Integrating Microscopic Simulation and Optimization: Application to Freeway Work Zone Traffic Control

RAGAB M. MOUSA, NAGUI M. ROUPHAIL, AND FARHAD AZADIVAR

This paper presents a methodology for optimizing performance of a traffic system on the basis of simulated observations of its microscopic behavior. The method integrates simulation and optimization submodels for describing traffic flow on urban freeway lane closures. The stochastic nature of traffic is accounted for in determining the true system response to traffic control variables. The simulation submodel has been validated at a series of work sites in the Chicago area expressway system. The optimization submodel optimizes a single objective function subject to a set of linear constraints. Preliminary model applications included the determination of an optimum merging strategy to be adopted by traffic entering the work zone in lanes to be closed for traffic. The model recommendation yielded the lowest average travel time in the work zone and, interestingly, did not incorporate many early merges; the latter is often viewed as a desired merging strategy. In addition, the optimum merging strategy varied with the traffic flow level entering the work zone and with the character of the objective function to be optimized.

Control solutions for complex traffic systems often require an explicit optimization of one or more system performance measures. Traffic signal system parameters such as cycle length, splits, and offsets are typically determined from an optimization of delays, queue lengths, etc. On urban freeways, the specification of ramp metering rates, priority lanes, and priority entry is formulated to optimize overall corridor performance using the FREQ model (*1*). The system of interest in this study concerns lane closure procedures on urban freeways. Several key decisions must be made for such a system to operate in a safe and efficient manner, such as the location of advance warning devices, the length and position of the construction taper, the distance between tapers for multilane closures, the location and layout of ramp tapers, etc. Although empirical guidelines exist for these parameters, as listed in the *Manual on Uniform Traffic Control Devices for Streets and Highways* (MUTCD) (*2*), they have yet to be evaluated in the context of overall optimal system performance.

A key hindrance toward integrating traffic simulation and optimization models lies in the nature of the traffic process. Traffic flow descriptors are generally stochastic in nature, ranging from headways and speeds to critical gaps and lane selection. This behavior complicates the evaluation of the system performance and consequently the process of reaching an optimal solution. Traffic models that represent the traffic system in terms of its microscopic components (individual drivers, vehicles, etc.), will usually require a large programming and debugging effort, exhibit more stringent storage requirements, and consume more computing time, while providing greater resolution and potentially more accuracy relative to alternative comparison (*3*). Although there is some consensus among traffic analysts as to the value of microscopic traffic simulation models in mimicking traffic behavior, their utility has been curtailed as a result of their inability to formally optimize system performance (short of trial-and-error procedures, which still would not guarantee an optimal solution). Examples of such evaluation models include NETSIM (*4*) for signalized networks, INTRAS/FREESIM (*3*) for freeway corridors, and FREECON (*5*) and ARTWORK (*6*) for freeway and arterial lane closures, respectively.

It is interesting to examine some of the popular design models that do contain some mechanism for optimizing system performance. Signal timing models such as SOAP (*7*) for isolated intersections, Passer II (*8*) or MAXBAND (*9*) for arterials, and TRANSYT (*10*) or SIGOP (*11*) for networks are all deterministic, macroscopic models in which traffic is represented in terms of average flow rates that occur consistently (or at least in a predetermined pattern) throughout the simulation exercise. It has been common practice among traffic engineers to use design models such as TRANSYT to optimize network signal settings and subsequently evaluate those settings in a more realistic traffic environment as in the NETSIM model.

This gap in current modeling capability has provided the impetus for the research effort described herein. The selection of the freeway lane closure traffic system was in part due to the authors' interest and experience in the topic, and also to fulfill the requirements of a research grant. In addition, the system provided a unique environment for testing some of the well-established traffic control procedures in the freeway work zone area.

RESEARCH APPROACH

The approach employed in this study consisted of the development of an integrated microscopic simulation and optimization model for urban freeway lane closures. A review of existing microscopic models indicated that (a) none of the models contained a formal optimization feature and (b) freeway work

R. M. Mousa, Civil & Environmental Engineering Department, California Polytechnic State University, St. Luis Obispo, Calif. 93407. N. M. Rouphail, CEMM Department and Urban Transportation Center, The University of Illinois, Chicago, Ill. 60680. F. Azadivar, Department of Industrial Engineering, Northern Illinois University, DeKalb, Ill. 60115-2854.

zone models did not account for the presence of ramps in modeling traffic flow. It was also important to limit the system size to maintain a viable optimization scheme that would require multiple simulation runs. Thus, the concept of using a large-scale system such as INTRAS/FREESIM, although noble in purpose, was beyond the computational limits and scope of a first-cut investigative study of this nature.

This paper presents the results of the integrated model development. The optimization submodel borrows an algorithm developed by Azadivar and Talavage (*12*), which is specifically designed to optimize stochastic systems. The simulation submodel is capable of representing an urban freeway environment with or without lane closures or ramps. The model was successfully interfaced with the optimization procedure to yield the preferred traffic control solutions. It must be stated that the optimization submodel is structurally independent of the simulation process, requiring only the specification of and periodic updating to observations of the system objective function. Thus, its interface with other type of models can be accomplished in a straightforward manner.

The remainder of the paper includes a brief description of the two submodels, validation studies of the simulation submodel, a definition and demonstration of the integrated model applications, and conclusions of the major findings along with considerations for potential model applications.

MODEL DESCRIPTION

Simulation Submodel

The simulation model developed in this study is coded in FORTRAN and uses the powerful capabilities of the SLAM II simulation language developed by Pritsker (*13*). The simulated freeway segment is represented in the simulation model by a system of finite length and width. The first point on the segment corresponds to the system entry point; the end of the freeway segment is represented by the system exit point. The model is designed to handle freeway sections of up to four lanes in each direction, including an entrance and exit ramp. The overall logic of the simulation model is shown in the flowchart in Figure 1. In this paper, only model attributes that differ from previous models are discussed.

Lane Assignment at Entry Point

Vehicles are generated (at the system entry point) from one stream of traffic and subsequently assigned a lane according to one of two criteria: (a) the lane that allows the highest entry speed or (b) a prespecified lane distribution (in which lane distribution must be specified by the user) at the system entry point. The results presented in this paper are based on the first criterion of lane assignment.

Desired Speed

A different desired speed definition was introduced that assumes the driver's desired speed to be somewhat dependent on traffic conditions, as opposed to the commonly used free flow speed.

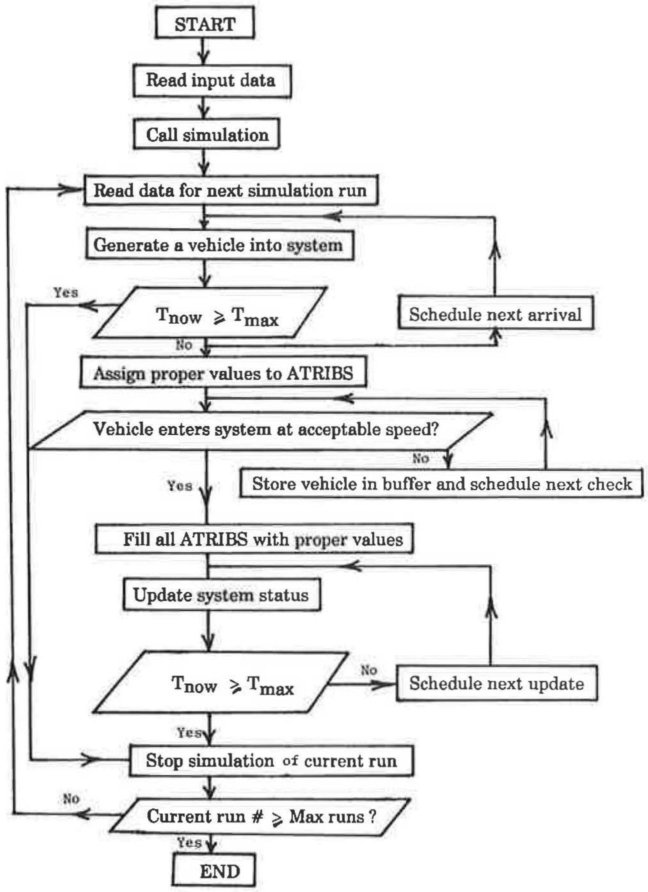

FIGURE 1 Overall simulation model logic.

The use of this new definition in the simulation has many advantages in terms of modeling convenience and quality of simulation results.

Vehicle Updating

The updating procedure was accomplished in a unique manner. Rather than updating all vehicles in the system at periodic intervals, only one vehicle is updated at a time at interval Δt (depending on the individual driver's reaction time τ). The main objective was to reduce the execution time to render the model feasible for optimization. While a vehicle's attributes are being updated, vehicles in the vicinity of that vehicle are accounted for. The model guarantees that all vehicles in the system are updated within 0.5 sec, which is very reasonable compared with 1.0-sec fixed-step sizes used in previous studies (*5, 14*) and to the 0.1 sec recommended in another study (*15*). The use of large updating step sizes without affecting the behavior of the model was achieved by using the proper car-following model. In this updating routine, all statistics are collected at both ends of the system as well as at different stations along the freeway segment.

Car-Following Model

The literature abounds with car-following models designed to predict the behavior of drivers in platoons (*3, 16, 17*). A review of the existing models revealed many drawbacks to their applications for work zone analysis. A new car-following model was developed in a manner suggested in the work by Gipps (*16*) with some modifications to make it amenable for use at lane closures. For example, in the original model Gipps assumes that 50 percent of the driver's reaction time is used as a safety margin to account for possible delays by the following driver in reacting to the lead driver's actions (hereafter termed θ). Thus, the distance $0.5\tau V$ is taken as the safety margin. In this study, the parameter θ is considered to vary depending on the traffic volume. A minimum distance was set in the model to guarantee adequate safety margins at the low speeds that are often experienced at construction zones. Unlike other car-following models, the outcome from this model is the maximum speed that can be achieved by the following driver after one reaction time, τ. Thus, an estimate of the required uniform acceleration or deceleration rate that can be applied by the following driver over the next time period τ is readily determined.

Driver's Critical Gap

The driver's gap is very critical in making lane-changing decisions and has been the subject of many studies (*18–20*). The driver's critical gap is stochastically assigned according to a probability distribution function that incorporates the following features:

- Drivers with higher desired speeds accept shorter gaps. This feature was previously used in a study by Rathi (*21*) of freeway lane closures.
- Drivers in closed lanes accept shorter gaps as they move toward the taper. This concept of nonstationary gap acceptance was originally developed by Abella (*22*).
- Critical gap for drivers in closed lanes decreases in value with an increase in gap searching time. This concept was reported in studies (*23, 24*) of left-turn traffic at intersections. In these studies, the driver's critical gap decreased with each rejected gap (i.e., waiting time for the adequate gap).

Lane Change Procedure

The lane change procedure is one of the essential components of the simulation model. Many variables and features have been included in the model under this element to mimic real-life lane change behavior. Two types of lane changes have been used in the model as defined in the INTRAS (*3*) model—nonessential and essential lane changes. Nonessential merges occur from one lane to another at any time and location over the highway segment. Some constraints, however, have been set to limit these merges to reasonable frequencies. On the other hand, essential merges are those to be made from closed lanes, and, as initiated in the model, the freeway section upstream of the construction taper is divided into N segments, each of length L_i (see Figure 2). The last segment ends at the beginning of the first taper. As mentioned earlier, traffic upstream of the first advance warning sign follows a prespecified lane distribution model. Let P_i denote the ratio of Segment i closed-lane traffic (that has not attempted the merge) to the total closed-lane traffic entering Segment 1. Thus, P_i is always 1, whereas P_{N+1} is 0. On the basis of this definition, these P's are subjected to a set of constraints in which $P_i \geq P_{i+1}$. For the given ratios of P_i and P_{i+1} at the beginning and the end of Segment i, the percentage of closed-lane traffic that would initiate the merge over that segment can be determined. For instance, if $P_2 = 0.80$ and $P_1 = 0.55$, then 20 percent and 25 percent of all closed-lane traffic (entering Segment 1) would initiate the lane change on Segments 1 and 2, respectively. It should be emphasized that these P_i's can be either input or, as discussed later, optimized on the basis of a given system measure of performance. In case of multiple lane closures, the same procedure has been applied with the exception that drivers in the outer closed lane attempt to perform the second merge immediately after they complete their first merge in a closed lane.

Having initiated the merge, the driver starts an attempt to move to an adjacent lane. The model first identifies the adjacent lane into which the driver is attempting to merge. Lane identification is based on the type of lane changing and the characteristics of gaps in adjacent streams. The model also identifies the effective gap in the target lane into which the driver will be merging. This gap is bounded by two vehicles: effective lead and lag vehicles. The effective lead or lag vehicle need not currently be in the target lane; therefore, the effective gap could be less than the apparent gap (see Figure 3).

Drivers involved in essential merges are assumed to be more alert and therefore can accept shorter gaps. Thus, in the merging process, the reaction time might be shorter than the normal reaction time under routine driving conditions. However, the model checks first to see if the gap is acceptable with no temporary reduction in reaction time for both merging and lag drivers. If the test fails, a second trial is made to see whether the driver will accept the gap with some temporary reduction in reaction time. This reduction may explain the shorter gaps observed in the field that would theoretically be unsafe under normal driving conditions. Interestingly, this concept was implemented indirectly in the INTRAS model

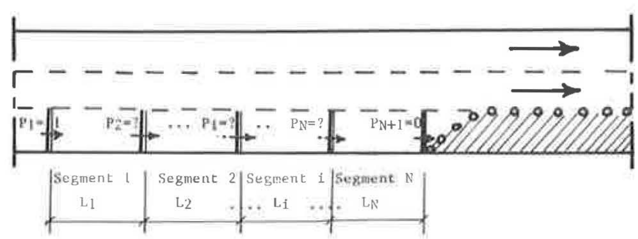

FIGURE 2 Segment definition for development of merge strategy.

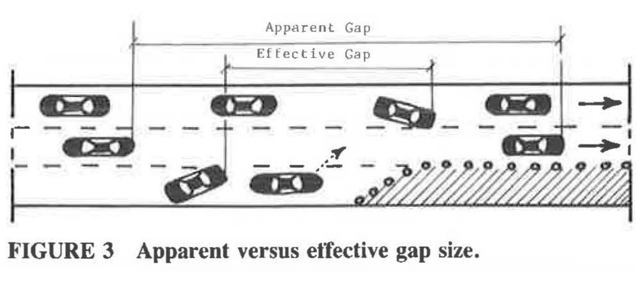

FIGURE 3 Apparent versus effective gap size.

(*3*) by allowing a temporary unsafe position of the merging vehicle during the finite period of the lane change. This allowance in INTRAS was set to enable the representation of forced lane changing, with a vehicle crowding into what might normally be considered an unavailable gap. The assumption of alertness for essential merges was also applied on a limited basis to some nonessential merges in the open lanes if the actual speed falls below a certain percentage of the driver's desired speed. While repeating this procedure, the drivers consider further actions, especially when the check fails to produce an acceptable gap in several trials. In this case the following options are pursued in the given order:

1. Accelerate if vehicle is not in the proper position to use the adjacent gap,
2. Decelerate if vehicle is not in the proper position to use the adjacent gap,
3. Accelerate to use the gap ahead of the adjacent gap, or
4. Decelerate to use the gap behind the adjacent gap.

Optimization Submodel

The optimization algorithm (SAMOPT) used in this study was developed for optimizing the response function of simulation models with stochastic behavior. Because of the stochastic nature of the simulation systems, the result of each evaluation by simulation is only a noisy observation of the true response. The algorithm uses these noisy responses to select values of the decision variables of the system such that the true response is optimized. Principles of stochastic approximation techniques have been used in developing this algorithm, which guarantees convergence to the optimum if a large number of observations are made. Even with limited sample sizes, the algorithm will yield reasonably accurate answers. The algorithm was further enhanced to incorporate decision variables that are subject to a set of linear constraints (*12*, *25*). Additional details about the source code and the validation of the algorithm can be found elsewhere (*25*).

In summary, the advantages of the optimization procedure are as follows:

- It can be interfaced with objective functions and decision variables obtained from microscopic simulation models;
- It can handle up to 10 decision variables (even more with some manipulations in the array dimensions);
- It accepts linear constraints on the values of decision variables; and
- It guarantees an optimum solution within a finite number of simulation runs; with a limited number of runs, the procedure still yields reasonable solutions.

On the other hand, working experience with the algorithm revealed some drawbacks:

- The procedure uses the number of simulation runs between $2 \cdot (2^n + 1)$ and $(N/2)$, where n is the number of the decision variables to be optimized and N is the maximum number of simulation runs available, to locate an initial point before applying the stochastic approximation technique. As many as half of all simulation runs are expended in finding that initial point. The procedure does not have the flexibility to use an initial point input by the user.
- Although economical in its overall use of simulation runs, the procedure performs only two simulation runs at each tested point in the initial runs, which may not be sufficient to distinguish noise from trend.
- Tolerance levels for the decision variables (which represent the convergence criteria) are dimensionless and must be small in magnitude regardless of the dimension of the decision variables.

Interfacing the Simulation and Optimization Submodels

Interfacing the submodels requires that one of the two models serves as a subroutine for the other. For example, if the optimization routine SAMOPT is to be used with N simulation runs, it first determines the values of the decision variables at the current run and calls the simulation submodel to pass the objective function on the basis of these decision variables. It then determines the decision variables to be used in the next simulation run, and so on. On the other hand, if the simulation submodel is to call SAMOPT, then it must retrieve from SAMOPT, on the basis of the current objective function, the values of the decision variables to be used in the next run, and so forth. The latter strategy was applied in this study. The interfacing logic of both submodels is shown in the flowchart in Figure 4.

In addition to the advantage of running both submodels interactively, some enhancements were performed on the optimization algorithm:

- An initial point can now be input by the user. This feature saves significant computer time, especially when the user has good information about a feasible initial point.
- In conjunction with the identification of an initial point, several system performance measures can be stored and used subsequently for optimization purposes.
- The optimization process usually requires a large number of simulation runs (typically about 80 runs for this application). With the enhancements, these runs may be divided into several small jobs rather than done in one job. Although this feature will not save CPU time, it allows performing the optimization at the user's convenience. This division is sometimes essential when there is a limitation on the maximum CPU time per job or in ranking jobs for execution.
- To economize on computer resources, a variable simulation time was employed. Longer simulations are performed in the process of determining the initial points as well as in analyzing the optimal solution.

MODEL VALIDATION

Because the optimization submodel was validated previously (*25*), this task is limited to the simulation submodel. Field data were collected from several freeway sites in the Chicago metropolitan area for validation purposes.

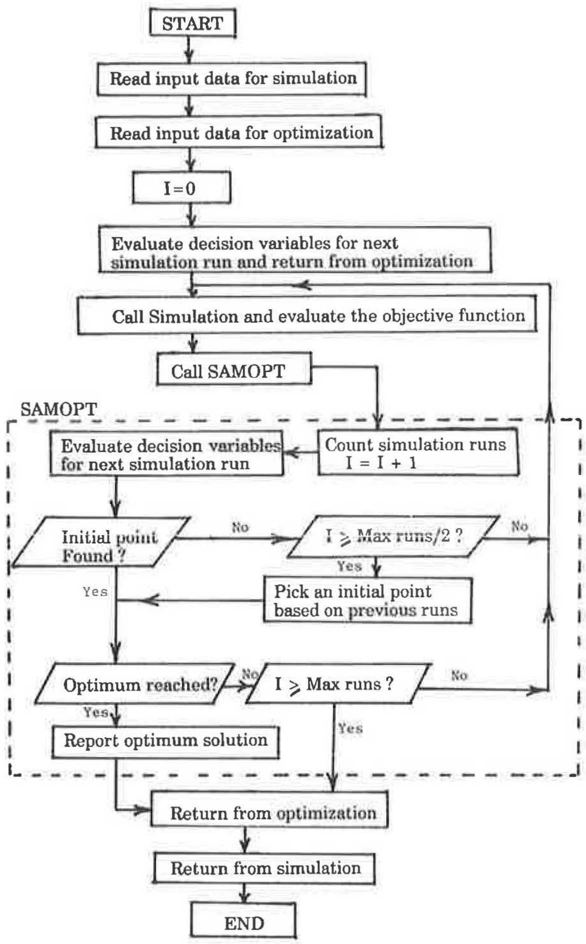

FIGURE 4 Interfacing of simulation model with the optimization routine.

Data Collection

Field Data Collection

Data were collected using two sets of video records. Each set consisted of a video camera with time feature and a portable VCR. The data incorporated most of the lane closure configurations (left versus right closure, single versus multiple closures). Filming was concentrated (when possible) at five key locations: (a) entry point (farthest point upstream of the taper where traffic is free flowing), (b) upstream of the taper (about 500 ft), (c) at taper, (d) downstream of the taper (about 500 ft), and (e) at exit point. Data collected included speeds, headways, and lane distribution of traffic at the key locations.

Data Reduction

Data were reduced in the laboratory using standard video reduction techniques. To ensure the synchronization of the data collected at both stations, individual vehicles were traced in both films; in most cases, it was possible to identify the same traffic at different locations. Headways and speeds were best fitted by lognormal and normal distributions, respectively. Moreover, a freeway lane at the bottleneck section carried up to 1,900 to 2,000 veh/hr. Furthermore, no significant reduction in speed due to construction was found when the demand volume was less than 1,700 to 1,800 veh/hr per lane in the bottleneck section. At higher rates, system breakdown (characterized by large speed variations and frequent stops) occurred in the taper area.

Validation Scenarios

Basic Segment

The model was first validated as a basic freeway segment. The parameters tested under this case were the car-following parameter θ and desired speed V_d. Several simulation runs were made for different combinations of V_d, θ, and flow levels. Lane capacity and speed obtained from each simulation experiment were compared with the corresponding values in the 1985 Highway Capacity Manual (HCM) (*26*) and with field observations. The calibrated car-following model and desired speed parameters (as a function of the traffic volume) are presented in Table 1. As indicated, the safety factor is inversely proportional to the flow rate. This relation may be explained by the fact that drivers are more sensitive to the gap length, expressed in terms of distance rather than time. Thus, if a safety distance is to be considered under all conditions, this distance (expressed in time) is shorter under low-volume conditions (because of higher speeds) than under heavy-volume conditions.

Space mean speeds from the calibrated model in comparison with field data as well as with the 1985 HCM values are shown in Figure 5. There was no significant drop in speed when the traffic volume was less than 1,600 veh/hr per lane; beyond this level, average speeds decreased substantially. Figure 5 also shows that the simulation model behaves consistently compared with the field data and the 1985 HCM. Hence, it was concluded that the model is a valid representation of basic freeway segments and ready for further testing with lane closures.

Lane Closures

A significant number of parameters were introduced into the model to account for the behavior of individual drivers at construction zones. The model parameters were first verified through an extensive sensitivity analysis on the basis of 90 simulation runs. The sensitivity analysis was performed using the more economical strategy 2^k factorial design (*27*). Subsequently, all model parameters were fixed except for the vector P_i, which was either provided as input or determined through optimization procedures, as discussed later in this paper.

For illustration purposes, the model results are compared herein with observations taken at two-lane closure sites in the Chicago area. A sample size of 300 observations was collected from the simulations for each site. Statistical testing consisted of a series of *t*-tests on the difference in mean speeds and headways (observed versus simulated) and the Kolmogorov-Smirnov test (*15*) for the lane distribution of traffic.

TABLE 1 CALIBRATION OF MODEL PARAMETERS WITH NO LANE CLOSURES

Parameter	Flow level (vphpl)								
	≤600	800	1000	1200	1400	1600	1800	1900	≥2000
θ (%)	10	10	10	10	10	10	10	20	40
V_d (mph)	60	58	57	56	55	53	48	43	35

θ Car-following model parameter.

V_d Driver's desired speed.

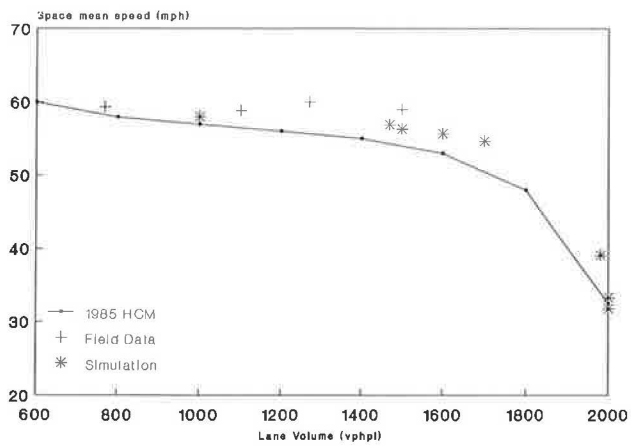

FIGURE 5 Space mean speed versus lane volume (without closures).

Figure 6a shows the layout of the first site on the Edens Expressway where construction took place on the right lane of the three-lane segment. Field data collected at that site and corresponding results from the simulation model are presented in Tables 2 and 3, respectively. Because information about the pattern of merge attempts or actual merges from the closed lane was difficult to collect in the field, the results from simulation runs were based on an experimental merge pattern that best fit the field observations.

The results in Tables 2 and 3 show an excellent statistical agreement between field and simulation data in terms of speed, flow, and proportion of flow in each lane. For instance, the average observed and simulated exit speeds were 58.5 and 58 mph, respectively. Statistically, these observations were not significantly different at the 5 percent significance level.

The second site was located on the I-88 Expressway where construction took place on the right lane of a three-lane freeway segment as shown in Figure 6b. Field data are presented in Table 4 and the corresponding simulation results are summarized in Table 5. The simulation model was again able to reproduce results that were quite comparable with its field counterparts.

The statistical agreement between field and simulation data consistently observed in most of the sites studied (including the two sites covered here) indicates that the microscopic simulation model is a valid representation of the lane closure traffic system.

MODEL APPLICATIONS

The microscopic simulation model developed in the course of this study has two significant applications:

1. Evaluation of existing traffic systems—to be done by providing the model with the proper input parameters as observed in the field and comparing field performance with optimum performance from the model.
2. Design of traffic systems—to be done by interfacing the simulation model with the optimization algorithm SAMOPT, to optimize the traffic system performance.

In this paper, the model is focused on the latter application. In a freeway lane closure system, the distribution of merges along the zone has a profound impact on system performance. A commonly held view among traffic engineers is that early merges are beneficial to traffic operations at work zones, because few vehicles are likely to be stranded at the taper. However, this strategy may result in evacuating the closed lane further upstream than is actually warranted, thus adding to the congestion on the through lanes. Obviously, the major drawback of a late merging strategy is the likelihood of conflicts occurring at the taper due to the high speed difference between vehicles in the through lanes and those attempting to merge.

Without a formal assessment of what constitutes an optimum merging strategy under various flow levels, it is difficult to justify a static placement of advance warning signs that are

TABLE 2 FLOW CHARACTERISTICS AT SITE 1 (EDENS EXPRESSWAY AT LAKE STREET, THREE LANES, RIGHT LANE CLOSED)

Station	Item	Lane[a] 1	Lane[a] 2	Lane[a] 3	Total/ Average	95% Conf. Interval
STATION I						
	Traffic volume (vph)	1224	1464	732	3420[b]	
	Speed (mph)	63.3	60.9	57.2	61.0[c]	60.2-61.7
	Lane distribution (%)	0.36	0.43	0.21		
	Cumulative lane dist.	0.36	0.79	1.00		
STATION II						
	Traffic volume (vph)	1800	1644	192	3636[d]	
	Speed (mph)	60.6	56.4	56.4	58.5	57.7-59.3
	Lane distribution (%)	0.49	0.45	0.05		
	Cumulative lane dist.	0.50	0.95	1.00		

[a] Lanes numbered from median to shoulder.

[b] Total flow rate on the freeway segment at station.

[c] Average speed on the freeway segment at station.

[d] Flow rate is higher than at station I due to ramp traffic.

TABLE 3 SIMULATED CHARACTERISTICS AT SITE 1 (EDENS EXPRESSWAY AT LAKE STREET, THREE LANES, RIGHT LANE CLOSED)

Station	Item	Lane 1	Lane 2	Lane 3	Total/ Average	95% Conf. Interval
STATION I						
	Traffic volume (vph)	1157	1432	853	3442	
	Speed (mph)	61.2	60.6	59.6	61.0	60.5-61.5
	Lane distribution (%)	0.34	0.42	0.25		
	Cumulative lane dist.	0.34	0.75	1.00		
	K-S difference[a]	0.02	0.03[b]	0.00		
STATION II						
	Traffic volume (vph)	1647	1801	215	3663	
	Speed (mph)	59.0	56.2	55.6	58.0	57.2-58.8
	Lane distribution (%)	0.45	0.49	0.06		
	Cumulative lane dist.	0.45	0.94	1.00		
	K-S difference	0.05[b]	0.01	0.00		

[a] Absolute difference between cumulative lane distribution (field vs. simulation).

[b] Not significant at the 5% level (max $K\text{-}S_{5\%,\,300} = 0.08$).

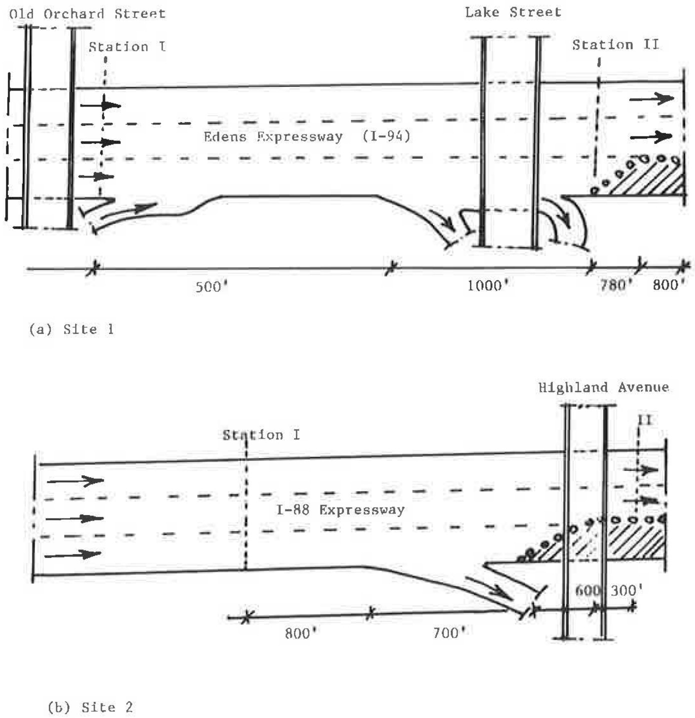

FIGURE 6 Layout of construction sites.

TABLE 4 FLOW CHARACTERISTICS AT SITE 2 (I-88 EXPRESSWAY AT HIGHLAND AVENUE, THREE LANES, RIGHT LANE CLOSED)

Station	Item	Lane 1	Lane 2	Lane 3	Total/ Average	95% Conf. Interval
STATION I						
	Traffic volume (vph)	1500	1309	957	3766	
	Speed (mph)	62.8	61.3	59.6	61.5	60.6-62.3
	Lane distribution (%)	0.40	0.35	0.25		
	Cumulative lane dist.	0.40	0.75	1.00		
STATION II						
	Traffic volume (vph)	1614	1353	-[a]	2967[b]	
	Speed (mph)	62.8	59.4	-	58.5	57.7-59.3
	Lane distribution (%)	0.54	0.46	-		
	Cumulative lane dist.	0.54	1.00	-		

[a] Lane closed in that section.

[b] Total flow rate lower than station I due to the exit ramp.

TABLE 5 SIMULATED CHARACTERISTICS AT SITE 2 (I-88 EXPRESSWAY AT HIGHLAND AVENUE, THREE LANES, RIGHT LANE CLOSED)

Station	Item	Lane 1	Lane 2	Lane 3	Total/ Average	95% Conf. Interval
STATION I						
	Traffic volume (vph)	1317	1615	816	3748	
	Speed (mph)	60.1	59.1	64.6	60.4	59.2-61.7
	Lane distributiôn (%)	0.35	0.43	0.22		
	Cumulative lane dist.	0.35	0.78	1.00		
	K-S difference	0.05[a]	0.04	0.00		
STATION II						
	Traffic volume (vph)	1507	1487	-[b]	2994	
	Speed (mph)	61.6	57.6	-	59.5	58.0-61.1
	Lane distribution (%)	0.50	0.50	-		
	Cumulative lane dist.	0.50	1.00	-		
	K-S difference	0.04[a]	0.00	-		

[a] Not significant at the 5% level.

[b] Lane closed in that section.

aimed at promoting an optimum strategy. This factor was considered to be the key application of the integrated model. In this paper, the results are reported for one performance measure, travel time in the work zone, with the qualification that other measures may yield entirely different optimum merging strategies. An added benefit of this exercise is the ability to use the model for evaluating field merging patterns against their derived optimal. Corrections can then be implemented in the field (i.e., in sign placement, taper length, etc.) aimed at bringing the observed merging pattern closer to its optimal.

In the simulation model, a merge pattern is established by randomly assigning each driver in the closed lanes a desired merge segment. When a driver enters a desired segment, the model automatically schedules a lane change attempt. Thus, the merge pattern is actually representative of merge attempts rather than merge executions. In any case, the model will output both distributions at the end of the simulation and optimization run. In the optimization mode, the model is essentially queried for the average percentage of drivers that should be assigned to the individual segments.

RESULTS

Before the optimization process was carried out, sensitivity runs were performed to determine whether merging strategies had a significant effect on the proposed system measure of performance, travel time. Two distinct patterns of merge initiations were studied at different flow levels. The early merge pattern assumes that all closed-lane traffic attempts to evacuate the lane within 0.7 mi from the taper. On the other hand, a late merge pattern assumes that all closed-lane traffic begins evacuating the closed lanes within 0.3 mi from the taper.

Results from this analysis are shown in Figure 7. The early merge produced results that were significantly superior to those under the late pattern at all flow levels. Furthermore, there appeared to be a range of flow over which the differences between the two merge strategies are more pronounced. This range varies from 1,500 to 2,000 veh/hr per lane. Below this range, there was no significant difference between the effect of the two merge strategies on travel time. Beyond that range, there was a breakdown in system performance with the late merge strategy, because of which a majority of closed lane traffic could not merge within the simulated time; thus the results were primarily representative of traffic in open lanes. It should be emphasized, however, that the results shown in Figure 7 do not imply that intermediate merge patterns should be expected to yield intermediate travel time values between the two extremes.

On the basis of this information, the optimization was performed within the flow range specified. Two flow rates were selected, 1,700 and 1,850 veh/hr per lane, respectively. The input and output of the optimization procedure at these two flow levels are discussed in some detail in the next section.

The integrated model was applied to determine the minimum travel time for a freeway work zone segment (3 lanes

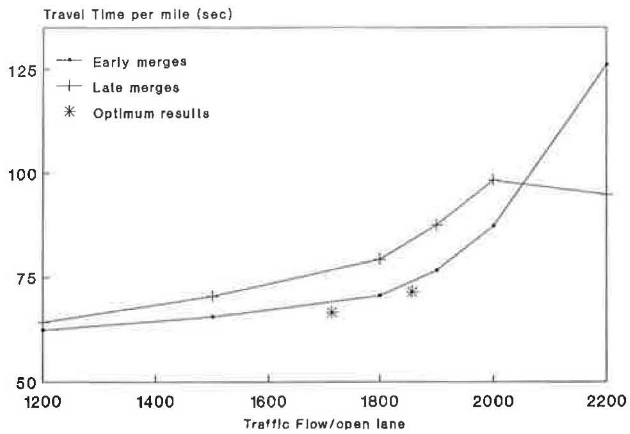

FIGURE 7 Travel time versus traffic volume.

and right lane closed) by determining the optimum merge pattern from the closed lane. In all runs (due to CPU limitation), the system was simulated for 15 min including a warm-up period of 5 min. The freeway segment upstream of the construction taper was divided into four equal segments, each ¼ mi in length. Thus, three decision variables were to be optimized (see Table 6 and Figure 2), namely the proportion of closed-lane traffic entering Segments 2, 3, and 4 to that entering Segment 1. Therefore, three decision variables and two inequality constraints were input to the model. A lower bound of 0, an upper bound of 1.0, and a tolerance of 0.025 were input for each variable. The two distinct constraints were $P_2 \geq P_3$ and $P_3 \geq P_4$. The typical maximum number of simulation and optimization runs required for flow rates of 1,700 and 1,850 veh/hr per lane were 70 and 80, respectively. A lengthy simulation run (of about 40 min) was performed after the optimal solution was reached to collect a sample of 1,500 observations. This large sample was needed to perform statistical analyses on the observations using the batch mean method (*28*). The advantage of the batch mean technique is that it can analyze data and construct confidence intervals for the system measures of performance even with the presence of correlation among the individual observations. With that in mind, performing one lengthy simulation run was more economical and beneficial than running several independent replications, each for a short period. The reason is that in the latter method a warm-up period for each replication is required during which no data are collected. In addition, with short simulation runs, the system may never reach a steady state condition, in which case the measure of performance obtained by averaging over a number of runs will not be representative of the system performance at the steady state condition.

Common output parameters of the optimization procedure are presented in Table 6. The optimum travel time is plotted on Figure 7 for both flow levels analyzed. As indicated in Table 6, exactly one-half of the maximum simulation runs specified for each flow rate were expended in determining the initial point. The optimum point under both flow levels was closer to the initial point, which implies that the model was quite successful in finding the initial point. The results also show that as the flow rate increased, a late merging strategy seemed to be more appropriate in terms of minimizing the travel time. Furthermore, as seen from Table 6 and Figure 7, the optimum (minimum) travel time resulted from neither the early nor late merging strategies. However, the results obtained from the early merge strategy were much closer to the optimum than those from the late merge strategy. Cumulative distributions of the optimum distribution of attempted and completed merges over the segments upstream of the taper for each flow level are shown in Figure 8. The horizontal distance between the merge attempt and merge completion curves is the distance traveled while searching for an acceptable gap. This value consistently increased as traffic approached the taper. At that point, the distance decreased because drivers were forced to merge in that zone. The search distances were rather long but this is not unexpected at the near-capacity flow levels that were tested. Unexpectedly, however, the average search distance decreased as the flow rate increased, specially at locations near the taper or further upstream of the taper. This observation does not necessarily mean that the

TABLE 6 OPTIMIZATION RESULTS FOR TWO FLOW RATES

Parameter	1700 vphpl		1850 vphpl	
	Initial	Optimal	Initial	Optimal
No. of simulation runs used	35	66	40	77
Optimal travel time (sec/mile)	66.5	66.5	72.0	71.5
Decision variable, P_2	0.44	0.46	0.94	1.00
Decision variable, p_3	0.19	0.19	0.31	0.34
Decision variable, p_4	0.06	0.02	0.19	0.25

(Note: P_i is the proportion of closed lane traffic that has not attempted the merge upstream of segment i).

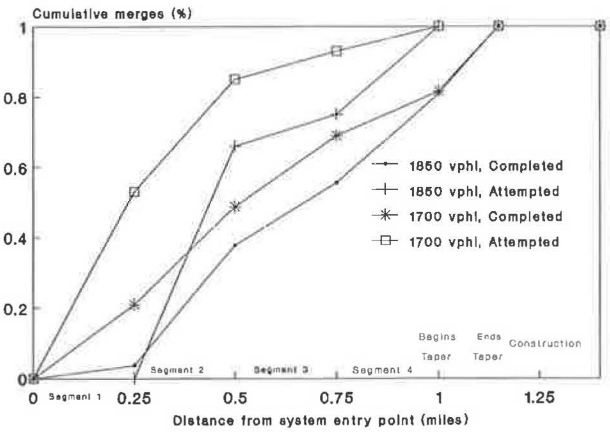

FIGURE 8 Recommended merge strategies.

search time decreased with an increase in flow rate but may simply be the consequence of speed reductions in both traffic streams. The figure also shows that as the flow level increased, completed merges occurred much further downstream (almost no merges occurred in Segment 1 with the 1,850 veh/hr per lane scenario). In addition, the figure shows that, under the optimal solution, up to 18 percent of closed-lane traffic, under both flow levels, merged along the construction taper. Interestingly, field observations at that flow rate revealed about the same percentage.

CONCLUSIONS

Because the work discussed in this paper is a first attempt at optimizing the performance of microscopic traffic systems, it has focused on the major optimization concepts within the integrated model. However, valuable conclusions can be drawn from the work zone applications of the model. These conclusions are limited to the specific lane closure configurations studied in this paper and should not be generalized without further validation effort.

- The integrated microscopic model can be used in optimizing freeway work zone traffic systems at a reasonable computational cost.
- The optimization model SAMOPT was enhanced in this study and consequently its efficiency in using computer CPU time has been considerably improved.
- The merge strategy from closed lanes at work zones had a significant impact on the system performance measured by travel time; this impact was more pronounced at flow rates ranging between 1,500 and 2,000 veh/hr per lane.
- The optimum merge strategy that resulted in minimum average travel time over the work zone was neither an early nor a late merge pattern.
- As the flow level increased, the optimum strategy recommends that the first attempt to merge from the closed lane be made further downstream compared to that for lighter flow rate.
- The model can be applied to designing other work zone elements such as the determination of the optimum length of taper, ramp work zone controls, and location of advance warning devices.

IMPLEMENTATION CONSIDERATIONS

Model Utilization

With the capability of the model to formally optimize a traffic system, the model can be applied to

- Determining an optimal merge strategy at freeway lane closures. This is an important step toward optimizing the traffic control plan in the field and ultimately providing the proper traffic control device to promote such merge pattern. This work, however, was beyond the scope of this study.
- Determining the optimum length of the construction taper as well as lengths of speed change lanes.
- Determining whether speed reduction is advisable at site for certain ranges of traffic volumes.

Model Limitations

Although the model applications are numerous, users must be aware of its underlying assumptions pertaining to the driver's critical gap and reaction time. Therefore, when field observations appear not to match the general scope of such assumptions, the results must be viewed with great caution.

RECOMMENDATIONS FOR FURTHER WORK

Following are some recommendations for future work in this area:

1. Broadening the optimization work to include different lane closure configurations (left versus right, single versus multiple closures, construction near ramps, etc.) to generalize the findings in this paper.
2. Interfacing the optimization algorithm with other existing traffic models to study and optimize other traffic problems (such as NETSIM model for intersections).
3. Testing the optimization of traffic systems by using other objective functions (such as acceleration noise) to see whether they yield the same optimum solution.
4. Investigating the use of multi-objective or utility functions.

ACKNOWLEDGMENTS

This study was sponsored by the National Science Foundation. Farhad Azadivar provided the source code of the optimization algorithm (SAMOPT) and valuable advice during the course of this study.

REFERENCES

1. T. Cilliers, A. D. May, and R. Cooper. *FREQ6PL—A Freeway Priority Lane Simulation Model, Final Report and Volume II.* California Department of Transportation, Sacramento, Calif., Sept. 1978.
2. *Manual on Uniform Traffic Control Devices for Streets and Highways (MUTCD).* FHWA, U.S. Department of Transportation, Washington, D.C., 1978.

3. *Development and Testing of INTRAS, A Microscopic Freeway Simulation Model.* Report FHWA/RD–80–106. National Technical Information Service, Springfield, Va., Oct. 1980.
4. *Traffic Network Analysis with NETSIM, A User's Guide, Implementation Package.* FHWA–IP–80–3. FHWA, U.S. Department of Transportation, Washington, D.C., 1980.
5. Z. A. Nemeth and N. M. Rouphail. Lane Closures at Freeway Work Zones: Simulation Study. In *Transportation Research Record 869*, TRB, National Research Council, Washington, D.C., 1982, pp. 19–25.
6. A. Sadegh, A. E. Radwan, and N. M. Rouphail. ARTWORK: A Simulation Model of Urban Arterial Work Zones. In *Transportation Research Record 1163*, TRB, National Research Council, Washington, D.C., 1988, pp. 1–3.
7. *SOAP 84 User's Guide, Implementation Package.* FHWA–IP–85–7. FHWA, U.S. Department of Transportation, Washington, D.C., Jan. 1985.
8. E. C. Chang, C. J. Messer, and B. G. Mardsen. *Analysis of Reduced-Delay Optimization and Other Enhancements to Passer II-80, Passer II-84.* Research Report 375–1F, Texas Transportation Institute, College Station, 1984.
9. J. D. C. Little and M. D. Kelson. *Optimal Signal Timing for Arterial Signal Systems, MAXBAND.* Summary Report, Operation Research Center, Massachusetts Institute of Technology, Cambridge, 1980.
10. D. I. Robertson. *TRANSYT: A Traffic Network Study Tool.* TRRL Report 253. Crowthorne, Berkshire, England, 1969.
11. E. B. Lieberman and J. L. Woo. *SIGOP II: A New Computer Program for Calculating Optimal Signal Timing Patterns.* In *Transportation Research Record 596*, TRB, National Research Council, Washington, D.C., 1976.
12. F. Azadivar and J. J. Talavage. Optimization of Stochastic Simulation Models. *Mathematics and Computers in Simulation*, Vol. 22, No. 3, 1980, pp. 231–241.
13. A. A. B. Pritsker. *Introduction to Simulation and SLAM II.* John Wiley and Sons, New York, 1986.
14. A. K. Rathi and Z. A. Nemeth. FREESIM: A Microscopic Simulation Model of Freeway Lane Closures. In *Transportation Research Record 1091*, TRB, National Research Council, Washington, D.C., 1986, pp. 21–24.
15. D. L. Gerlough and M. J. Huber. In *Special Report 165: Traffic Flow Theory*, TRB, National Research Council, Washington, D.C., 1975.
16. P. G. Gipps. A Behavioral Car-Following Model for Computer Simulation. *Transportation Research*, Vol. 15B, 1981, pp. 105–111.
17. D. C. Gazis, R. Herman, and R. W. Rothery. Non-Linear Follow The Leader Models of Traffic Flow. *Operations Research*, Vol. 9, 1961, pp. 545–567.
18. J. Pahl. Gap-Acceptance Characteristics in Freeway Traffic Flow. In *Highway Research Record 409*, HRB, National Research Council, Washington, D.C., 1972, pp. 57–63.
19. O. Adebisi. Driver Gap Acceptance Phenomena. *Journal of Transportation Engineering, ASCE*, Vol. 108, No. TE6, Nov. 1982, pp. 667–688.
20. R. Miller. *Gap Acceptance and Gap Estimation in a Freeway Driving Situation.* Master's thesis. The Ohio State University, Columbus, 1980.
21. A. Rathi. *Development of a Microscopic Simulation Model for Freeway Lane Closures.* Ph.D dissertation. The Ohio State University, Columbus, 1983.
22. R. Abella. *Analysis and Design of Freeway Lane Drops using Continuous Time Semi-Marbou Process.* Ph.D dissertation. The University of Toledo, Ohio, 1975.
23. H. Mahmassani and Y. Sheffi. Using Gap Acceptance to Estimate Gap Acceptance Functions. *Transportation Research*, Vol. 15B, 1981, pp. 143–148.
24. G. Weiss and A. Maradudin. Some Problems in Traffic Delay. *Operations Research*, Vol. 10, 1962, pp. 74–104.
25. F. Azadivar. *Optimization of Stochastic Systems Through Simulation Using Stochastic Approximation Method.* Ph.D. dissertation. Purdue University, Ind., 1979.
26. *Special Report 209: Highway Capacity Manual.* TRB, National Research Council, Washington, D.C., 1985.
27. A. M. Law and W. D. Kelton. *Simulation Modeling and Analysis.* McGraw-Hill, Inc., New York, 1982.
28. G. S. Fishman. *Principles of Discrete Event Simulation.* John Wiley and Sons, New York, 1978.

Publication of this paper sponsored by Committee on Traffic Safety in Maintenance and Construction Operations.