Distinguishing Between Incident Congestion and Recurrent Congestion: A Proposed Logic

ANA I. GALL AND FRED L. HALL

A key element of freeway traffic management systems (FTMSs) is the detection of incidents. The problem with most incident-detection algorithms is that they do not detect incidents as such; rather they detect congestion, whether it is caused by an incident (incident congestion) or by a recurrent bottleneck situation (recurrent congestion). The purpose of this paper is to present a logic for distinguishing between incident congestion and recurrent congestion. The logic uses 30-sec volume and occupancy summaries at each FTMS detector station to classify traffic operations into one of four states. If congestion is detected at one detector station, the cause of this congestion is defined on the basis of the traffic state at the downstream detector station. Results from a preliminary evaluation of the proposed logic are promising.

Freeway traffic management systems (FTMSs) have been in operation for more than 20 years. A key element of such systems is the detection of incidents. Incidents, including accidents, spilled truck loads, and stalled cars (*1*), can be defined as random events that may disrupt the orderly flow of freeway traffic. Incidents can be detected through a variety of methods. One method that has become increasingly important to the effective management of freeway facilities is the automatic detection process. This process uses computer algorithms to monitor data from presence detectors at regular time intervals to evaluate the nature of traffic operations and to identify the presence of a capacity-reducing incident.

Several incident-detection algorithms are in use. Differences among the algorithms are due either to the different underlying logics or to the different detection criteria. The detection criteria refer specifically to the rules used to declare the occurrence of an incident. Despite these inherent differences, most algorithms share a common problem: they do not detect incidents as such; rather they detect congestion, whether it is caused by an incident (incident congestion) or by a recurrent bottleneck situation (recurrent congestion). Consequently, false alarms are a prevalent problem. What is needed is a means to distinguish between recurrent and incident congestion. In this paper, a proposed logic to achieve this is presented, and the results from a feasibility test of the logic are provided. The proposed logic would complement current incident-detection algorithms and thus improve their performance.

Included in this paper are a description of the logic, a description of the study site and of the data base for the feasibility test, a discussion of the calibration process, the results from the feasibility test, and conclusions.

DESCRIPTION OF PROPOSED LOGIC

The proposed logic grew largely out of the current research in incident detection at McMaster University (*2*). The core of the logic is the realization that traffic operation downstream of a permanent bottleneck differs from that downstream of an incident-caused (or temporary) bottleneck. This realization is not new. Wattleworth and Berry (*3*, p. 2) noted that "two types of freeway operation result from these two types of bottlenecks." Levin and Krause (*4*) and Levin, Krause, and Budrick (*5*) recommended that traffic behavior at permanent bottlenecks be investigated during incident conditions to distinguish between incident and incident-free shockwaves.

For purposes of this discussion, it is assumed that congestion has already been detected at a station, and further, that it has been detected on the basis of the single-station logic described in the previous reference, although this is not an essential assumption. The focus in this paper is solely on the problem of deciding whether that congestion has been caused by an incident or whether it is recurrent congestion. The presentation of the logic is divided into two parts: the first deals with classifying traffic operations into traffic states on the basis of variables describing the traffic stream, and the second uses this information to distinguish the cause of the congestion.

The first step of the logic, shown in Figure 1, classifies traffic operations on the freeway facility into one of four possible traffic states on the basis of two variables—volume and occupancy. These variables are obtained from electronic detectors, located at intervals along the facility. Occupancy, a measure of concentration, is defined as the percentage of time a detector is occupied by a vehicle (or vehicles) during the reporting interval.

Figure 1 was developed using the understanding of traffic operations relationships discussed by Persaud and Hall (*2*). As part of that discussion, they found that uncongested operations on a flow-occupancy (or volume-occupancy) plot tend to cluster tightly about a line, the lower bound of which can be established fairly clearly, in Figure 1. Athol (*6*) also found the same pattern. The maximum uncongested occupancy is defined as OCMAX. A volume-occupancy data pair located to the left of the boundary line, and left of OCMAX is classified as State 1, or uncongested. If the volume-occupancy data pair lies to the right of (or below) the boundary line and

Traffic Research Group, Department of Civil Engineering and Engineering Mechanics, McMaster University, Hamilton, Ontario, L8S, 4L7, Canada.

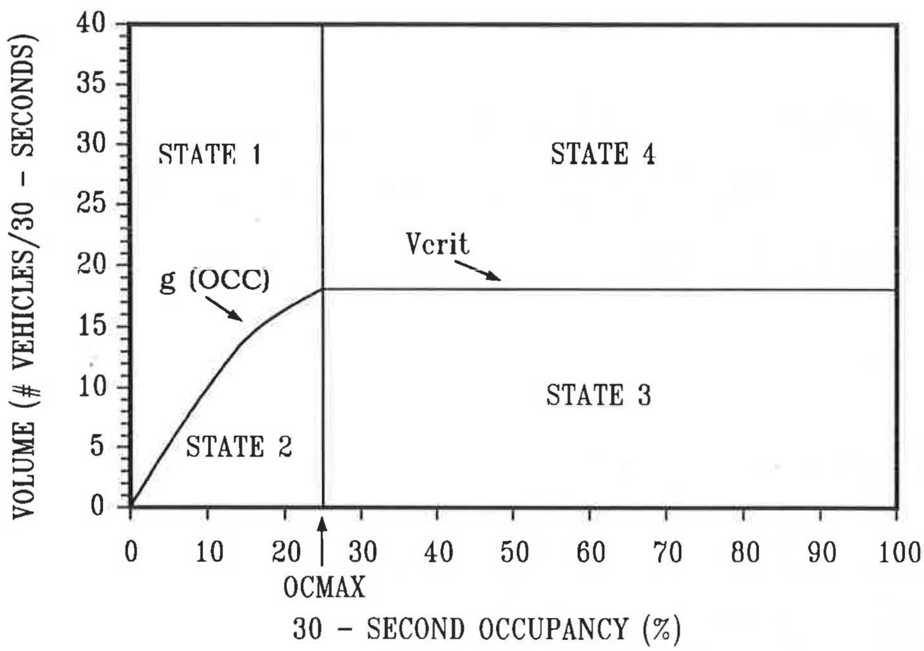

FIGURE 1 An illustration of the volume-occupancy template for traffic state classification.

left of OCMAX, it is classified as State 2, one type of congested operations. To the right of OCMAX are two states: State 3, the second type of congested operations, and State 4, which reflects traffic operations downstream of a permanent bottleneck in a section of roadway operating at or near capacity with accelerating speeds. State 3 can be distinguished from State 4 by Vcrit (which is defined later). In view of this, only stations downstream of an entrance ramp will include a State 4 region in their station-specific volume-occupancy template.

In Figure 2, the first part of the logic is depicted in decision tree form. One additional traffic state, State -1, has been included in Figure 2. This state identifies those sampling intervals for which detector data are missing (as denoted by either volume or occupancy recorded as -1). For the FTMS data we have used, each detector station contains two detectors to measure speed data as well. Hence data from the downstream detector are used whenever data from the upstream loop detector are missing. In the first portion of the decision tree, this screening process is depicted. Only in the event that both detectors fail to record data will the traffic state be classified as -1.

If congested operations are detected (identified by States 2 or 3) at a detector station, i, then the logic shown in Figure 3 is used. This second part of the logic focuses on evaluating the traffic operations at Detector Station $i + 1$, to identify the cause of the congestion detected at Station i.

It is important to note that although the logic uses freeway data from adjacent detector stations to establish the cause of congestion, the logic is not similar to the standard comparative algorithms. Instead, the logic relies on the freeway data from a single detector station to characterize the traffic operations there and looks downstream of the congested detector station to find the cause of this congestion. Also, the logic need be applied only to those freeway sections known to be bottlenecks. At these locations false alarms can arise because of recurrent congestion, and thus at these locations distinguishing the cause of congestion is required.

It also is important to note that the main focus of this discussion is not congestion detection, so only a very simple test for that is provided here (identification of either State 2 or State 3 at a detector station). Incident detection at a single station, discussed briefly by Persaud and Hall (*2*), is the focus of other ongoing research. The focus in this paper is identifying the cause of the congestion once it has been detected.

Before discussing the second step of the logic, it is necessary to review the patterns defining both incident congestion and recurrent congestion within the context of Figure 1. In the case of incident congestion, the typical pattern is as follows: an incident will reduce roadway capacity, causing traffic to queue upstream of the incident. Therefore, traffic operations upstream of the incident will be State 3 (after perhaps a brief move into State 2). Downstream of the incident site, however, the volume is reduced but the roadway capacity is normal. Hence, traffic conditions will either be in State 1 if the detector is located sufficiently downstream of the incident to allow vehicles to resume desired speeds or in State 2 if vehicles are still accelerating back to the desired speed (*7,8*). Even if there is an entrance ramp between the incident site and the downstream detector, operations downstream will still most likely

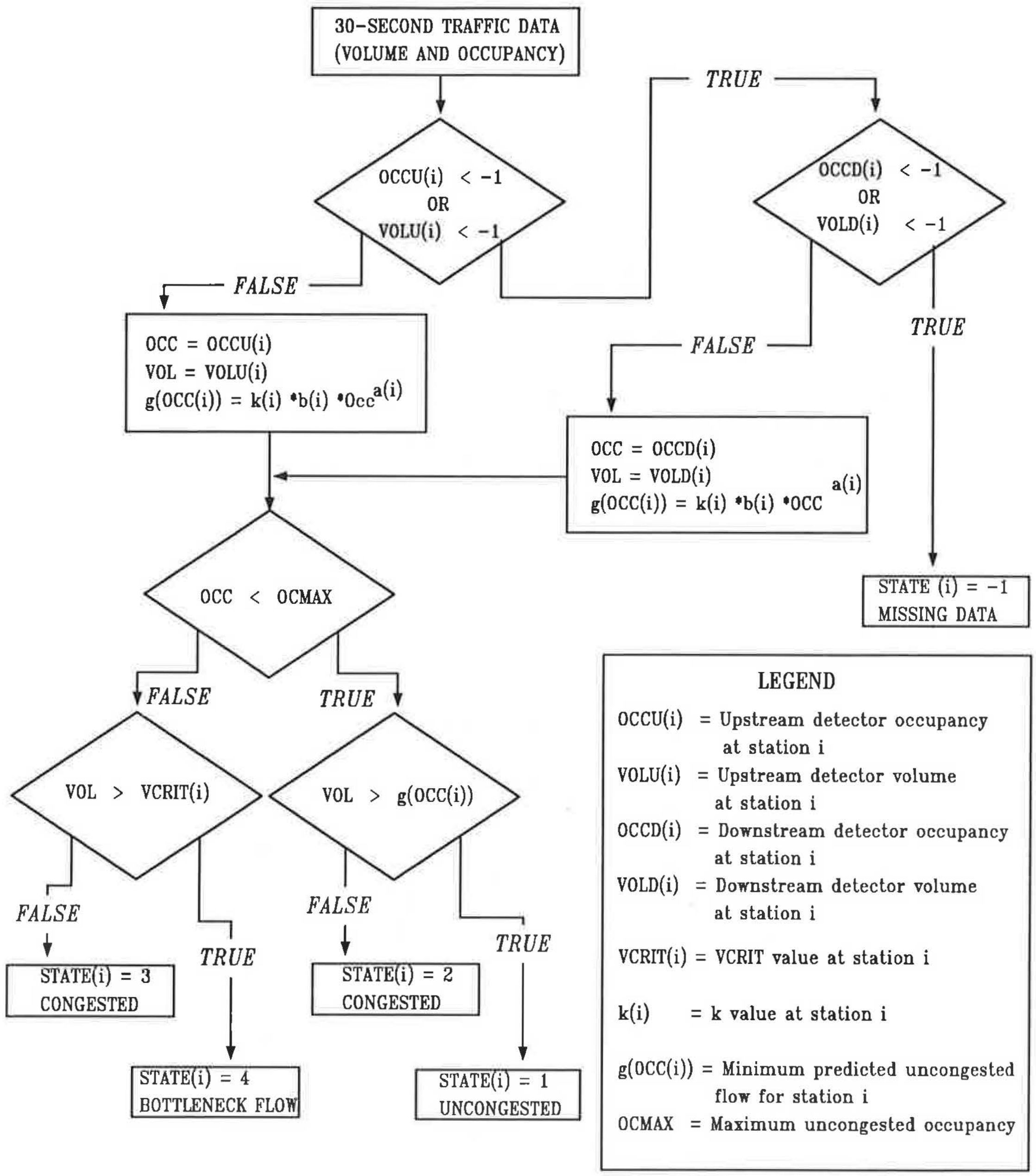

FIGURE 2 Traffic state classification decision tree.

be in States 1 or 2. Only if the ramp feeds more traffic than the reduction in capacity caused by the incident would this statement not be true.

In the case of recurrent congestion, the pattern is different from the previous one. The simplest mental picture of this situation is a high-volume entrance ramp merging with a roadway that is already near capacity. The volume arriving at the bottleneck (i.e., the section of roadway immediately downstream of the entrance ramp) exceeds its capacity, causing traffic to queue upstream. Traffic operations upstream of the bottleneck site will be in State 3 (or perhaps briefly in State 2) identical to the incident congestion pattern. However, downstream of the merge point (the point where ramp traffic merges with mainline traffic), traffic flow will be at or close to capacity (State 4). Once again depending on the distance between the downstream detector and the merge point, vehicles may or may not be back to the desired speed. If they are, operations will be near the left edge of State 4. If they are not, the occupancies will be increased (for any given flow rate) by the reduced speeds, leading operations to be toward the right of State 4.

The logic depicted in Figure 3 is based on these two patterns. Beginning at the first detector station, i, and moving in the direction of flow, the traffic state at each detector station is evaluated on the basis of the volume and occupancy values. If the traffic state is 1, proceed to the next detector station, $i + 1$, and repeat the procedure. If the traffic state is either 2 or 3, traffic operations are congested. It is now

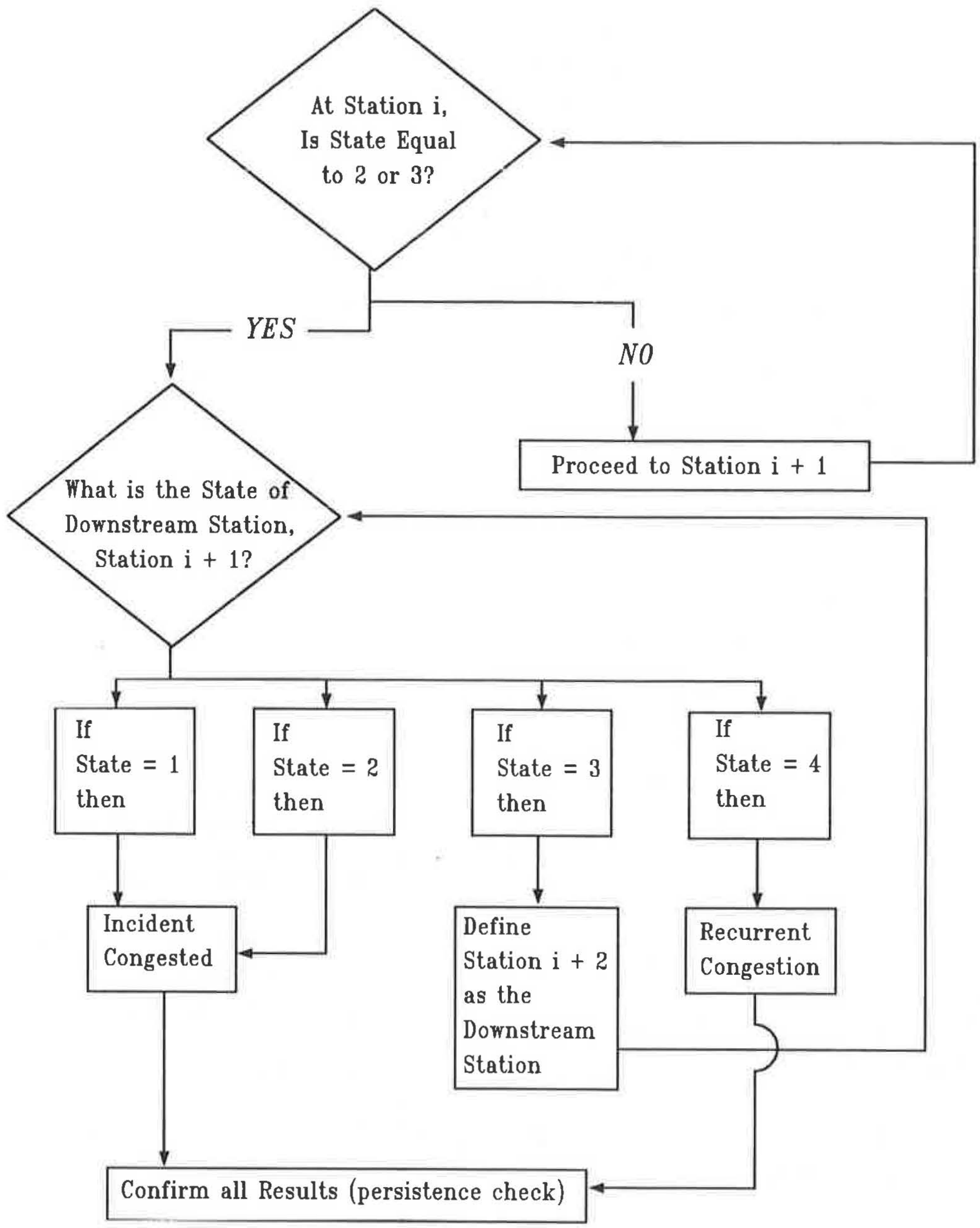

FIGURE 3 Flow chart for distinguishing between recurrent congestion and incident congestion.

necessary to evaluate the traffic state at the downstream detector station, $i + 1$. Three conditions at the downstream station are possible. If the downstream state is either 1 or 2, then it is highly likely that a capacity-reducing incident has occurred between Stations i and $i + 1$. If the downstream state is 4, then it is highly likely that the congestion is due to either an input of extra volume or a lane drop located between Stations i and $i + 1$. If the downstream state is 3, then the cause of the congestion is further downstream; proceed to Station $i + 2$. If the logic is used to complement a current incident-detection algorithm, then once congestion has been detected at a detector station, this logic would be used to evaluate only the downstream detector stations to identify the cause of the congestion.

DESCRIPTION OF THE STUDY SITE AND DATA BASE

The study site selected for a feasibility test of the proposed logic is a portion of the eastbound Queen Elizabeth Way (QEW) in Mississauga, Ontario (Figure 4). The prime reason for the selection of this portion of the QEW was its geometrics. The QEW is fairly flat, three lanes are maintained throughout the section, and the entrance ramps at Highway 10 and Cawthra Road cause recurrent congestion during the morning peak period. This portion of the QEW is approximately 3.2 km long and includes five detector stations. Each detector station is comprised of a pair of inductance loop detectors in each lane. Three traffic variables—average speed

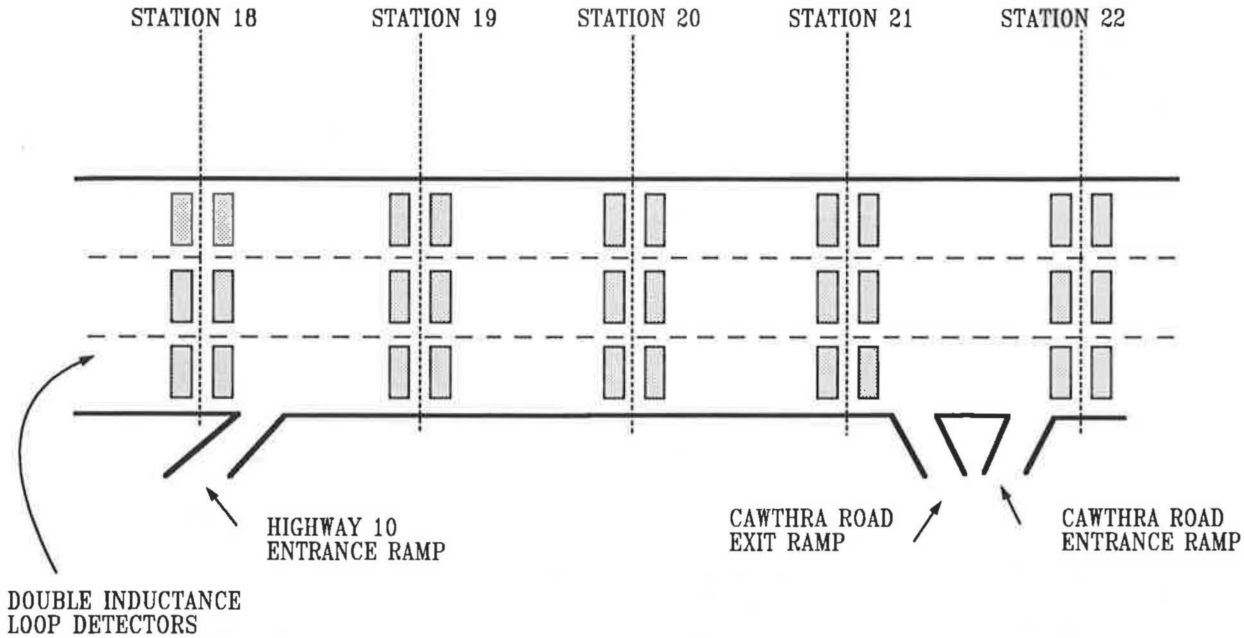

FIGURE 4 Schematic of study site—portion of eastbound Queen Elizabeth Way in Mississauga, Ontario.

(km/hr), volume (number of vehicles), and occupancy (percent)—are summarized at 30-sec intervals 24 hr a day at each of the three lanes.

Although the FTMS facility on the QEW measures freeway data daily, it does not regularly store these data. Before commencing the research for this paper, a few days of data had been stored on magnetic tapes for related research, and these form the basis for this test. The data base includes 2 incident-free days, September 30, 1987, and October 1, 1987, and 1 day, November 19, 1987, in which two incidents were recorded. Each data set contains 24 hr of 30-sec summaries of average speed, volume, and occupancy. Only the median lane data were used in the analysis, primarily because trucks are prohibited from using the median lane. Consequently, volume counts were used directly without conversion to passenger car units. Therefore this test depends strongly on the assumption that any incident or bottleneck will cause a queue in the median lane that grows at least as fast as queues in the other lanes. The use of only the median lane relies on the concept referred to as "lane sympathy." Dudek and Messer (*9*) found that

> although there is a degree of sympathy of speed between lanes regardless of volume, stoppage waves do not necessarily move in unison on each lane of a freeway . . . for the incidents studied, stoppage waves were first detected on either the median or the middle lanes, or both, in 98 percent of the cases.

In light of Dudek and Messer's observation, the assumption does not seem to be unreasonable.

CALIBRATION

To use the proposed logic, a volume-occupancy template (as in Figure 1) is required for each detector station in the study site. The calibration of the station-specific template involves the calibration of the function $f(\text{occ})$ and the parameters k, OCMAX, and Vcrit. As previously mentioned, $f(\text{occ})$ is a function that defines the line along which uncongested volume-occupancy data tend to lie. The parameter k is a value (between 0 and 1) that will produce the boundary line defined by $g(\text{occ})$. OCMAX is defined as the maximum occupancy for uncongested operations. Vcrit refers to the critical volume by which State 3 and State 4 are distinguished.

The function $g(\text{occ})$, which defines the minimum uncongested volume threshold in Figure 1, is defined as

$$g(\text{occ}) = k * f(\text{occ})$$

where $f(\text{occ}) = b * \text{occupancy}^a$

Depending on whether a multiplicative or additive error structure is assumed for the model, linear or nonlinear estimation techniques will be appropriate. The additive form was assumed, that is

$$f(\text{occ}) = b * (\text{occupancy})^a + e$$

This form of the model is intrinsically nonlinear in the parameters. Hence, the parameters, a and b, must be estimated using nonlinear techniques.

To estimate the parameters of $f(\text{occ})$, a sample of volume-occupancy pairs reflecting uncongested traffic operations is required. Because 2 days of incident-free data were available, it was decided to use these data in the estimation procedure (treating each day separately). The purpose of using the 2 days of data was to evaluate the sensitivity of the parameters. From each 24-hr period, 20 hr of data were used. The morning peak period from 6 a.m. to 10 a.m. was excluded because most of these data reflected congested traffic operations. The data were screened first for missing values (that is, any variables recorded as -1) and then screened to remove congested data. Congested data were defined as data in which occupancies exceeded a maximum uncongested occupancy threshold, OCMAX, and speeds were less than some minimum speed threshold. A visual inspection of volume-occupancy plots indicated that an OCMAX value of 25 percent was appropriate for all five detector stations. The threshold speed was different for each station, but generally about 65 km/hr. The values used for OCMAX and the minimum speed threshold are preliminary. More work will be done to derive station-specific values for these parameters.

The resulting uncongested data sets were then used as input to a nonlinear parameter estimation program. Each of the two uncongested data sets had between 1,500 and 1,800 volume-occupancy pairs. Generally, smaller data sets could be used, but larger data sets are preferred to reduce the variance of the residuals and thus improve the precision of the parameter estimates. As shown in Table 1, the two estimates for each parameter differed. These differences were tested using a standard statistical test with a confidence level of 95 percent, and it was found that generally the probability of the occurrence of this difference exceeded 5 percent, and therefore, could not be attributed to chance. This result suggests there are significant day-to-day changes in the traffic characteristics, implying that some updating technique for the parameters is required. In practice, estimates of the parameters could be obtained from available uncongested data, but the parameters should be updated on-line. In the feasibility test, no updating technique was employed; hence, the parameters of the function $f(\text{occ})$ were treated as fixed values. The parameter estimates used in the feasibility test were arbitrarily selected as those derived using Data Set 2.

The second part of the calibration process dealt with determining an appropriate value for k. Different values of k (where $0 < k < 1$) were tested such that the resulting line, $k * f(\text{occ})$ would be a lower bound for 95 percent of all volumes observed at a given occupancy value. Generally, a k value of 0.8 was found to be appropriate for all five detector stations. It appears that results are not particularly sensitive to k. This parameter will likely not need to be calibrated individually for each station.

In finding an appropriate value for OCMAX, the data from the two incident-free days were examined and volume-occupancy plots produced. Using these plots, a maximum occupancy value of 25 percent was set for OCMAX. At present, the parameter OCMAX is not station-specific. This may change as more work is done.

The focus of the final part of the calibration process was to determine an appropriate value for Vcrit. State 4 is applicable only to stations located immediately downstream of an entrance ramp; hence, the definition of Vcrit as a minimum discharge volume. The mean discharge volume corresponded to 19 vehicles/30 sec (2,280 vph), so the minimum discharge flow, Vcrit, was set at 16 vehicles/30 sec (1,920 vph) to provide a lower bound.

FEASIBILITY TEST OF PROPOSED LOGIC

The long-term objective for an evaluation of the logic is to compare the identification of the types of congestion made by the proposed logic with the FTMS operator's perception of the traffic conditions along the facility. This comparison could not be accomplished at present, however; but a feasibility test was conducted with the available data.

The best method for achieving the long-term objective is an on-line evaluation, but it was not possible to schedule such an evaluation at this time for several reasons: (a) the FTMS communications system is currently being upgraded; (b) traffic flow during the summer months is lighter than normal; (c) due to bridge construction upstream of the study site, traffic patterns and volumes have been altered; and (d) during the summer months, the FTMS facility is staffed only from 6 a.m. to 9 a.m. For these reasons, it was also impractical to collect more data at this time.

An off-line test was not possible because of insufficient data. The available data represent only 2 days of incident-free operation and 1 day with two recorded incidents. Further, for the incident data, only a minimal incident log is kept by the FTMS operators as part of the daily FTMS operations record. The operators record an incident only if the incident required a response. Hence, not all incidents are recorded. As part of the daily FTMS operations record, the operators identify the time congestion appeared and dissipated. From conversations with one operator, it was confirmed that this congestion period refers specifically to recurrent congestion that appears upstream of the Highway 10 bottleneck. The time period recorded is subjective as each operator may define "congestion" differently. Given the amount of stored data available and the limited information about these data, a full off-line evaluation was not possible. Therefore, only a pre-

TABLE 1 SUMMARY OF ESTIMATED PARAMETERS FOR THE FLOW MODEL

Data Set	Station 18	Station 19	Station 20	Station 21	Station 22
#1	a=0.8400 b=2.2410	a=0.8436 b=1.6921	a=0.7882 b=1.8360	a=0.8492 b=1.5950	a=0.7800 b=2.4170
#2	a=0.8350 b=2.5070	a=0.8325 b=1.6900	a=0.8344 b=1.6950	a=0.8108 b=1.7570	a=0.8155 b=2.2810

liminary off-line evaluation with the available data was performed, in the nature of a feasibility test.

The two objectives for this feasibility test are (a) to determine whether the logic can correctly identify the recurrent congestion that occurs daily, and (b) to see whether the logic can identify the incident congestion in the third day's data. The limited information provided by the operators was used to evaluate the logic. With respect to recurrent congestion, the logic was deemed successful if it could identify the operator-labeled congestion as recurrent. With respect to incident congestion, the logic was deemed successful if the incident congestion identified corresponded to a recorded incident. Two time periods from each data set were selected for evaluation, a morning peak period (6:00 a.m. to 10:00 a.m.) and an afternoon period (2:00 p.m. to 6:00 p.m.). A tentative persistence check (of three consecutive 30-sec intervals) was set to confirm identifications made using the logic.

The results (Table 2) show that the recurrent congestion portion of the logic was a clear success. For the 2 incident-free days, the congestion detected was identified as recurrent. Also, recurrent congestion was identified only at the two bottleneck locations and only during the morning peak period (6 a.m. to 10 a.m.). The time of day (not summarized in Table 2) was comparable to the time period in which recurrent congestion is known to be present along the study site.

The incident congestion portion of the logic did not meet with similar success. Although several short periods of incident congestion were identified, an average of one short period every 4 hr, these identifications did not correspond to any recorded incident. Two issues arise: the first is a possible explanation of why the recorded incidents did not cause any identifiable congestion, and the second deals with the incident congestion that was identified.

To understand why the recorded incidents were not found, we must examine the incident data more closely. With respect to the second recorded incident, its location is ambiguous and may have been upstream of the study area.

As indicated in Table 3, the first incident was very short. It seems likely that this short-duration incident did not impede traffic operations enough to cause congestion to reach the

TABLE 2 RESULTS OF PRELIMINARY EVALUATION

Data Set	Time Period	Recurrent Congestion Identifications	Incident Congestion Identifications
30 09 87	6:00 am to 10:00 am	At both bottlenecks	1 at Station 20 See Note 1
	2:00 pm to 6:00 pm	None	See Note 2
01 10 87	6:00 am to 10:00 am	At both bottlenecks	1 at Station 20 See Note 1 1 at Station 21 See Note 1
	2:00 pm to 10:00 am	None	See Note 2
19 11 87	6:00 am to 10:00 am	At both bottlenecks	1 at Station 19 See Note 3
	2:00 pm to 6:00 pm	None	1 at Station 20

Note 1: These identified incident congestion periods were of short duration and were declared after recurrent congestion was identified at both bottlenecks.

Note 2: Several periods of congestion were detected at Station 20, but due to missing data at Station 21, it was not possible to identify the type of congestion.

Note 3: Incident congestion was identified at Station 19 prior to the identification of recurrent congestion at the Highway 10 bottleneck. An incident was logged 4 minutes prior to the identification of incident congestion but the incident was recorded as occurring downstream of Station 19 at Cawthra Road.

TABLE 3 INCIDENT INFORMATION

Data Set	Start Time Logged	Location	End Time Logged
19 11 87	6:35 am	E/B left lane west of Cawthra	6:37 am
19 11 87	8:53 am	E/B left lane at Hwy. 10	9:05 am

detectors. Manual review of the data revealed that the pattern defining incident congestion was present, but did not persist beyond the persistence requirement, and was followed immediately by the pattern defining recurrent congestion.

Because the first recorded incident occurred at the onset of recurrent congestion, the data from the 2 incident-free days were also manually reviewed at the same location at approximately the same time. The incident-free data revealed a similar incident congestion pattern before recurrent congestion was firmly established. Thus, the result of the first manual check is not as positive as it might seem.

As previously mentioned, the incident congestion identifications did not correspond to any recorded incident. As summarized in Table 2, incident congestion was identified at Station 20 and Station 21 during the morning peak period of both incident-free days. These stations are located upstream of the Cawthra Road bottleneck. The queue from this bottleneck can, and often does, extend beyond Station 20. The incident congestion identified here may have been produced by the stop-and-go nature of operations within a queue. During the second time period (2 p.m. to 6 p.m.), it was not possible to identify the cause of congestion identified at Station 20 due to missing data at Station 21. Recorded incidents are incidents that require a response. Consequently, these recorded incidents do not constitute all the incidents that may have occurred at the study site during the 3 days. It is possible that the incident congestion identifications corresponded in fact to incidents that required no response.

The results from the feasibility test are positive but far from complete. The long-term objective can best be realized through an extensive on-line evaluation. Such an evaluation was not possible but will be performed in the near future.

CONCLUSIONS

The proposed logic, as described in the paper, makes it possible to distinguish between congestion due to an incident and congestion due to a bottleneck. The core of the logic is extremely simple but, we believe, quite accurate. The simplicity of the logic makes it feasible to implement.

The empirical results are promising, but, as a result of lack of field validation and limited data, are not conclusive. A more rigorous test of the proposed logic is needed and is planned.

It is important to note that although this logic uses data from adjacent stations to establish the cause of congestion, the logic is not similar to the standard comparative algorithms. Instead, this algorithm relies on the data from a single station to identify the state of traffic there and looks downstream of a congested situation to find the cause.

ACKNOWLEDGMENT

Most of the research for this paper was performed under a contract with the Ontario Ministry of Transportation. The authors wish to acknowledge this financial support as well as the assistance of Colin Rayman, Mike Young, Hoi Wong, Emmanuel Morala and others in the Freeway Traffic Management Section in making the freeway traffic data available. The authors also acknowledge the support of the Natural Sciences and Engineering Research Council of Canada.

REFERENCES

1. J. T. West. California Makes Its Move. *Traffic Engineering*, Vol. 41, No. 4, 1971, pp. 12–18.
2. B. N. Persaud and F. L. Hall. Catastrophe Theory and Patterns in 30-Second Traffic Data—Implications for Incident Detection. In *Transportation Research*, Vol. 23A, 1989, pp. 103–113.
3. J. A. Wattleworth and D. S. Berry. Peak Period Control of a Freeway System—Some Theoretical Investigations. In *Highway Research Record 89*, HRB, National Research Council, Washington, D.C., 1965, pp. 1–25.
4. M. Levin and G. M. Krause. Incident Detection Algorithms, Part 1. Off-Line Evaluation. In *Transportation Research Record 722*, TRB, National Research Council, Washington, D.C., 1979, pp. 49–58.
5. M. Levin, G. M. Krause, and J. A. Budrick. Incident Detection Algorithms, Part 2. On-Line Evaluation. In *Transportation Research Record 722*, TRB, National Research Council, Washington, D.C., 1979, pp. 58–64.
6. P. Athol. Interdependence of Certain Operational Characteristics within a Moving Traffic Stream. In *Highway Research Record 72*, HRB, National Research Council, Washington, D.C., 1965, pp. 58–87.
7. B. N. Persaud. *Study of Freeway Bottlenecks to Explore Some Unresolved Traffic Flow Issues*. Ph.D. dissertation, University of Toronto, Ontario, Canada, 1986.
8. B. N. Persaud and V. F. Hurdle. Some New Data That Challenge Some Old Ideas about Speed-Flow Relationships. In *Transportation Research Record 1194*, TRB, National Research Council, Washington, D.C., 1988, pp. 191–198.
9. C. L. Dudek and C. J. Messer. Detecting Stoppage Waves for Freeway Control. In *Highway Research Record 469*, HRB, National Research Council, Washington, D.C., 1973, pp. 1–15.

Publication of this paper sponsored by Committee on Freeway Operations.