Point Queue-X: A Macroscopic Demand and Supply Model for Oversaturated Dynamic Transportation Queueing Systems

Presenter: Xuesong Zhou, PhD (xzhou74@asu.edu)
School of Sustainable Engineering and the Built Environment, Arizona State University

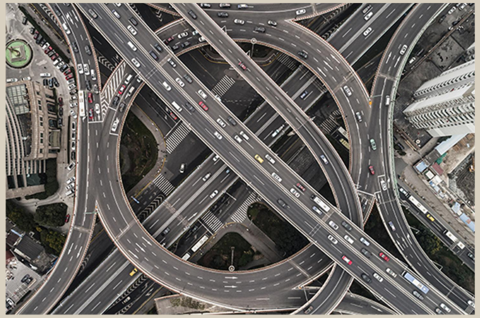
Outline

- Background and introduction
- Problem description
- Polynomial approximation for the inflow rate
- Calibration and results
- Discussions and applications
- Summary

1. Background and introduction

❑ Manage and control traffic congestion

Capacity building

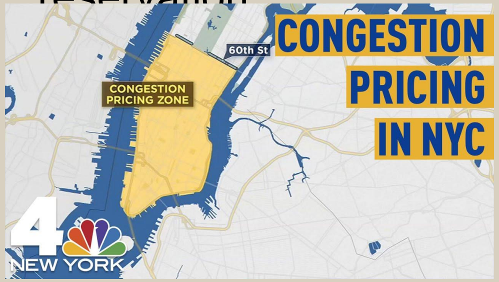


Land-use development

Infrastructure construction

Variable congestion toll

Flexible travel reservation



Demand regulation

Traffic congestion in USA costs:

$1348 per driver in 2018

or $87 billion in total in 2018

or 97 hours/driver in 2018

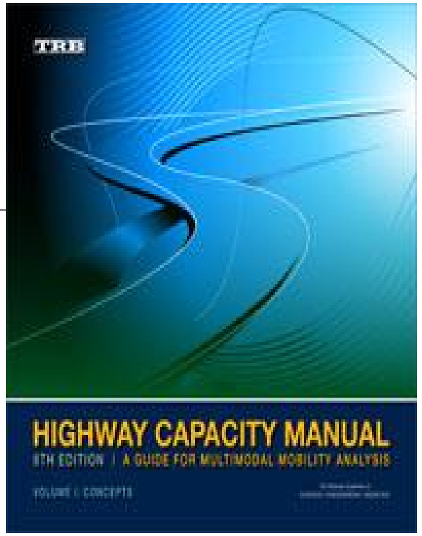
We need a system performance evaluation model, which can be used to optimize the system performance jointly through the **demand regulation policies** and **infrastructure capacity building** efforts, rather than adopting locally oriented congestion reduction strategies in isolation.

1. Background and introduction

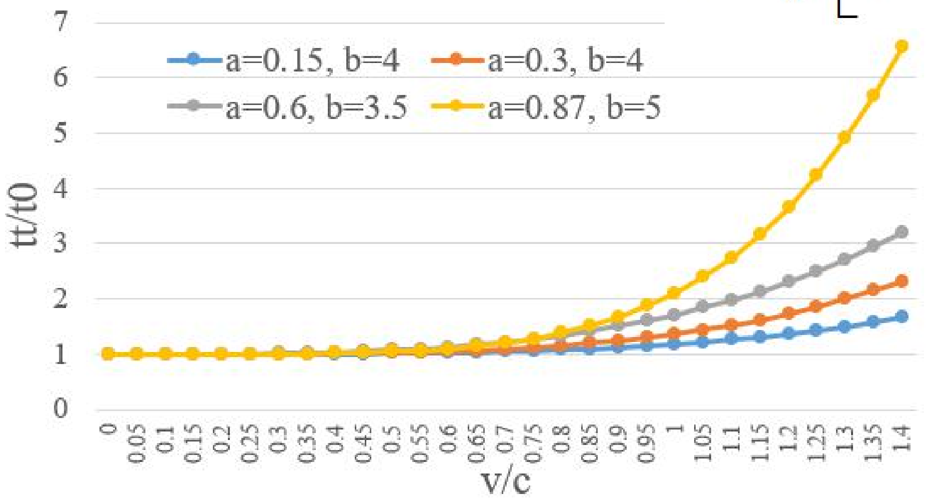
☐ Typical volume-delay function

Literatures	Mathematical forms	tt/t_0 when $V=C$	tt/t_0 when $V=0$
CATS (1960); Muranyi (1963)	$tt=t_0\cdot 2^{(V/C)}$	2	1
Smock (1962, 1963)	$tt=t_0\cdot\exp\left(\frac{V}{C}-1\right)$	1	0.37
BPR (1964)	$tt=t_0\cdot\left[1+0.15\cdot\left(\frac{V}{C}\right)^4\right]$	1.15	1
Davidson (1966)[1]	$tt=t_0\cdot\left(1+\frac{J_D\cdot V}{C-V}\right),\ V<C$	$+\infty$	1
Akçelik (1978)[2]	$tt=\begin{cases}t_0\cdot\left(1+\frac{J_D\cdot V}{C-V}\right), \text{ if } V\le mC,\ 0<m<1\\ t_m+K_m(V-mC), \text{ if } V>mC,\ 0<m<1\end{cases}$	Related to J_D and m	1
Spiess (1990)	$tt=t_0\cdot\left[2-\beta-\alpha\left(1-\frac{V}{C}\right)+\sqrt{\alpha^2\left(1-\frac{V}{C}\right)^2+\beta^2}\right]$, where $\beta=\frac{2\alpha-1}{2\alpha-2},\ \alpha>1$	2	1
Tisato (1991)[3]	$tt=\begin{cases}t_0\cdot\frac{1-mx}{1-x},\ x\le x_c\\ t_0\cdot\frac{1-mx_c}{1-x_c}+30T(x-x_c),\ x>x_c\end{cases}$ wherre $x=\frac{V}{C},\ x_c=1-\sqrt{\frac{t_0}{30T}(1-m)}$	Related to T and m	1
Akçelik (1991)	$tt=t_0+0.25T\cdot\left(\frac{V}{C}-1+\sqrt{(\frac{V}{C}-1)^2+\frac{8J_AV}{TC^2}}\right)$	Related to J_A and T	1

tt: travel time
t_0: free flow travel time
V: traffic volume
C: capacity
J_A, J_D, α, β: parameters



BPR Function $$tt=t_0\cdot\left[1+a\left(v/c\right)^b\right]$$



$$\log(tt-t_0)=\log t_0+\log a+b\log(v/c)$$

1. Background and introduction

BPR volume-delay function

Speed Limit (mph)	Practical Capacity (vehicle per hour)	Model Parameters	
		a	b
0 - 30	0 - 240	0.7312	3.6596
0 - 30	249 - 499	0.6128	3.5038
0 - 30	500 - 749	0.8774	4.4613
0 - 30	750 - 999	0.6846	5.1644
0 - 30	1000+	1.1465	4.4239
31 - 40	250 - 499	0.6190	3.6544
31 - 40	500 - 749	0.6662	4.9432
31 - 40	750 - 999	0.6222	5.1409
31 - 40	1000+	1.0300	5.5226
41 - 50	500 - 749	0.6609	5.0906
41 - 50	750 - 999	0.5423	5.7894
41 - 50	1000+	1.0091	6.5856
50+	500 - 749	0.8776	4.9287
50+	750 - 999	0.7699	5.3443
50+	1000+	1.1491	6.8677

Reference: Mannering et al. (1990)

Facility Type	Free-Flow Speed	"a"	"b"
6-Lane Freeway	70 mph	0.88	9.8
	60 mph	0.83	5.5
	50 mph	0.56	3.6
4-Lane Multilane Highway	70 mph	1.00	5.4
	60 mph	0.83	2.7
	50 mph	0.71	2.1

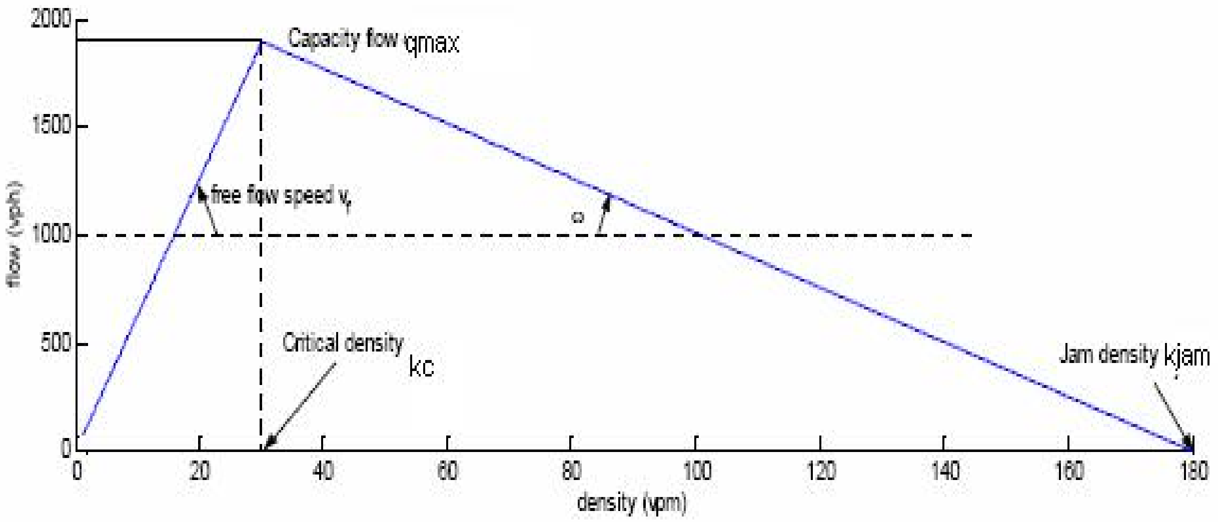
Reference: Horowitz (1991)

Both for highway and urban road.
Related to speed limit and practical capacity.

1. Background and introduction

BPR volume-delay function

- Widely used in static traffic assignment
- Fail to capture queue evolution, building-up, propagation, dissipation
- Fail to represent a bottleneck with low flow but high travel time



1. Background and introduction

Functional requirements of VDF

- Capture queue building-up, propagation, dissipation
  - ✓ Key: how to detect queue spills back upstream link(s)
- Represent multiple origin-destination flow
  - ✓ Freeway merge and diverge
  - ✓ Arterial intersections (signal, stop signs, yield signs, no control)

Evaluation Criteria

- Theoretically rigorous
- Numerically reliable
- Computationally efficient

1. Background and introduction

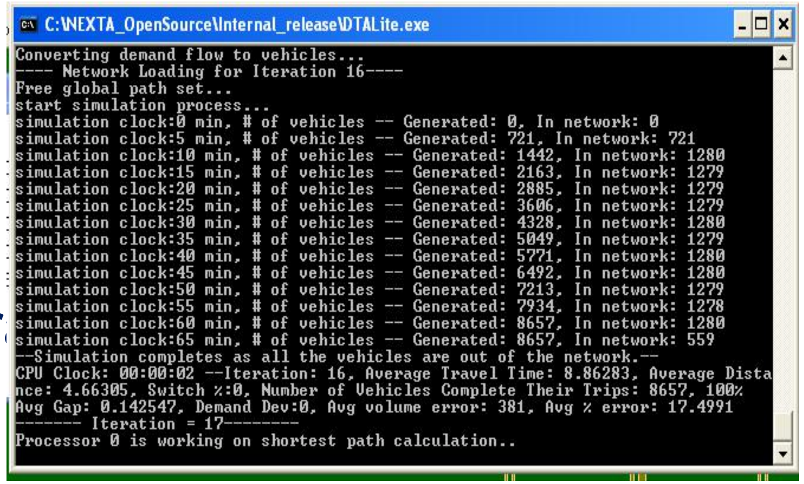
Existing simulation-based DTA models

- Integration (Vertical queue)
- DynaSMART (Link density function, vertical queue)
- DynaMIT (Link density function, moving queue)
- VISTA (Cell Transmission Model)
- Vissim (Car following)
- Transims (Cellular Automata)
- Transmodeler (Car following)
- DTALite (Simplified kinematic wave model)

1. Background and introduction

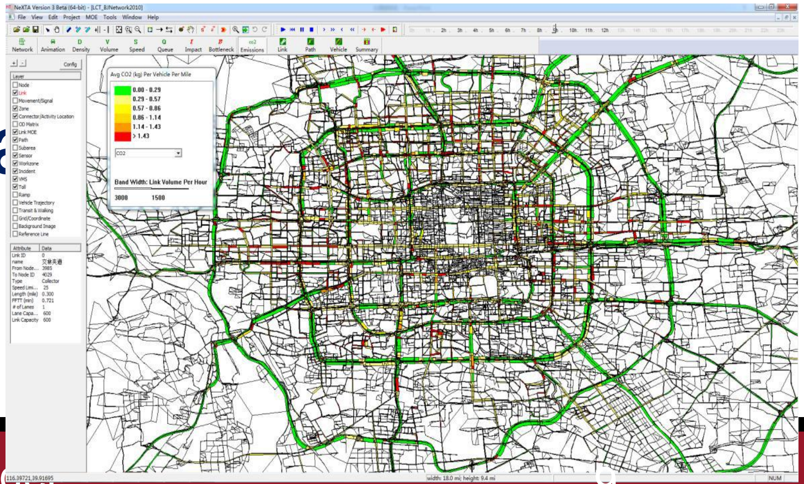
◘ DTALite: Open-source computational engine (C++)

- ➢ Light-weight and agent-based DTA
- ➢ Simplified kinematic wave model (Newell)
- ➢ Simplified car follow modeling (Newell)
- ➢ Built-in OD demand matrix estimation (ODME) progr
- ➢ Emission prediction (light-weight MOVES interface)



◘ NEXTA: front-end Graphical User Interfa

- ➢ https://github.com/xzhou99/dtalite_software_release



1. Background and introduction

Discussion in TMIP

TMIP FMIP
Better Methods. Better Outcomes.

HOME FORU

Thu, 03/28/2019 - 2:33pm #4 ster

nmars

Thu, 03/28/2019 - 4:58pm #6

John Gibb

Norman's article points out the absurdity of volume greater than capacity, but it's a little more tolerable to consider it demand rather than volume per se. Of course, we know that's not what actually flows through or downstream of a bottleneck; the rest of the problems are well known.

Impetus 1 does not in itself require microsimulation, nor even HCM-style intersection analysis. Norman's study used DTALite, a kinematic-wave DTA with link capacities. Two separate questions that can follow are:

(1) Is there a non-microsimulation DTA that's better than STA? (Some DTAs have theoretical problems of their own, e.g. unreasonable FIFO violation.)

(2) Is it good enough to be a replacement for STA as part of a full-region demand-equilibrium model?

Since "all models are wrong but some are useful," what's "better" or "good enough" depends; more detail is not always better. A model can be utterly faithful to the effects of every geometric and operational detail scientifically examined, but a high input-detail burden upon humans under limited budgets and schedules can make such a model less useful and more wrong than a simpler model, particularly for future scenario evaluation.

regio

packages.

- "… I argue that static assignment should not be used in freeway capacity planning studies - either separately or in combination with microsimulation…"
- "… it doesn't work to extract over-capacity subarea trip tables and then input them into capacity-constrained DTA or microsimulation …"
- "Is there a non-microsimulation DTA that's better than STA? (Some DTAs have theoretical problems of their own, e.g. unreasonable FIFO violation.)"
- "Is it good enough to be a replacement for STA as part of a full-region demand-equilibrium model?"
- "Since *all models* are wrong but *some* are useful, what's *better* or *good enough* depends; more detail is not always better."
- "… but a high input-detail burden upon humans under limited budgets and schedules can make such a model less useful and more wrong than a simpler model, particularly for future scenario evaluation…"

1. Background and introduction

□ Discussion in TMIP

TMIP FMIP
Better Methods. Better Outcomes.

HOME FORU

Thu, 03/28/2019 - 2:33pm #4

nmars

Thu, 03/28/2019 - 4:58pm #6

John Gibb

Norman's article points out the absurdity of volume greater than capacity, but it's a little more tolerable to consider it demand rather than volume per se. Of course, we know that's not what actually flows through or downstream of a bottleneck; the rest of the problems are well known.

Impetus 1 does not in itself require microsimulation, nor even HCM-style intersection analysis. Norman's study used DTALite, a kinematic-wave DTA with link capacities. Two separate questions that can follow are:

(1) Is there a non-microsimulation DTA that's better than STA? (Some DTAs have theoretical problems of their own, e.g. unreasonable FIFO violation.)

(2) Is it good enough to be a replacement for STA as part of a full-region demand-equilibrium model?

Since "all models are wrong but some are useful," what's "better" or "good enough" depends; more detail is not always better. A model can be utterly faithful to the effects of every geometric and operational detail scientifically examined, but a high input-detail burden upon humans under limited budgets and schedules can make such a model less useful and more wrong than a simpler model, particularly for future scenario evaluation.

packages.

□ Our perspectives:

- ➢ Bottleneck discharge rate
- ➢ Highway capacity
- ➢ Queue evolution: Deterministic fluid approximation vs stochastic m/m/1
- ➢ Fluid-queue-oriented travel time performance function

1. Background and introduction

Static traffic assignment

- computational efficiency
- easy for implementation
- cannot capture traffic flow dynamics
- cannot describe queue evolution
- cannot be used for oversaturated traffic conditions
- volume over capacity or demand over capacity?

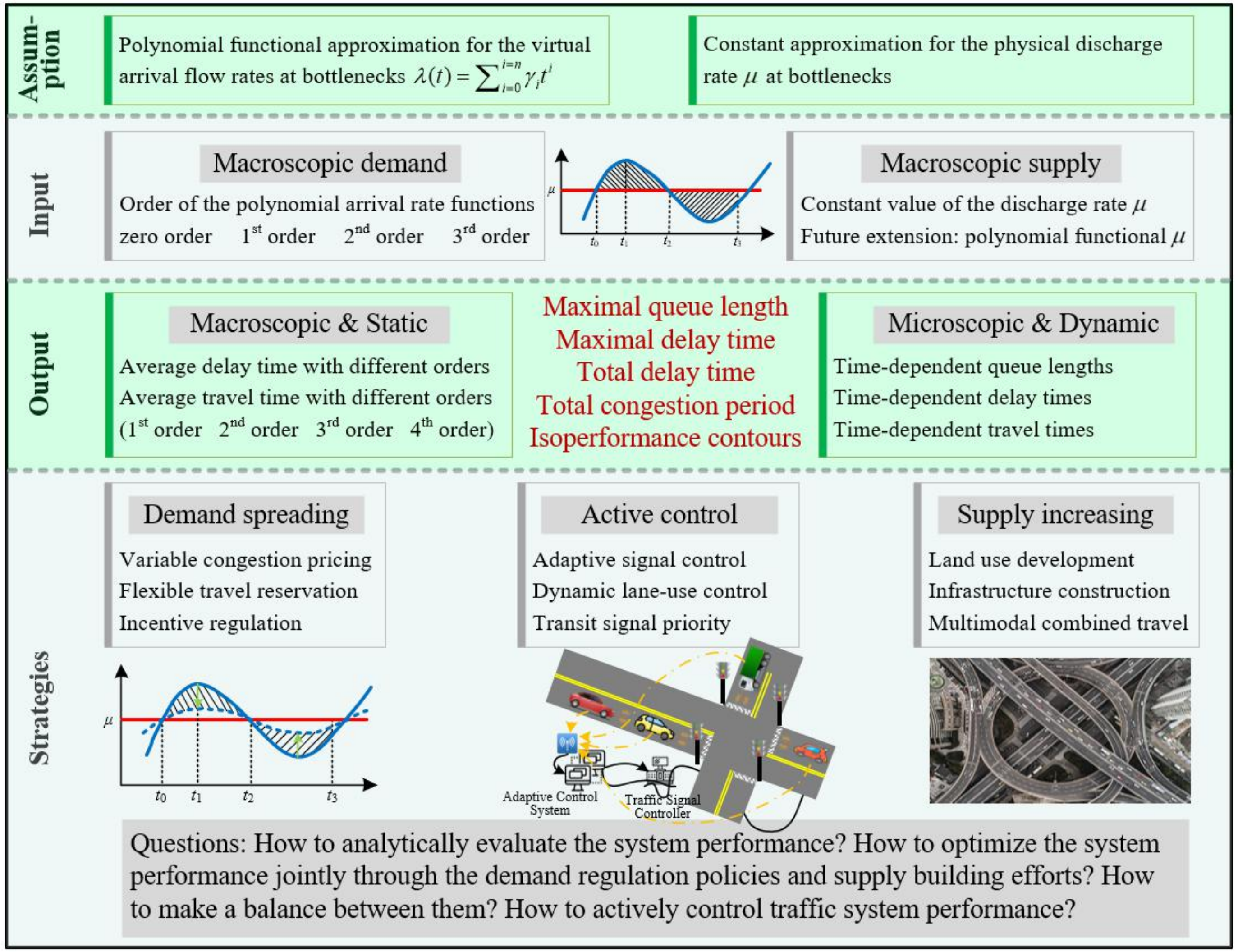
$$t = t_0 \left[1 + 0.15 \cdot \left(\boxed{V / C} \right)^4 \right]$$

- the used values of alpha and capacity are on the high side

Dynamic traffic assignment

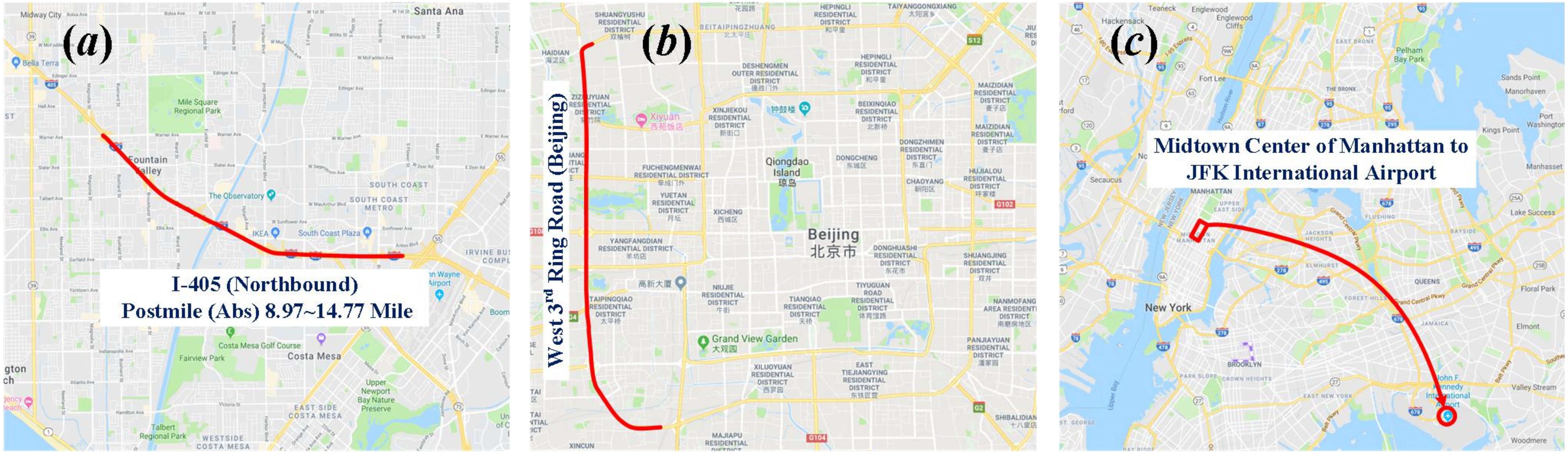
- capture traffic flow dynamics
- describe queue evolution
- point queue or physical queue
- CTM- or LTM-based model
- high input-detail burden
- computational inefficiency

1. Overview



Assumption
Polynomial functional approximation for the virtual arrival flow rates at bottlenecks $\lambda(t)=\sum_{i=0}^{i=n}\gamma_i t^i$
Constant approximation for the physical discharge rate μ at bottlenecks
Input
Macroscopic demand
Order of the polynomial arrival rate functions
zero order 1st order 2nd order 3rd order
μ
t_0 t_1 t_2 t_3
Macroscopic supply
Constant value of the discharge rate μ
Future extension: polynomial functional μ
Output
Macroscopic & Static
Average delay time with different orders
Average travel time with different orders
(1st order 2nd order 3rd order 4th order)
Maximal queue length
Maximal delay time
Total delay time
Total congestion period
Isoperformance contours
Microscopic & Dynamic
Time-dependent queue lengths
Time-dependent delay times
Time-dependent travel times
Strategies
Demand spreading
Variable congestion pricing
Flexible travel reservation
Incentive regulation
Active control
Adaptive signal control
Dynamic lane-use control
Transit signal priority
Supply increasing
Land use development
Infrastructure construction
Multimodal combined travel
μ
t_0 t_1 t_2 t_3
Adaptive Control System
Traffic Signal Controller
Questions: How to analytically evaluate the system performance? How to optimize the system performance jointly through the demand regulation policies and supply building efforts? How to make a balance between them? How to actively control traffic system performance?

1. Areas of the queueing system



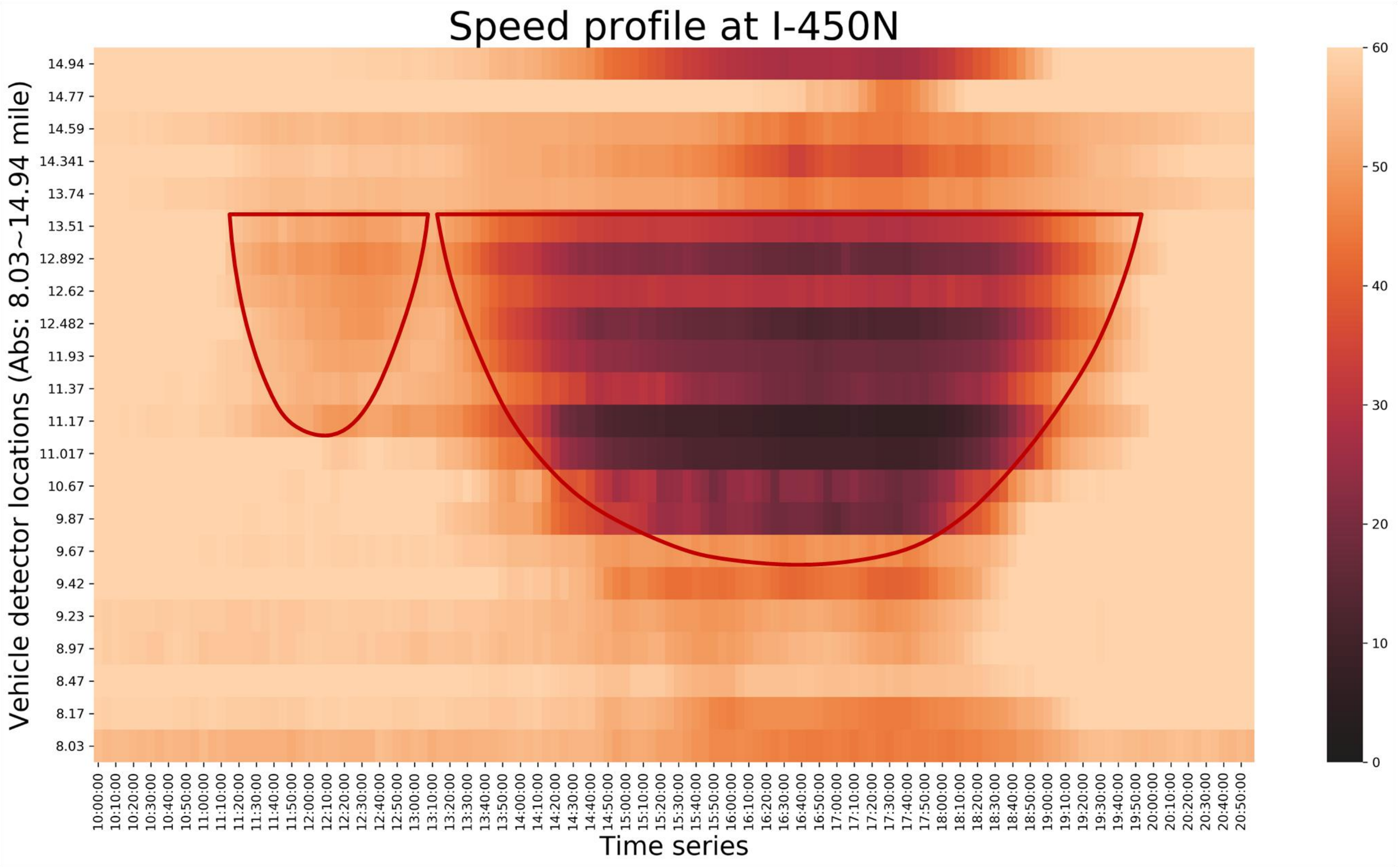
DS1: Los Angeles

DS2: Beijing

DS3: New York

1: DS1 Los Angeles

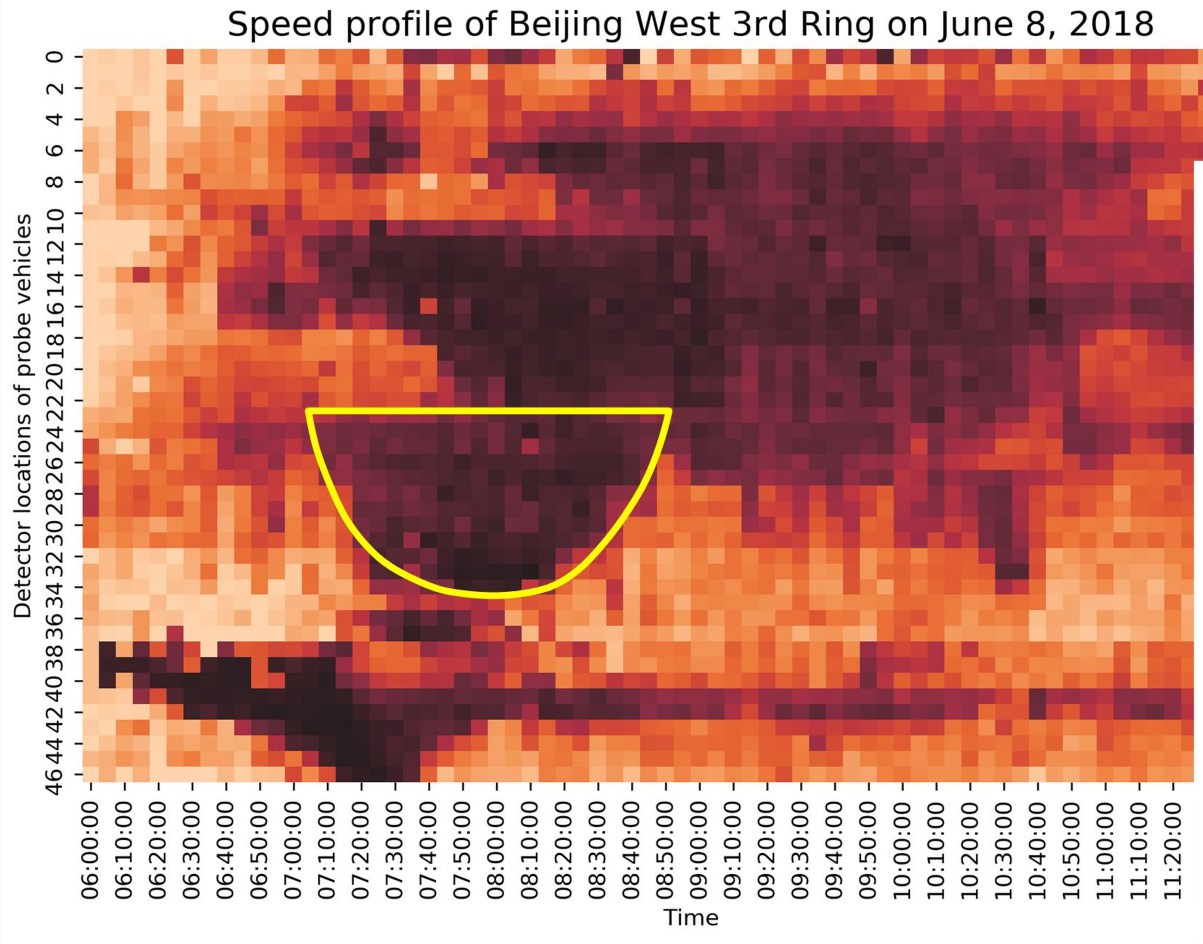
❑ Bottleneck detection with speed contour map



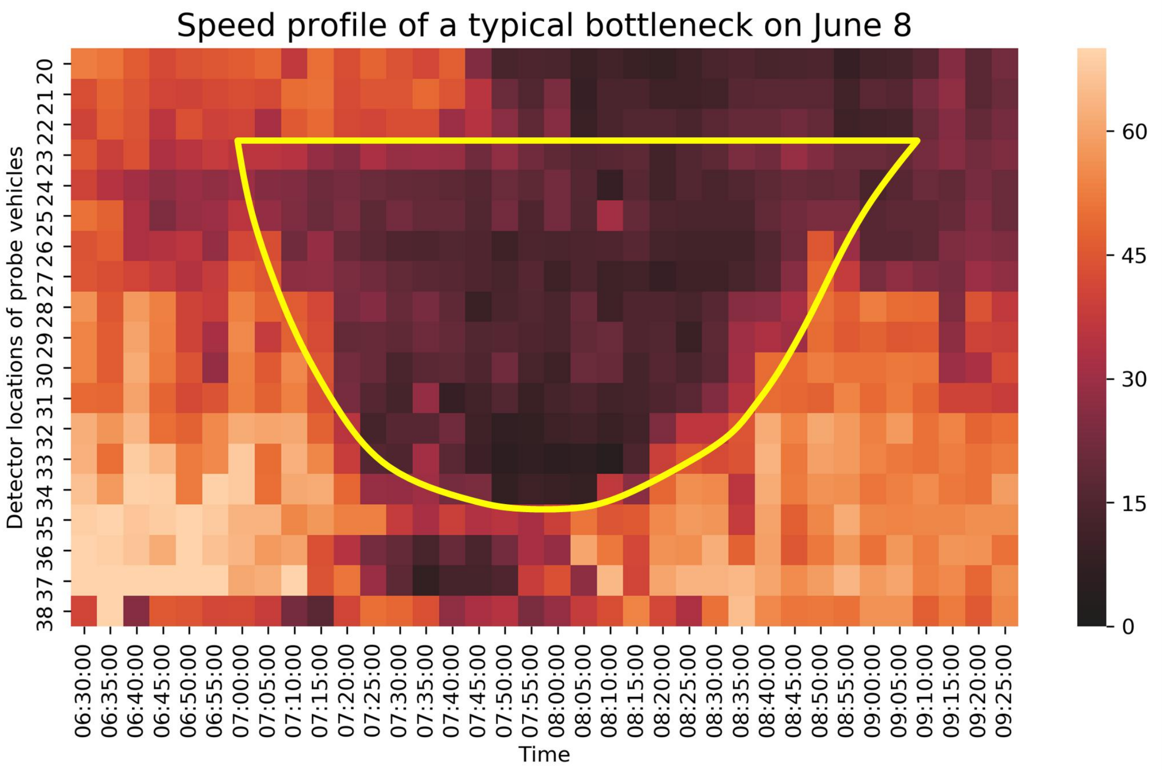
It is clear that the bottleneck is located at Abs=13.51 mile. In this case, we only analyze one single bottleneck with one peak period from t_0 = 13:10 to t_3 = 19:45.

1. DS2 Beijing

□ Bottleneck detection with speed contour map

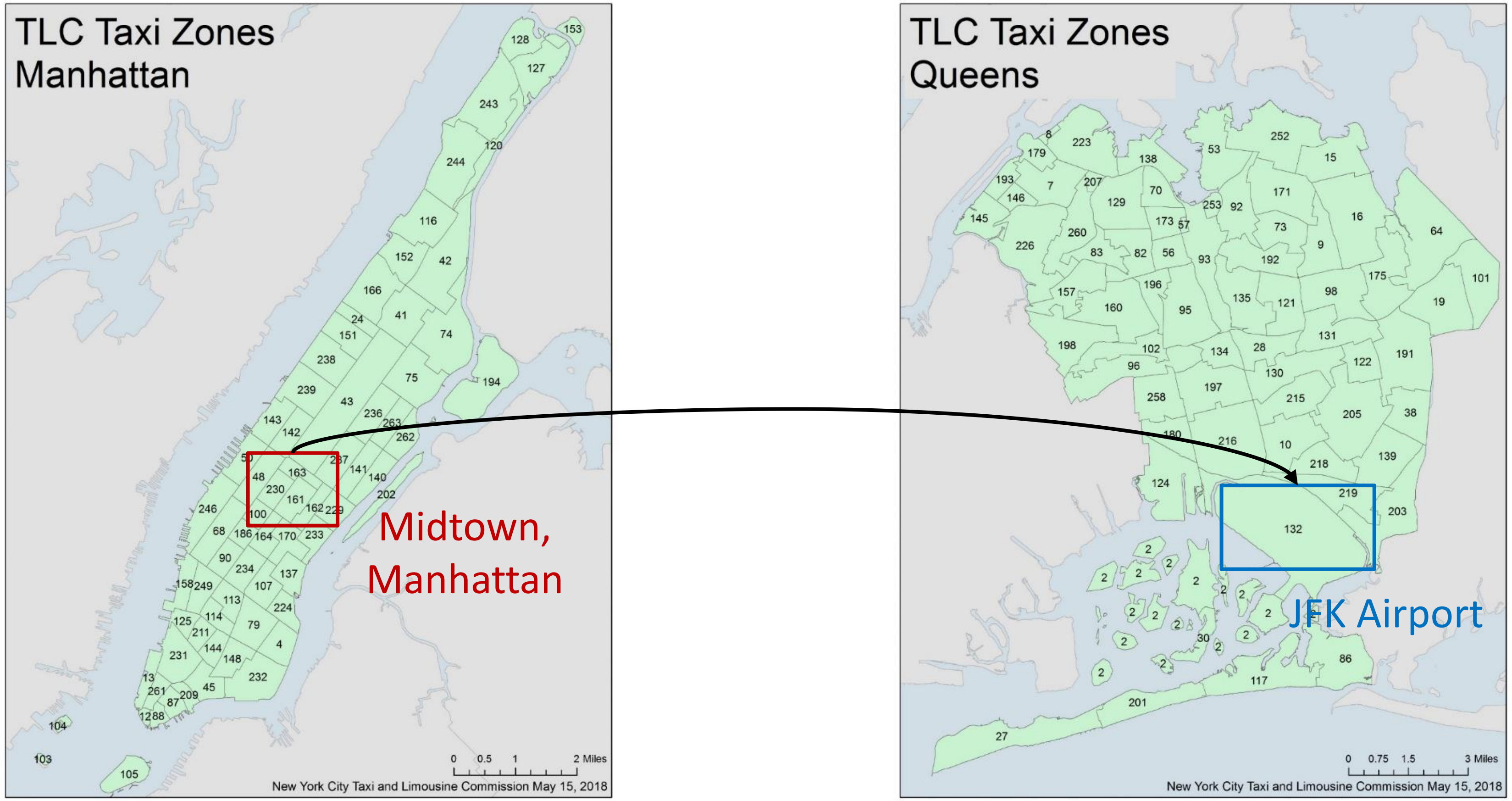


It exists the spatial queue spillback and the temporal queue connection phenomena.



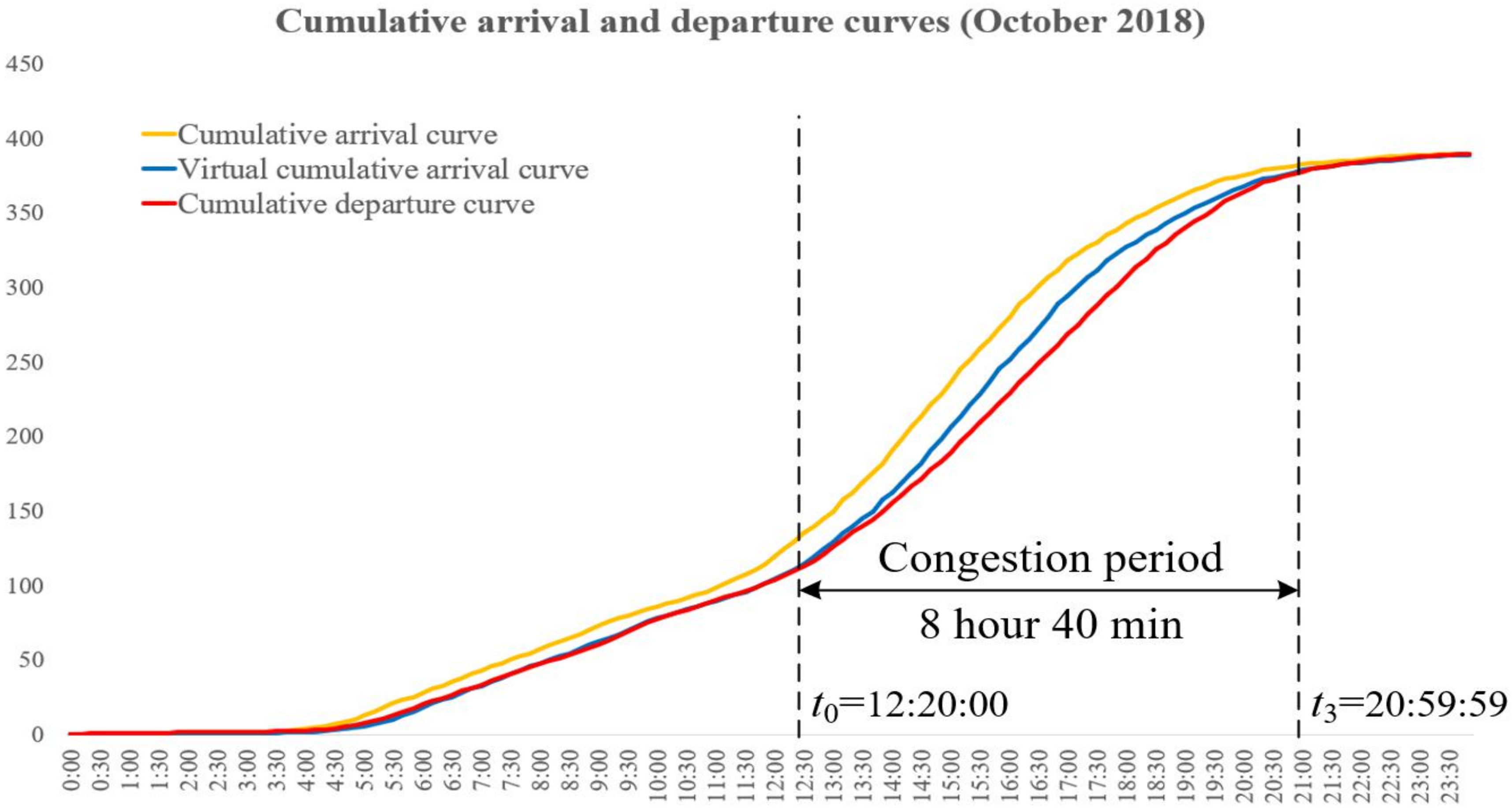
1: DS3 New York

◻ Bottleneck detection



DS3 New York

- Cumulative arrival and departure curves



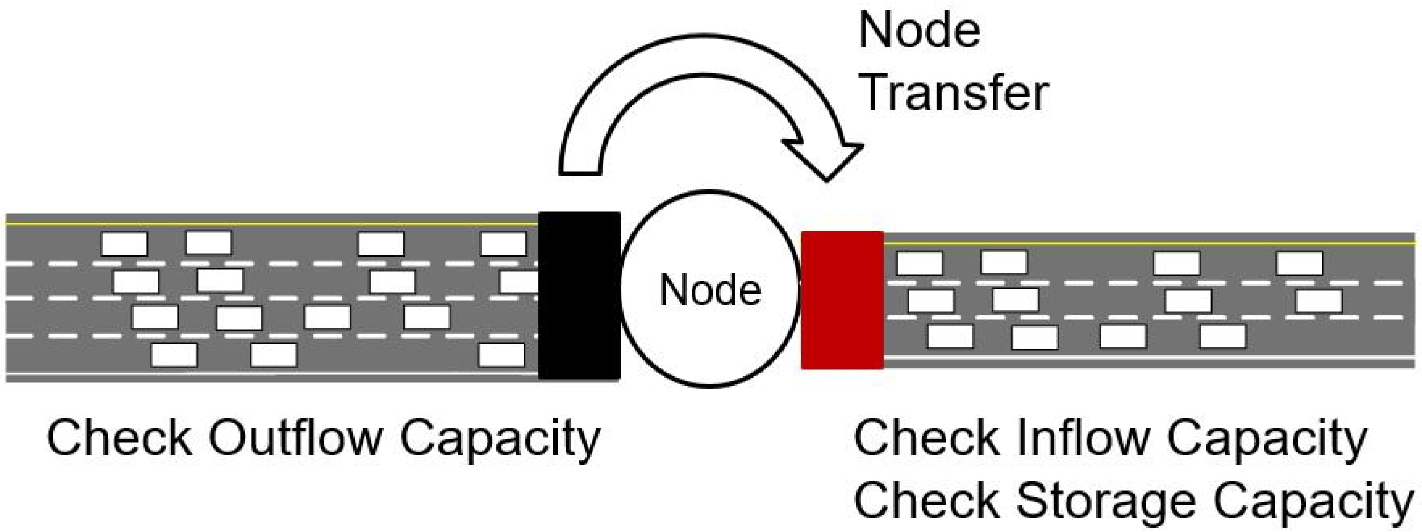
Description of empirical data sets

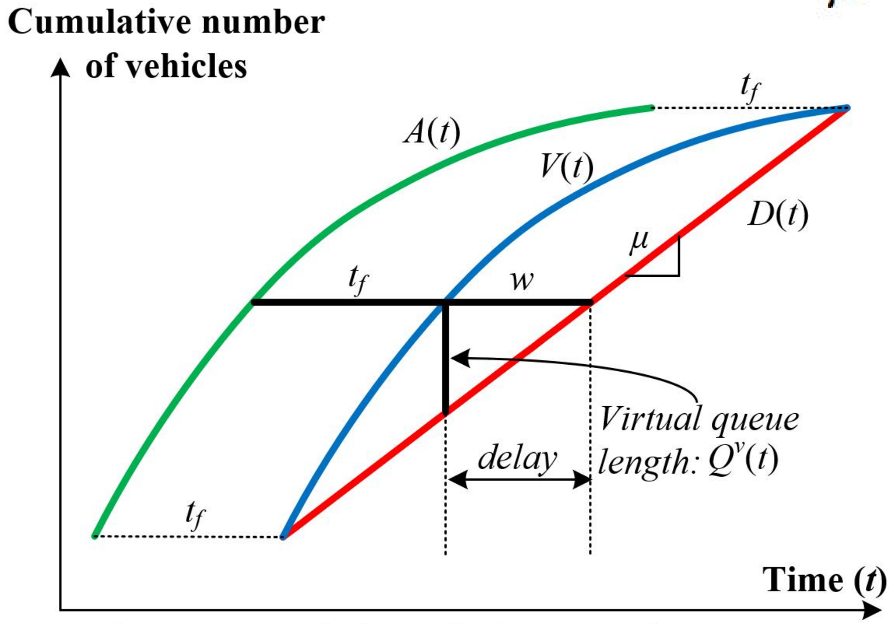
Data set	Queueing system	Object	Data type and information	Collected time and location
DS1	Traffic system	Vehicle	Freeway detector data with traffic flow, speed, and occupancy, etc.	Collected from the Northbound direction of I-405 freeway between absolute postmile 8.97 to 14.77 mile in Los Angeles on April, 2019, from 11:00 a.m. to 20:00 p.m.
DS2	Traffic system	Vehicle	Remote traffic microwave sensor data with the traffic flow, and probe vehicle data with averaged traffic speed, etc.	Collected from the west-third-ring of Beijing City on June 8, 2018, from 6:00 a.m. to 12:00 a.m.
DS3	Transportation system	Taxi	Taxi trip record data with trip ID, pickup time and location, drop off time and location, etc.	Collected from the Midtown Center of Manhattan to the John F. Kennedy International Airport in New York City on October 2018, provided by the New York City Taxi and Limousine Commission

2. Problem description

- How to evaluate the system performance for a single oversaturated bottleneck?

$$w(t) = \frac{Q^v(t)}{\mu}$$





$A(t)$ cumulative arrival curve at time t

$V(t)$ cumulative virtual arrival curve at time t

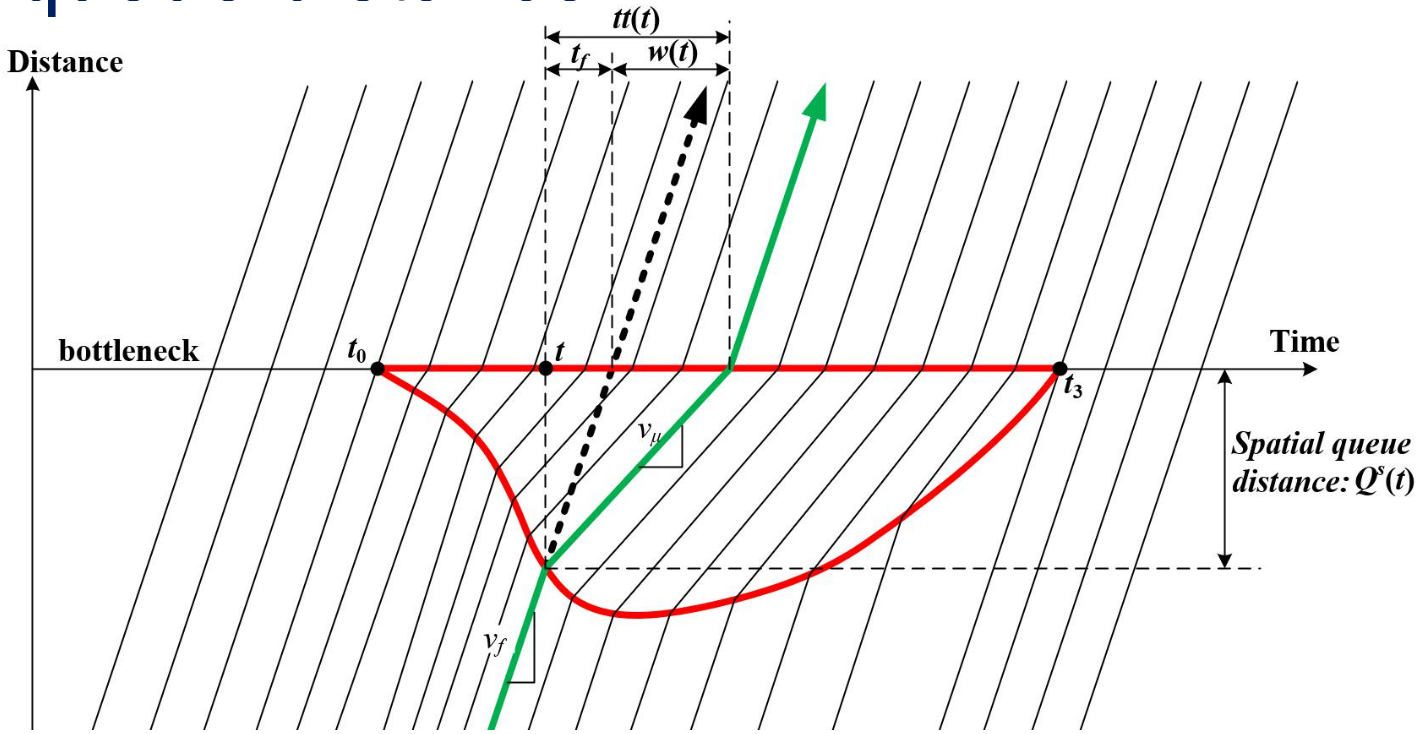
$D(t)$ cumulative departure curve at time t

2. Problem description

- How to evaluate the system performance for a single oversaturated bottleneck?
  - Point queue model
    - William Vickrey、ADL (Arnott, De Palma, Lindsey)、Kenneth Small、Michael Zhang、Erik Verhoef、Hai-Jun Huang、Wenlong Jin、Yu Nie、Xuegang Jeff Ban、W Y Szeto、Jiancheng Long、Ke Han、et al.

2. Problem description

- Virtual queue length vs. Physical queue length vs. Spatial queue distance



$$travel\ time\ in\ queue = \frac{Q^s(t)}{v_\mu}$$

$$= t_f + w(t) = \frac{Q^s(t)}{v_f} + w(t)$$

$$w(t) = \frac{Q^v(t)}{\mu}$$

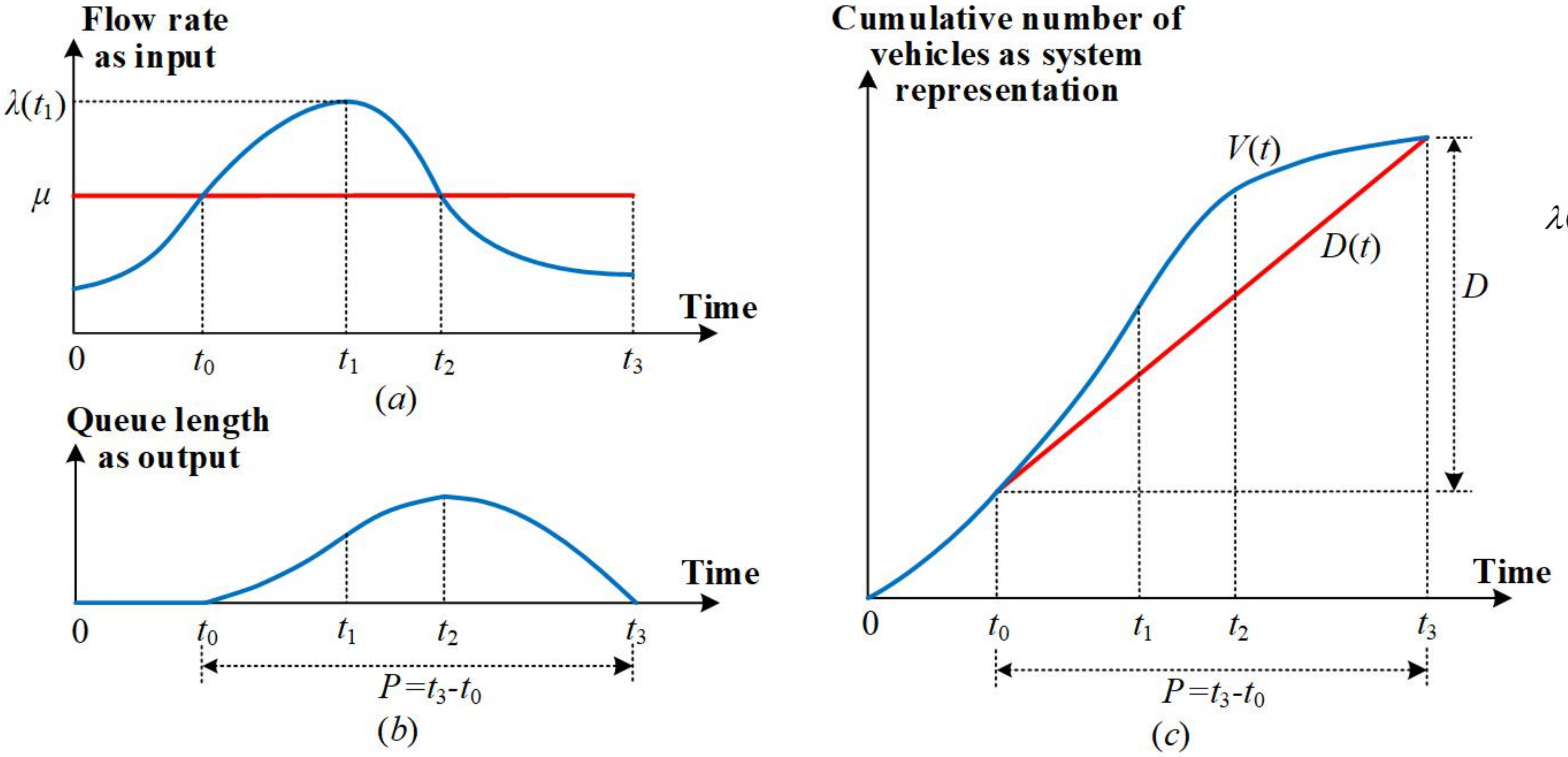
$$\boxed{Q^p(t) = \frac{Q^v(t)}{1 - \frac{v_\mu}{v_f}}}$$

$$\boxed{Q^s(t) = \frac{Q^v(t)}{\mu \cdot \left(\frac{1}{v_\mu} - \frac{1}{v_f}\right)}}$$

Lawson, T. W., Lovell, D. J., & Daganzo, C. F. (1997). *Transportation Research Record*, *1572*(1), 140-147.

2. Problem description

Deterministic queueing theory



t_0 start time of congestion period

t_1 time with maximum inflow rate

t_2 time with maximum queue length

t_3 end time of congestion period

$\lambda(t)$ inflow rate function at time t

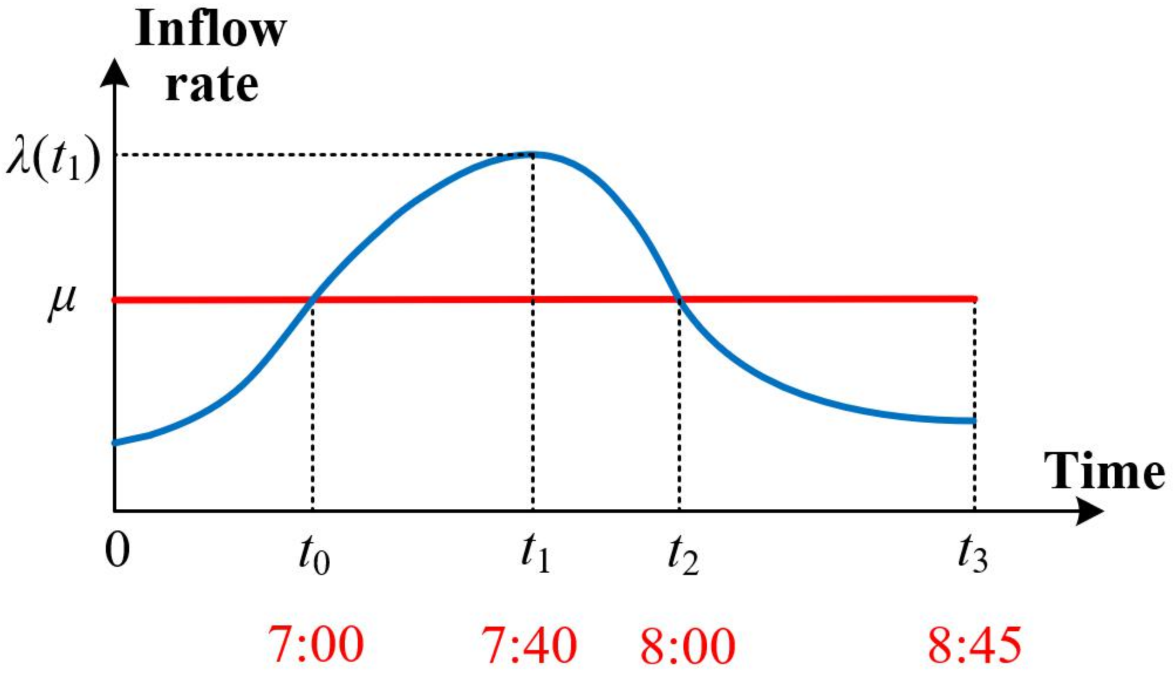
μ capacity (or discharge rate), assumed to be a constant value

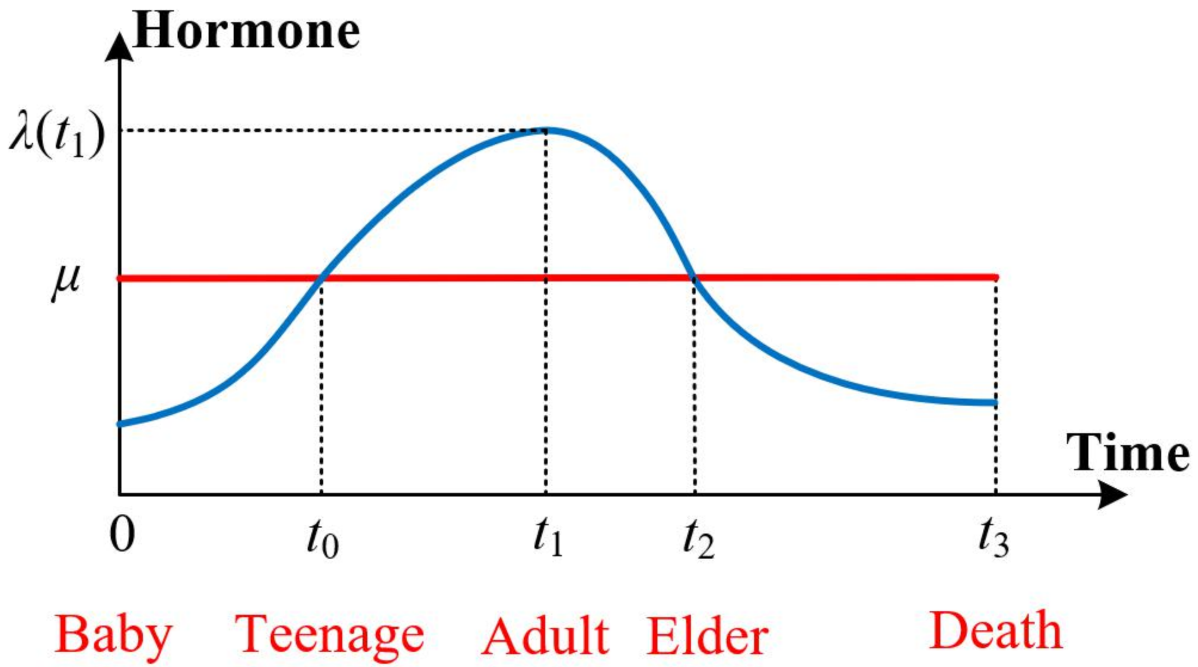
Newell, G.F. (1968). Queues with time-dependent arrival rates. III—A mild rush hour. Journal of Applied Probability, 5(3), 591-606.

Newell, G.F. (1982). Applications of queueing theory, 2nd ed. Chapman and Hall Ltd, New York.

2. Problem description

Examples





2. Problem description

Deterministic queueing theory

- Time-dependent virtual queue length at time t

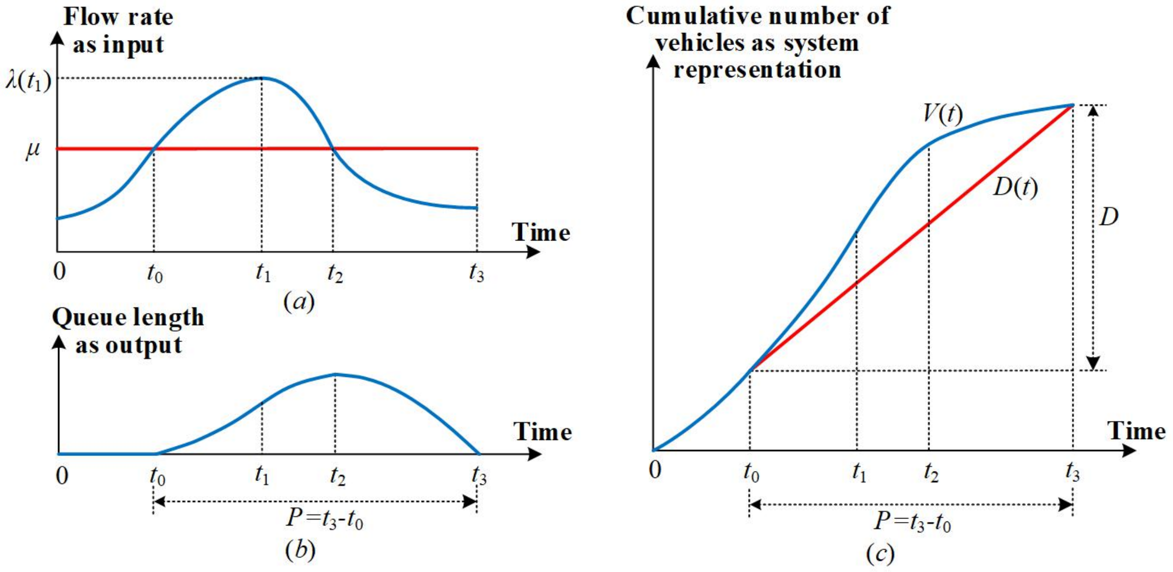
$$Q^v(t) = A(t) - D(t) = \int_{t_0}^{t} [\lambda(\tau) - \mu] d\tau$$

- Time-dependent delay at time t

$$w(t) = \frac{Q^v(t)}{\mu}$$

- Time-dependent travel time at time t

$$tt(t) = t_f + w(t)$$



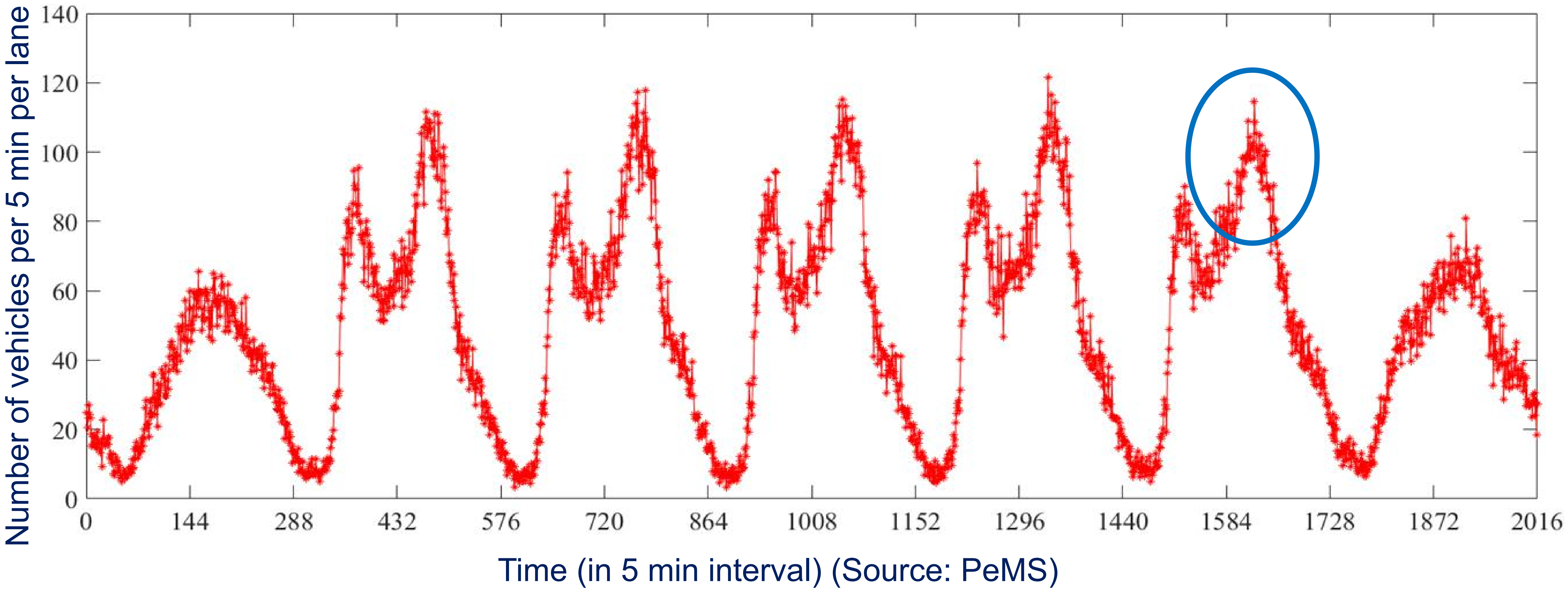
- Total delay $W = \int_{t_0}^{t_3} Q^v(\tau) d\tau$

- Average delay $w = \frac{W}{D} = \frac{W}{\mu P}$

- Average travel time $tt = t_f + w$

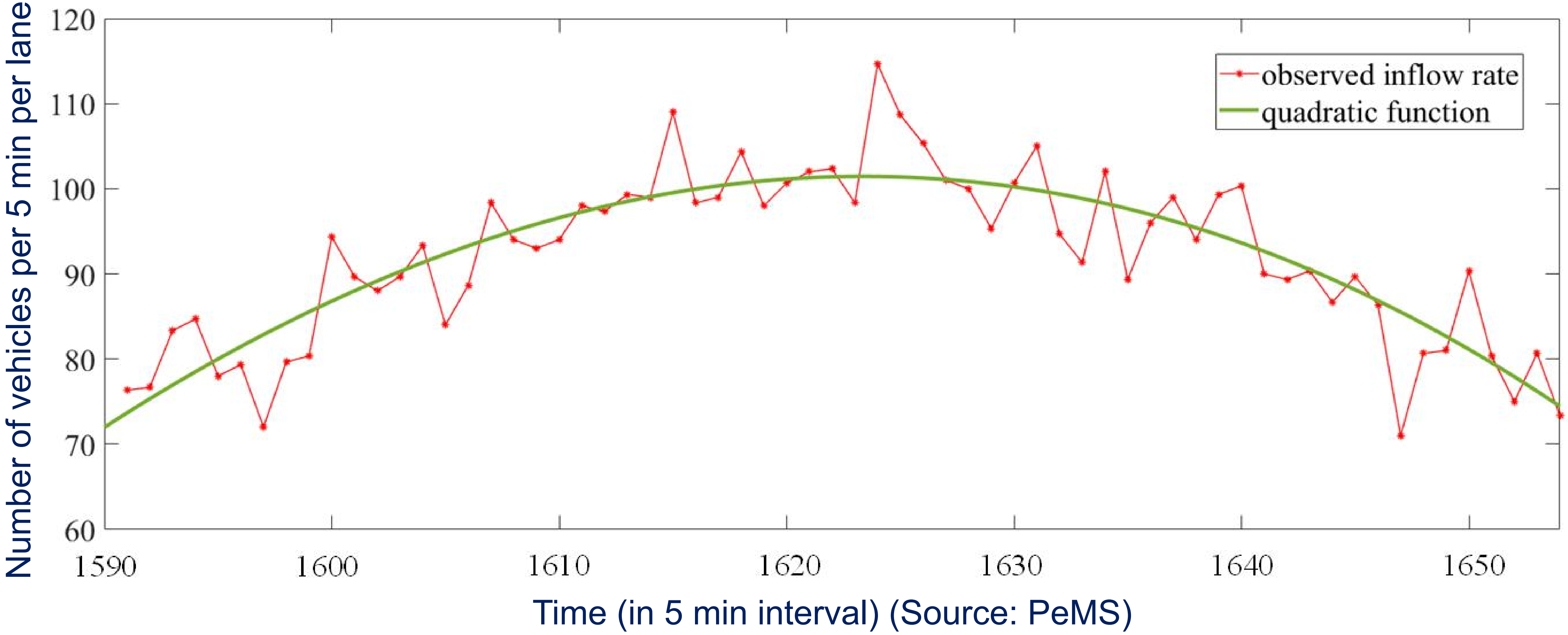
3. Polynomial approximation for the inflow rate

- Observation



3. Polynomial approximation for the inflow rate

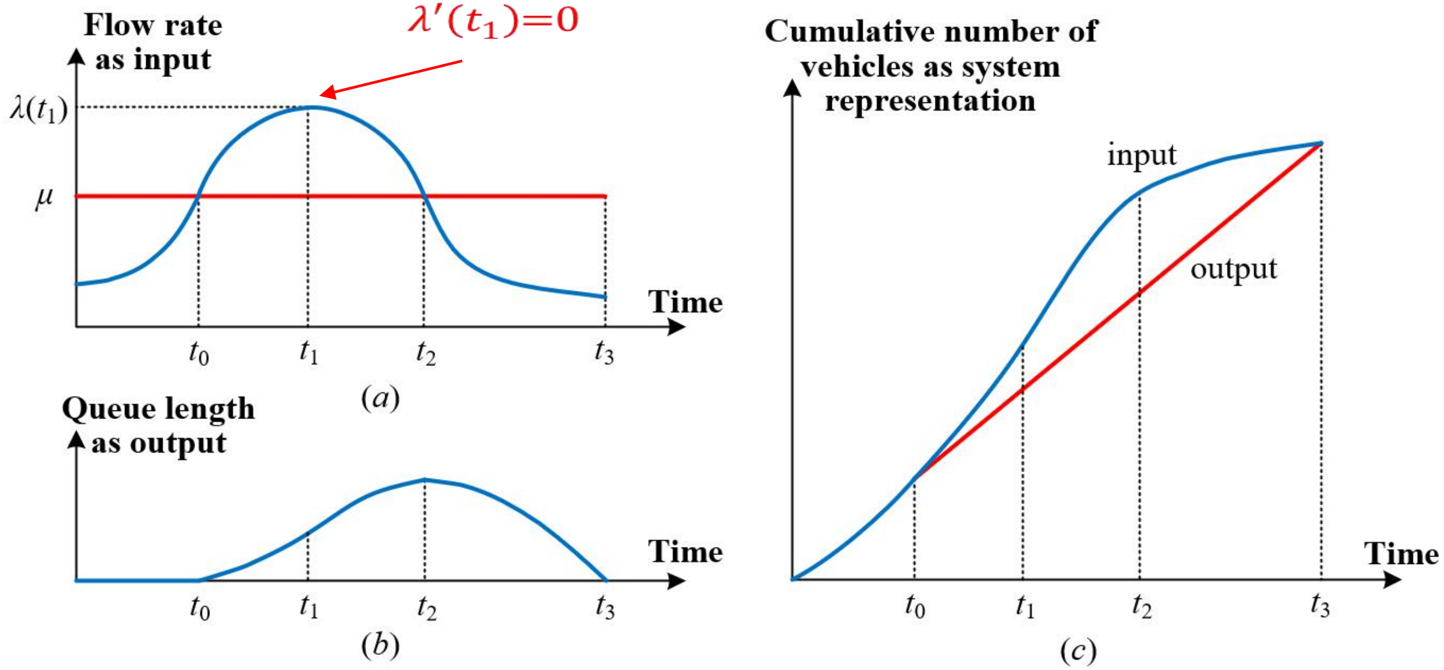
- Observation



3. Polynomial approximation for the inflow rate

□ Newell (1968, 1982)

- ➢ Assumption: quadratic inflow rate $\lambda(t) = \lambda(t_1) + \overset{=\,0}{\boxed{\lambda'(t_1)}} \cdot (t - t_1) + \overset{=\,-\rho}{\boxed{\frac{\lambda''(t_1)}{2}}} (t - t_1)^2$



$\lambda(t)$: inflow rate at time t
$\lambda'(t_1)$: first-order derivative of the inflow rate function at time t_1
$\lambda''(t_1)$: second-order derivative of the inflow rate function at time t_1
t_0: start time of congestion period
t_1: time index with maximum inflow rate
t_2: time index with maximum queue length
t_3: end time of congestion period
μ: discharge rate (or capacity)

3. Polynomial approximation for the inflow rate

◻ Newell (1968, 1982)

- $\lambda(t) = \lambda(t_1) - \rho(t - t_1)^2$

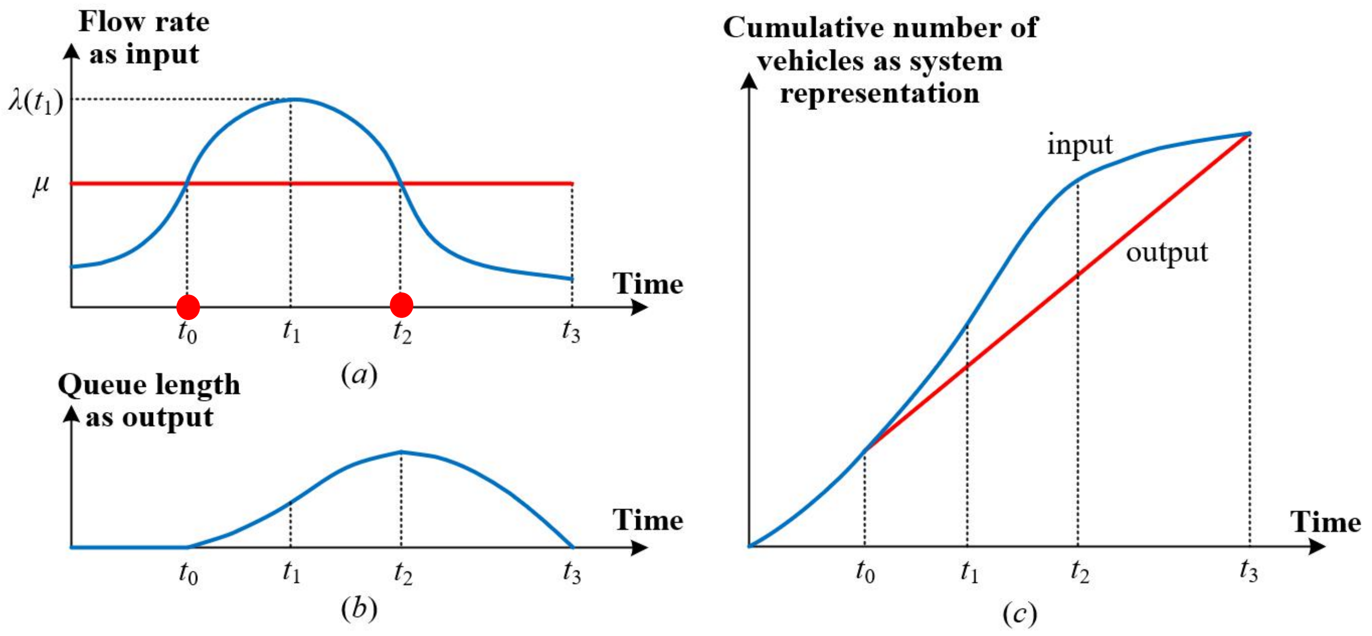
$\because \mu = \lambda(t_0) = \lambda(t_2)$ $\quad\quad \therefore \mu = \lambda(t_1) - \rho(t_0 - t_1)^2 = \lambda(t_1) - \rho(t_2 - t_1)^2$

$$\therefore t_0 = t_1 - \left[\frac{\lambda(t_1) - \mu}{\rho}\right]^{1/2}$$

$$t_2 = t_1 + \left[\frac{\lambda(t_1) - \mu}{\rho}\right]^{1/2}$$

- Factored form of $\lambda(t) - \mu$:

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



3. Polynomial approximation for the inflow rate

□ Newell (1968, 1982)

- Queue length: $A(t) - D(t)$

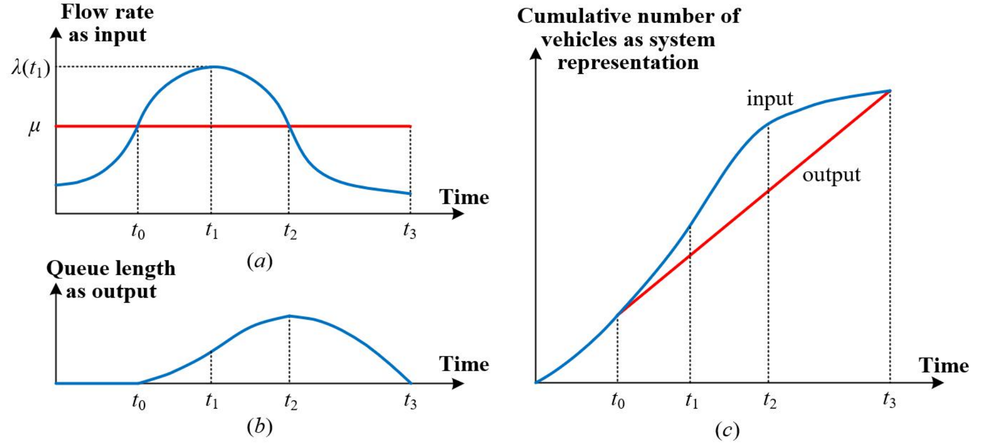
$$Q(t) = A(t) - D(t) = \int_{t_0}^{t} [\lambda(\tau) - \mu]\, d\tau$$

$$= \int_{t_0}^{t} [\rho(\tau - t_0)(t_2 - \tau)]\, d\tau$$

Integration by substitution

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

$$\boxed{Q(t) = \frac{\rho}{3}(t - t_0)^2(t_3 - t)}$$



- Maximum queue length at time t_2:

$$Q(t_2) = \frac{\rho}{6}(t_2 - t_0)^3 = \frac{4[\lambda(t_1) - \mu]^{3/2}}{3\rho^{1/2}}$$

- Queue dissipating time t_3:

$$\boxed{t_3 = t_0 + \frac{3}{2}(t_2 - t_0)} = t_0 + 3(t_1 - t_0)$$

3. Polynomial approximation for the inflow rate

- Newell (1968, 1982)
  - Total delay: W

$$W = \int_{t_0}^{t_3} Q(\tau)\, d\tau = \frac{\rho}{36}(t_3 - t_0)^4 = \frac{9[\lambda(t_1) - \mu]^2}{4\rho}$$

$$\boxed{\text{Only two parameters: } \rho, \mu}$$

  - Average delay:

$$w = \frac{W}{d_3 - d_0} = \frac{\rho}{36} \cdot \frac{P^4}{D} = \frac{\rho}{36\mu} \cdot \left(\frac{D}{\mu}\right)^3$$

$$\boxed{\mu = \frac{d_3 - d_0}{t_3 - t_0} = \frac{D}{P}}$$

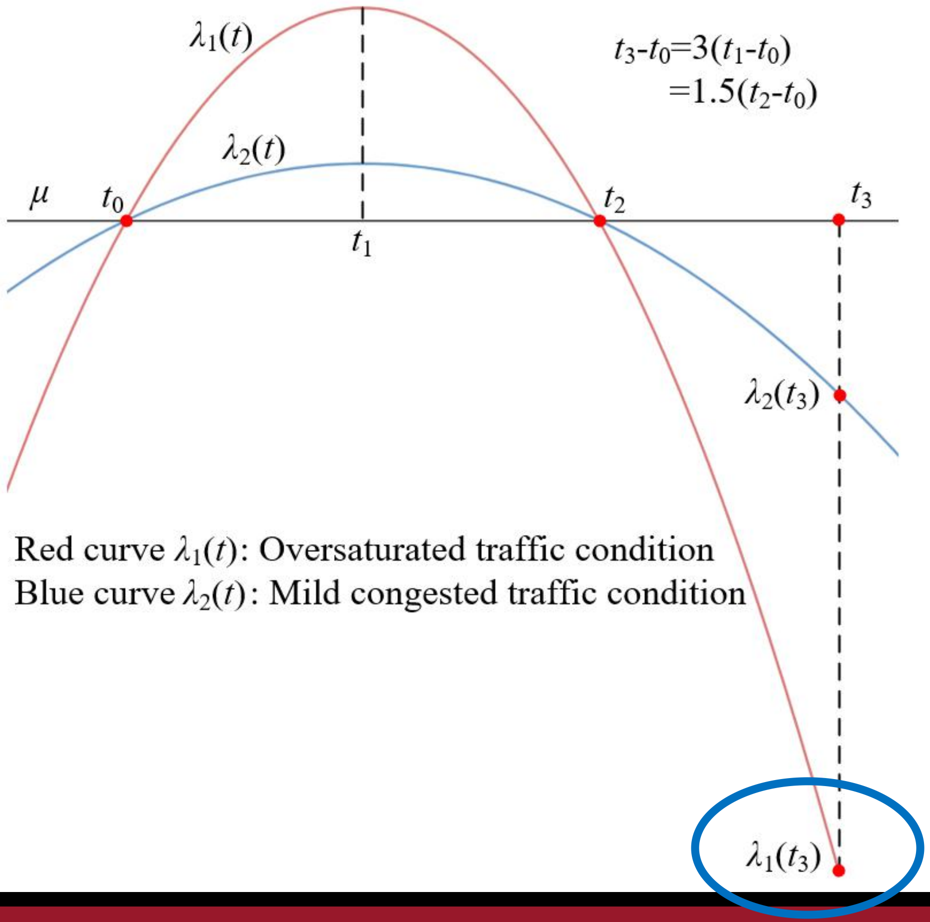
  - Average travel time function:

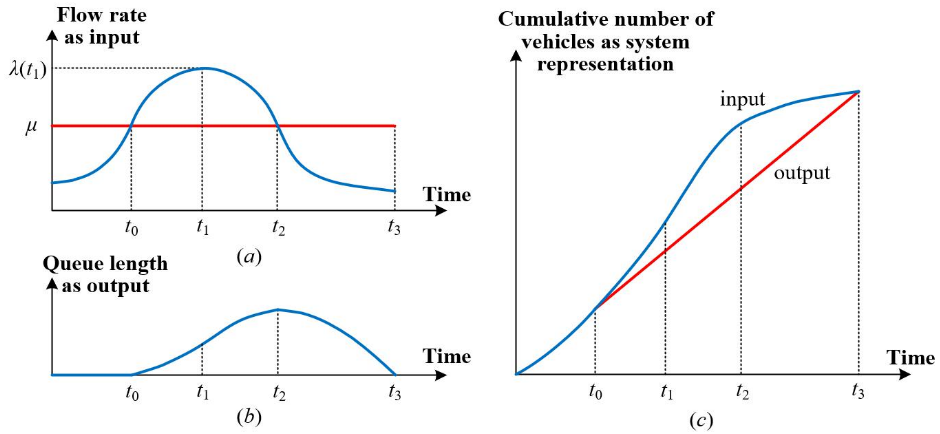
$$t = t_f + w = t_f \left[1 + \frac{\rho}{36\mu \cdot t_f} \cdot \left(\frac{D}{\mu}\right)^3\right]$$

$$\boxed{\text{Only two parameters: } \rho, \mu}$$

3. Polynomial approximation for the inflow rate

◻ Problems in the quadratic form





Only applicable for mild traffic congestion!

It even appears to be a negative inflow rate here!

3. Polynomial approximation for the inflow rate

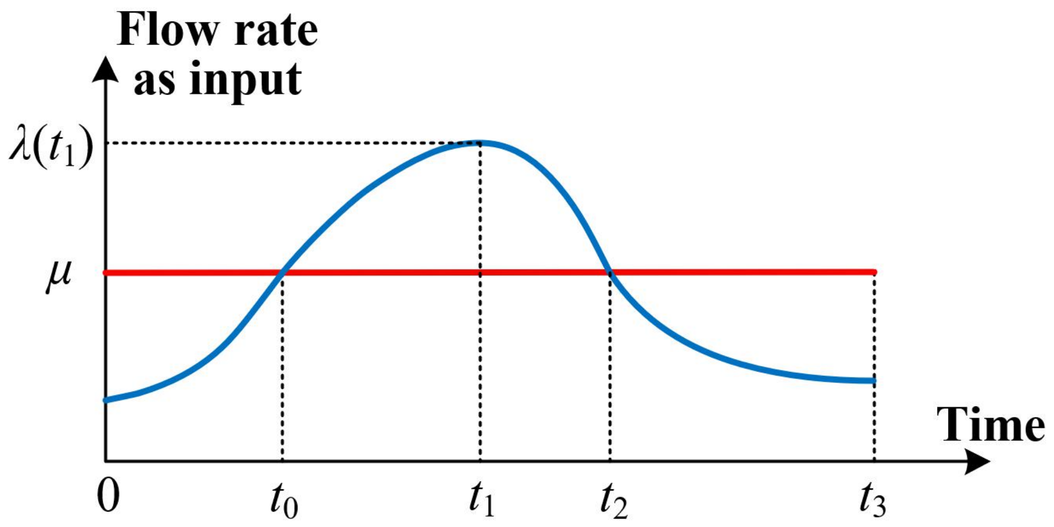
☐ Cubic form

- Assumption 1: The inflow rate can be approximated by a third-order polynomial (cubic) function, i.e.,

$$\lambda(t) = \gamma(t-t_0)(t-t_2)(t-\overline{t}) + \mu \quad or \quad \lambda(t) - \mu = \gamma(t-t_0)(t-t_2)(t-\overline{t})$$

- Three properties:

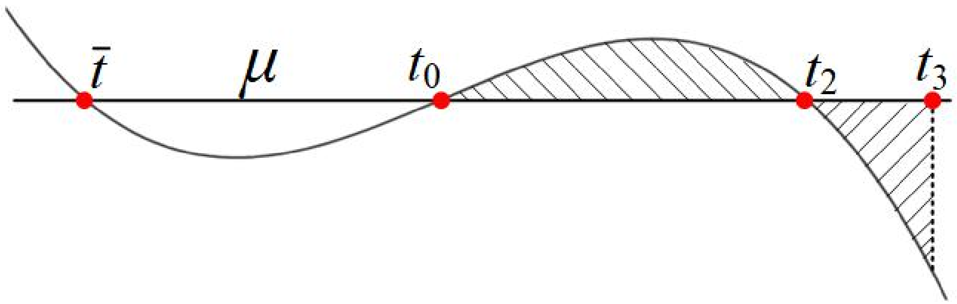
1) $\lambda(t) - \mu > 0$ when $t_0 < t < t_2$;
2) $\lambda(t) - \mu < 0$ when $t_2 < t < t_3$;
3) $\lambda(t_3)$ should not be very small.



3. Polynomial approximation for the inflow rate

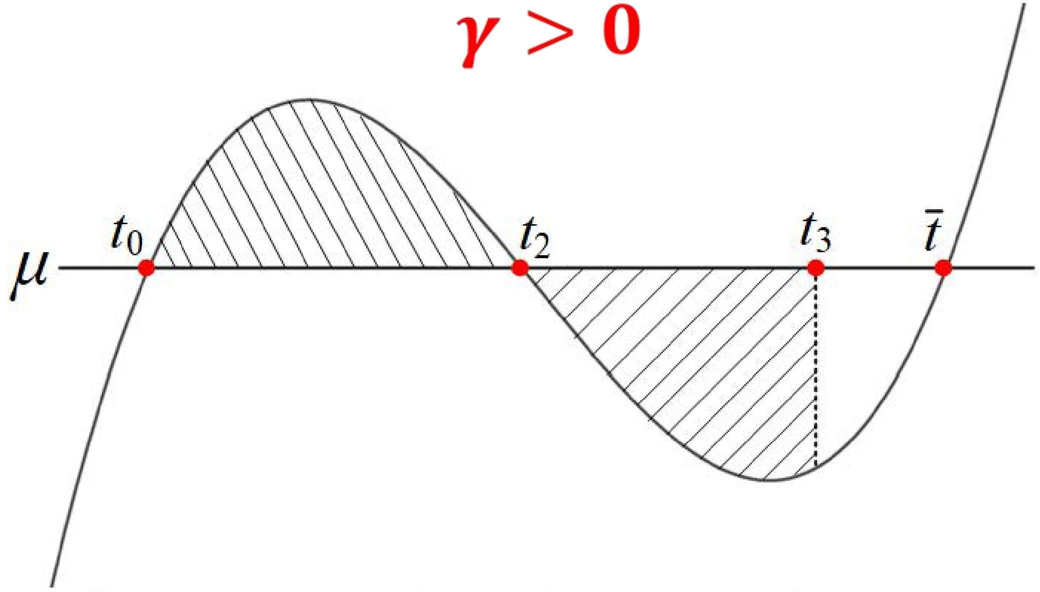
- Cubic form
  - Only two possible types

$\gamma < 0$



$$\lambda(t) - \mu = \gamma(t - t_0)(t - t_2)(t - \bar{t}),\ \gamma < 0,\ \bar{t} \le t_0 < t_2$$

$\gamma > 0$



$$\lambda(t) - \mu = \gamma(t - t_0)(t - t_2)(t - \bar{t}),\ \gamma > 0,\ t_0 < t_2 < \bar{t}$$

3. Polynomial approximation for the inflow rate

◻ Cubic form

- Assumption 2: The discharge rate (or capacity) of the bottleneck is a constant value (which equals μ).
- Assumption 3: The ratio between the time duration from the start of congestion to the time with maximum queue length and the whole congestion duration is *m*, i.e.,

$$t_2 - t_0 = m(t_3 - t_0),\ 0 < m < 1$$

3. Polynomial approximation for the inflow rate

☐ Cubic form with negative third-order shape parameter

- Input: net flow function $\lambda(t)-\mu=\gamma(t-t_0)(t-t_2)(t-\overline{t}),\ \gamma<0,\ \overline{t}\le t_0<t_2$
- Output: virtual queue length

$$Q^v(t)=\int_{t_0}^{t}\left[\gamma(\tau-t_0)(\tau-t_2)(\tau-\overline{t})\right]d\tau$$

$$=\gamma\cdot(t-t_0)^2\cdot\left[\frac{1}{4}(t-t_0)^2+\frac{1}{3}(2t_0-t_2-\overline{t})(t-t_0)+\frac{1}{2}(t_0-t_2)(t_0-\overline{t})\right]$$

$$Q^v(t_3)=0 \quad \Rightarrow \quad \overline{t}-t_0=\frac{3(t_3-t_0)^2-4(t_2-t_0)(t_3-t_0)}{4(t_3-t_0)-6(t_2-t_0)}$$

3. Polynomial approximation for the inflow rate

☐ Cubic form with negative third-order shape parameter

- Input: net flow function $\lambda(t)-\mu=\gamma(t-t_0)(t-t_2)(t-\bar{t}),\ \gamma<0,\ \bar{t}\le t_0<t_2$
- Output: virtual queue length

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

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

- Output: maximum virtual queue length $Q^v(t_2)=\gamma\cdot\dfrac{m^3(m-1)^2}{8-12m}\cdot(t_3-t_0)^4$

3. Polynomial approximation for the inflow rate

Cubic form with negative third-order shape parameter

- Input: net flow function $\lambda(t)-\mu=\gamma(t-t_0)(t-t_2)(t-\overline{t}),\ \gamma<0,\ \overline{t}\le t_0<t_2$
- Output: time-dependent delay

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

- Output: time-dependent travel time

$$tt(t)=t_f+\frac{\gamma\cdot(t-t_0)^2}{\mu}\cdot\left[\frac{1}{4}(t-t_0)^2-\frac{1}{3}\cdot\left(\frac{3-4m}{4-6m}+m\right)(t_3-t_0)(t-t_0)+\frac{1}{2}\cdot\frac{(3-4m)m}{4-6m}(t_3-t_0)^2\right]$$

3. Polynomial approximation for the inflow rate

- Cubic form with negative third-order shape parameter
  - Input: net flow function $\lambda(t)-\mu=\gamma(t-t_0)(t-t_2)(t-\bar{t}),\ \gamma<0,\ \bar{t}\le t_0<t_2$
  - Output: total delay between time t_0 and t_3

$$W=\int_{t_0}^{t_3} Q(t)dt=\gamma\cdot g(m)\cdot(t_3-t_0)^5$$

g(m): conversion factor

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

3. Polynomial approximation for the inflow rate

- Cubic form with negative third-order shape parameter
  - Input: net flow function $\lambda(t)-\mu=\gamma(t-t_0)(t-t_2)(t-\bar{t}),\ \gamma<0,\ \bar{t}\le t_0<t_2$
  - Output: average delay function

$$w=\frac{W}{D}=\frac{\gamma\cdot g(m)}{\mu}\cdot\left(\frac{D}{\mu}\right)^4 \quad \Leftarrow \quad P=t_3-t_0=D/\mu$$

  - Output: average travel time function

$$tt=t_f\left[1+\frac{\gamma\cdot g(m)}{\mu\cdot t_f}\cdot\left(\frac{D}{\mu}\right)^4\right]$$

3. Polynomial approximation for the inflow rate

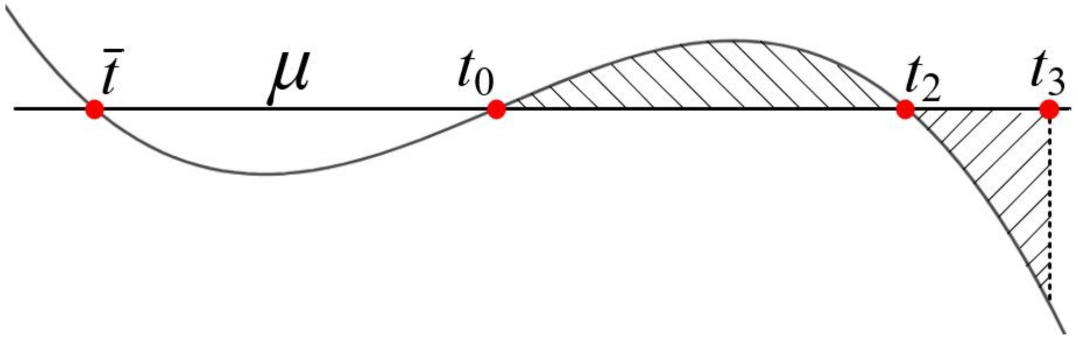
☐ Cubic form with negative third-order shape parameter

- Input: net flow function $\lambda(t)-\mu=\gamma(t-t_0)(t-t_2)(t-\bar{t}),\ \gamma<0,\ \bar{t}\le t_0<t_2$
- Discussion of *m*

$$Q^{v}(t_3)=0 \quad \Rightarrow \quad \bar{t}-t_0=\frac{3(t_3-t_0)^2-4(t_2-t_0)(t_3-t_0)}{4(t_3-t_0)-6(t_2-t_0)}$$

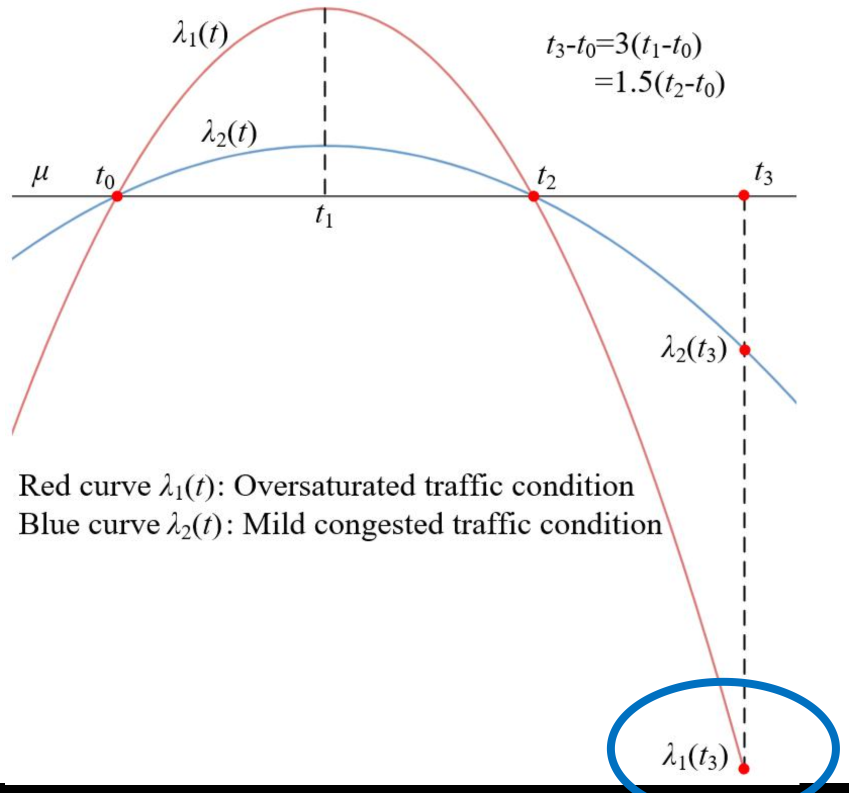
In a cubic form with $\gamma<0$: $\bar{t}-t_0\le 0$

$$\Rightarrow \quad m\in\left(\frac{2}{3},\frac{3}{4}\right] \qquad g(m)\in\left(-\infty,-\frac{1}{80}\right]$$



3. Polynomial approximation for the inflow rate

- Cubic form with negative third-order shape parameter
  - Note: only applicable for mild traffic congestion



It even appears to be a negative inflow rate here!

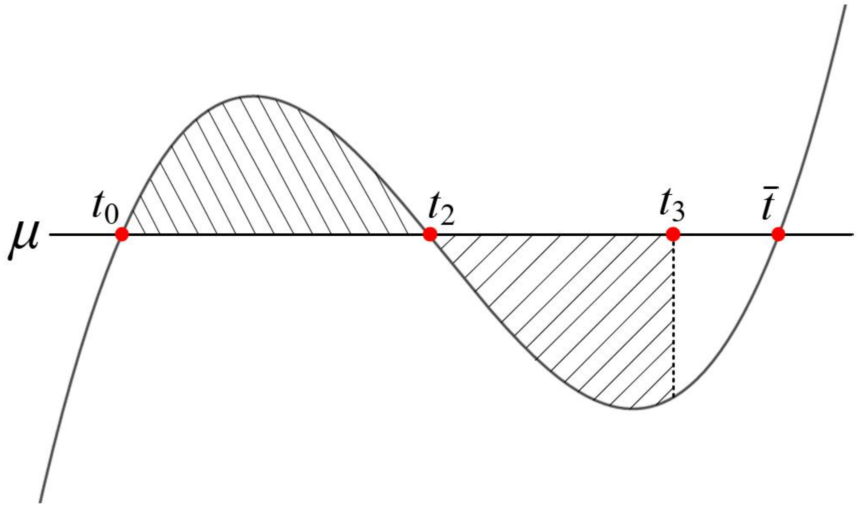
3. Polynomial approximation for the inflow rate

□ Cubic form with positive third-order shape parameter

- Can be used for both mild congestion and oversaturation!
- Input: net flow function $\lambda(t)-\mu=\gamma(t-t_0)(t-t_2)(t-\bar{t}),\ \boxed{\gamma>0,\ t_0<t_2<\bar{t}}$
- All the output formulas are the same with $\gamma<0$, except for the m

$\bar{t}\ge t_3>t_0 \quad \text{for} \quad \gamma>0$

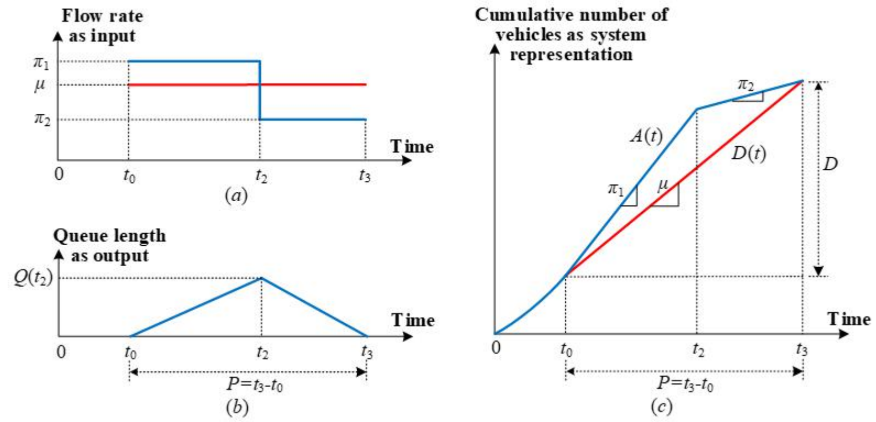
$\Rightarrow \quad m\in\left[\frac{1}{2},\frac{2}{3}\right) \qquad g(m)\in\left[\frac{1}{120},+\infty\right)$



3. Polynomial approximation for the inflow rate

A family of polynomial-function-based formulations

(1) Constant form for inflow rates



Zero-flow-first-queue (0F1Q)

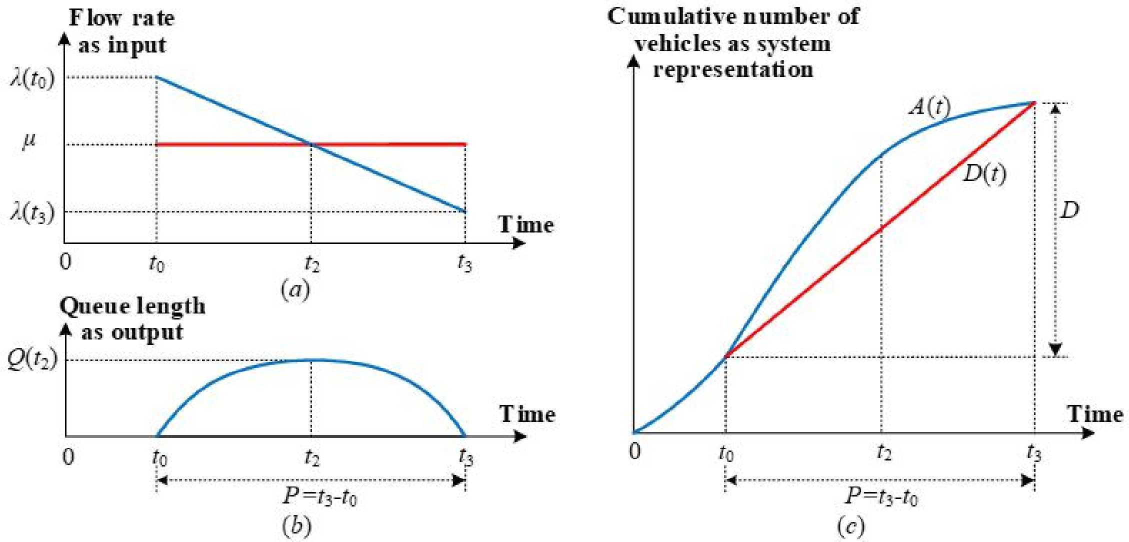
Vickrey, ADL, et. al.

Inflow rate	$\lambda(t)=\begin{cases}\pi_1>\mu,\ t_0\le t<t_2\\ \pi_2<\mu,\ t_2\le t\le t_3\end{cases}$
Timestamp relationship	$t_3=t_2+\dfrac{(t_2-t_0)(\pi_1-\mu)}{\mu-\pi_2}$
Queue length	$Q(t)=\begin{cases}(\pi_1-\mu)(t-t_0),\ t_0\le t<t_2\\ (\mu-\pi_2)(t_3-t),\ t_2\le t\le t_3\end{cases}$
Time-dependent delay	$w(t)=\begin{cases}\dfrac{(\pi_1-\mu)(t-t_0)}{\mu},\ t_0\le t<t_2\\ \dfrac{(\mu-\pi_2)(t_3-t)}{\mu},\ t_2\le t\le t_3\end{cases}$
Average delay	$\bar{w}=\dfrac{(\pi_1-\mu)(\mu-\pi_2)}{2\mu(\pi_1-\pi_2)}\cdot\left(\dfrac{D}{\mu}\right)$
Average travel time	$tt=t_f\cdot\left[1+\dfrac{(\pi_1-\mu)(\mu-\pi_2)}{2\mu(\pi_1-\pi_2)\cdot t_f}\cdot\left(\dfrac{D}{\mu}\right)\right]$

3. Polynomial approximation for the inflow rate

- A family of polynomial-function-based formulations

(2) Linear form for inflow rate



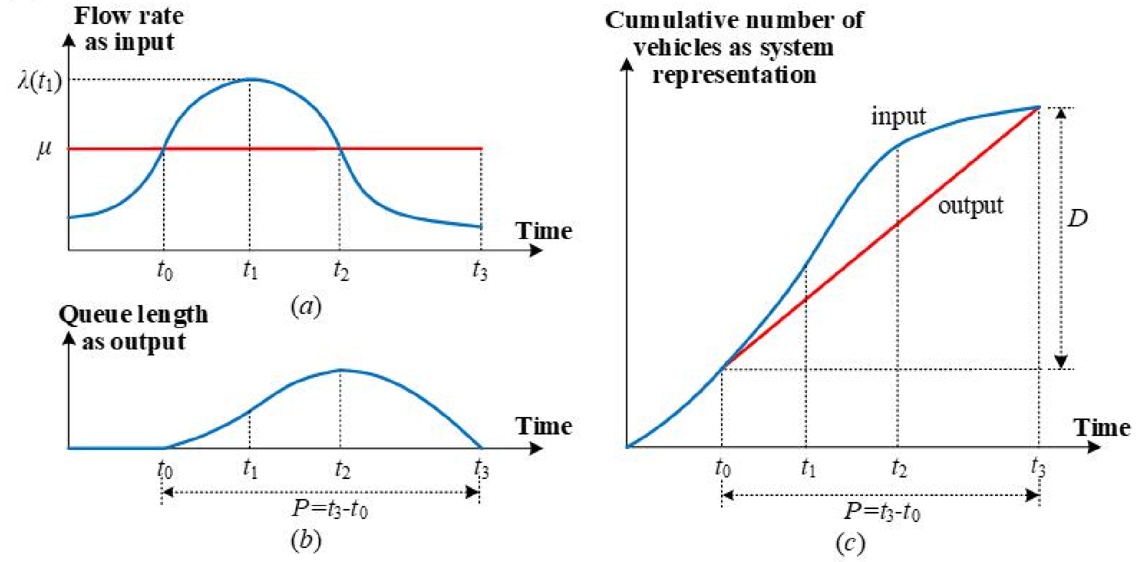
First-flow-second-queue (1F2Q)

Inflow rate	$\lambda(t) = -\kappa(t - t_2) + \mu,\ \kappa > 0$
Timestamp relationship	$t_3 = 2t_2 - t_0$
Queue length	$Q(t) = \frac{\kappa}{2}(t - t_0)(t_3 - t)$
Time-dependent delay	$w(t) = \frac{\kappa}{2\mu}(t - t_0)(t_3 - t)$
Average delay	$\bar{w} = \frac{\kappa}{12\mu} \cdot \left(\frac{D}{\mu}\right)^2$
Average travel time	$tt = t_f \cdot \left[1 + \frac{\kappa}{12\mu \cdot t_f} \cdot \left(\frac{D}{\mu}\right)^2\right]$

3. Polynomial approximation for the inflow rate

- A family of polynomial-function-based formulations

(3) Quadratic form for inflow rate



Second-flow-third-queue (2F3Q)

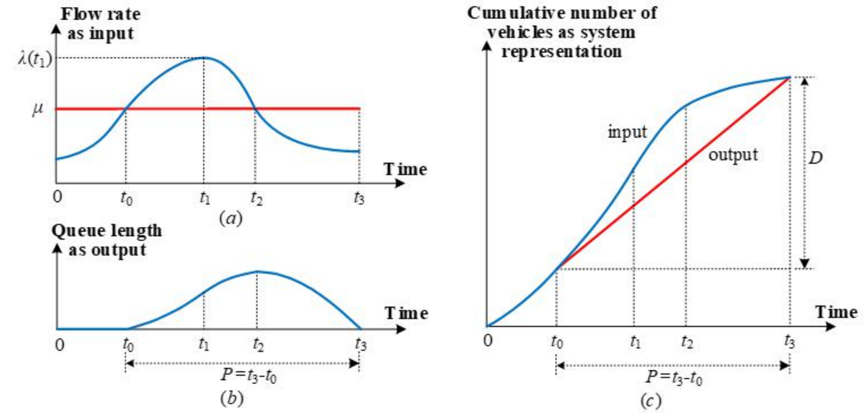
Newell (1982)

Inflow rate	$\lambda(t) = -\rho(t-t_0)(t-t_2)+\mu,\ \rho>0$
Timestamp relationship	$t_3 = t_0 + \frac{3}{2}(t_2-t_0) = t_0 + 3(t_1-t_0)$
Queue length	$Q(t) = \frac{\rho}{3}(t-t_0)^2(t_3-t)$
Time-dependent delay	$w(t) = \frac{\rho}{3\mu}(t-t_0)^2(t_3-t)$
Average delay	$\bar{w} = \frac{\rho}{36\mu}\cdot\left(\frac{D}{\mu}\right)^3$
Average travel time	$tt = t_f \cdot \left[1+\frac{\rho}{36\mu\cdot t_f}\cdot\left(\frac{D}{\mu}\right)^3\right]$

3. Polynomial approximation for the inflow rate

◻ A family of polynomial-function-based formulations

(4) Cubic form for inflow rate



Third-flow-fourth-queue (3F4Q)

Our model

β-flow-($\beta+1$)-queue

Inflow rate	$\lambda(t)=\gamma(t-t_0)(t-t_2)(t-\bar{t})+\mu$
Timestamp relationship	$t_2-t_0=m(t_3-t_0),\ 0<m<1$
Queue length	$Q(t)=\gamma\cdot(t-t_0)^2\cdot\left[\frac{1}{4}(t-t_0)^2-\frac{1}{3}\cdot\left(\frac{3-4m}{4-6m}+m\right)(t_3-t_0)(t-t_0)+\frac{1}{2}\cdot\frac{(3-4m)m}{4-6m}(t_3-t_0)^2\right]$
Time-dependent delay	$w(t)=\frac{\gamma\cdot(t-t_0)^2}{\mu}\cdot\left[\frac{1}{4}(t-t_0)^2-\frac{1}{3}\cdot\left(\frac{3-4m}{4-6m}+m\right)\cdot(t_3-t_0)(t-t_0)+\frac{1}{2}\cdot\frac{(3-4m)m}{4-6m}(t_3-t_0)^2\right]$
Average delay	$w=\frac{W}{D}=\frac{\gamma\cdot g(m)}{\mu}\cdot\left(\frac{D}{\mu}\right)^4$
Average travel time	$tt=t_f\left[1+\frac{\gamma\cdot g(m)}{\mu\cdot t_f}\cdot\left(\frac{D}{\mu}\right)^4\right]$

3. Polynomial approximation for the inflow rate

- A family of polynomial-function-based formulations

Arrival rate form	Arrival rate function	Average travel time function
Constant form	$\lambda(t) = \begin{cases} \pi_1 > \mu, \ t_0 \le t < t_2 \\ \pi_2 < \mu, \ t_2 \le t \le t_3 \end{cases}$	$tt = t_f \cdot \left[1 + \frac{(\pi_1 - \mu)(\mu - \pi_2)}{2\mu(\pi_1 - \pi_2) \cdot t_f} \cdot \left(\frac{D}{\mu} \right) \right]$
Linear form	$\lambda(t) = -\kappa(t - t_2) + \mu, \ \kappa > 0$	$tt = t_f \cdot \left[1 + \frac{\kappa}{12\mu \cdot t_f} \cdot \left(\frac{D}{\mu} \right)^2 \right]$
Quadratic form	$\lambda(t) = -\rho(t - t_0)(t - t_2) + \mu, \ \rho > 0$	$tt = t_f \cdot \left[1 + \frac{\rho}{36\mu \cdot t_f} \cdot \left(\frac{D}{\mu} \right)^3 \right]$
Cubic form	$\lambda(t) = \gamma(t - t_0)(t - t_2)(t - \overline{t}) + \mu$	$tt = t_f \cdot \left[1 + \frac{\gamma \cdot g(m)}{\mu \cdot t_f} \cdot \left(\frac{D}{\mu} \right)^4 \right]$

Related to free flow travel time, capacity, demand, shape parameter, (m).

3. Polynomial approximation for the inflow rate

A family of polynomial-function-based formulations

- Further extension to arbitrary order for the inflow rate function
- Assumption: $\lambda(t)-\mu=\rho(t-t_0)^k(t_2-t)^{n-k}$
- Queue length:

$$Q(t)=\sum_{i=k+1}^{n+1}\frac{\rho k!(n-k)!}{i!(n+1-i)!}(t-t_0)^i(t_2-t)^{n-i+1},\ t_0<t_2<t_3$$

with the boundary conditions of

$$Q(t_0)=0,\ Q(t_3)=0$$

$$Q(t)>0,\ t\in(t_0,t_3)$$

- We can analytically solve $Q(t)=0$ to obtain t_0 and t_3, and then obtain other traffic performance measurements step by step.

3. Polynomial approximation for the inflow rate

- A family of polynomial-function-based formulations
  - Further extension to arbitrary order for the inflow rate function
  - For example: $n = 2.5, k = 1$, i.e.,

$$\lambda(t) - \mu = \rho(t - t_0)(t_2 - t)^{1.5}$$

$$= \frac{\rho(t - t_0)(t_2 - t)^3}{(t_2 - t)^2}$$

Speed Limit (mph)	Practical Capacity (vehicle per hour)	Model Parameters	
		a	b
0 - 30	0 - 240	0.7312	3.6596
0 - 30	249 - 499	0.6128	3.5038
0 - 30	500 - 749	0.8774	4.4613
0 - 30	750 - 999	0.6846	5.1644
0 - 30	1000+	1.1465	4.4239
31 - 40	250 - 499	0.6190	3.6544
31 - 40	500 - 749	0.6662	4.9432
31 - 40	750 - 999	0.6222	5.1409
31 - 40	1000+	1.0300	5.5226
41 - 50	500 - 749	0.6609	5.0906
41 - 50	750 - 999	0.5423	5.7894
41 - 50	1000+	1.0091	6.5856
50+	500 - 749	0.8776	4.9287
50+	750 - 999	0.7699	5.3443
50+	1000+	1.1491	6.8677

Reference: Mannering et al. (1990)

4. Calibrations

Calibration model

$$\min_{\mathbf{x}=(\mu,\gamma,m)} f(\mathbf{x})$$

$$= \frac{1}{2}\sum_{t=1}^{|P|}\left\{\left[\frac{N(t)-N_{\min}}{N_{\max}-N_{\min}}-\frac{\widehat{N}(t)-\widehat{N}_{\min}}{\widehat{N}_{\max}-\widehat{N}_{\min}}\right]^2+\left[\frac{Q(t)-Q_{\min}}{Q_{\max}-Q_{\min}}-\frac{\widehat{Q}(t)-\widehat{Q}_{\min}}{\widehat{Q}_{\max}-\widehat{Q}_{\min}}\right]^2+\left[\frac{w(t)-w_{\min}}{w_{\max}-w_{\min}}-\frac{\widehat{w}(t)-\widehat{w}_{\min}}{\widehat{w}_{\max}-\widehat{w}_{\min}}\right]^2\right\}$$

subject to

$$\mu > 0$$

$$Q(t) \geq 0 \qquad m \in \begin{cases} \left[\frac{1}{2},\frac{2}{3}\right), \gamma > 0 \\ \left(\frac{2}{3},\frac{3}{4}\right], \gamma < 0 \end{cases}$$

$$\lambda(t) \geq 0$$

$N(t)$, $Q(t)$, and $w(t)$ are the theoretical values of the cumulative number of objects, the queue length, and the delay time at time interval t, respectively; $\hat{N}(t)$, $\hat{Q}(t)$, and $\hat{w}(t)$ are the observed values of the cumulative number of objects, the queue length, and the delay time at time interval t, respectively; $N_{\max}(N_{\min})$, $Q_{\max}(Q_{\min})$, and $w_{\max}(w_{\min})$ are the theoretical maximal (or minimal) values of the corresponding measurements, while $\hat{N}_{\max}(\hat{N}_{\min})$, $\hat{Q}_{\max}(\hat{Q}_{\min})$, and $\hat{w}_{\max}(\hat{w}_{\min})$ are the observed maximal (or minimal) values; other parameters follow the definitions aforementioned before.

4. Calibrations: DS1 Los Angeles

Data description

- Data: flow (*q*), occupancy (*occ*), speed (*v*), detector location, number of lanes
- Measurements:

1) Time-dependent physical queue length

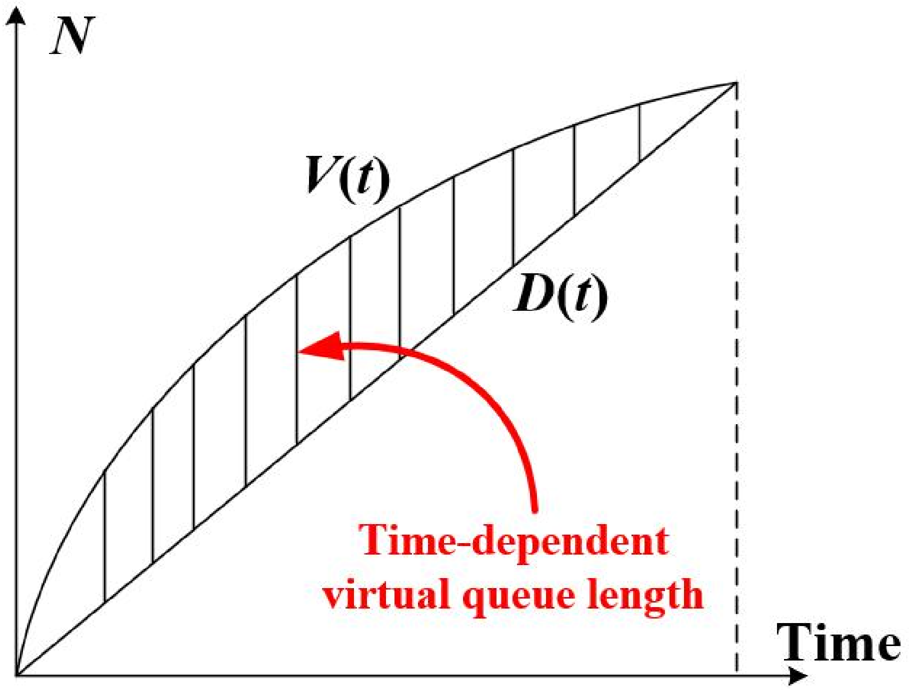
$$Q^p(t) = \frac{Q(t)}{1 - \frac{v_\mu}{v_f}} = Observed\ queue(t) + \varepsilon_{queue}$$

$$Observed\ queue(t) = \sum_{i=1}^{n} [k_{it} - k_c, 0]^+ * A_i$$

2) Time-dependent delay

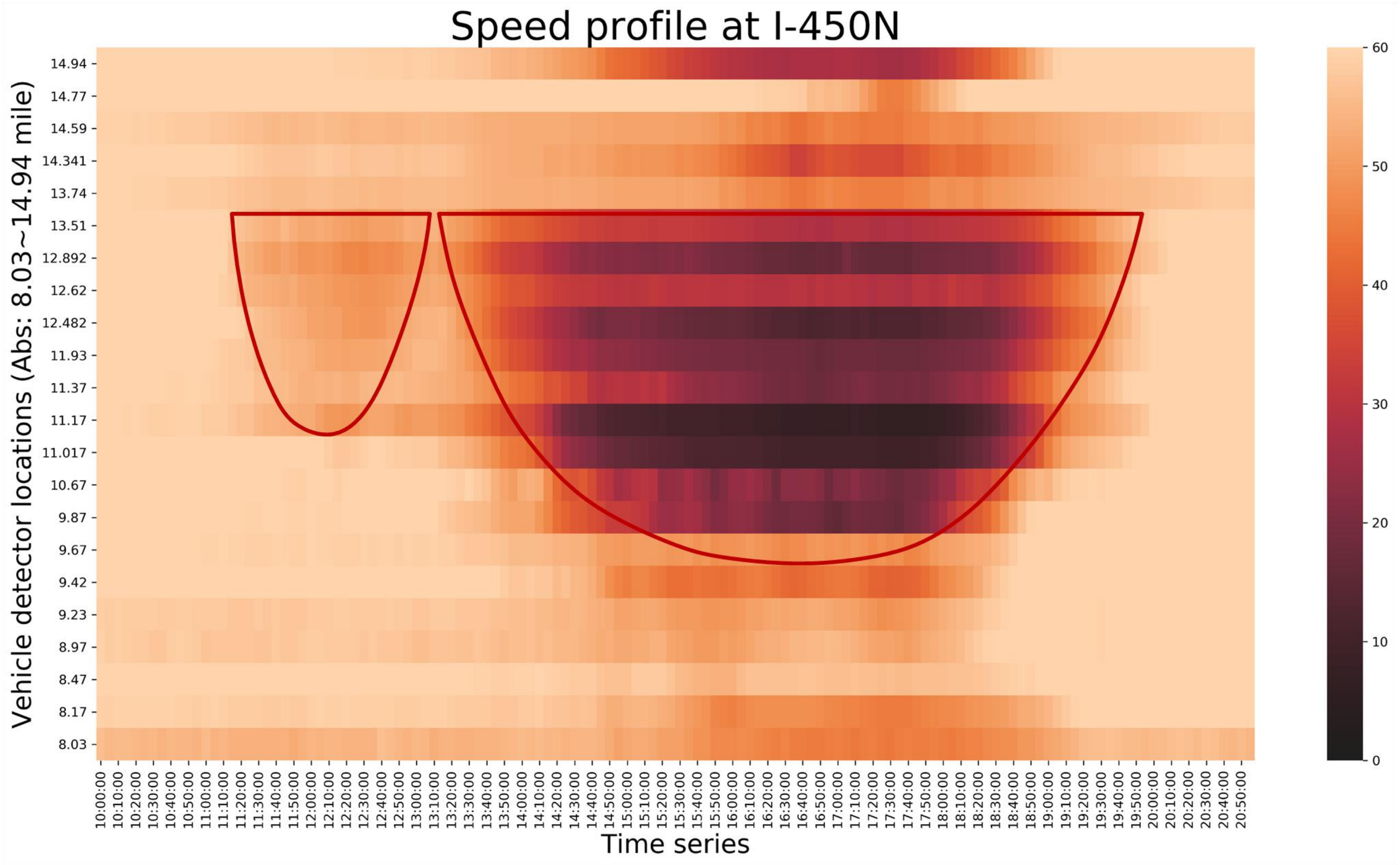
$$Q(t)/\mu = Observed\ delay(t) + \varepsilon_{delay}$$

3) Time-dependent flow rate $\mu = q(\tau) + \varepsilon_{flow}$



4. Calibrations: DS1 Los Angeles

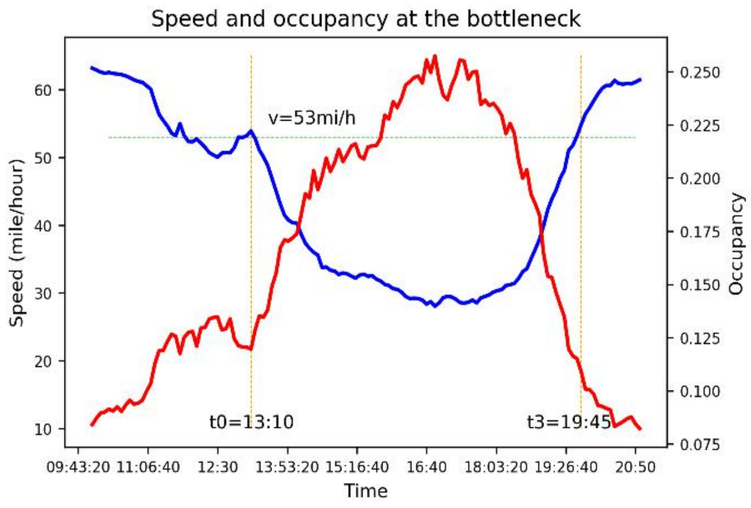
◻ Bottleneck detection with speed contour map

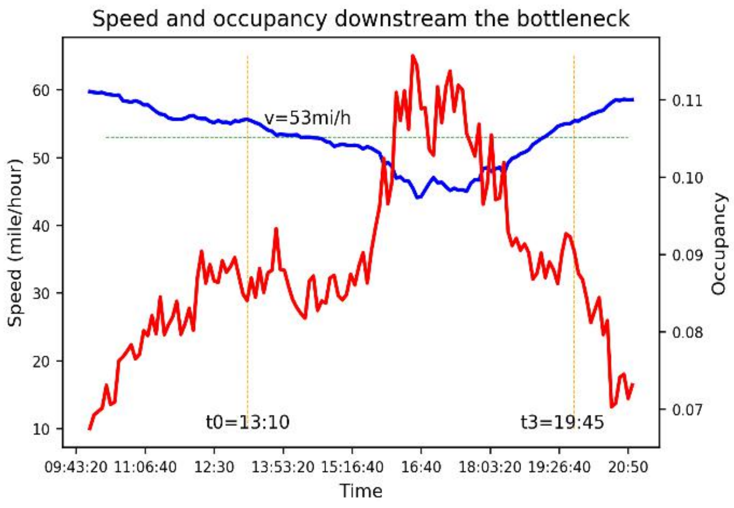


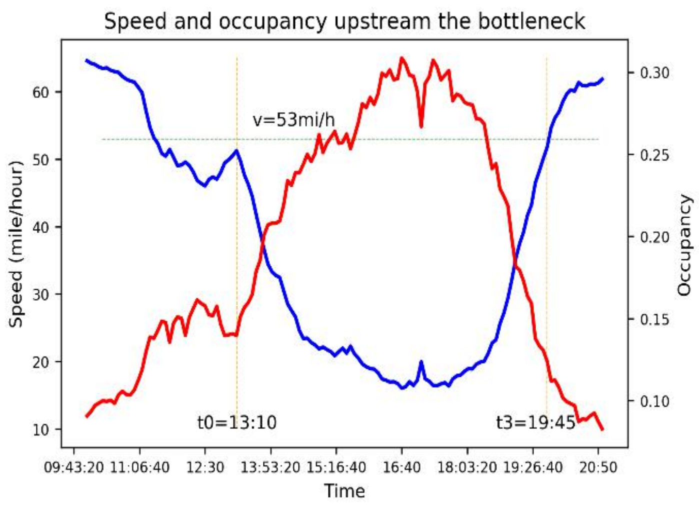
It is clear that the bottleneck is located at Abs=13.51 mile. In this case, we only analyze one single bottleneck with one peak period from t_0 = 13:10 to t_3 = 19:45.

4. Calibrations: DS1 Los Angeles

◻ Speed and occupancy



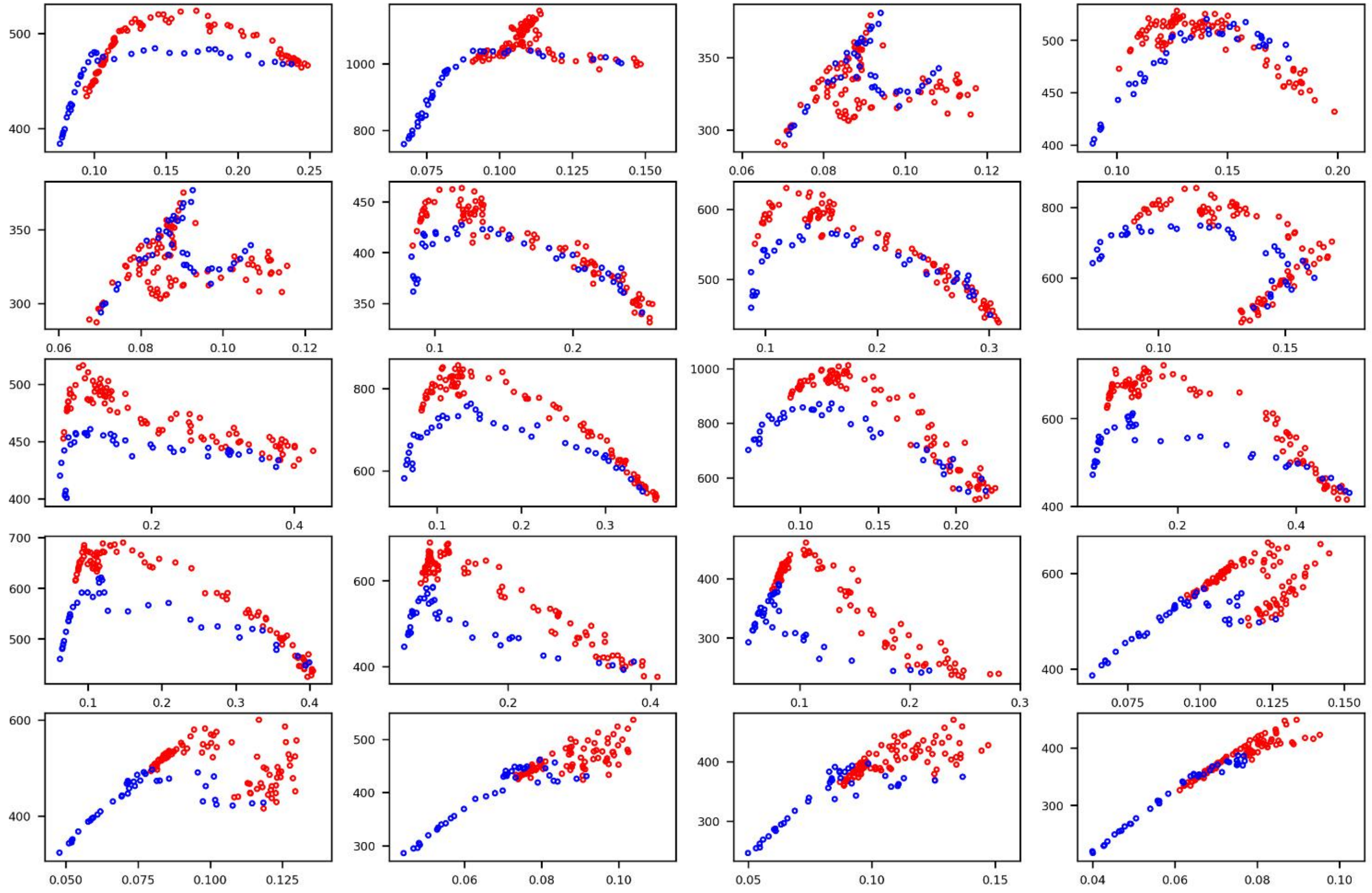




The blue curves depict the **speed**, while the red curves depict the **occupancy**. We can see that the speed downstream the bottleneck almost keeps stable over 45mile/hour during the peak period, while the speed at the bottleneck and upstream the bottleneck is reduced sharply during the peak period. Similarly, the occupancy downstream the bottleneck almost below 0.11 during the peak period, while the occupancy at the bottleneck and upstream the bottleneck reaches up to 0.25 during the peak period.

4. Calibrations: DS1 Los Angeles

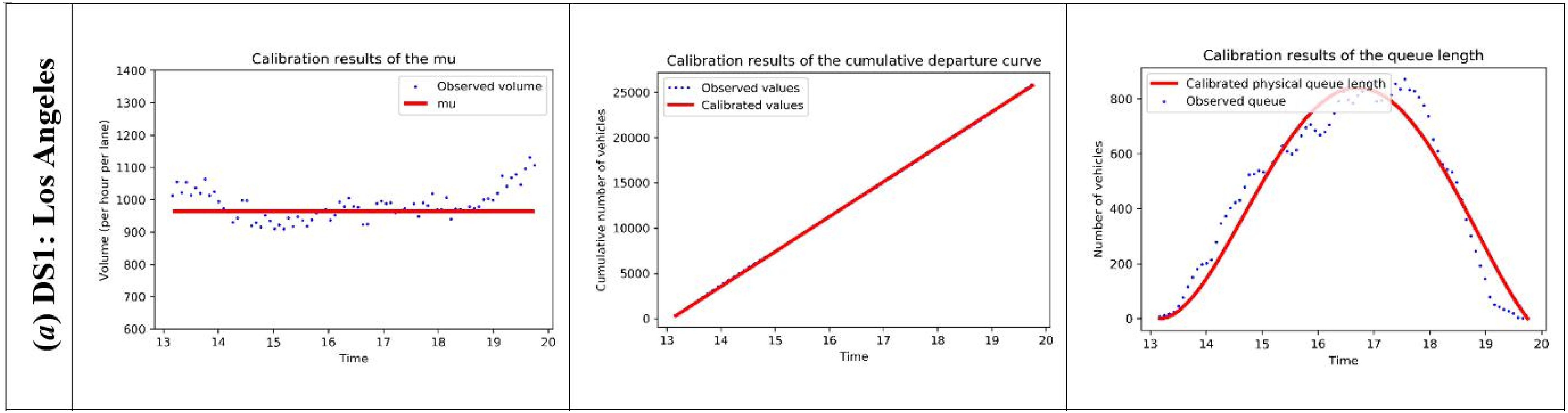
Flow-occupancy plot



The horizontal axis is the **occupancy**, while the vertical axis is the **flow**. The red dots represent the traffic loading process, while the blue dots represent the traffic unloading process. The bottleneck is located on the second row and second column. We can see that the critical occupancy of the queueing system closes to 0.13, which is used to calculated the observed physical queue length for the parameter calibration.

4. Calibrations: DS1 Los Angeles

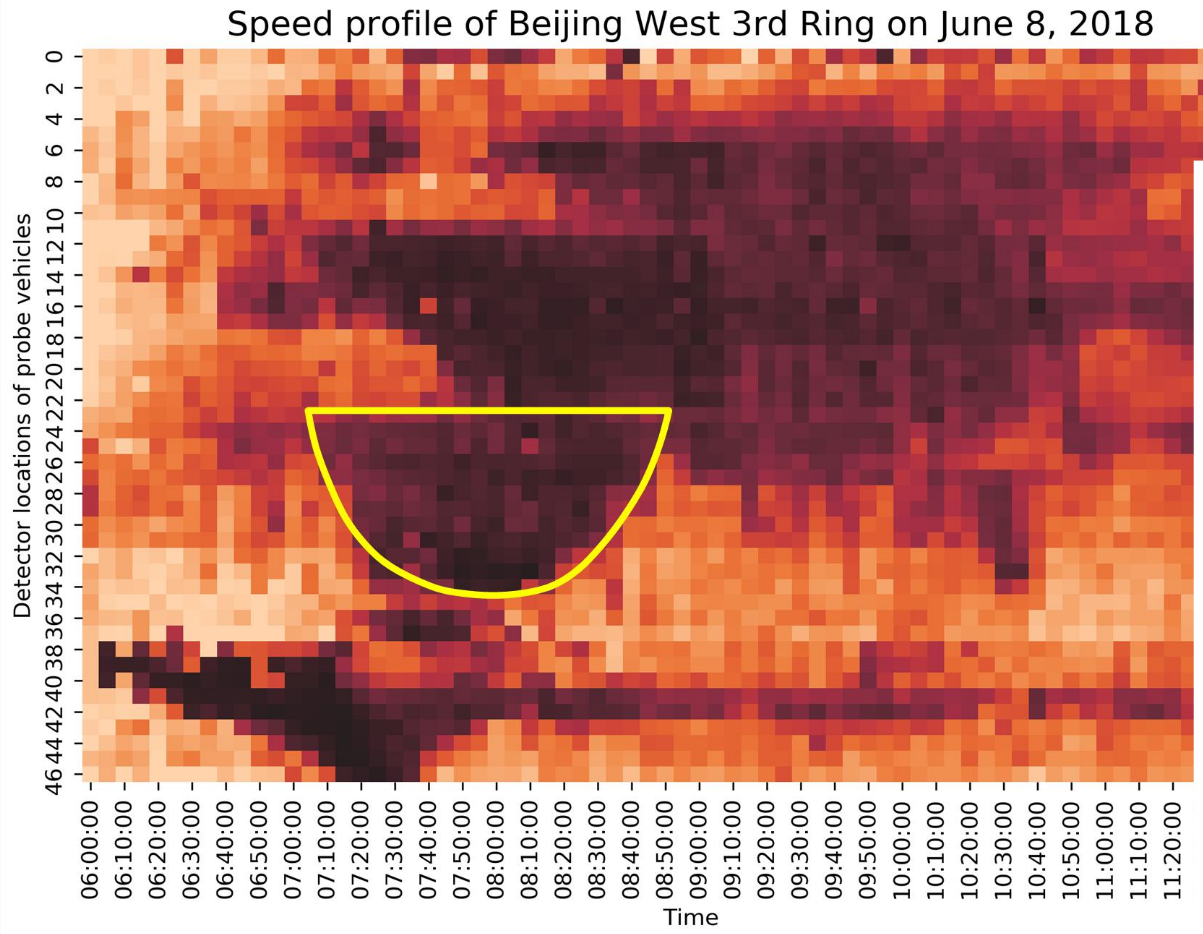
☐ Results

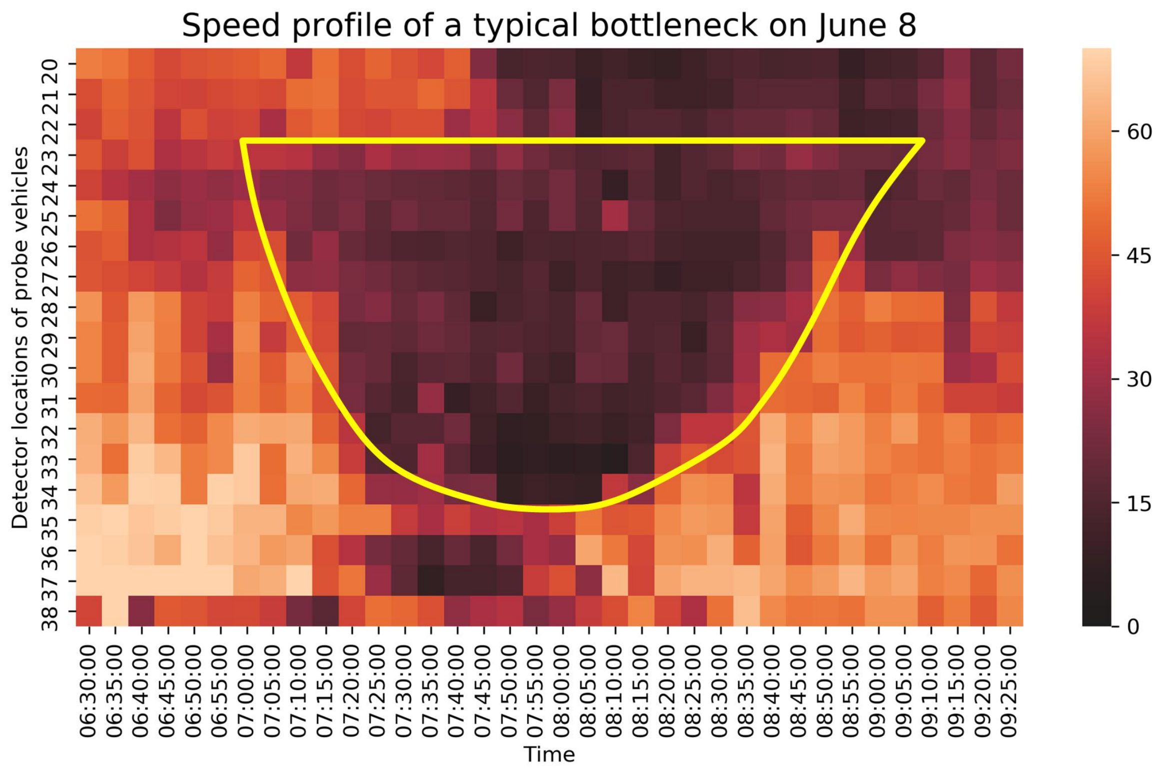


$$\mu_{DS1} = 3860\,veh/hour \quad \text{or} \quad \mu_{DS1} = 965\,veh/hour/lane\,,\quad \gamma_{DS1} = 13veh/hour^4\,,\quad m_{DS1} = 0.537$$

4. Calibrations: DS2 Beijing

- Bottleneck detection with speed contour map

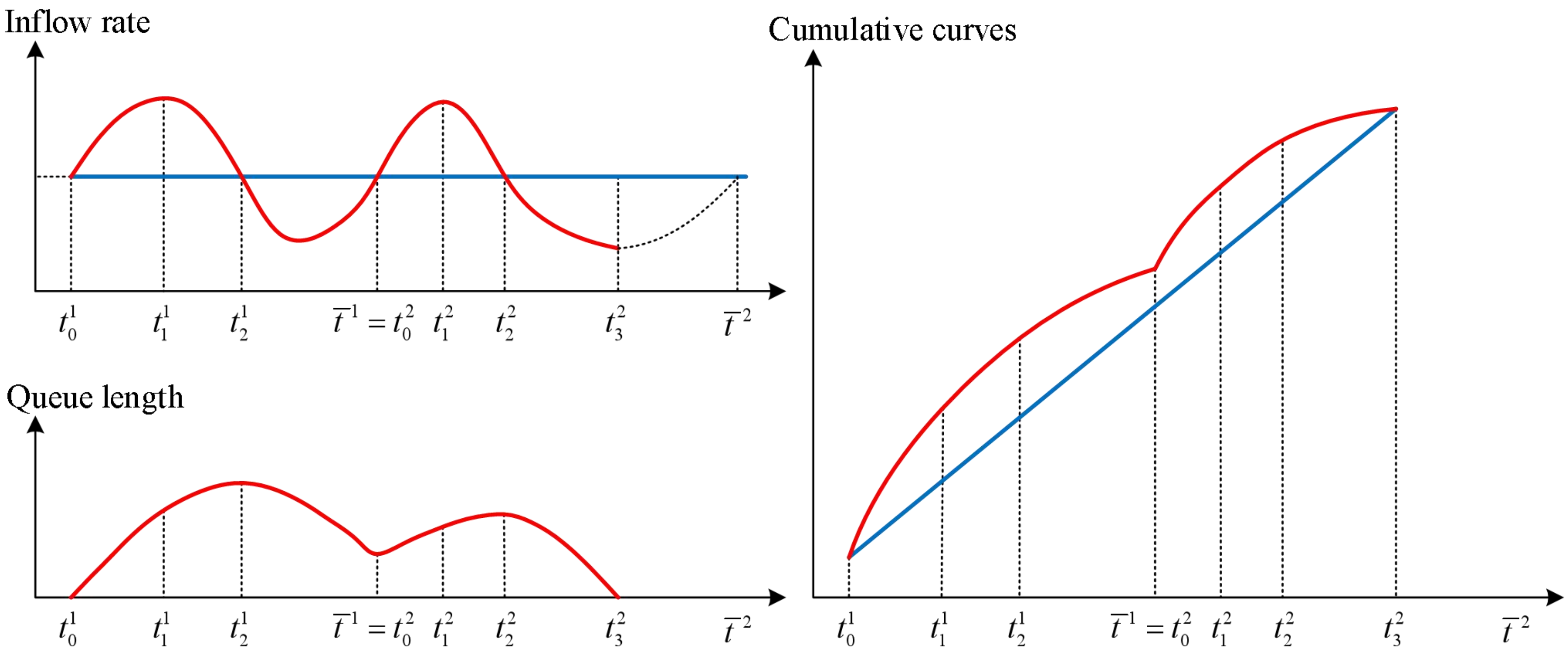




It exists the spatial queue spillback and the temporal queue connection phenomena.

4. Calibrations: DS2 Beijing

◻ Illustration of the two peak period model



It can be used to solve the temporal queue connection problem.

4. Calibrations: DS2 Beijing

Illustration of the two peak period model

$$\lambda(t) = \begin{cases} \gamma_1(t-t_0^1)(t-t_2^1)(t-\bar{t}^1)+\mu_1,\ t\in[t_0^1,\bar{t}^1] \\ \gamma_2(t-t_0^2)(t-t_2^2)(t-\bar{t}^2)+\mu_2,\ t\in[t_0^2,t_3^2] \end{cases}$$

When $\mu_1 = \mu_2 = \mu$

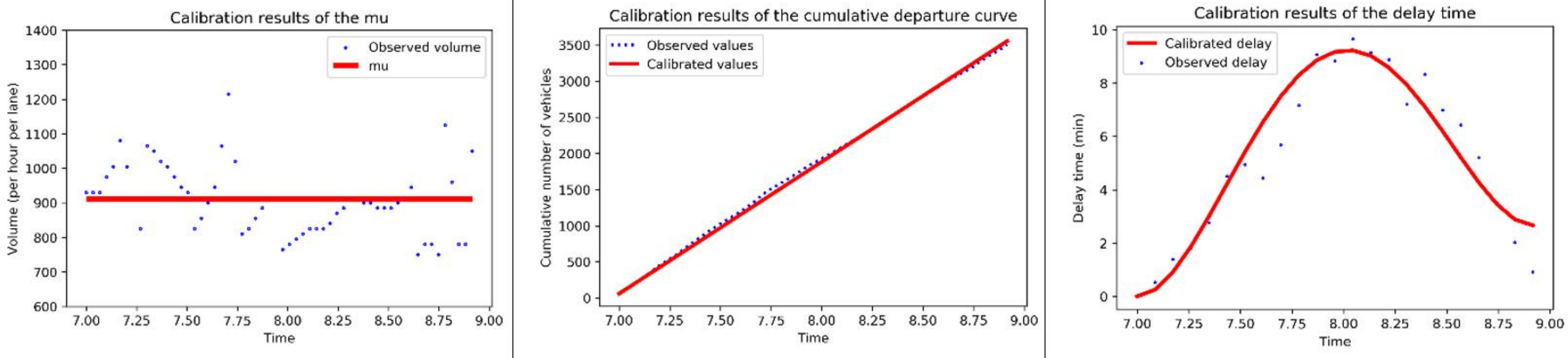
$$\lambda(t)-\mu = \begin{cases} \gamma_1(t-t_0^1)(t-t_2^1)(t-\bar{t}^1),\ t\in[t_0^1,\bar{t}^1] \\ \gamma_2(t-t_0^2)(t-t_2^2)(t-\bar{t}^2),\ t\in[t_0^2,t_3^2] \end{cases}$$

$$Q(t) = \begin{cases} \gamma_1(t-t_0^1)^2\left[\frac{1}{4}(t-t_0^1)^2+\frac{1}{3}(2t_0^1-t_2^1-\bar{t}^1)(t-t_0^1) + \frac{1}{2}(t_0^1-t_2^1)(t_0^1-\bar{t}^1)\right],\ t\in[t_0^1,\bar{t}^1] \\ Q(\bar{t}^1)+\gamma_2(t-t_0^2)^2\left[\frac{1}{4}(t-t_0^2)^2+\frac{1}{3}(2t_0^2-t_2^2-\bar{t}^2)(t-t_0^2) + \frac{1}{2}(t_0^2-t_2^2)(t_0^2-\bar{t}^2)\right],\ t\in[t_0^2,t_3^2] \end{cases}$$

4. Calibrations: DS2 Beijing

☐ Results

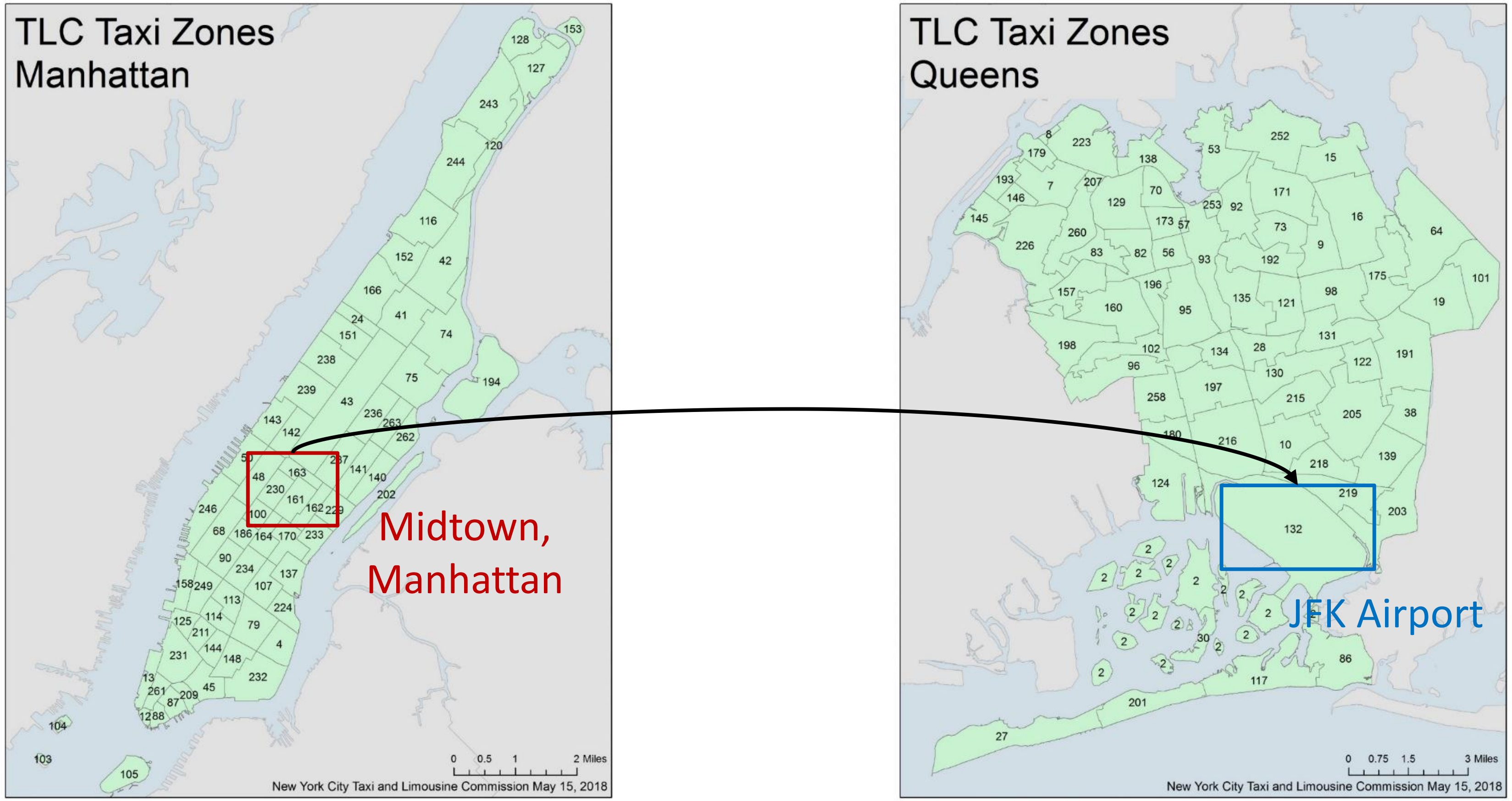
(b) DS1: Beijing



$$\mu_{DS2} = 1824\ veh/hour \quad \text{or} \quad \mu_{DS2} = 912\ veh/hour/lane, \quad \gamma_{DS2} = 1130\ veh/hour^4, \quad m_{DS2} = 0.470$$

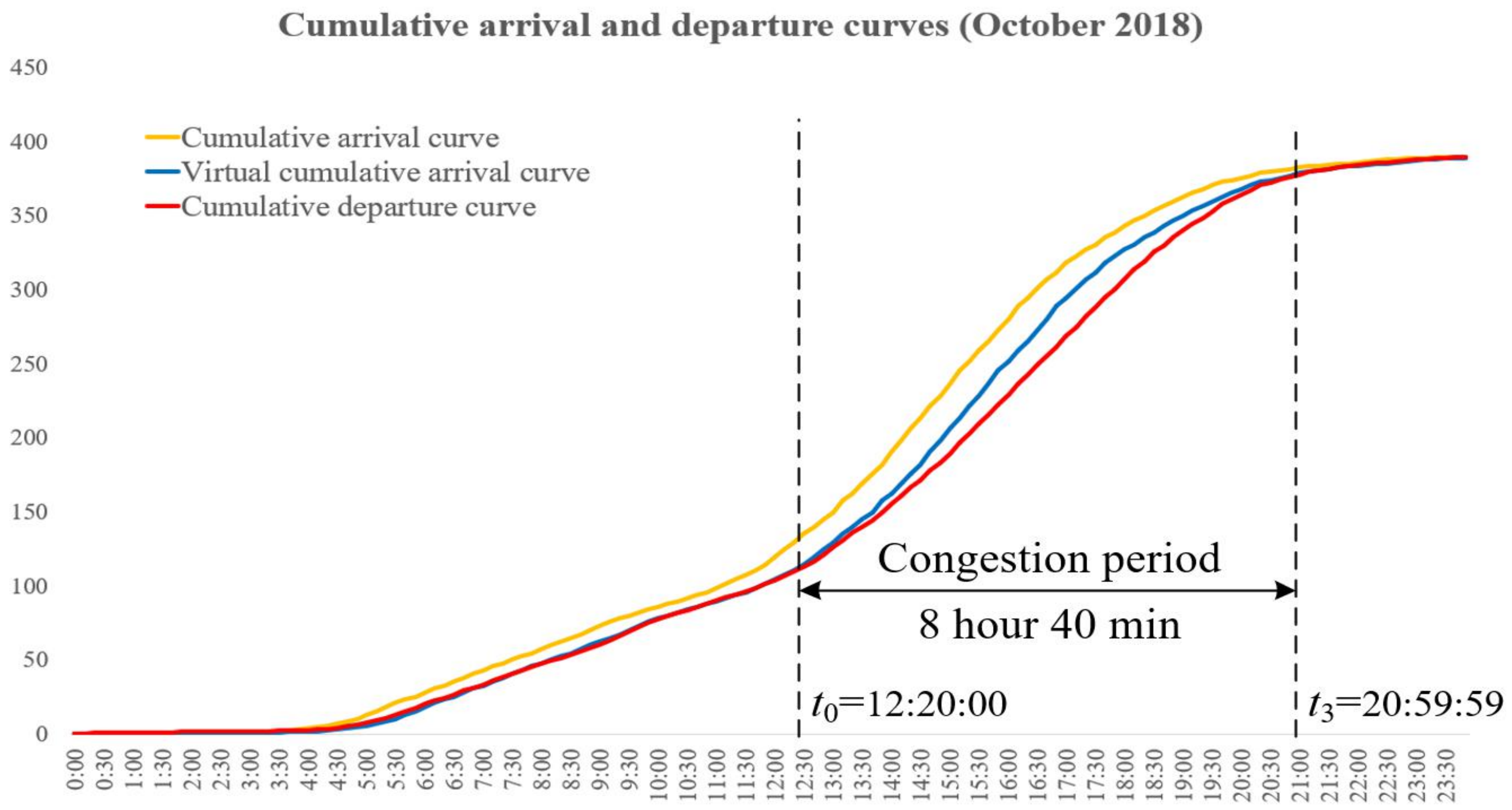
4. Calibrations: DS3 New York

- Bottleneck detection



4. Calibrations: DS3 New York

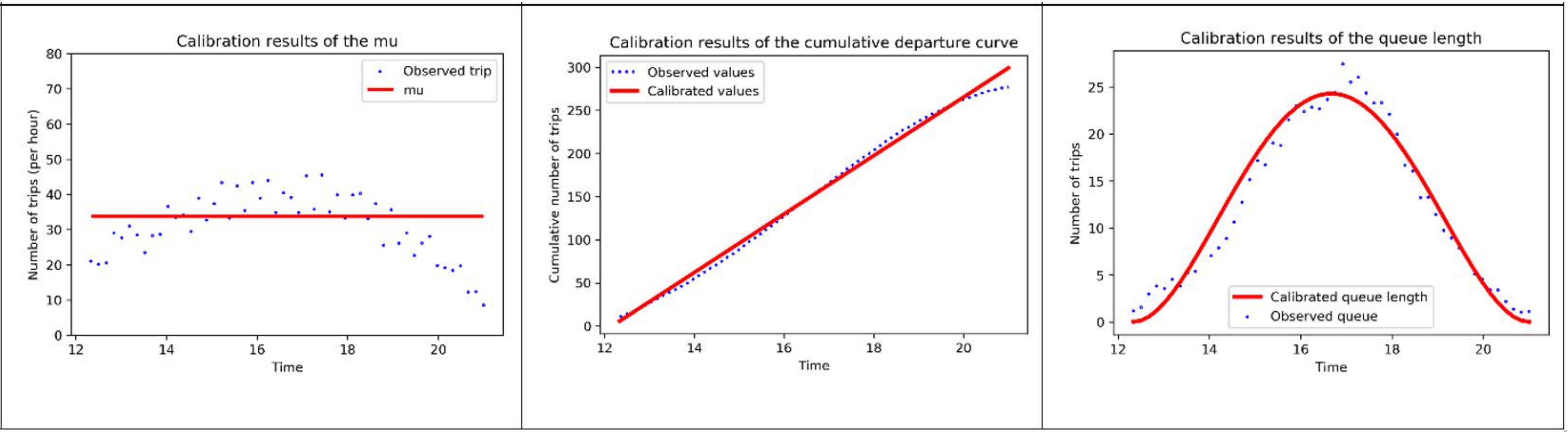
☐ Cumulative arrival and departure curves



4. Calibrations: DS3 New York

◻ Results

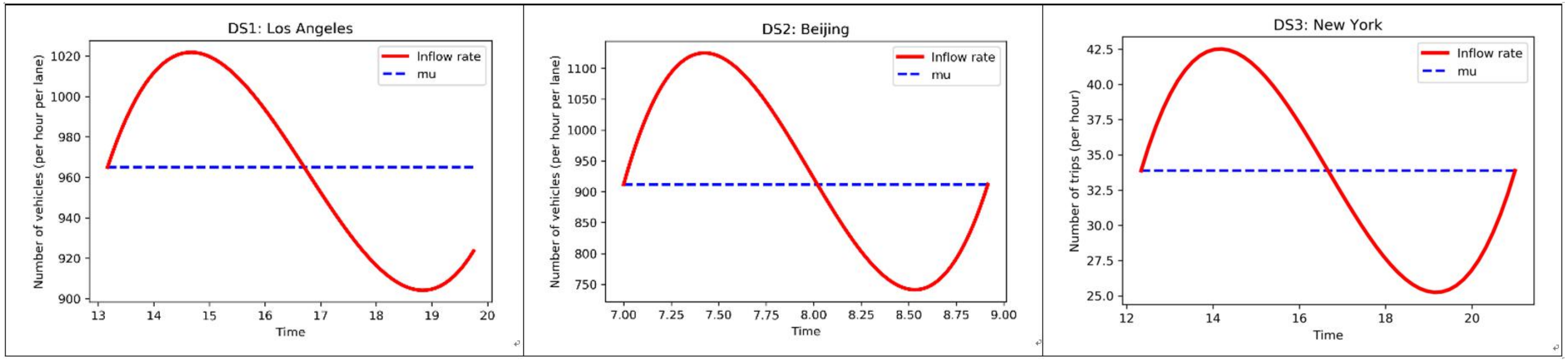
(c) DS3: New York



$$\mu_{DS3} = 34\ trip/hour\ , \quad \gamma_{DS3} = 0.276\ trip/hour^4\ , \quad m_{DS3} = 0.500$$

4. Calibrations: Comparison

☐ Results

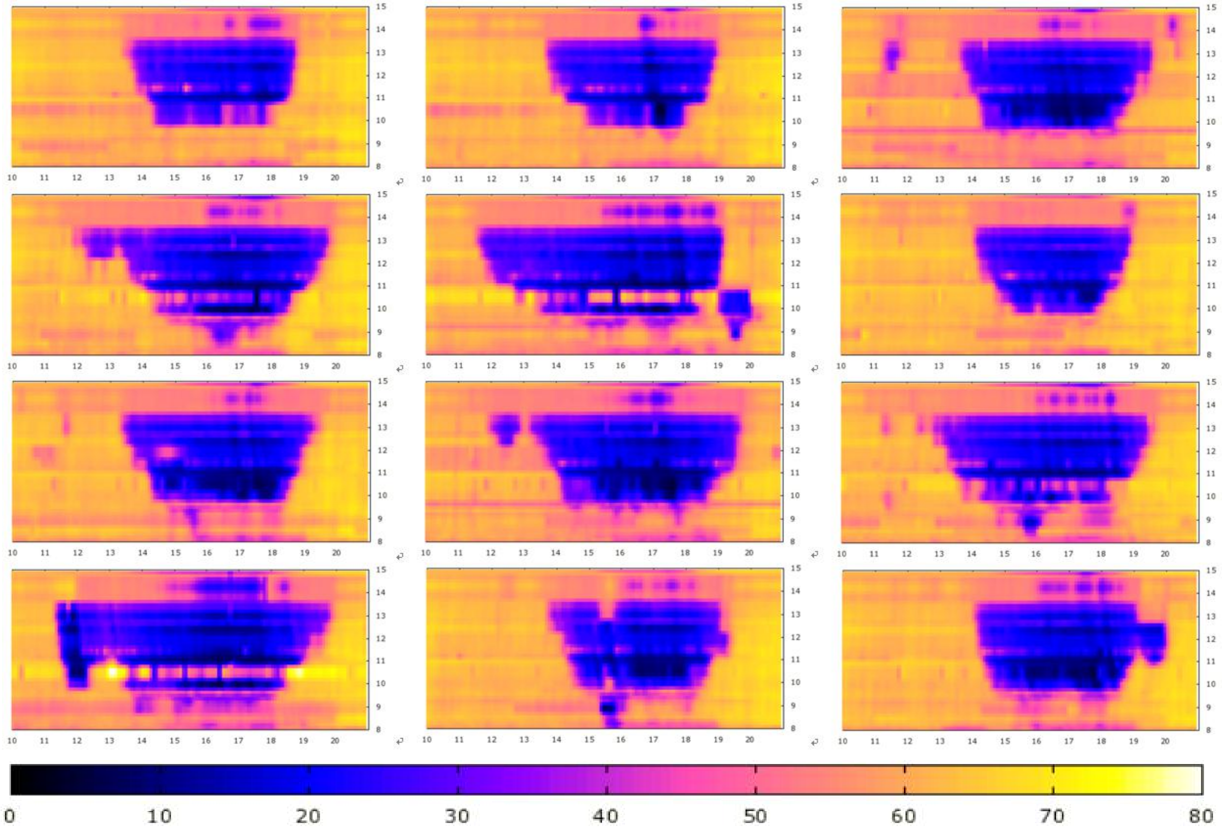


The ratio of λ/μ indicates the variation of the demand over supply.

The results of λ/μ are: $0.937 \le (\lambda/\mu)_{DS1} \le 1.059$, $0.813 \le (\lambda/\mu)_{DS2} \le 1.234$, $0.745 \le (\lambda/\mu)_{DS3} \le 1.255$.

5. Discussions and applications

- How to calibrate it with multiple days' data?

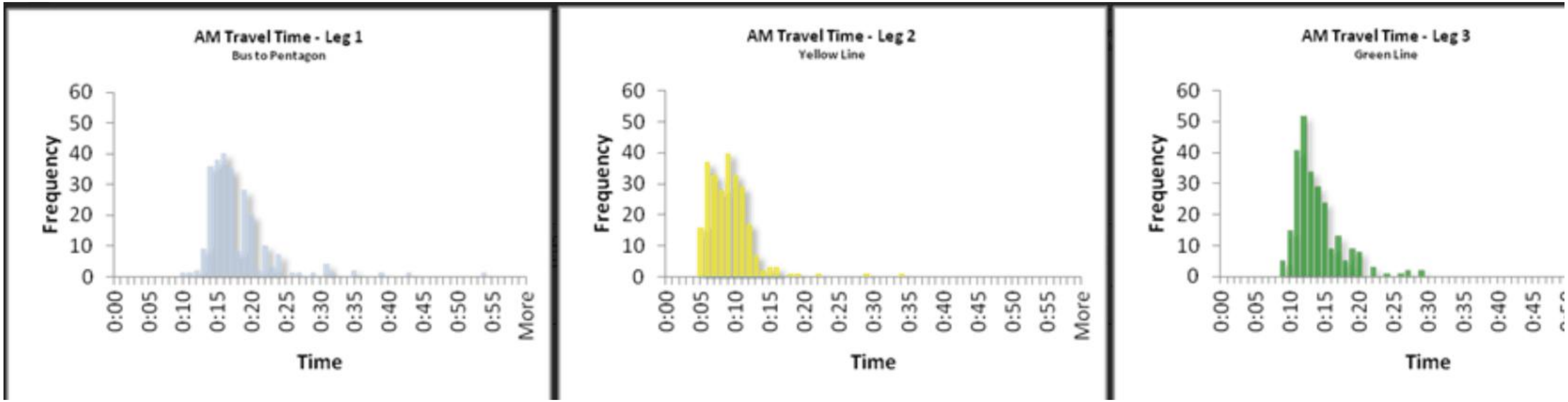


Speed evolution process from 10:00 am to 21:00 pm

Day *N*	t_0	t_3	P
Day 1	15:00	18:00	3 hour
Day 2	15:25	18:35	3.17 hour
Day 3	14:45	18:10	3.42 hour
Day 4	15:40	18:20	2.67 hour
Day 5	15:30	18:30	3 hour
Day 6	15:10	18:00	2.83 hour
…	…	…	…

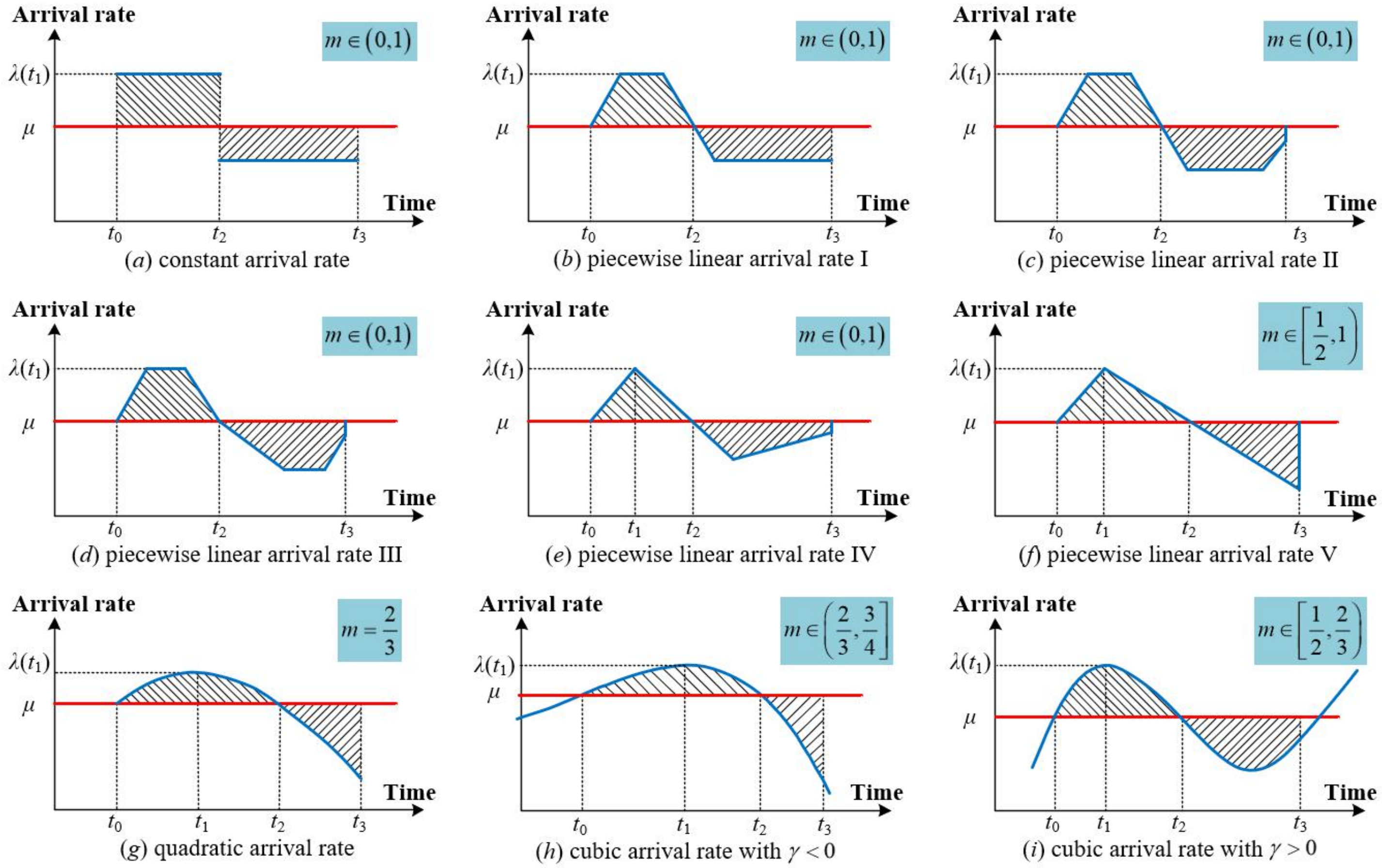
5. Discussions and applications

- What's the relationship to the transportation reliability?



5. Discussions and applications

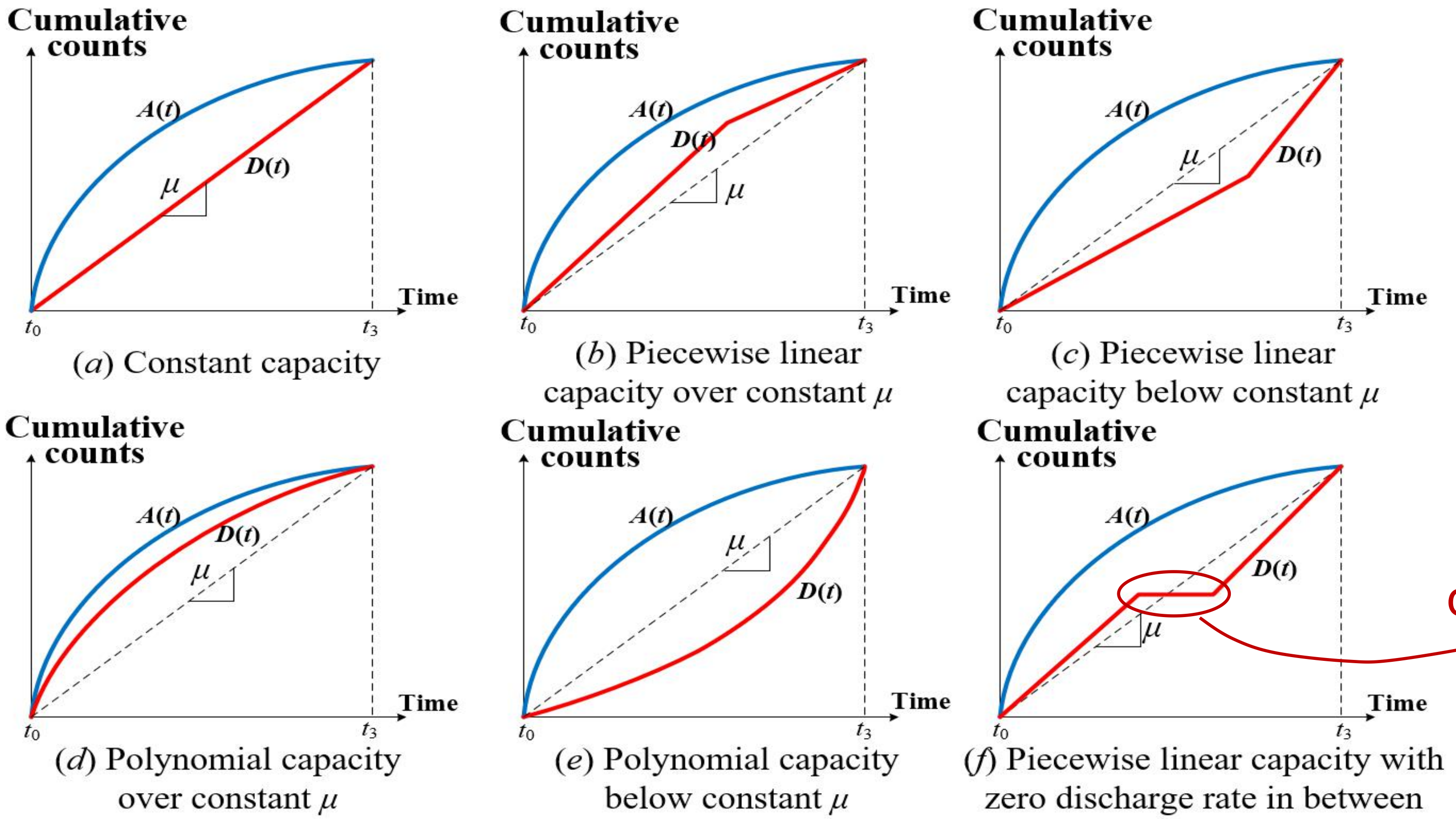
❑ Demand pattern



Different patterns of the arrival rate function in dynamic oversaturated queueing systems. The red horizontal line is the constant discharge rate μ, and the blue line or curve is the arrival rate function $\lambda(t)$ with diverse patterns. The area of the shadow between t_0 and t_2 is the maximal queue length, thus the shadow area before and after t_2 should be equal, indicating a flow conservation condition.

5. Discussions and applications

Supply pattern



Different patterns of the cumulative departure curve in dynamic oversaturated queueing systems. The red line or curve $D(t)$ is the cumulative departure curve, the blue curve $A(t)$ is the cumulative arrival curve, and the dashed line is an imaginary cumulative departure curve with a slope of μ .

due to incident

5. Discussions and applications

◻ Activity profile and queue profile

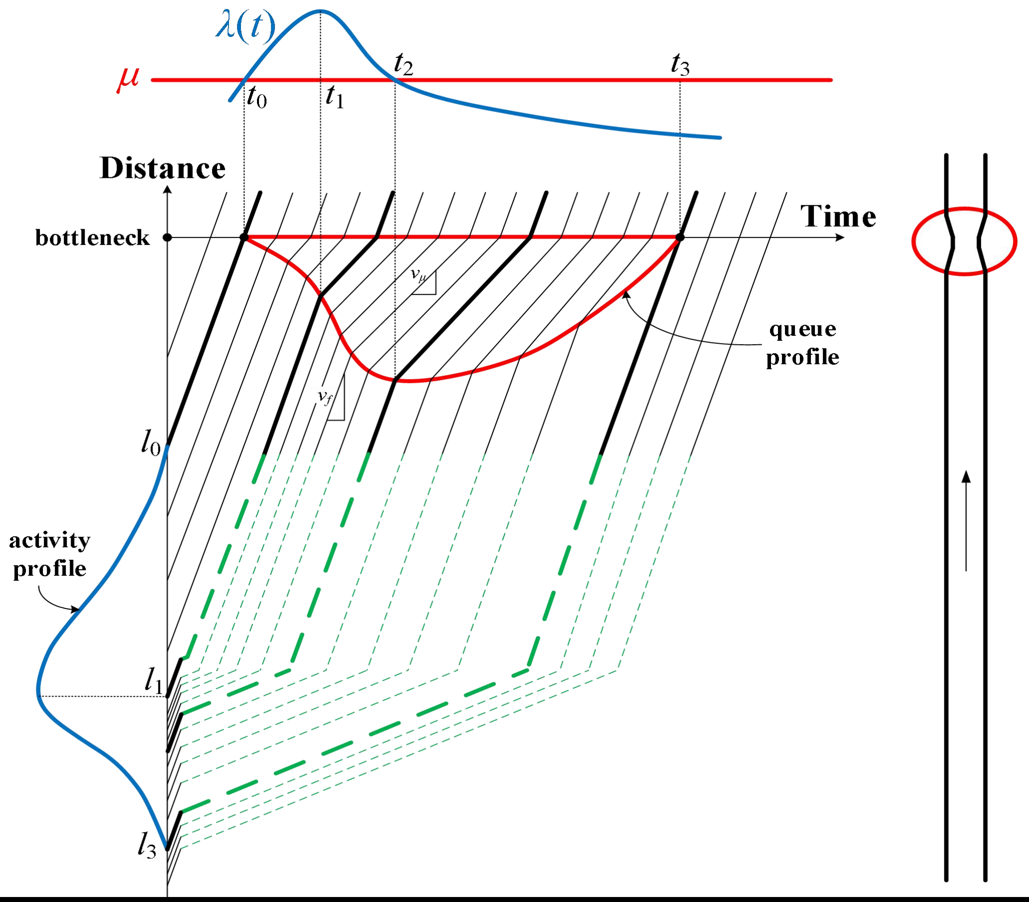
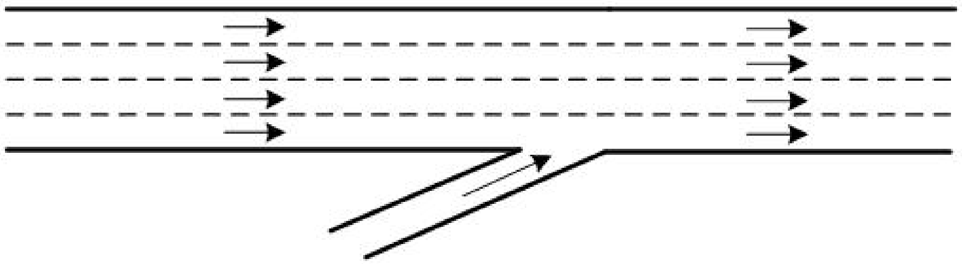


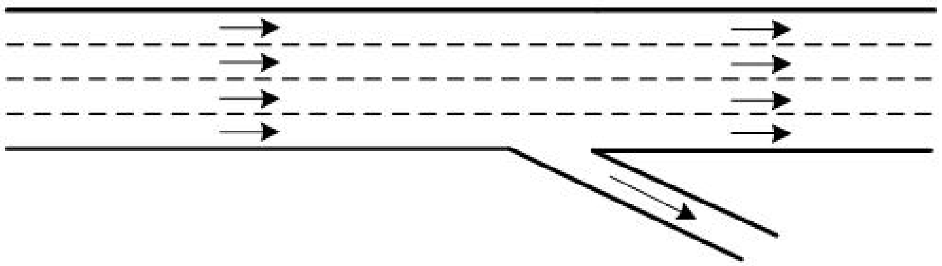
Illustration of the activity profile and queue profile in an oversaturated queueing system. Vehicle trajectories shown with blue colors are illustrated by the broken lines due to space limitation; however, they should be straight lines before they encounter with the back of the queue. It reveals that traffic demand of $\lambda(t)$ and the queue are the consequences of the travel activity upstream the bottleneck.

5. Discussions and applications

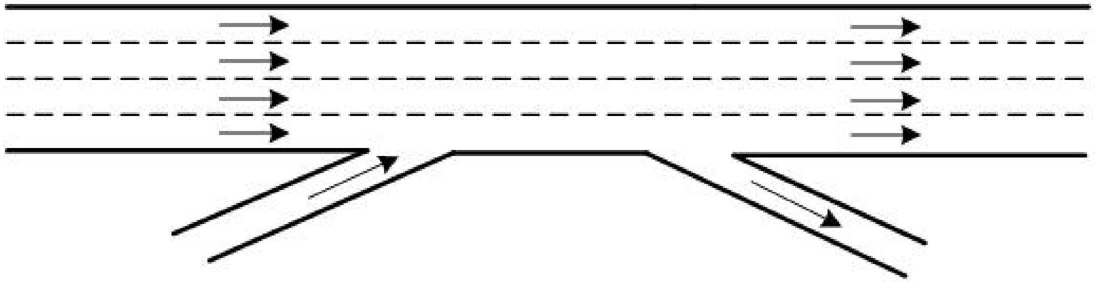
- Merge and diverge, for ramp controlling



(a) Freeway merge bottleneck



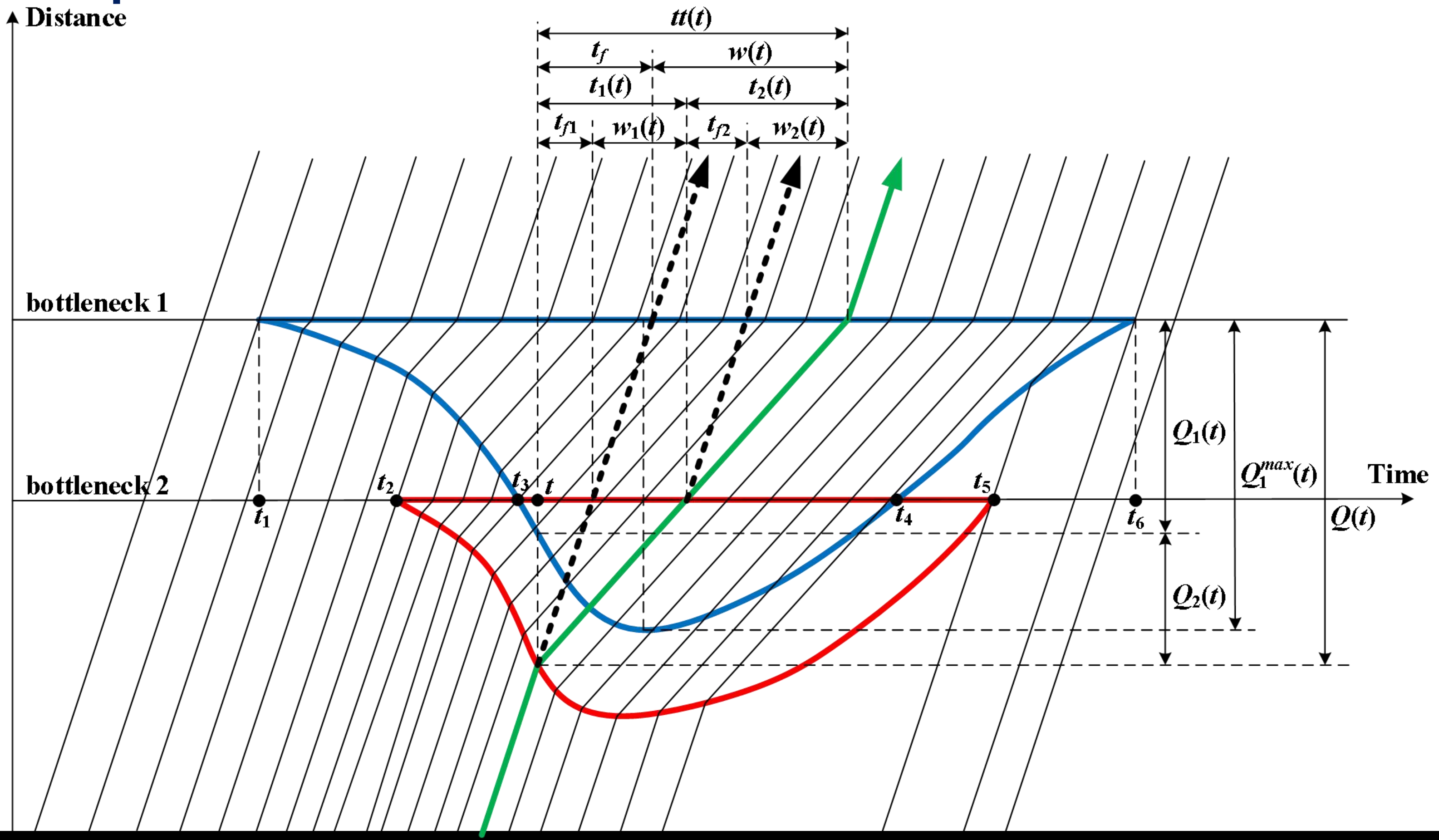
(b) Freeway diverge bottleneck



(c) Freeway merge and diverge bottlenecks

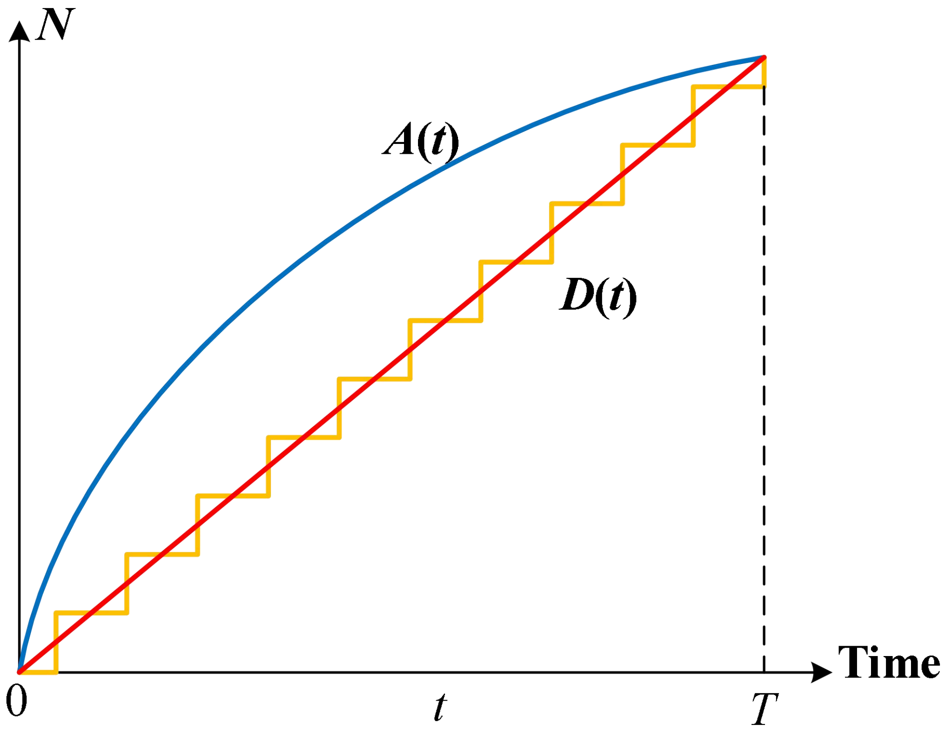
5. Discussions and applications

☐ Queue spillback



5. Discussions and applications

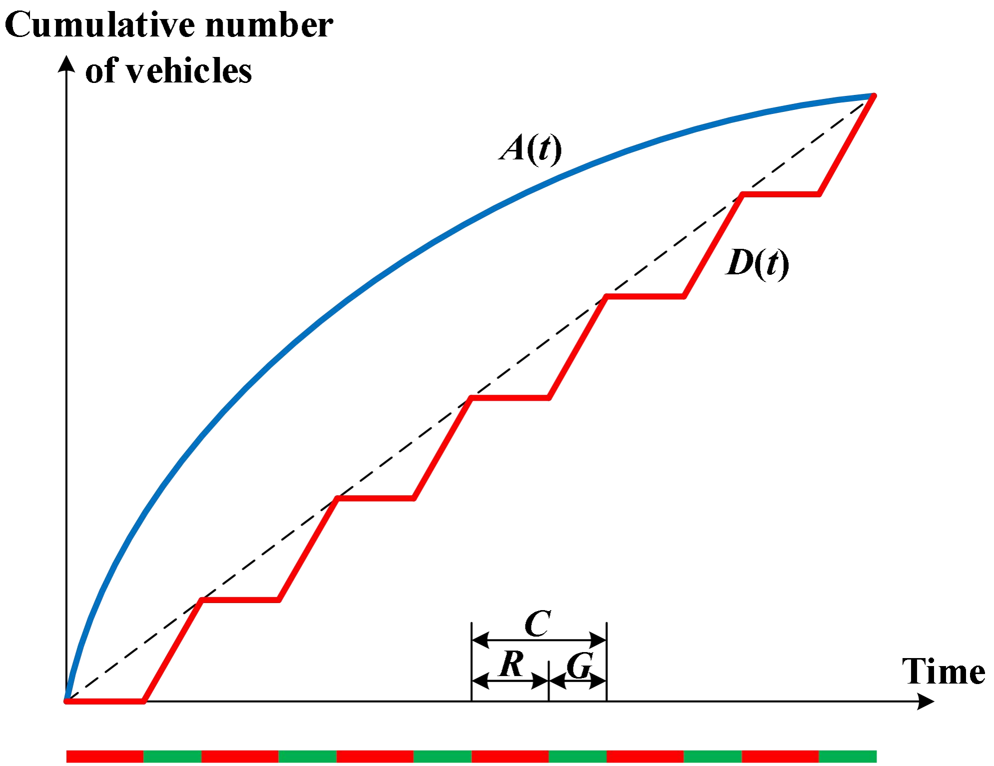
☐ Bus scheduling problem

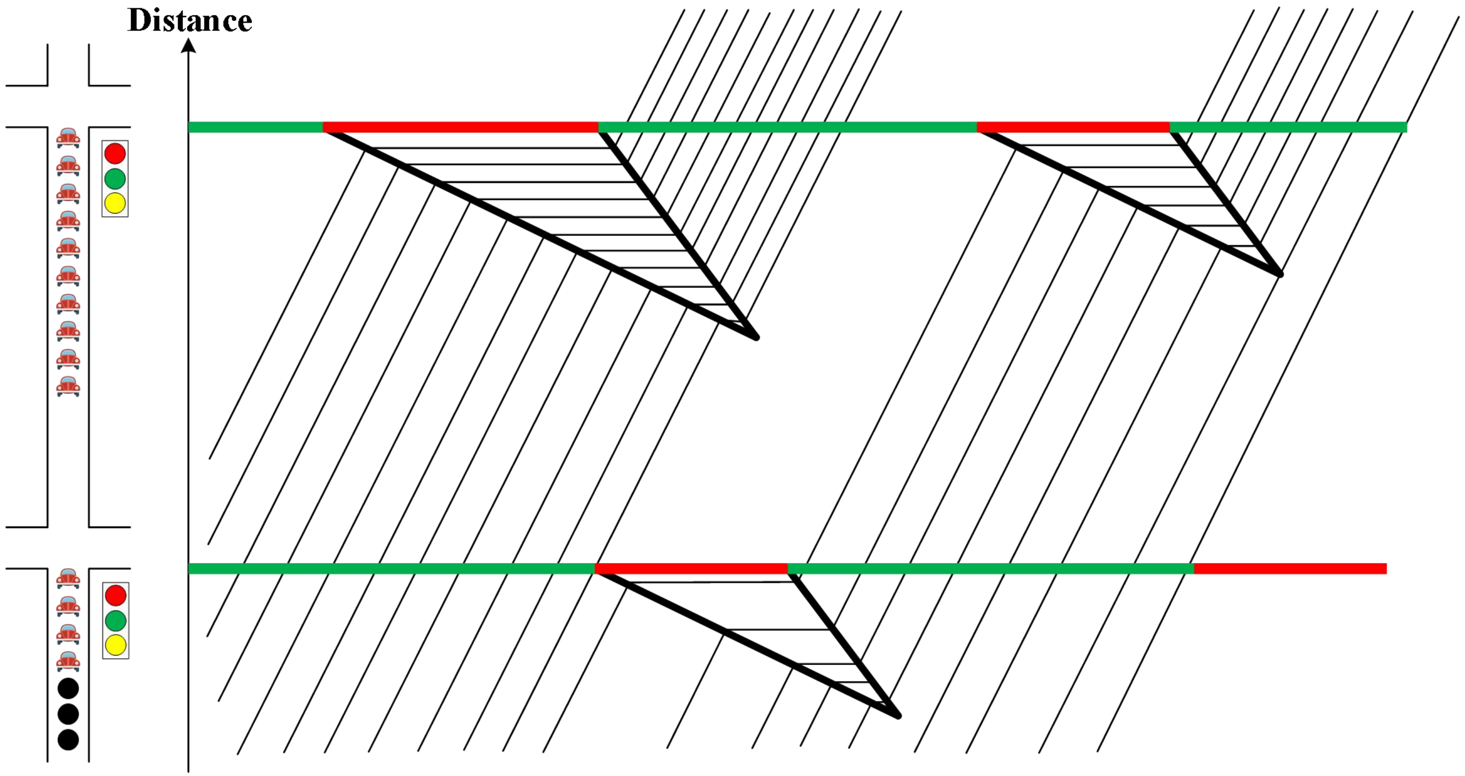


Blue curve: cumulative arrival curve
Orange step curve: physical cumulative departure curve
Red line: virtual cumulative departure curve

5. Discussions and applications

☐ Dynamic signal control under overcongestion





5. Discussions and applications

Open questions

- What are the real demand and inflow rate, and how to measure them?
- What are the stochastics and dynamics in the traffic system?
- How to implement it from a single bottleneck to connected bottlenecks and the network-wide system?