Assignment Topic: Queueing-based Volume-Delay Function (QVDF) to capture flow dynamics and queue evolution process

# Problems:

Before addressing the problems, please read the “reading required” section at the end of this document.

Problem 1: In this problem you have to address the definitions and key assumptions in developing the QVDF.

1. Derive the Eq (10) (hint: use the time when queue would dissipate as a boundary point in Eq 6)

Answer: The queue will dissipate at time , i.e., , then we can obtain :

Using this and the fact that in Eq 6 we can easily derive Eq 10

1. Derive Eq 11 (hint: refers to the definition of delay based on cumulative arrival and departure and use Fig. 1(c) to calculate the total delay)

Answer: The total delay between the time and can also be calculated by the area between and in Fig. 1(c) through integrating Eq. (10).

1. Explain the difference between “cut-off-speed” and “free flow speed”.

Answer: the cut-off speed can be used more systematically to distinguish “congested” vs. “uncongested” states of traffic bottleneck. However, the free flow speed (also known as desired speed) is the speed of a vehicle in a free flow condition of a road segment which is higher than cut-off-speed

1. Consider the following observations for a specific bottleneck in PM period (3 hours). Calculate the Volume-to-capacity, inflow demand-to-capacity ratios, and percentage of congested flows within the entire analysis period (data units are mph and veh/hour).

Free flow speed: 60 (assume cut-off-speed is 0.7 of Free speed), Lane Capacity = 1850, lanes = 4,

Observed speed = 27.5, observed discharge rate = 1300, observed congestion duration = 2.5

1. To derive Eq(25b), we use the fact that the average discharge rate should be less than the capacity, i.e., . What additional constraint for congestion duration must be satisfied based on the aforementioned fact? (hint: use the relation between average discharge rate and congestion duration to derive the constraint).

Answer: To ensure the capacity constraint, the following condition should be satisfied: It implies that the D/C ratio is the lower bound of congestion duration .

1. Using Eq(25b), derive the relation between average discharge rate and D/C ratio. Explain what additional condition should be satisfied.

Answer:

1. Show that (hint: use Eqs 28b and 32)

Answer:

1. Derive an equation for time dependent speed (hint: use Eq(33) and time-dependent delay)

Answer:

1. Derive an explicit form for relation between the lowest speed () and D/C ratio. (hint: use Eq(27))

Answer:

Consider we have given the following two data to analyze both traffic states and queue evolution process. Please answer the following question?

Assume that we already calibrated the S3 shape FD with the given data and the results is as follows:

Graphical user interface, chart, scatter chart

Description automatically generated

The data is provided in excel files. For example, a part of the data are given as follows:

A screenshot of a computer

Description automatically generated with low confidence

# Reading Required:

Introduction and Benefits of QVDF

The macroscopic volume-delay function (VDF) which has been used widely in static traffic assignment for transportation planning has deficiencies in capturing traffic flow dynamics and queue evolution process.

The most essential functions used in static traffic assignment (STA) are the volume-delay functions (VDF). Among a variety of VDFs, the Bureau of Public Roads (BPR) function, created by the US Bureau of Public Roads in 1964, plays an important role in system-wide performance evaluation. The BPR function is a simple polynomial equation to relate the relationship between demand and delay. Although the macroscopic volume-delay function (VDF) has been widely used in static traffic assignment for transportation planning, the planning community has long recognized its deficiencies as a static function in capturing traffic flow dynamics and queue evolution process.

On the other hand, Dynamic Traffic Assignment (DTA) models calibration can be a challenging task in its own right, especially for real-world congested networks with complex traffic dynamics. By extending the fluid-based polynomial arrival queue (PAQ) model with quadratic inflow rates proposed by Newell (1982) and cubic inflow rates by Cheng et al. (2022), Zhou et al (2022) proposed a cross-resolution Queueing-based Volume-Delay Function (QVDF) to explicitly establish a coherent connection between (a) the macroscopic average travel delay performance function in a long-term planning horizon and (b) the mesoscopic dynamic queuing model during a single oversaturated period. By introducing two types of elasticity functional forms, they developed a relationship from the macroscopic inflow demand-to-capacity (D/C) ratio to the congestion duration of a bottleneck, from the congestion duration to the magnitude of speed reduction. The QVDF can be directly utilized to provide closed-form expressions for both average travel delay performance and the time-dependent speed profiles.

It should be remarked that, if a simulation-based DTA tool is used at the mesoscopic link level, we need to clearly understand the complexity in modeling the time-dependent queue discharge rates and inflow patterns so as to reliably reproduce the queue evolution process.

The ultimate hourly capacity from the macroscopic model, if adopted as the default queue discharge rate, could be a significant overestimate of its true value. If the affected queue discharge rate is modeled internally through the spatial queue propagation, then major efforts are still needed to (1) calibrate the queue discharge rate at the downstream location of each bottleneck, and (2) obtain precise time-dependent inflow patterns at different incoming links upstream of a bottleneck. The latter is in turn mainly determined by the complex time-varying OD matrix and route choice behaviors.

The ultimate goal in developing a QVDF model can be summarized as follows:

1. Develop a simplified (cross-resolution) link performance function, which not only can be used in evaluating macro-level traffic assignment tasks but also has the capability of reflecting meso-level traffic characteristics under typical arrival flow patterns and congestion levels.
2. Efficiently derive meso-level (time-dependent) speed and queue length profiles which are consistent with the macroscopic link volume and average link performance from STA

## Recall from quadratic inflow rates proposed by Newell (1982):

Chart, line chart

Description automatically generated

Newell assumed that the inflow rate at time could be approximated by the quadratic Taylor expansion (for notation refer to Table 1):

Given , let , then Eq. (1) can be transformed to:

Since describes the curvature or shape of the time-dependent inflow arrival rates, we term it as the inflow curvature parameter. By assuming a constant discharge rate within a single queue duration, the queue discharge rate or service rate ( where ):

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

Now we can write in a factored form:

The virtual queue length at time which equals and can be obtained:

The maximum queue length achieved at time is:

Eq. (6) can also be written as follows:

The total delay between the time and can also be calculated as follows:

Above is the introduction of Newell’s method based on the assumption of the quadratic inflow rate. With the total delay in Eq. (11), we can further derive the average delay during the congestion period from to :

where is the *total* *inflow demand volume* from to . Denote the peak period as (i.e., ), then the discharge rate (or effective capacity) can be represented by and :

Substituting Eq. (13) into Eq. (12) leads to the following average delay between and :

The quadratic model is only applicable to analyze mild traffic conditions. When the system is oversaturated, the arrival rate from the quadratic inflow model might be a (counterintuitive) negative value. Recently, Cheng et al. (2022) revisited Newell’s model and described queueing systems with cubic time-dependent arrival rates. This cubic model is ideal for efficient dynamic modeling and management as it can analytically calculate the time-dependent queue length, delay, and travel time.

Table 2 summarizes the functional forms and key parameters of travel delay functions under different arrival rate patterns. **Fig. 2** illustrates the flow rates, queue length, and cumulative flow rate evolution process over time in the cubic model (Cheng et al., 2022). Given an arrival rate and constant discharge rate, the time-dependent queue length and the time-dependent delay can be determined in the same as illustrated in **Fig. 2** for cubic arrival rate.

**Table 2**: Average travel delay functions for bottlenecks based on constant discharge rate (Cheng et al., 2022).

|  |  |  |
| --- | --- | --- |
| Forms | Arrival rate functions | Average travel delay functions for bottlenecks |
| Constant form |  |  |
| Linear form |  |  |
| Quadratic form |  |  |
| Cubic form |  |  |

Chart, line chart

Description automatically generated

## Basic concepts of Queueing-based Volume-Delay Function (QVDF):

The critical point to develop a QVDF is to connect two important variables: inflow demand-to-capacity (D/C) ratio and inflow curvature parameter in PAQ models, via a set of intermediate variables (e.g., discharge rate and congestion duration).

Understanding this critical point requires key insights to distinguish “volume” and “demand” and to understand “Congestion duration”, “cut-off-speed”, and “average discharge rate”.

Volume-to-capacity (V/C) ratio vs. inflow demand-to-capacity (D/C) ratio:

1. Period volume () is the total lane-based volume loaded on a road link during an analysis period.
2. Inflow demand () is the queued volume or queued demand which represnts the total volume with travel speed under a specific cut-off speed ().

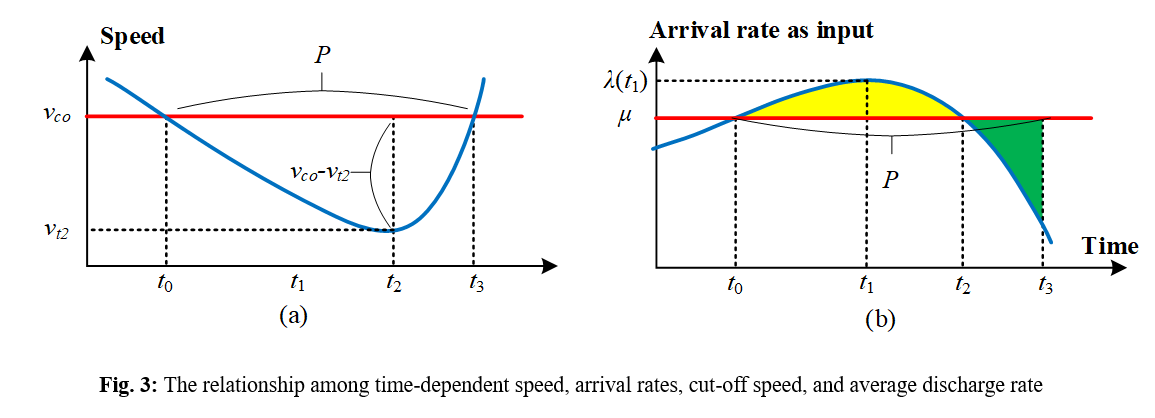
The time period when speed is slower than is then defined as “congestion duration”. For instance, if , the inflow demand *D* is the total volume within the “congestion duration” with a speed lower than 45 miles/hour. We introduce a queued demand factor (QDF) to covert to , which represents the percentage of congested flows within the entire analysis period:

where and .

The hourly maximum flow rate per lane when the level of service is under E (Branston, 1976; HCM, 2010) is used in this study as the “ultimate hourly capacity” in D/C.

Congestion duration, cut-off-speed, and average discharge rate:

One critical concept shared by both the D/C ratio and PAQ models is “congestion duration”. In PAQ models, congestion duration indicates the peak period from to (i.e., ), where arrival rate is higher than the average discharge rate . According to the definition, should be lower than the ultimate capacity *C*. Noted that congestion duration is dependent on two correlated factors: the time-dependent arrival rate and the average discharge rate. As shown in Fig.3, and D/C can be connected , i.e., where is assumed to be an exogenous parameter in our model.



## Linking Polynomial Arrival Queue (PAQ) with Macroscopic Volume-Delay Function

This section elaborates step 1 through step 4 derivations as illustrated in Fig 4. Zhou et al. (2022) introduced two elasticity factors and the cubic PAQ model to systematically link the temporal queue evolution with the traffic performance measures in a long-term planning process.

A picture containing graphical user interface

Description automatically generated

### **Step 1: Link congestion duration and ratio**

In general, the elasticity term of a function shows the relative percentage change of a dependent variable due to a relative percentage change from an independent variable (e.g., demand as a function of ticket price). We assume a power function between congestion duration and traffic intensity in terms of the D/C ratio.

where is the oversaturation-to-duration elasticity and is a congestion duration constant in response to D/C changes. The formula implicitly assumes the relationship between ultimate capacity and average discharge rate . When , represents the baseline value of congestion duration. When , and .

Furthermore, the average discharge rate should be less than the capacity, i.e., . which leads to Eq. (25b):

### **Step 2: Linking magnitude of speed reduction and congestion duration**

As shown in Fig. 3 (a), we attempt to connect and the speed reduction in the PAQ model (i.e., ). Therefore, we first define the magnitude of speed reduction (MSR). We further establish a relationship between congestion duration and . Note that and in Eq. (27) are directly observable.

Where is the duration-to-speed reduction elasticity factor and is an MSR reduction constant in response to changes of congestion duration. The boundary condition is satisfied for as and . When , represents the baseline value for speed reduction magnitude, which corresponds to the lowest speed ratio of .

### **Step 4: Time-dependent delay and average speed during congestion duration**

We have the ordinary differential equation within congested space-time regimes for each link. By assuming dynamic arrival rates, departure rates, queue length process, and introducing elasticity parameters and , we can capture the relationship between the congestion duration and the average discharge rate . Besides, we correlate congestion duration and maximum virtual queues (at the lowest speed) through regression analysis.

If the cubic model is adopted with , we will have the following time-dependent queue and delay.

We have and according to Cheng et al. (2022). Then the longest time-dependent delay at becomes available as follows:

Then the average delay during a congestion duration can be calculated by the following equation (see Cheng et al., 2022):

This leads to an important property that the ratio of average waiting time and longest waiting time (i.e., ) is a constant value. Given , dividing Eq. (32) by Eq. (28b) gives us

We have the average speed during the congestion duration as:

Eq. (33) can be transformed in terms of speed reduction factor, , which is in turn a function of ratio below. This leads to a BPR-like link performance function.

Where and . An alternative form can be written as .

## Notations:

**Table 1:** Symbols and definitions used in this study.

|  |  |
| --- | --- |
| Symbols | Definitions |
|  | link length |
|  | start time of congestion period |
|  | time index with maximum inflow rate |
|  | time index with maximum queue length |
|  | end time of congestion period |
|  | capacity (or discharge rate), assumed to be a constant value |
|  | total in-flow demand during the whole peak period |
|  | lane-based ultimate hourly capacity |
|  | total lane-based volume loaded on a road link during an analysis period (i.e., AM, MD, PM, or NT) |
|  | free-flow travel time |
|  | congestion speed |
|  | cut-off speed[[1]](#footnote-1) |
|  | lowest speed on a link |
|  | inflow curvature parameter used in polynomial form |
|  | inflow rate function at time |
|  | cumulative inflow rate at time |
|  | cumulative discharge rate at time |
|  | queue length at time |
|  | traffic delay departing at time |
|  | average delay during the whole peak period |
|  | average travel time during the whole peak period |
|  | parameters in the BPR-form link performance function |
|  | constant in elasticity function for mapping D/C ratio to congestion duration |
|  | elasticity coefficient of congestion duration in response to D/C changes, i.e., oversaturation-to-duration elasticity |
|  | constant in elasticity function for mapping congestion duration to the magnitude of speed reduction |
|  | elasticity coefficient of speed reduction magnitude in response to congestion duration changes, i.e., duration-to-speed reduction elasticity |

1. As the cut-off speed can be used more systematically to distinguish “congested” vs. “uncongested” states of traffic bottleneck (Hale, et al. 2016). [↑](#footnote-ref-1)