sample size for this example should be less than one-half of the total sample size. This appears to suggest that, in general, the initial random sample can be small, regardless of the total sample size. The size of this sample may in fact be dictated by the requirements on the estimation of the distribution of the explanatory variables in all the groups. This point was not addressed in this paper, which assumed that this distribution is known.

## CONCLUSIONS

The two major conclusions from the work described here may be stated as follows:

1. The SO procedure can introduce a significant increase in parameter estimation accuracy, and
2. This optimization need not be based on accurate initial parameter guesses; only a small pilot sample is needed to produce sufficiently accurate guesses.

It should be emphasized, however, that these conclusions result from a specific set of tests performed on prespecified models. Even though these models were chosen without any regard to the final results, these results can be generalized only with caution. The results are, however, encouraging in that the SO procedure appears to be worthwhile in cases where it can be applied. It requires nonlinear optimization software, which may not be easily used in many environments.

## REFERENCES


1. W.G. Cochran. Sampling Techniques. Wiley, New York, 1963.
2. C.F. Daganzo. Optimal Sampling Strategies for Statistical Models with Discrete Dependent Variables. Transportation Science, Vol. 14, No. 4, Nov. 1980.
3. Y. Sheffi and Z. Tarem. Experiments with Optimal Sampling for Discrete Choice Models. Office of University Research, U.S. Department of Transportation, Aug. 1982.
4. S.R. Lerman and C.F. Manski. Sample Design for Discrete Choice Analysis of Travel Behavior: The State of the Art. Transportation Research, Part A, Vol. 13A, No. 1, Feb. 1979.
5. C.F. Daganzo. On the Uniqueness and Globality of Optimal Data Gathering Strategies. Transportation Science, Vol. 16, No. 2, May 1982.
6. E. Polak. Theory of Optimal Control and Mathematical Programming. McGraw-Hill, New York, 1979.
7. T. Domeneich and D. McFadden. Urban Travel Demand: A Behavioral Approach. North-Holland, Amsterdam, 1975.


*Publication of this paper sponsored by Committee on Passenger Travel Demand Forecasting.*

# Procedure for Predicting Queues and Delays on Expressways in Urban Core Areas

THOMAS E. LISCO

**A procedure that predicts morning inbound and evening outbound queuing delays on express highway facilities in downtown areas is discussed. The procedure is based on the relationships among hourly traffic capacities at bottleneck points, daily volumes at those points, and associated queues and delays. The need for such a procedure arose from difficulties in using traffic assignment or other existing analysis techniques to predict queues and delays associated with alternative highway plans. Empirical delay data for developing the procedure came from nearly 600 speed runs conducted on the express highway system in and near downtown Boston. Fourteen queuing and potential queuing situations were analyzed. The relationships derived appear to be generalizable, and the specific results from the Boston area should apply to other urban areas of comparable size.**

A procedure that predicts peak-period queuing and delays on express highway facilities in downtown areas is discussed. The procedure is based on the relationships among hourly traffic capacities at bottleneck points, daily volumes at those points, and associated peak-period queues and delays. (In this paper the term daily volume refers to average weekday traffic.) The procedure was developed by comparing observed bottleneck capacities with empirical delay data for traffic upstream of the bottlenecks. Capacities were derived from traffic counts at bottleneck locations. The delay data were from almost 600 speed runs conducted on express highway facilities in and near downtown Boston, mostly during 1978 and 1979. The procedure was developed for use in detailed evaluations of potential traffic impacts and benefits of alternative highway investments in downtown areas.

The need for such a procedure arises initially from difficulties in using the output from traffic assignment models to predict peak-period operating conditions and cost-benefit statistics associated with alternative highway plans. The basic problem is that the regional traffic assignment process derives speeds for individual links separately based on their individual volume/capacity (v/c) ratios and does not consider the queuing effects of bottleneck locations. Thus in typical downtown area queuing situations, where one bottleneck highway segment can create queues stretching into many other segments, traffic assignments cannot indicate the locations and extents of queues or the delays associated with them. Because queuing can be of major importance in peak-period expressway operations in downtown areas, the assignments can be grossly inaccurate in predicting peak-period operating speeds. Similarly, the associated cost-benefit statistics can miss much of the phenomenon they are intended to measure.

A potential solution to this problem would be to attempt a queuing analysis based on peak-period traffic assignment results. Such an analysis would fail for two reasons. First, by its very nature a traffic assignment is balanced, with all highway links clearing all traffic assigned to them for the time period of analysis. Thus there is no possibility of an assignment producing for a bottleneck link the different vehicle arrival and service rates necessary to perform a queuing analysis. Second, a well-calibrated traffic assignment will indicate all bottleneck links operating exactly at capacity during peak periods, with no indication of which are major and minor bottlenecks. In some cases the assignments will indicate volumes greater than actual capacities at bottlenecks, but the degree to which such volumes are indicated is related more to the nature of the capacity constraint in the assignment program than to the queuing phenomenon. Therefore, these greater-than-capacity volumes are not particularly helpful in predicting the extents of potential queues.

An alternative solution would be to perform a queuing analysis based on daily traffic assignment volumes with given fractions of daily traffic assigned to peak hours. The traffic assigned to peak hours would be compared with capacities at bottleneck points. Again there would be severe problems. One problem is that different bottlenecks process different fractions of daily traffic during the peak periods, with lower fractions being handled by severe bottlenecks. Thus a given fraction applied to all bottlenecks would underestimate the effects of small bottlenecks and overestimate the effects of large ones. A more important consideration is that queues rarely contain more than several hundred vehicles at one time. Thus any procedure that attempts to predict queues through calculating differences between arrival and service rates must project flows with a great deal of accuracy. Certainly, this cannot be done by allocating fractions of daily traffic to hourly flows at bottleneck points. As before, the delays calculated will relate far more to the assumptions used in the allocation than to the queuing phenomenon.

Because of the difficulties involved in predicting vehicle arrival and service rates from traffic assignments and, more generally, the problems of accurately predicting these rates by any method (1-3), the procedure documented in this paper follows an approach that predicts queuing delays directly without calculating the difference between arrival and service rates. Specifically, the analysis approach assumes that there is a consistent relationship between daily traffic volume at a bottleneck point compared with capacity, and typical peak-period delays upstream of the bottleneck.

To search for such a relationship, an extensive analysis was conducted of the complex expressway queuing phenomenon in and near downtown Boston. Delay data were compared with volumes and capacities at bottleneck points, and a set of rules was developed that operates in the formation of queues and appears to explain the interrelationships among them. Ultimately, a procedure was developed that predicts morning inbound queues and evening outbound queues for downtown area expressways. The procedure is in two parts. In the first part the average maximum peak-period delays are predicted by using a comparison of daily bottleneck volumes with hourly capacities. In the second part queue speeds are derived from hourly v/c ratios of queue sections, and queue lengths are calculated from queue speeds and delays.

In this analysis no attempt has been made to predict outbound morning delays or inbound evening delays, or delays on highways that are not downtown oriented. Also, no consideration has been given to predicting delays caused by heavy stop-and-go traffic with no explicit bottleneck points. Such circumstances were not adequately represented in the data. Further, the procedure as presented does not include any consideration of the variation of queue lengths during the peak period. Patterns of within-peak variations tend to be similar among queues and can be adjusted as circumstances require.

## BASIC RELATIONSHIPS GOVERNING MORNING AND EVENING PEAK-PERIOD QUEUING DELAYS

The basic relationships between average maximum peak-period delays and daily traffic related to hourly bottleneck capacity are shown in Figures 1 and 2 for morning and evening peak periods. The relationships shown are manually fitted curves from the Boston speed-run data. Six data points are for the morning peak period, and eight data points are for the evening peak period. The data in the figures indicate that peak-period queues and delays begin to materialize when daily traffic volumes reach the vicinity of 8 to 10 times the hourly capacity at bottleneck points. Evening peak-period delays are greater than morning delays for any given daily volume relative to hourly bottleneck capacity because evening peak-period traffic tends to be heavier than morning peak-period traffic. Similarly, evening delays increase more quickly than morning delays for given increases in daily volumes relative to bottleneck capacities.

In evaluating the curves shown in Figures 1 and 2, it can be seen that their shapes are quite regular and sensible. Also, the relationships between the fitted curves and the data points are close. In no case does the predicted delay from the curves

Figure 1. Daily traffic volume as a multiple of hourly capacity at bottleneck versus average maximum morning peak-period delay.

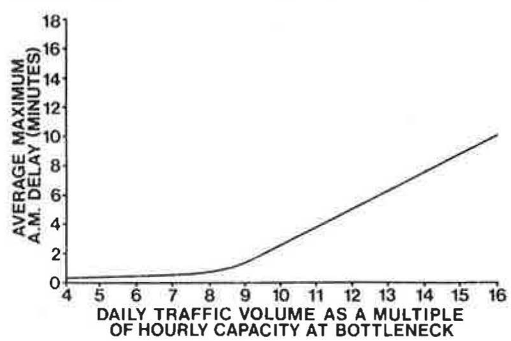
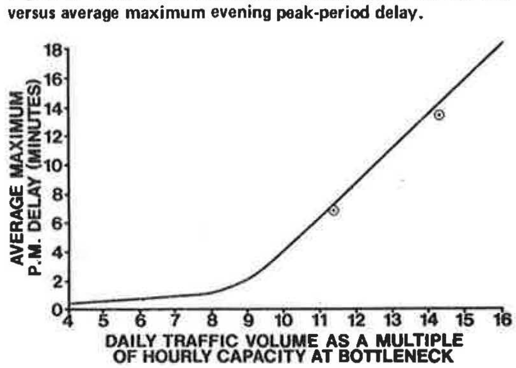



Figure 2. Daily traffic volume as a multiple of hourly capacity at bottleneck versus average maximum evening peak-period delay.



differ from the experienced average delay of the speed runs by more than 1 min. This difference represents less than 15 vehicles per lane in a typical queue.

Of the 14 data points, only 2 are irregular in their derivation. These data points are circled in Figure 2. The circled point with the greater delay is for travel from Logan Airport to downtown Boston through the Sumner Tunnel, an inbound rather than an outbound route. This data point was included in the evening outbound statistics because Logan Airport is a major traffic generator in the Boston core area, and because evening peak-period traffic from Logan Airport can be considered to be outbound, regardless of its direction.

The second irregular data point, which shows less delay, is the data point for I-93 and the Boston Central Artery southbound during the evening peak period. In the derivation of delay data, the segment of this route considered is assumed to have one long queue, even though it has an intermediate section that does not become solidly queued every evening. Because this section is quite short, it was not considered to substantially affect the validity of the data point.

There is one major drawback in the data: there are so few data points; i.e., a total of 14 to fit two curves. Boston has only a few explicit bottleneck points on its express highway system in and near the downtown area; thus data were taken for all of them.

## CALCULATING QUEUE LENGTHS FROM DELAYS

To calculate queue lengths from the delay curves shown in Figures 1 and 2, it is necessary to compare queue speeds on highway segments with speeds on the same segments under uncongested conditions. When the delay per unit distance that the difference between queue speed and congested speed implies is known, as well as the total delay in the queue, queue length can be determined by calculating the distance of travel necessary to accumulate the total delay.

Information on queue speeds is shown in Figure 3. The data in this figure relate queue speeds to conventional hourly v/c ratios and also indicate what is, in effect, a level-of-service F curve for queues. The input speed data for the figure were actual speeds from speed runs for all segments of all morning and evening queues on the highway system in the downtown Boston area. As the data in this figure reveal, almost all of the observed speeds are within 1 or 2 mph of what would be predicted by the estimated curve.

Also shown in Figure 3 is the level-of-service F curve from the 1965 Highway Capacity Manual (4, p. 264). It is interesting to note that the estimated curve for queues in the downtown Boston area has speeds less than those of the curve in the Highway Capacity Manual. Although the reason for this is not clear, it appears that the level-of-service F curve in the Highway Capacity Manual was derived from statistics for stop-and-go conditions, with no explicit bottleneck points and no explicit queues. There is support for this notion because the only three Boston data points that are near the curve in the Highway Capacity Manual (the points circled in Figure 3) are those for I-93 and the Central Artery (southbound) in the evening. As noted previously, this section of highway has a segment that is not solidly queued every evening. Thus average speeds are higher. In any case, the fitted Boston curve is appropriate for estimating existing and future queue speeds and lengths.

The following is a hypothetical delay and queue-length calculation. Suppose an expressway has three travel lanes inbound, each of which has a capacity of 2,000 vehicles per hour. Total inbound capacity of the highway is 6,000 vehicles per hour. At one point there is the constriction of a lane being dropped. Beyond this point two lanes remain with a total capacity of 4,000 vehicles per hour. Suppose also that the average weekday traffic inbound at the bottleneck is 50,000 vehicles, or 12.5 times the hourly capacity at that point. Finally, suppose that the highway operates at 55 mph during uncongested periods.

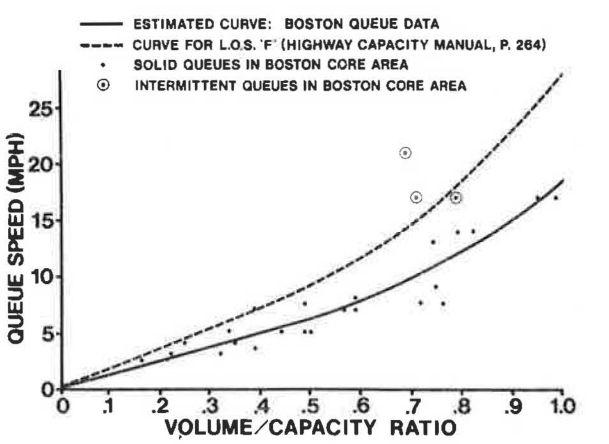
The questions to be answered are as follows: (a) What will be the average maximum morning delay upstream of the bottleneck? and (b) How long will the average maximum morning queue be in which that delay will be experienced? The answer to the first question comes directly from Figure 1. With an average daily traffic volume 12.5 times the hourly bottleneck capacity, the average maximum morning delay will be about 5.7 min.

The calculation of queue length is a little more complicated. In the queue area the v/c ratio is 0.67 (4,000 vehicles per hour traveling on three lanes that could handle 6,000 vehicles if it were not for the bottleneck). This corresponds with a queue speed of 9.5 mph (as shown in Figure 3). At this speed it takes 6.316 min to travel a mile (1/9.5 x 60). In uncongested conditions it takes 1.091 min to travel a mile (1/55 x 60). Thus a vehicle traveling 1 mile in the queue will incur 5.225 min of delay (6.316 - 1.091). Because the total delay in the queue was calculated to be 5.7 min, the average maximum queue length will be 1.091 miles (5.700/5.225), or 5,760 ft.

Clearly, the procedure for calculating queue delays and lengths is quite simple. A little more work is required if there are on-ramps and off-ramps or variations in capacity within the queued section. In such cases v/c ratios and speeds must be calculated separately by segments of the highway section (moving upstream from the bottleneck) and delays added up by segment until the total queue delay is achieved.

One final note is appropriate concerning the application of the model. In determining hourly bottleneck capacity for the determination of delay, the actual peak-period capacity of the bottleneck should be used, including vehicle mix, weaves, and geometrics. Alternatively, counts may be used. However, for determining queue length, capacities should be considered to be approximately 2,000 vehicles per lane per hour because vehicle mix, weaves, and geo-

**Figure 3. Relationship between hourly v/c ratio and queue speed: morning and evening queues.**



metrics become largely irrelevant when vehicles are waiting in line.

DETERMINING LOCATIONS OF BOTTLENECK POINTS

Before the procedure for predicting queue delays and lengths can be carried out, the exact locations of the bottleneck points relevant to the given queues must be identified. This task can be more complex than the application of the procedure. During the course of the development of the basic model in this study, a number of methods of selecting bottleneck points were tested in an attempt to develop consistent relationships between peak-period delays and daily volumes relative to hourly capacities at bottlenecks. Ultimately, the best relationships were established by using data that resulted from defining and selecting bottlenecks according to the rules set forth in the following sections. In performing the queuing analysis, the same rules should be used for determining the locations of the bottleneck points.

Simple Queue

When a queue forms on an express highway with heavy traffic, the location of the queue will be upstream of the point with the highest daily volume relative to capacity, which point is the bottleneck point. Such a point may be at a constriction, such as a bridge or a lane drop, or at a merge or diverge of a major flow of traffic.

A simple queue is shown in Figure 4, which shows a bottleneck point and the queue upstream of it. Also shown in Figure 4 are areas upstream of the queue and downstream from the bottleneck where free flows of traffic are maintained.

Two Queues in Succession

In some circumstances a highway may have two bottleneck points in succession. Such a circumstance is shown in Figure 5, which depicts an upstream bottleneck A and a downstream bottleneck B. Here queues and delays depend primarily on which is the greater bottleneck (higher daily volume relative to capacity). If bottleneck A is the greater bottleneck, a queue will develop upstream of bottleneck A but no queue will develop at bottleneck B, because bottleneck A will meter traffic to bottleneck B, so that no queue can develop there. Similarly, if bottleneck B is the greater bottleneck, a queue will form there but none will form at bottleneck A, because traffic will meter itself in anticipation of the queue downstream.

The only circumstance in which queues will develop at both locations will be where the bottlenecks are relatively far apart and substantial volumes of traffic enter and leave the highway between them. In this circumstance traffic at the two bottlenecks is mostly composed of different vehicles, and delays at the two bottlenecks should be predicted separately by using the volume relative to capacity at each.

Split at Head of Queue

Where a highway divides at the head of a queue, three potential bottleneck points may be considered for predicting queue length and delay. This circumstance is shown in Figure 6, which shows bottleneck A before the diverge point and bottlenecks B and C to the left and right after the diverge point. Hypothetical queues predicted from the bottlenecks are shown in the figure, where each queue is based on the daily volume relative to capacity of the given bottleneck.

In the case shown, bottleneck A would generate the smallest queues and delays, bottleneck B would generate the largest queues and delays, and bottleneck C would generate queues and delays of intermediate length and duration. Because it produces the largest queues and delays, bottleneck B should be used for prediction. Potential queues formed by bottlenecks A and C would simply be submerged in the bottleneck B queue.

Split Near Head of Queue

A somewhat similar circumstance to that of a split at the head of a queue is that of a major diverge point near the head of a queue, with the diverging traffic entering a bottleneck itself shortly after the diverge point. This circumstance is shown in Figure 7, which again shows the potential bottlenecks for use in queue and delay prediction. As shown in the figure, bottleneck A is on the main line just before the diverge point, bottleneck B is

Figure 4. Simple queue.

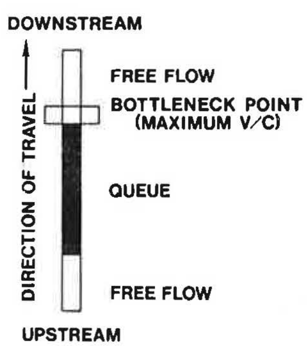



Figure 5. Two queues in succession.

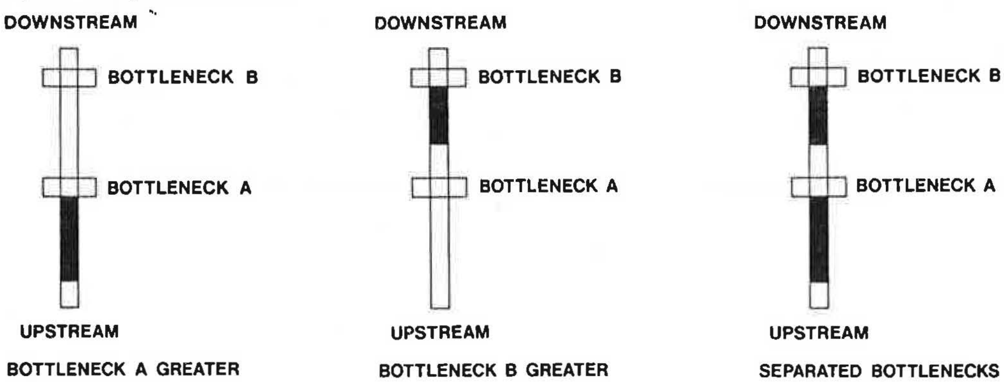


Figure 6. Split at head of queue.

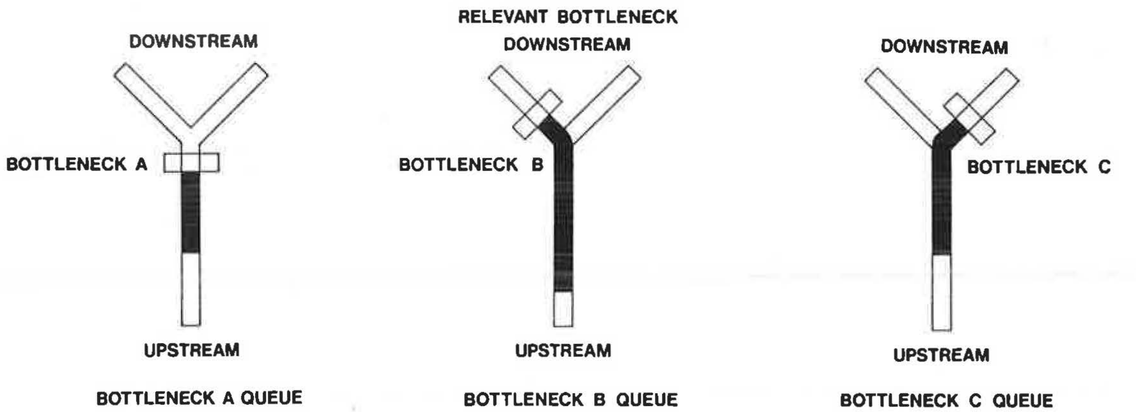
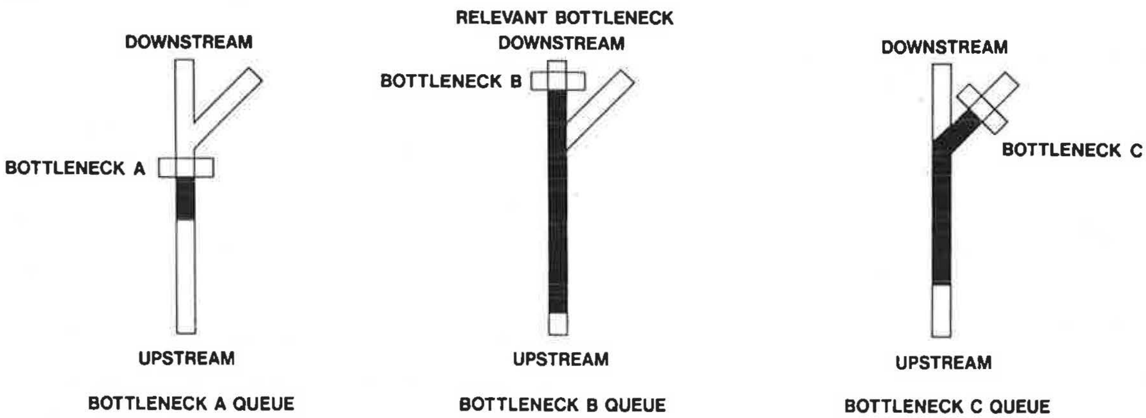



Figure 7. Split near head of queue.




on the main line downstream of the diverge point, and bottleneck C is on the route used by the diverging traffic.

Also shown in Figure 7 are hypothetical queue lengths implied by the three bottlenecks individually. Bottleneck A would generate the shortest queue, bottleneck B the longest queue, and bottleneck C a queue of intermediate length. In this case it is the bottleneck that produces the queue that stretches to the point farthest upstream that should be used for prediction. In the example the relevant queue is from bottleneck B. As before, potential queues from the other bottlenecks would simply be submerged in the bottleneck B queue.

### Two Queues Joining at Bottleneck

Yet another circumstance is that of two major highway flows joining and encountering a bottleneck at the merge point. Such a situation is shown in Figure 8. In this case the question is whether the daily volume relative to capacity of the joined flow at bottleneck A should be used to predict equivalent queues and delays for the two merging flows of traffic, or whether the two flows at bottlenecks B and C should be considered separately. In this circumstance the flows should be considered separately. The daily volumes to be used are those at bottlenecks B and C. The capacities to be used, however, are not those at bottlenecks B and C, but the fractions of the capacity at bottleneck A available through channelization to the traffic flows from bottlenecks B and C.

### Queue Joining Queue Near Bottleneck

A final circumstance is that of two major highway flows joining upstream of a bottleneck on one of them. This circumstance is shown in Figure 9 by three hypothetical cases. In all three cases a main line queue is generated from bottleneck A. Three different possible queues are illustrated from bottleneck B, which is upstream from bottleneck A and applies to the merging traffic where it enters the main flow.

In case 1 bottleneck B creates a small queue for the entering traffic. This is the circumstance in which the entering traffic is a relatively small fraction of the traffic on the main line and can merge into the main flow without difficulty. Presumably, the relationship between daily traffic and potential merge capacity at bottleneck B would create only a minor queue. In case 2 a queue is formed upstream of bottleneck B equal in length to that on the main line. Here both flows are determined effectively by bottleneck A, and there is really one queue with two equivalent tails. In case 3 bottleneck B creates a queue longer than that of the main line upstream of the merge point. Here the queues are probably separate in cause and operation.

Which of these three cases applies in any given situation is difficult to determine because the general circumstance is, in part, equivalent to two queues in succession. The following guidelines, however, may help determine which case applies. If the traffic flows through bottlenecks A and B are largely composed of different vehicles, the queuing

Figure 8. Two queues joining at bottleneck.

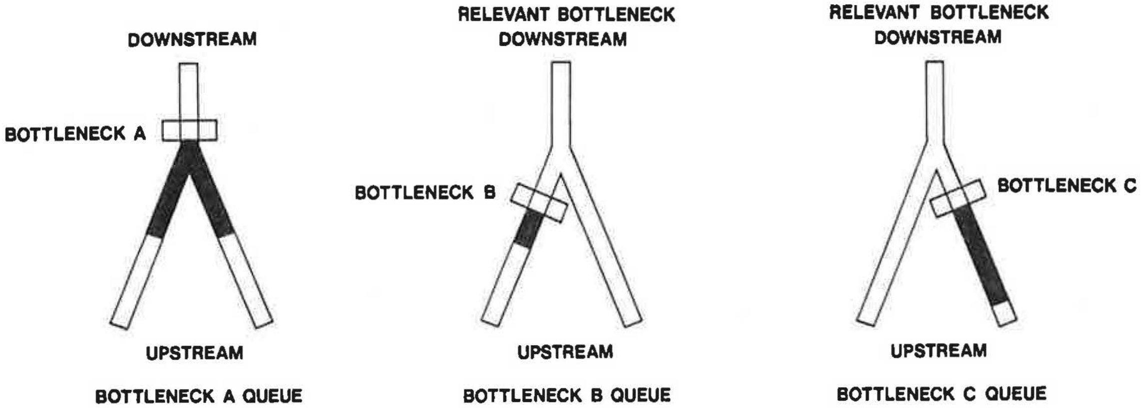
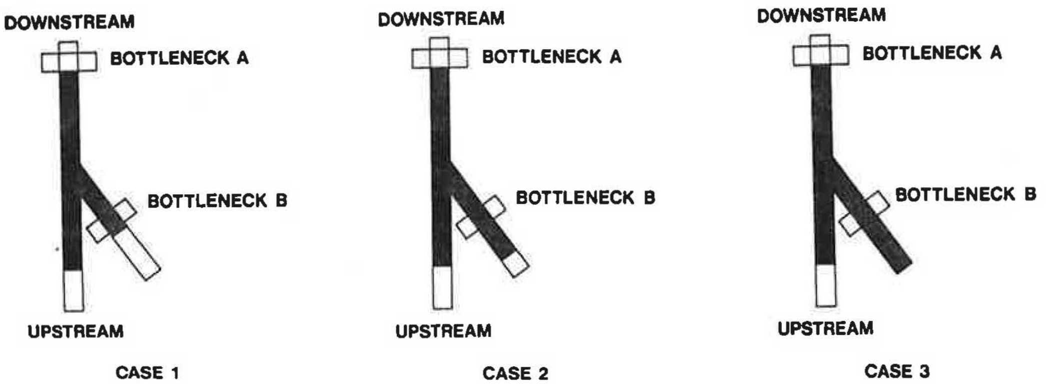



Figure 9. Queue joining queue near bottleneck.




prediction can probably be accomplished separately for the two bottlenecks, as in cases 1 and 3. If most vehicles from both routes are destined for bottleneck A, however, the queuing should probably be predicted by assuming one queue with equivalent tails from bottleneck A, as in case 2.

### Summary

The rules just discussed for bottlenecks would indicate that

1. The relationship between queue delays and daily volumes compared with hourly capacities pertains only to unbroken stretches of congested traffic;
2. The ratio to apply is that of the point with the highest daily volume compared with hourly capacity, the point of which will be at the head of the queue; and
3. The delay to apply is that to the most distant end of the queue.

There are qualifications, and the rules need to be applied with careful attention paid to actual circumstances. But with adequate consideration of geometrics and traffic flows, following the rules previously described yields clear relationships between queue delays and daily volumes relative to hourly capacities at bottleneck points.

## STRENGTHS AND LIMITATIONS OF THE PROCEDURE

The procedure described in this paper has a number of strengths. Primary among these are its ability to use traffic assignment data as input, its simplicity, and its generally reasonable and consistent results. The procedure appears to solve successfully the extremely difficult problem of predicting vehicle arrival and service rates. At the same time, however, the relationships developed for the procedure are based on data collected for only a few queues. Only six data points for morning inbound queues and eight data points for evening outbound queues could be derived from observations of traffic in the Boston core area. Further, some of these data points are subject to question.

An additional limitation of the procedure is its narrow range of applicability: morning inbound and evening outbound queues in the cores of urban areas about the same size as Boston. No attempt was made to calibrate procedures for queues in reverse flows or in nondirectional flows (such as on circumferential routes), for temporary queues where construction projects are under way, or for queues in urban areas of different sizes. Nevertheless, the basic approach appears to be applicable to these circumstances, and analogous procedures could be derived for them with further data collection and analysis.

Certainly, addressing problems of queuing is central to improving the operations of many urban expressway systems. To the extent that the basic approach can be applied to other cities and circumstances, the prediction of queuing from relationships between daily traffic and bottleneck capacities may provide a powerful analysis tool. It could enhance considerably the analyst's ability to predict and evaluate the potential impacts of urban expressway projects.

ACKNOWLEDGMENT

The preparation of this paper was financed through a contract with the Massachusetts Department of Public Works and with the cooperation of the FHWA, U.S. Department of Transportation.

REFERENCES

1. M. Wohl. Notes on Transient Queuing Behavior, Capacity Constraint Functions, and Their Relationship to Travel Forecasting. In The Regional Science Association Papers (M.D. Thomas, ed.), Regional Science Association, The Wharton School, Philadelphia, 1968, pp. 191-202.
2. D.R. Drew. Traffic Flow Theory and Control. McGraw-Hill, New York, 1968.
3. W.A. Stock, R.C. Blankenhorn, and A.D. May. The "FREQ3" Freeway Model. Institute of Transportation and Traffic Engineering, Univ. of California, Berkeley, Rept. 73-1, 1973.
4. Highway Capacity Manual. HRB, Special Rept. 87, 1965, 411 pp.

*Publication of this paper sponsored by Committee on Passenger Travel Demand Forecasting.*