Consistency of Input-Output Model and Shockwave Analysis in Queue and Delay Estimations

YI Ping1,*, TIAN Zongzhong2, ZHAO Qiang1

1 Department of Civil Engineering, University of Akron, Akron, OH 44325-3905, USA
2 Department of Civil and Environmental Engineering, University of Nevada, Reno, NV 89557, USA

Abstract: This paper investigates the consistency between the simple input-output model and the shockwave analysis method as applied to traffic bottleneck problems. Mathematical derivations considering changes in arriving demand and discharge rate have shown that these models are consistent in queue and delay estimations, as both methods produce identical results at anytime during the congestion propagation and dissipation processes. Earlier findings on the inconsistency of these two methods are attributed to the lack of consideration of the baseline factor, which is dependent on the level of background traffic. Numerical examples are presented to illustrate the importance of the baseline factor, to ensure consistency regardless of the use of different flow-density relationships.

Key Words: Input-output model; shockwave; bottleneck; queue; delay

1 Introduction

Accurate estimation of traffic congestion and delay is an important part of studying traffic flow characteristics and developing effective traffic control strategies. A good example is that delays and other congestion-related measures have been widely used to assist transportation agencies in devising system-wide traffic signal and ramp metering operations. Traffic delay is also a key parameter when selecting management strategies at work zones and under incident conditions.

Two methods are commonly used to estimate queues and delays: one is to use the simple input-output model and the other is based on shockwave analysis. The simple input-output model determines traffic queues based on the difference between the total arrivals and the total departures of a traffic facility. The shockwave analysis method describes traffic queues through changes in traffic density along a roadway section. Unlike the simple input-output model that estimates “vertical queues” without considering the space dimension, shockwave analysis includes both time and space in its state equations.

Because of the above-mentioned differences between the two methods, there have been a significant number of discussions in the literature over the consistency of these two methods. Michalopoulos and Pishardy [1] have discussed the limitations of the input-output model and, through a numerical example, demonstrated that the input-output model and shockwave analysis yield different delays. Daganzo [2] has shown that such differences are due to a glitch in specifying the length of the jam density in Michalopoulos’s example and claimed that the two methods should provide the same answer if discrepancies due to approximation of numerical calculations are allowed. McShane and Roess [3], through a numerical example concluded that the two methods can produce different results. Chin [4] in his example points out that the numerical discrepancies between the two methods can be minimized to a negligible level (e.g., 5%), through a graphical compensation. However, Nam and Drew [5] claim that the two methods are fundamentally different, and that the input-output model always underestimates queue size and delay. In their view, shockwave analysis is more realistic as it considers vehicle interactions due to drivers’ reactions to congestion.

Other similar discussions not included in the earlier summary have also contributed to the debate, but most of the conclusions have been drawn primarily by comparing the
numerical results derived from the two methods independently. To be sure, an effort that looks into the intrinsic connection of the two methods is lacking. Although some researchers tried to investigate this problem graphically, this paper studies the interrelationship between the input-output model and shockwave analysis by mathematical derivations. Through numerical examples involving variable arriving and departure flows at a roadway bottleneck, the paper demonstrates that the two methods are exactly the same in queue and delay estimations, and that the result from one method can be easily converted to that of the other. By means of this paper, the authors wish to further help clarify the inconsistency problem between the input-output model and shockwave analysis, thereby promoting correct understanding and application of those two methods in engineering practice and education.

2 Input-output model and shockwave analysis

The simple input-output model (also called deterministic queuing model or cumulative arrival and departure model) is commonly used to describe traffic congestion in a bottleneck. By convention, the queue size at anytime is represented by the difference between the cumulative arrival curve and the cumulative departure curve, and the total travel time (including delay) experienced by all the vehicles going through the bottleneck, is measured as the area enclosed by the arrival and the departure curves. The shockwave queuing analysis keeps track of the queue propagation and dissipation. Multiplying the length of the queue by its corresponding density at any time gives an estimate of the queue size. Similarly, the total travel time can be represented by the product of traffic density and the area enclosed by the trajectory of queue length in a time-space domain. Fig. 1 illustrates the concept of the two methods, where \( q_a \) and \( q_c \) are the arrival demand, \( q_d \) is the queue discharging rate, and \( u_c \) and \( u_{\text{sh}} \) are the speeds of shockwave during queue propagation and dissipation, respectively.

In a pipeline setting (no entering or exiting flow in the midsection), a downstream bottleneck results in a queue which can be determined by the two methods, as

**Input-output analysis:**

\[
\text{Queue} = (q_a - q_d) \times t
\]

(1)

**Shockwave analysis:**

\[
\text{Queue} = u_c \times t \times (q_a - q_d)/(k_c - k_s)
\]

(2)

where \( k_s \) is the corresponding density of the arriving flow and \( k_c \) the discharging flow. Clearly, Eq. (1) differs from Eq. (2) by a factor of \( k_c/(k_c - k_s) \). Furthermore, because \( k_s > k_c \) and \( k_c/(k_c - k_s) > 1 \), the estimated queue size from Eq. (2) is larger than from Eq. (1)\(^{[6]} \). This observation has been the basis of earlier discussions on the inconsistency of the input-output model and shockwave analysis.

3 Consistency of the two methods

The input-output model and shockwave analysis are based on the principle of conservation of mass, and both methods are deterministic (vs. statistical) without using numerical approximations. What has contributed to the differences when both are in fact used to describe the same physical phenomenon? Is that indeed because of certain limitations of the methods themselves, or is it simply due to the lack of understanding or misinterpretation of the problem?

A more detailed study of the queuing dynamics at a bottleneck was made to search for answers to the above-mentioned questions. Assume two density regions exist at time \( t_1 \) in Fig. 2 due to a downstream bottleneck in a roadway section, \( L \). The flow rate and corresponding density are \( q_a \) and \( q_c \) in the arriving demand and \( k_c \) in the congested region, respectively. As congestion propagates upstream, the length of each region has changed at time \( t_2 \), where \( t_2 - t_1 = \Delta t \). According to the input-output model, the total number of vehicles delayed in the system will be

\[
N = (q_a - q_d) \Delta t
\]

(3)

As input-output analysis does not explicitly consider the space dimension, Eq. (3) is often said to produce a “vertical queue”. It is important to realize that if there had been no changes in the downstream flow and density, the original \( q_a \) and \( k_c \)—the “baseline condition”—would not have been affected throughout section \( L \), therefore, no queuing and delay would have been considered.

---

**Fig. 1** Traffic queue propagation and dissipation

**Fig. 2** Congestion propagation at bottleneck
Eq. (3) describes the net change of vehicles during over the baseline should be considered in delay estimation. The baseline factor calculates the number of vehicles as the background traffic on the roadway segment before density experiences changes. Without the baseline factor Eq. (9) would be the same as Eq. (2). As the earlier analysis applies to a pipeline situation and the authors intend to explore the effectiveness of the baseline factor in situations closely representing real world traffic conditions. A few additional scenarios are investigated next, which involve changes in the arriving demand and the queue discharging rate. For simplicity, in each case the queue size is assumed to be zero at \(t=0\) before the queue propagation starts.

### 3.1 Change in arriving demand

From 0 to \(t_1\) (Fig. 3), congestion propagates upstream and the queue size at anytime \(t\) can be obtained through Eqs. (4) or 7. Beginning at \(t_1\), the arrival flow rate is reduced (due to demand diversion, for example) to \(q_a\), where \(q_a<q_c\). The cumulative queue size starts to shrink until \(t_d\), where the queue is completely dissolved and the entire roadway section returns to the normal operation condition with flow \(q_c\).

The rate at which the queue size is decreasing depends on the magnitude of the dissipation wave, \(u_w\), which moves downstream toward the bottleneck. It is seen that starting from \(t_1\), the background flow has changed to \(q_a\), which represents the new baseline condition. At any time between \(t_1\) and \(t_d\), the queue size in the system, according to input-output analysis, can be estimated as

\[
N(t) = (q_a - q_c)(t - t_1) \quad (t_1 \leq t \leq t_d) \tag{10}
\]

The estimated queue size from shockwave analysis would be

\[
N(t) = (q_a - q_c)(t - t_1)(k - k_a) 
\]

\[
= (q_a - q_c)(t - t_1)(k - k_a) \quad (t_1 \leq t \leq t_d) \tag{11}
\]

Equations (10) and (11) show that at any time during queue dissipation the estimated queue size (and therefore the delay) is the same from either the input-output model or shockwave analysis.

### 3.2 Change in queue discharging rate

When a change in queue discharging rate is introduced in the process, the queueing dynamics become more complex. Because of different queue propagation speeds from the bottleneck before and after the change of queue discharging rate, specific steps must be taken to treat multiple queues. Fig. 3 shows comparative sketches involving changing queue discharging rates between the input-output model and shockwave analysis.

At any time between 0 and \(t_1\) (before the discharging flow rate changes from \(q_c\) to \(q_a\)), the estimated queue size in the system from both the input-output model and the shockwave analysis is the same as shown in Eqs. (4) or (7).

\[(a)\] Queue propagation

Beyond \(t_1\), as the farthest end of the queue continues moving upstream due to arriving vehicles, the change in discharging flow at the bottleneck point (due to partial removal of incident, for example) results in a releasing wave \(u_r\), which traverses upstream from \(t_1\) to \(t_2\), where the maximum queue spillback distance \(L_s\) is reached. At any time between \(t_1\) and \(t_d\), multiple density segments exist between the farthest end of the queue and the bottleneck point. The queue size in the system, according to input-output analysis, can be...
estimated as
\[ N_{\text{est}}(q_c, q_r)_{t=0} + (q_c - q_r)(t - t_0) \quad (t \leq t_2) \quad (12) \]
Similarly, the queue size based on shockwave analysis will be
\[ N_{sw} = -[u_r(t - t_1)](k_r - k_c) + [-u_c(t - t_1)](k_c - k_r) \]
\[ -\frac{q_c - q_r}{k_r - k_c} - \frac{q_r - q_c}{k_c - k_r} (t - t_1) + \frac{q_r - q_c}{k_c - k_r} (t - t_2) \]
\[ -\frac{q_r - q_c}{k_c - k_r} (t - t_1) - (q_c - q_r) \cdot t + \frac{q_r - q_c}{k_c - k_r} (t - t_2) \]
\[ = (q_c - q_r) \cdot t_1 + (q_r - q_c) (t - t_1) \quad (t \leq t_2) \quad (13) \]
Therefore, \( N_{\text{est}} = N_{sw} \).

(b) Queue dissipation
When the releasing wave up reaches the farthest end of the queue at \( t_r \), another wave \( u_r \) is formed due to the discharging flow \( q_r \) and the arriving flow \( q_c \). This new wave moves downstream until complete removal of the queue at \( t_d \) when the normal operation condition of the roadway section returns. At \( t \geq t_d \), the queue size estimated from the two methods will be
\[ N_{\text{est}}(q_c, q_r)_{t=0} + (q_c - q_r)(t - t_2) \quad (t \leq t_2) \quad (14) \]
\[ N_{sw} = -[u_r(t - t_1)](k_r - k_c) - \frac{q_r - q_c}{k_c - k_r} (t - t_1) + \frac{q_r - q_c}{k_c - k_r} (t - t_2) \quad (t \leq t_2) \quad (15) \]
Notice that the first term on the right-hand side of Eq. (15) is the queue size at \( t_r \), which is the same as Eq. (13) when \( t = t_r \), therefore,
\[ N_{sw} = (q_c - q_r) \cdot t_1 + (q_c - q_r)(t - t_1) \quad (t \leq t_2) \quad (16) \]
Again, \( N_{\text{est}} = N_{sw} \).

3.3 Change in both arriving and discharging flows
Often, when a bottleneck is formed due to an incident both the arriving demand and the discharging flow may change as shown in Fig. 4. The authors have analyzed the formation and dissipation of queues under such conditions in a similar manner to those in previous discussions. The entire time period from \( t = 0 \) to \( t = t_4 \) were broken into smaller time intervals, \( (t_1) \), \( (t_2 - t_1) \), \( (t_3 - t_2) \), and \( (t_4 - t_3) \), where Eqs. (4) to (16) were used to compare the queue size estimated from the input-output model and the shockwave analysis. Variations in the relative values of \( q_a, q_r, q_r', q_r '' \), and their respective durations were also made in the analysis. The results demonstrated that both the input-output model and shockwave analysis gave exactly the same queue size estimate in each of the time intervals.

4 Case studies
Two numerical examples are presented in this section to demonstrate the consistency of the input-output model and shockwave analysis. The first example is based on Chin[4] and the second adopted from Roess[7].
the results will become

\[ R_2 = \left( \frac{q_2 - q_1}{k_2 - k_1} \right) (T_{as} - t_2) k_2 + \frac{500 - 1200}{90 - 22} (T_{as} - 1) \times 22 \]

=97.058(T_{as}−1) veh (corresponding to \( q_2 \))

Maximum queue size:

\[ N_{m,N} = \left( \frac{q_2 - q_1}{k_2 - k_1} \right) k_1 t_1 - R_1 = 551 - 251 = 300 \text{ veh} = N_{I/O} \]

Time of congestion:

\[ \left( \frac{q_2 - q_1}{k_2 - k_1} \right) k_1 t_1 - R_1 = \left( \frac{q_2 - q_1}{k_2 - k_1} \right) (T_{as} - t_1) - R_2 \]

Simplify and solve to get

\[ T_{sw} = 2 \text{ h} = 120 \text{ min} = T_{I/O} \]

Maximum individual travel time through the bottleneck:

\[ d_{m,N} = \frac{u_w k_1 t_1 - R_1}{q_c} \]

=\[ \frac{1500 - 1800}{90 - 44} (90 \times 1 - 251)/1500 \]

=0.2 h=12 min=d_{I/O}

### 4.2 Example 2

This example deals with a three-lane roadway, where a flow-density relationship is known as \( q = 0.4734(k)(130 - k) \). A bottleneck exists with a capacity \( q_c = 4000 \text{ vph} \). The normal traffic demand is \( q_1 = 5000 \text{ vph} \), and is reduced to \( q_2 = 3000 \text{ vph} \) after two hours until complete removal of the congestion.

Similar calculations were carried out for Example 1. In order to save space, the results are summarized in Table.

### 5 Conclusions

This paper demonstrated the consistency of the input-output model and the shockwave analysis in queue and delay estimation. In a variety of traffic and boundary conditions it showed that at any time during the congestion period, both the methods yielded the same queue size at a roadway bottleneck. Therefore, other statistics were generated, such as time of congestion, total number of affected vehicles, and maximum queue size, and so on, which were also found to be consistent.

The consistency between the input-output model and the shockwave analysis is maintained by the baseline factor. The mathematical derivations in the paper show that the lack of consideration of the baseline factor leads to inconsistent estimates of queues and delays. The numerical examples demonstrate that the effectiveness of the baseline factor is scenario independent and is not affected by the use of different flow-density relationships. This paper helps clarify the conceptual confusion of inconsistency between the two models. Its findings are important not only to college education but also to engineering applications of these methods.

### References


