_3 - t_0)^5 = \frac{\gamma g(m)}{\mu}\left(\frac{D}{\mu}\right)^4 \tag{16b}$$

The average total travel time is equal to the travel time at free flow speed and the average delay:

$$t_t = t_f + \overline{w} \tag{17b}$$

$$t_t = t_f\left[1 + \frac{\gamma g(m)}{\mu t_f}\left[\frac{D}{\mu}\right]^4\right] \tag{18b}$$

References

[1] F. Dion, H. Rakha, Y.S. Kang, Comparison of delay estimates at under-saturated and over-saturated pre-timed signalized intersections, Transp. Res. Part B 38 (2) (2004) 99–122.
[2] C. Cheng, Y. Du, L. Sun, Y. Ji, Review on theoretical delay estimation model for signalized intersections, Transp. Rev. 36 (4) (2016) 479–499.
[3] F.V. Webster, Traffic Signal Settings 39 (1958).
[4] R. Akçelik, 1980, R. Akçelik, Time-dependent expressions for delay, stop rate and queue length at traffic signals, (1980).
[5] X.S. Zhou, Q. Cheng, X. Wu, P. Li, B. Belezamo, J. Lu, M. Abbasi, A meso-to-macro cross-resolution performance approach for connecting polynomial arrival queue model to volume-delay function with inflow demand-to-capacity ratio, Multimodal Transp. 1 (2) (2022) 100017.
[6] HCM, Highway capacity manual, Transportation Research Board, 2020.
[8] T. Nagatani, Modified KdV equation for jamming transition in the continuum models of traffic, Phys. A: Stat. Mech. Appl. 261 (3-4) (1998) 599–607.
[7] B.S. Kerner, Failure of classical traffic flow theories: Stochastic highway capacity and automatic driving, Phys. A Stat. Mech. Appl. 450 (2016) 700–747.
[11] A.D. May, Traffic flow theory-the traffic engineer's challenge, Proc. Inst. Traffic Eng. (1975) 290–303.
[9] R. Akcelik, A queue model for HCM 2000, ARRB Transportation Research Ltd. Vermont South Australia, 1999.
[12] A. Sharma, D.M. Bullock, J.A. Bonneson, Input–output and hybrid techniques for real-time prediction of delay and maximum queue length at signalized intersections, Transp. Res. Rec. 2035 (1) (2007) 69–80.
[13] G. Vigos, M. Papageorgiou, Y. Wang, Real-time estimation of vehicle-count within signalized links, Transp. Res. Part C. *16* (1) (2008) 18–35.
[10] B. Belezamo, Data-driven methods for characterizing transportation system performances under congested conditions: a phoenix study, Doctoral dissertation Arizona State University, 2020.
[15] M.J. Lighthill, G.B. Whitham, On kinematic waves. I: flood movement in long rivers. II: a theory of traffic flow on long crowded roads, Proc. R. Soc. Lond. Ser. A. Math. Phys. Sci. 229 (1178) (1955) 281–345.
[17] P.I. Richards, Shock waves on the highway, Oper. Res. 4 (1) (1956) 42–51.
[19] G. Stephanopoulos, P.G. Michalopoulos, Modeling and analysis of traffic queue dynamics at signalized intersections, Transp. Res. Part A 13 (5) (1979) 295–307.
[18] A. Skabardonis, N. Geroliminis, Real-time monitoring and control on signalized arterials, J. Intell. Transp. Syst. 12 (2) (2008) 64–74.
[16] H.X. Liu, X. Wu, W. Ma, H. Hu, Real-time queue length estimation for congested signalized intersections, Transp. Res. Part C. 17 (4) (2009) 412–427.
[14] X. Ban, P. Hao, Z. Sun, Real time queue length estimation for signalized intersections using travel times from mobile sensors, Transp. Res. Part C. 19 (6) (2011) 1133–1156.
[23] T.Q. Tang, Z.Y. Yi, J. Zhang, T. Wang, J.Q. Leng, A speed guidance strategy for multiple signalized intersections based on car-following model, Phys. A: Stat. Mech. Appl. 496 (2018) 399–409.
[20] T. Nagatani, Chain reaction of traffic breakdowns in coupled-cycle networks, Phys. A: Stat. Mech. Appl. 587 (2022) 126549.
[21] T. Nagatani, Macroscopic traffic flow in multiple-loop networks, Phys. A: Stat. Mech. Appl. 609 (2023) 128324.
[22] T. Nagatani, K. Tobita, Vehicular motion in counter traffic flow through a series of signals controlled by a phase shift, Phys. A: Stat. Mech. Appl. 391 (20) (2012) 4976–4985.
[24] W.X. Zhu, J.Y. Zhang, Z.R. Song, Study on braking process of vehicles at the signalized intersection based on car-following theory, Phys. A: Stat. Mech. Appl. 523 (2019) 1306–1314.
[25] R. Jiang, Q.S. Wu, A stopped time dependent randomization cellular automata model for traffic flow controlled by traffic light, Phys. A: Stat. Mech. Appl. 364 (2006) 493–496.
[26] G.F. Newell, Traffic flow on a network: an unsolved problem, Oper. Res. 13 (1) (1965) 45–52.
[27] W.T. Wong, S.C. Wong, C.O. Tong, Sheared delay formulae for the TRANSYT traffic model: a review and calibration, Transp. Rev. 23 (1) (2003) 1–20.
[28] V. Zuylen, H. Viti, A probalistic model for queues, delays and waiting time at controlled intersections, Pap. Presente 86th Annu. Meet. Transp. Res. Board Wash. DC (2007).
[29] A.D. May, H.M. Keller, A deterministic queuing model, Transp. Res. 1 (2) (1967) 117–128.
[30] R. Kimber, E. Hollis, Traffic queues and delays at road junctions, TRRL Lab. Rep. UK (1979) 909.
[31] E.B. Lieberman, J. Chang, E. Shenk, Formulation of real-time control policy for oversaturated arterials, Transp. Res. Rec. 1727 (1) (2000) 77–88.
[35] X. Wu, H.X. Liu, D. Gettman, Identification of oversaturated intersections using high-resolution traffic signal data, Transp. Res. Part C. 18 (4) (2010) 626–638.
[33] B. Park, C.J. Messer, T. Urbanik, Traffic signal optimization program for oversaturated conditions: genetic algorithm approach, Transp. Res. Rec. 1683 (1) (1999) 133–142.
[32] H. Liu, K.N. Balke, W.H. Lin, A reverse causal-effect modeling approach for signal control of an oversaturated intersection, Transp. Res. Part C. 16 (6) (2008) 742–754.

[34] H. Wang, X. Peng, Coordinated control model for oversaturated arterial intersections, IEEE Trans. Intell. Transp. Syst. 23 (12) (2022) 24157–24175.
[36] D.C. Gazis, Optimum control of a system of oversaturated intersections, Oper. Res. 12 (6) (1964) 815–831.
[39] M. Sasaki, T. Nagatani, Transition and saturation of traffic flow controlled by traffic lights, Phys. A 325 (3-4) (2003) 531–546.
[37] P.G. Michalopoulos, G. Stephanopoulos, Oversaturated signal systems with queue length constraints—II: systems of intersections, Transp. Res. 11 (6) (1977) 423–428.
[38] P.G. Michalopoulos, G. Stephanopoulos, Optimal control of oversaturated intersections: theoretical and practical considerations, Traffic Eng. Control 19 (5) (1978) 216–221.
[42] G.F. Newell, Queues with time-dependent arrival rates I—the transition through saturation, J. Appl. Probab. 5 (2) (1968) 436–451, 1968.
[40] G.F. Newell, Queues with time-dependent arrival rates: III. A mild rush hour, J. Appl. Probab. 5 (3) (1968) 591–606, 1968.
[41] G.F. Newell, Queues with time-dependent arrival rates: II. The maximum queue and the return to equilibrium, J. Appl. Probab. 5 (3) (1968) 579–590.
[43] G.F. Newell. Applications of queueing theory, 2nd ed., Chapman and Hall Ltd New York, 1982.
[45] J. Zhao, V.L. Knoop, M. Wang, Microscopic traffic modeling inside intersections: Interactions between drivers, Transp. Sci. 57 (1) (2023) 135–155.
[44] J. Zhao, V.L. Knoop, M. Wang, Two-dimensional vehicular movement modelling at intersections based on optimal control, Transp. Res. Part B: Methodol. 138 (2020) 1–22.
[46] C.F. Daganzo, In traffic flow, cellular automata= kinematic waves, Transp. Res. Part B: Methodol. 40 (5) (2006) 396–403.
[47] M. Hadi, X. Zhou, D. Hale, Multiresolution Modeling for Traffic Analysis: Guidebook (No. FHWA-HRT-22-055), Federal Highway Administration, United States, 2022.
[48] J. Lu, X.S. Zhou, Virtual track networks: a hierarchical modeling framework and open-source tools for simplified and efficient connected and automated mobility (CAM) system design based on general modeling network specification (GMNS), Transp. Res. Part C: Emerg. Technol. 153 (2023) 104223.
[49] Liu, H.X., Ma, W., Wu, X. and Hu, H. Development of a real-time arterial performance monitoring system using traffic data available from existing signal systems. 2008. Accessible via ⟨https://www.lrrb.org/pdf/200901.pdf⟩.
[50] National Academies of Sciences, Engineering, and Medicine, Signal Timing Manual - Second Edition, Second Edition, The National Academies Press, Washington, DC, 2015, https://doi.org/10.17226/22097 [Online]. Available:.
[51] National Academies of Sciences, Engineering, and Medicine, Operation of Traffic Signal Systems in Oversaturated Conditions, Volume 1 Practitioner Guidanc, The National Academies Press, Washington, DC, 2014, https://doi.org/10.17226/22290 [Online]. Available:.
[52] A. Wang, Measuring the Quality of Arterial Traffic Signal Timing – A Trajectory-based Methodology, University of Nevada, Reno, August 2020. ⟨http://hdl.handle.net/11714/7617⟩.
[54] FHWA, "Automated Traffic Signal Performance Measures," [Online]. Available: https://ops.fhwa.dot.gov/arterial_mgmt/performance_measures.htm#:~:text=Automated%20Traffic%20Signal%20Performance%20Measures%20(ATSPMs)%2C%20included%20in%20the,operations%2C%20maintenance%2C%20management%20and%20design.
[53] AASHTO Innovation Initiative, Hand-out of Automated Traffic Signal Performance Measures. 2014. [Online]. Available: http://aii.transportation.org/Documents/ATSPMs/atspms-handout-press.pdf.
[55] FHWA, "EDC-4 ATSPMs," [Online]. Available: www.fhwa.dot.gov/innovation/everydaycounts/edc-4.cfm.
[56] National Academies of Sciences, Engineering, and Medicine, Performance-Based Management of Traffic Signals, The National Academies Press, Washington, DC, 2020, https://doi.org/10.17226/25875 [Online]. Available:.
[57] R. Dowling. *Traffic analysis toolbox volume vi: Definition, interpretation, and calculation of traffic analysis tools measures of effectiveness*. United States, Federal Highway Administration, Office of Operations, 2007.
[58] L. Ambühl, M. Menendez, Data fusion algorithm for macroscopic fundamental diagram estimation, Transportation Research Part C: Emerging Technologies 71 (2016) 184–197.
[59] S.M.S. Mahmud, L. Ferreira, M.S. Hoque, A. Tavassoli, Micro-simulation modelling for traffic safety: A review and potential application to heterogeneous traffic environment, IATSS Res. 43 (1) (2019) 27–36, https://doi.org/10.1016/j.iatssr.2018.07.002. Online]. Available:.
[60] X. Wu, J. Guo, K. Xian, X. Zhou, Hierarchical travel demand estimation using multiple data sources: A forward and backward propagation algorithmic framework on a layered computational graph, Transportation Research Part C: Emerging Technologies 96 (2018) 321–346.
[61] A. Emami, M. Sarvi, S. Asadi Bagloee, Using Kalman filter algorithm for short-term traffic flow prediction in a connected vehicle environment, J. Mod. Transport. 27 (2019) 222–232, https://doi.org/10.1007/s40534-019-0193-2 [Online]. Available:.
[62] R.E. Kalman, A new approach to linear filtering and prediction problems, J. Basic Eng. 82 (1) (1960) 35–45.
[63] S.V. Kumar, Traffic Flow Prediction using Kalman Filtering Technique, Procedia Eng. 187 (2017) 582–587.
[64] A. Stathopoulos, M.G. Karlaftis, A multivariate state space approach for urban traffic flow modeling and prediction, Transp. Res. C Emerg. Technol. 11 (2) (2003) 121–135.
[66] J. Lee, B. Park, I. Yun, Cumulative Travel-Time Responsive Real-Time Intersection Control Algorithm in the Connected Vehicle Environment, J. Transp. Eng. 139 (2013) 1020–1029, https://doi.org/10.1061/(ASCE)TE.1943-5436.0000587 [Online]. Available:.
[65] G. Welch and G. Bishop, "An introduction to the Kalman filter," TR 95-041, Dept. of Computer Science, Univ. of North Carolina at Chapel Hill, Chapel Hill, NC, Updated: Monday, July 24, 2006. [Online]. Available: http://www.cs.unc.edu/~welch/kalman/kalmanIntro.html.
[67] S. Papadopoulou, C. Roncoli, N. Bekiaris-Liberis, et al., Microscopic simulation-based validation of a per-lane traffic state estimation scheme for highways with connected vehicles, Transp. Res. C Emerg. Technol. 86 (2018) 441–452.
[68] T.W. Lawson, D.J. Lovell, C.F. Daganzo, Using input-output diagram to determine spatial and temporal extents of a queue upstream of a bottleneck, Transp. Res. Rec. 1572 (1) (1997) 140–147.
[69] F. 42. Viti, H.J. Van Zuylen, Probabilistic models for queues at fixed control signals, Transp. Res. Part B: Methodol. 44 (1) (2010) 120–135.
[70] Rouphail, N., Tarko, A., Li, J., (2000). Traffic flow at signalized intersections. In: Lieu, H. (Ed.) Revised Monograph of Traffic Flow Theory, Update and Expansion of the Transportation Research Board (TRB) Special Report 165. “Traffic Flow Theory”, Published in 1975 (chapter 9).
[71] Q. Cheng, Z. Liu, J. Guo, X. Wu, R. Pendyala, B. Belezamo, X. Zhou, Estimating key traffic state parameters through parsimonious spatial queue models *Transp*, Res. Part C. 137 (2022) 103596 (Article).
[72] City of Chandler Transportation Master Plan 2019 Update. https://www.chandleraz.gov/sites/default/files/Chandler-TMP_Final-Report_01-09-20_lr.pdf.
[73] P. Li, P. Mirchandani, X. Zhou, Solving simultaneous route guidance and traffic signal optimization problem using space-phase-time hypernetwork, Transp. Res. Part B: Methodol. 81 (2015) 103–130 (Li, P.).
[74] P.T. Li, X. Zhou, Recasting and optimizing intersection automation as a connected-and-automated-vehicle (CAV) scheduling problem: a sequential branch-and-bound search approach in phase-time-traffic hypernetwork, Transp. Res. Part B: Methodol. *105* (2017), 479-50.
[75] Y. Pan, J. Guo, Y. Chen, Calibration of dynamic volume-delay functions: a rolling horizon-based parsimonious modeling perspective, Transp. Res. Rec. *2676* (2) (2022) 606–620.
[76] Y.A. Pan, J. Guo, Y. Chen, S. Li, W. Li, Incorporating traffic flow model into a deep learning method for traffic state estimation: a hybrid stepwise modeling framework, J. Adv. Transp. (2022) 1–17.
[77] V. Potluri, Developing a cyber-physical system for real-time proactive traffic control and management of mixed fleet of vehicles with various levels of autonomy, Doctoral dissertation Arizona State University, 2023.
[78] W.L. Reynolds, X. Zhou, N.M. Rouphail, M. Li, Estimating sustained service rates at signalized intersections with short left-turn pockets: mesoscopic approach, Transp. Res. Rec. *2173* (1) (2010) 64–71.