mula and Resource Allocation Model. Report DOT-TSC-FRA-82-1. FRA, U.S. Department of Transportation, Sept. 1982.
3. J. Hitz and M. Cross. Rail-Highway Crossing Resource Allocation Procedure User's Guide. Report FHWA-IP-82-7. FRA, U.S. Department of Transportation, Dec. 1982.
4. J. Hitz, ed. Summary Statistics of the National Railroad-Highway Crossing Inventory for Public at Grade Crossings. Report FRA-RPD-78-20. FRA, U.S. Department of Transportation, Sept. 1978.
5. P. Mengert. Rail-Highway Crossing Hazard Prediction Research Results. Report FRA-RRS-80-02. U.S. Department of Transportation, March 1980.
6. E. Farr. Rail-Highway Crossing Resource Allocation Model. Report FRA-RRS-81-001. U.S. Department of Transportation, April 1981.
7. Rail-Highway Crossing Accident/Incident and Inventory Bulletin No. 4. FRA, U.S. Department of Transportation, 1981.
8. Risk Concepts in Dangerous Goods Transportation Regulations. National Transportation Safety Board, U.S. Department of Transportation, 1971.
9. D.R. Cox. Analysis of Binary Data. Halstead Press/John W. Stey, New York, 1970.
10. J. Coleman and G.R. Stewart. Investigation of Accident Data for Railroad-Highway Grade Crossing. In Transportation Research Record 611, TRB, National Research Council, Washington, D.C., 1976, pp. 60-67.

Publication of this paper sponsored by Committee on Railroad-Highway Grade Crossings.

Stability and Other Considerations in Simulation Analysis of Signal Control

FENG-BOR LIN

ABSTRACT

Lack of understanding of the nature of simulation and the characteristics of a system to be simulated can result in misuse of simulation models and simulation results. To promote better applications of simulation models to the evaluation of signal controls, three problems related to the generation and interpretation of simulation data are discussed in this paper. These problems include the stability of simulation results, the use of seed numbers for generating probabilistic events, and the aggregation of input data. Using simple examples of signal control, several fallacies in the application of signal simulation models are illustrated. Suggestions for avoiding these fallacious applications are presented.

Simulation models are increasingly used to aid in the design and evaluation of signal control systems. Some of these models, such as UTCS-1 (1) and NETSIM (2), are intended for general application in the evaluation of traffic control alternatives. These models require microscopic simulation of traffic flow characteristics to approximate the real world.

Experience with existing microscopic simulation models has produced a wealth of information on the potential and limitations of applying such models (3). Current concerns appear to focus on model enhancement, user needs and constraints, resource requirements for model application, and promotion and implementation of application by the traffic engineering community. The problem of experimental design for simulation analysis has also drawn some attention.

With increased use of simulation models for evaluation purposes, the risk of misuse and misinterpretation of simulation results can be expected to increase. A reason for this is that simulation models require substantial user interactions. An evaluation model is essentially a tool for data collection. Consequently, simulation results should be treated as a sample of observations. Estimates obtained from such a sample should be subjected to statistical tests for interpretation. It follows that experimental design should be an important part of simulation analysis. At issue is how, within the capability of a model, a user can apply the model efficiently to obtain statistically valid estimates.

The experimental design for simulation analysis is a profound subject. It requires a comprehensive understanding of the characteristics of a system to be simulated and the nature of simulation. At the present time, such an understanding is nonexistent. This is due in part to the large number of different systems a model has to accommodate. High costs and the reliability issue associated with the use of a model are also contributing factors. Nevertheless, there are several aspects of simulation application that concern experimental design and can be readily discussed to benefit users of simulation models. These include the stability characteristics of simulation outputs, use of seed numbers for generating probabilistic events, and data aggregation. The purpose of this paper is to discuss the nature of these features.

STABILITY OF SIMULATION OUTPUTS

The operation of a signal control can be characterized by a set of measures of performance. These measures are in fact random variables because of the probabilistic nature of signal operation. A common purpose of simulation studies is to estimate the true value of a measure of performance. In this undertaking, users of a simulation model will have to deal with the stability of simulation outputs either directly or indirectly.

There are two primary features that determine the stability of simulation outputs. One is the dependence or independence of an estimate on the length of the simulation period. The other is the variation of estimates from true means. These features influence the length of a simulation run and the number of replicated runs needed to obtain reliable estimates. A discussion of these stability-related features based on vehicle delays follows.

Stable, Metastable, and Unstable States

Control performance can be classified in three states: stable state, metastable state, and unstable state. There are no clear-cut boundaries between stable state and metastable state nor between metastable state and unstable state. Nevertheless, these states have distinct characteristics that render them identifiable from simulation outputs.

A stable state usually exists when traffic volumes are light or moderate. Under this circumstance, the average delay of a traffic flow is governed primarily by the flow rate. For a given arrival pattern (e.g., random arrival), the sequence of the arriving headways or that of an event (e.g., gap acceptance) has little influence on the estimated mean value of a measure of performance. Using arriving headways as an example, this implies that rearranging the sequence of arriving headways will produce only slight changes in the performance of a control. In a simulation analysis, rearranging the sequence of arriving headways can be accomplished by using different seed numbers in replicated runs. It follows that estimated measures of performance from replicated runs will have only small variations. Another characteristic of a stable state is that estimated mean values of measures of performance can reach stabilized values as the simulation process is advanced. In other words, the estimates are time independent. An example of the stable state is the average delay of a flow of 400 vehicles per hour (vph) under a pretimed control shown in Figure 1.

In a metastable state, average delay depends not only on the flow rate but also on the sequence of arriving headways. A headway sequence of 3.2, 1.6, 5.3, ..., 10.2, 3.7, 2.1 sec, for example, can produce an estimate significantly different from that produced by a reversed sequence of 2.1, 3.7, 10.2, ..., 5.3, 1.6, 3.2 sec. Therefore, replicated simulation runs with different seed numbers can result in large variations in the estimated mean values of measures of performance. However, these estimates are still time independent (i.e., they can still reach stabilized values as the simulation process

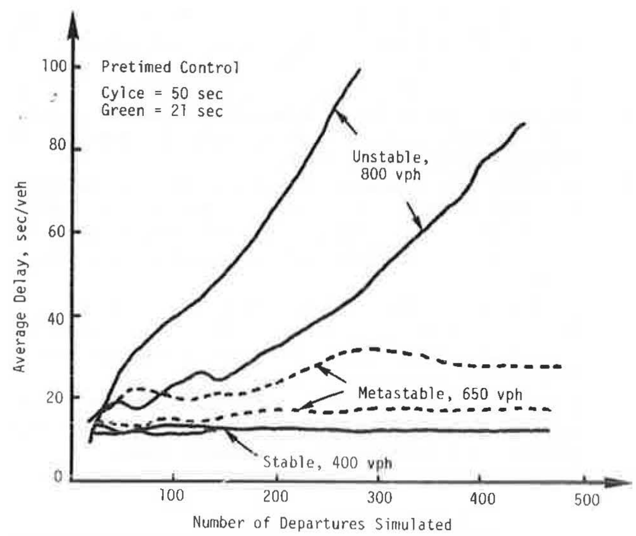

FIGURE 1 States of performance of signal controls.

continues). These characteristics are shown in Figure 1 by two simulation runs for a flow of 650 vph.

A metastable state may emerge when traffic flows are heavy enough to induce occasional carry-over of queuing vehicles from one cycle to the next. As the flows increase further, the queue length in a lane may grow with time if the same flows persist. Consequently, the performance of a control may become unstable. In an unstable state average delay depends not only on the flow rate and the sequence of arriving headways but also on the length of the simulation period. The longer the simulation period, the longer the average delay becomes. The estimated mean values of measures of performance are time dependent, and variations in the estimates from replicated runs can be expected to be large. The characteristics of the unstable state are also shown in Figure 1 by two replicated simulation runs.

The time-dependent features of a queuing system are rarely treated in the context of queuing theory because of mathematical complexities. In classic queuing theory, the operation of a system is usually assumed to be in a steady state. This implies that the average performance characteristics of a system do not change with time. This approach creates confusion when attempts are made to compare the output of a steady-state queuing model with simulation results. One example of such confusion involves Webster's delay formula (4). The approximate form of this formula is

$$D = 0.9 \left\{ [C(1-x)^2/2(1-xy)] + [y^2/2Q(1-y)] \right\} \quad (1)$$

where

D = average delay of vehicles in a traffic lane,
C = cycle length,
x = effective green-to-cycle length ratio,
y = saturation ratio, and
Q = flow rate.

This formula is a steady-state queuing model for a flow pattern with random arrivals and a uniform flow rate of Q. It assumes that the flow rate will persist indefinitely. As a result, when the saturation ratio y approaches 1.0 the estimated average delay approaches infinity. For most signal operations, this phenomenon cannot be observed in the field because heavy flows do not last forever or even for a very long time. When heavy flows induce an unstable state, shorter simulation periods will result in shorter estimated average delays. This is shown in Figure 2. For this reason a comparison of simulation output with the output of a steady-state model becomes meaningless unless a very long simulation period is used. This also underscores the importance of determining whether simulation results are time dependent before they are used for comparative analyses.

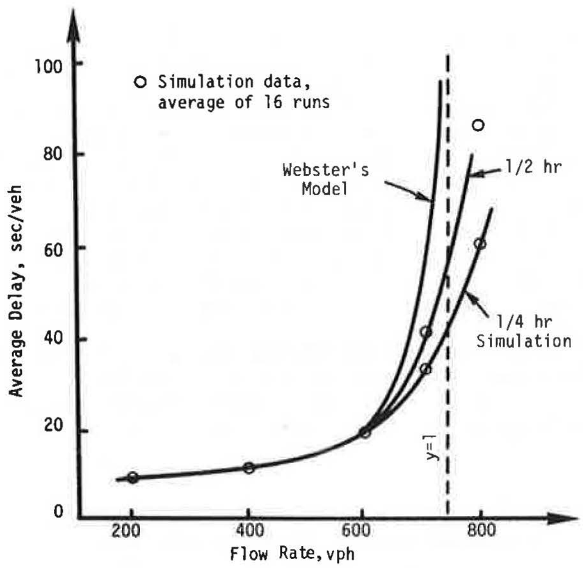

FIGURE 2 Comparison of simulated delays with Webster's estimates under a pretimed control.

To facilitate identification of this time-related stability feature, the simulation period beyond a nonrecording transient interval may be divided into time blocks of equal length. Each block should be at least equal to the expected maximum cycle length. Let x_{ij} represent the value of the jth observation (e.g., delay of a vehicle) in the ith time block. For each block the estimated mean $\bar{x}_i$ and standard deviation S_i can be determined from x_{ij} based on m_i observations. In addition, the grand mean $\bar{x}$ and the grand standard deviation S for all the time blocks simulated can be determined.

If variable flow rates are used as inputs to approximate an actual flow pattern, the simulation period should coincide with the actual duration of the flow pattern. In this case, simulation results may become time dependent for a short time and then stabilize. At the end of the simulation, the grand mean $\bar{x}$ should be printed out as a function of the number of time blocks simulated. This information can be used to determine whether and when the operation of a control is time dependent.

If uniform flow rates are used instead, users may be allowed to specify a minimum simulation period and a maximum simulation period beyond the transient interval. The minimum period is for the purpose of obtaining a sufficient number of observations before the stability check begins. Ten minutes of real time is probably a reasonable period to use. The maximum simulation period can be based on cost consideration. Between the minimum and the maximum, an algorithm incorporated into a simulation model may be used to determine whether simulation results have reached stabilized values or whether the operation of a control has become unstable.

A simple approach to detecting an unstable state is to examine the trend of $\bar{x}_i$ for several consecutive blocks (e.g., 4 or 5) from block a to block b. Block b represents the block that has just been simulated. If $\bar{x}_i$ increases monotonically from block a to block b, there is reason to suspect that an unstable state exists. In such a case, the estimate $\bar{x}_b$ of the current block may be compared with the estimate $\bar{x}_a$ several blocks earlier. The purpose is to determine whether $\bar{x}_b$ is significantly greater than $\bar{x}_a$ by a specified amount (e.g., 5 sec/vehicle). This can be carried out with a simple t-test.

In choosing a t-test, one should realize that the true standard deviations of the observations in two time blocks are unknown and are not necessarily the same. Therefore, a proper test is to determine the following statistics (5):

$$t = (\bar{x}_b - \bar{x}_a - \delta)/[(S_a^2/m_a) + (S_b^2/m_b)^{1/2}] \quad (2)$$

and

$$f \simeq [(S_a^2/m_a) + (S_b^2/m_b)]^2 / \{[(S_a^2/m_a)^2/(m_a - 1)] + [(S_b^2/m_b)^2/(m_b - 1)]\} \quad (3)$$

where

t = t-value of the difference $\bar{x}_b - \bar{x}_a$;
f = approximate degrees of freedom;
m_a, m_b = sample sizes of block a and block b, respectively;
S_a, S_b = sample standard deviations of block a and block b, respectively; and
δ = specified level of difference.

Given a level of significance α and degrees of freedom f, a critical t-value, denoted as t_c, can be estimated from a statistical table of t-distribution. If the t-value obtained from Equation 2 is greater than t_c, one can conclude that $\bar{x}_b$ differs significantly from $\bar{x}_a$ by the specified amount of δ. Consequently, one may assume that an unstable state exists.

This approach is not infallable. Figure 1 shows an example. One of the simulation runs for the metastable operation reveals that, between a total departure of about 140 to about 280 vph, the average delay increases monotonically. The equivalent time is about 12 min. The operation of the control in this period may be labeled unstable and the simulation process is subsequently terminated. If the process is allowed to continue, however, the estimate of the average delay stabilizes. Such an incorrect identification will likely occur only when the operation of a control falls in a gray area between a metastable state and an unstable state. Therefore, it will in fact force users to interpret the simulation results more cautiously.

Whether simulation results have reached stabilized values may be checked in a similar manner on the basis of a few decision rules. First, the block mean $\bar{x}_i$ should not display a monotonic increase or decrease from block a to block b. When this condition exists, the average value of the block means of block a through block b can be determined. Next, the proportion of observations having values greater than this average can be determined for each block. If the proportion is about 50 percent in each block and it does not show a monotonic increase or decrease from block a to block b, the performance of the control may have stabilized. A more stringent test may then be performed to determine whether the proportion in block b is significantly different from that a block a. If the difference is not significant, the simulation results may be assumed to have reached stabilized values. The simulation process may be terminated if the estimated means of measures of performance have satisfied a desired level of accuracy.

If a model does not have an internal algorithm to check time-related stability, users will have to perform at least two replicated runs with different simulation periods. When an unstable state exists, the longer run will give a much higher estimate of average delay. In some cases, the number of runs may be reduced to one if the grand means at the end of each time block or the block means are printed out as a function of the number of blocks simulated. This information allows users to visually examine the performance of a control over time.

Within-Run and Between-Run Variations

In a simulation run, individual vehicle delays will vary. These within-run variations affect the accuracy of estimated average delays. For a given flow pattern, simulated average delays can also vary from one run to another. These between-run variations result from changes in operating conditions without a change in the average characteristics of a simulated flow pattern. Changes in the operating conditions can be induced by using different seed numbers in replicated runs for generating probabilistic events. For example, a seed number may lead to a headway sequence of 4.2, 1.5, ..., 10.2 sec. A new seed will certainly produce a different sequence.

In general, within-run variations can be expected to increase with flow rate. This is inherent to the nature of a queuing system. For example, assume that there is a single channel queuing system with random vehicle arrivals and random service time. Let λ represent the arrival rate and β the service rate. It can be shown (6) that, if λ and β remain unchanged, the cumulative probability density function for the waiting times is

$$f(w \leq t) = (\lambda/\beta) - [(\lambda/\beta)\, e^{-(\beta-\lambda)t}] \qquad (4)$$

where $f(w \leq t)$ represents the probability of a waiting time w less than or equal to t.

Based on this function, the average waiting time becomes

$$\mu = \lambda/\beta(\beta-\lambda) \qquad (5)$$

and the standard deviation of the waiting times is

$$\sigma = [2\lambda/\beta(\beta-\lambda)]^{1/2} \qquad (6)$$

Equation 6 reveals that σ increases with the arrival rate λ. This characteristic is undesirable from the viewpoint of simulation. It means different signal control problems require different simulation periods in a single run to achieve a given level of accuracy in estimates. It also implies that heavier flows may need a much longer simulation period.

To determine whether simulation results have reached a specified level of accuracy, one has to have estimates of the true mean (e.g., μ) and the true standard deviation (e.g., σ) of a measure of performance. Let $\bar{x}$ and S be the estimates of mean and standard deviation, respectively. Also let n be the number of observations used in determining $\bar{x}$ and S. Regardless of the sample size n, the confidence interval of $\bar{x}$ can be determined as (5)

$$R = \bar{x} \pm t_c[S/(n-1)^{1/2}] \qquad (7)$$

where

R = confidence interval of $\bar{x}$, and
t_c = critical t-value corresponding to n - 1 degrees of freedom and a level of significance α.

If the service rate β in Equation 6 is 400 vph, the value of σ given by the same equation equals 4.2 sec/vehicle for an arrival rate λ = 200 vph and 7.4 sec/vehicle for λ = 300 vph. At a level of significance α = 5 percent, the value of t_c approaches 1.96 for large n. Using the σ values for S in Equation 7, one can see that the sample size required to reduce the confidence interval to within 1 sec of the estimated mean is about n = 68 for λ = 200 vph and about n = 211 for λ = 300 vph. These sample sizes are equivalent to 0.4 hr of observations for λ = 200 vph and 0.7 hr for λ = 300 vph.

The between-run variations are smaller than the within-run variations because they involve the mean of each run. Nevertheless, such variations can be too large to ignore. Table 1 gives an example of

TABLE 1 Between-Run Variations in Average Delays Resulting from a Fully Actuated Control

Simulation Run No.	Average Delay (sec/veh) Lane Flow (vph) 200	400	600	700
1	4.0	9.6	20.1	49.7
2	5.1	10.1	22.3	59.6
3	4.8	9.7	23.1	55.6
4	4.3	11.3	22.9	52.3
5	5.2	9.8	28.3	64.5
6	4.6	8.8	22.4	51.3
7	5.2	9.1	22.4	69.9
8	4.7	9.0	22.3	53.8
9	4.2	10.1	23.1	63.5
10	4.9	9.9	22.6	50.2
$\bar{x}$	4.7	9.7	23.0	57.0

Note: Simulation period = 0.5 hr; extension interval = 0 sec.

the between-run variations in the average delays of vehicles. The vehicles are subjected to a two-phase fully actuated control that uses presence detectors. Each phase has two lanes. The flow rates are the same in all the lanes.

When the major phase flows are under 400 vph/lane, the table shows that any of the 10 runs results in an estimated mean within about 1 sec/vehicle of the overall mean $\bar{x}$. At higher flow rates, estimated means of individual runs begin to differ more and more from the overall mean. This example reveals the need to perform replicated runs and the potential risk of using the results of a single run for comparative analysis.

When replicated runs are performed, the estimation of the mean and standard deviation of a measure of performance becomes more complicated. The analysis of variance techniques described by Brownlee (5) may be used under the circumstance. These techniques require that the individual observations (e.g., vehicle delays) obtained in each run be stored in one way or another as a data file. The results of several runs would form a matrix of individual vehicle delays for analysis.

Obviously, the need to treat and analyze such a matrix after each additional run would be a heavy burden on most users. A more practical approach is to obtain the same (or approximately the same) number of observations in each run. As an approximation, this can be accomplished by using equal simulation periods for all the replicated runs. In this case, users have to deal only with the estimated mean of each run. Let $\bar{x}_r$ represent the mean estimated from the rth replicated run and k the number of runs. Then, the true mean can be estimated as

$$\bar{x} = \sum_{r=1}^{k} \bar{x}_r/k \tag{8}$$

and the standard deviation of the estimated mean, denoted as $S_{\bar{x}}$, becomes

$$S_{\bar{x}} = \left\{ \left[\sum_{r=1}^{k} (\bar{x}_r - \bar{x})^2 \right] /k(k-1) \right\}^{1/2} \tag{9}$$

For runs with equal numbers of observations, it can be shown analytically that $\bar{x}$ and $S_{\bar{x}}$ are the same as those obtainable from a more complicated analysis of variance technique. These estimates can be used in several ways. For example, they allow users to determine the confidence interval of $\bar{x}$. In this application the term $S/(n - 1)^{1/2}$ in Equation 7 can be replaced by $S_{\bar{x}}$, and the critical t-value (i.e., t_c) has k - 1 degrees of freedom. For comparative analysis of estimated $\bar{x}$, it is advisable that the confidence intervals of these estimates be approximately the same. Otherwise one would be comparing estimates of different levels of accuracy. Given this understanding, Equations 2 and 3 can be modified to test the significance of the difference between the operations of alternative controls as represented by $\bar{x}$. For this purpose, δ of Equation 2 can be set to zero; S_a^2/m_a and S_b^2/m_b are equivalent to the values of $S_{\bar{x}}$ of two alternative controls; and m_a and m_b represent, respectively, the numbers of replicated runs performed for the alternatives.

So far no reliable methods are available to introduce between-run variations into a single run and thus eliminate the need to perform replicated runs. To reduce the number of replicated runs, one will have to find a way to rapidly obtain narrow confidence intervals of $\bar{x}$. This can be achieved only if one can produce a large reduction in $S_{\bar{x}}$ from each additional run. One approach to this problem is to use negatively correlated random numbers in replicated runs (7). This approach can be easily implemented by using a random number R ($0 \leq R \leq 1.0$) in one run and 1^{-R} in another.

USE OF SEED NUMBERS

Ideally, separate events (e.g., headways in a lane or lane changes) in a simulation run should be generated from different seed numbers. This is not difficult to do because a single seed number can be used to generate randomly a string of seed numbers. These seed numbers can then be used to generate different events. In the simulation of a large system, however, numerous events are encountered. If an event is assigned a separate seed number, memory storage requirements may be increased substantially. Central processing unit time will also increase. To simplify the task, a simulation model may rely on a single string of random numbers generated from a seed number to simulate various events. Questions have been raised about the desirability of such a practice (3).

The use of a single string of random numbers for different events will certainly destroy the ability of a simulation model to perform controlled analyses. This is indeed undesirable if one is to compare alternative controls based on a single run for each alternative. However, one should also realize that in such comparative analyses the only event that a simulation model can control precisely is input flow patterns in the form of arriving headways. When vehicles are processed downstream and subjected to influences by control strategies and by conflicts between traffic flows, not many events can remain the same under alternative controls. Therefore, the real issue is whether a single string of random numbers can produce distributions of various events that conform to predetermined distributions.

To provide an insight to this problem, consider the generation of headways for vehicles in a traffic lane. With a random arrival pattern, the headways follow a shifted negative exponential function:

$$f(h \geq t) = e^{[(t-\tau)/(T-\tau)]} \tag{10}$$

where

$f(h \geq t)$ = probability that a headway h is greater than or equal to t;
T = average headway of vehicles; and
τ = minimum headway, taken to be 1 sec.

Assume that the arrival of vehicles in that lane is only one of 15 events to be generated with a single string of random numbers. An event may not have any sampling unit for observation during a particular period of time in the simulation process. Therefore, in one case two or more successive random numbers may be used consecutively for generating headways. In another, 14 successive random numbers may be used to generate other events before the next one is used again for the headways. Of the 80 distributions of headways randomly generated in this manner for a flow rate of 600 vph, 3 do not conform to the shifted negative exponential function at a level of significance of 5 percent (chi-square test). Another 4 have mean headways significantly different from the intended mean of 6 sec. Therefore, there is approximately a 9 percent chance that the use of single-string random numbers will fail to produce the desired arrival pattern for a flow of 600 vph. A similar experiment for a flow of 200 vph also results in a failure rate of 9 percent. In contrast, when headways are determined sequentially from every random number generated, the failure rate is 6 percent for a flow of 600 vph and 9 percent for a flow of 200 vph.

These brief analyses show that, from the viewpoint of generating predetermined distributions of events, the use of a single string of random numbers is as good as the use of multiple strings. The analyses also reveal that a generated distribution may be quite different from what is desired. The results of a simulation analysis may be made more reliable if seed numbers are chosen carefully to ensure that generated distributions conform to intended distributions.

AGGREGATION OF INPUT DATA

The use of aggregated data results in a loss of useful information. For simulation analysis of signal controls, it may prompt a model to produce estimates quite different from those obtained from the use of disaggregated data.

One type of data aggregation commonly encountered in the use of a signal simulation model is average flow rates. The impact of such practice on estimated average vehicle speed has been examined in two previous studies based on UTCS-1 and TRANSYT (8-10). The time aggregation of flow rates was found to have insignificant impact on estimated average vehicle speeds.

It should be cautioned, however, that different measures of performance may have different sensitivities to the aggregation of input data. Table 2 gives an example of this possibility in terms of average delays resulting from a two-phase, fully actuated control. This control employs 50-ft-long presence detectors. The extension intervals are set at 1.5 sec and the maximum greens at 50 sec. Each phase has two lanes with equal traffic volumes. The actual flow rates per 5-min interval over a 1-hr period are assumed to have the following relative values: 0.85, 0.90, 0.95, 1.0, 1.1, 1.2, 1.2, 1.1, 1.0, 0.95, 0.90, and 0.85. The average flow rate for the entire hour has a relative value of 1.0.

TABLE 2 Effects of Time Aggregation of Flow Rates on Estimated Average Delays

	Average Delay (sec/veh)			
	Uniform Rate Simulation		Variable Rate Simulation	
	vph/Lane		vph/Lane	
Run No.	400	600	400	600
1	10.7	20.6	11.3	20.4
2	11.0	22.1	11.6	27.6
3	9.5	21.7	10.1	24.3
4	10.6	20.0	10.9	21.2
5	10.9	20.2	11.5	22.5
6	9.6	21.0	9.9	26.7
7	10.1	19.3	10.5	20.0
8	9.6	21.3	10.0	23.5
9	9.8	21.0	10.5	21.7
10	10.5	22.2	10.9	28.9
$\bar{x}$	10.2	20.9	10.7	23.7
S	0.6	0.9	0.6	2.8

Note: Simulation period = 1 hr; extension interval = 1.5 sec.

The performance of the control is simulated in two ways for average flow rates of 400 and 600 vph. One is based on the variable flow rates and the other uses the average flow rates as input. In both cases, arriving headways are assumed to distribute according to the shifted negative exponential function of Equation 10.

The results given in this table reveal that the use of average flow rate leads to smaller between-run standard deviations. This is a desirable feature if the estimated average delays are insensitive to the data aggregation. Unfortunately, statistical tests based on Equations 2 and 3 show that data aggregation tends to produce lower estimated averages. The differences are significant at a level of significance of 5 percent for flow rates of 600 vph. The difference is insignificant for a flow of 400 vph. Thus, data aggregation may significantly bias estimates under heavier flow conditions.

This simple example shows the potential risks of using average flow rates. However, one should not conclude that average flow rates should not be used for simulation. What is important is that the use of simulation results should be consistent with the simulation conditions. The purpose of a simulation analysis should dictate the choice of the input data. For simulating the actual operation of a signal control for the purpose of evaluation, it is advisable not to use average flow rates. For comparative analyses of alternative controls, uniform as well as variable flow rates may be used if it is known that the data aggregation will not alter the estimated relative merits of the controls.

Another common practice of data aggregation is the use of average headways. For example, the average discharge headway of queuing vehicles in a given queuing position may be used in place of a probability distribution of individual headways. This type of data aggregation can provide adequate estimates under a variety of circumstances, but there is always a possibility that it may produce a distorted picture of signal operation.

Consider the operation of a fully actuated control based on presence detectors. Under this mode of control a vehicle can extend the green if it actuates a detector before a vehicle ahead departs from the detection area and the extension interval expires. The arrival time of a vehicle at the upstream edge of a detector and the departure time from the downstream edge of the same detector can be measured in the field. Table 3 gives the average values of these flow characteristics in relation to a 50-ft detector as observed on Almond Street, Syracuse, New York. These averages indicate that a vehicle in the third queuing position waiting upstream of the detector at the onset of the green needs an average of 5.2 sec to reach the detector. The vehicle ahead in the second queuing position requires 5.6 sec to depart from the detection area. Vehicles in the queuing positions farther upstream have average arrival times longer than the average departure times of the vehicles immediately ahead.

TABLE 3 Average Arrival and Departure Times of Queuing Vehicles in Relation to a 50-ft Detector

Queue Position	Departure Time (sec)	Arrival Time (sec)
1	3.1	-
2	5.6	-
3	7.9	5.2
4	10.1	8.1
5	12.2	10.6
6	14.4	12.9
7	16.4	15.1
8	18.5	17.2

Note: Data gathered for Almond Street, Syracuse, New York.

Based on the average arrival times and departure times, one can conclude that, if no extension is given to the green after each departure and only one lane flow is associated with a signal phase, vehicles in the back of a long queue will face a certain premature termination of the green. In reality, arrival times and departure times are probabilistic. For example, the observed departure times of vehicles in the first queuing position range from 1.2 to 5.8 sec. The field data show that under the same conditions vehicles in the fourth queuing position have a 10 percent chance of facing a premature termination of the green. For those in the seventh queuing position the chance is 40 percent.

The effect of using the averages in estimating average delays is shown in Figure 3. This figure is based on a two-phase control with two lanes in each phase. The lane flow Q in the major phase is twice as heavy as that in the minor phase. The extension

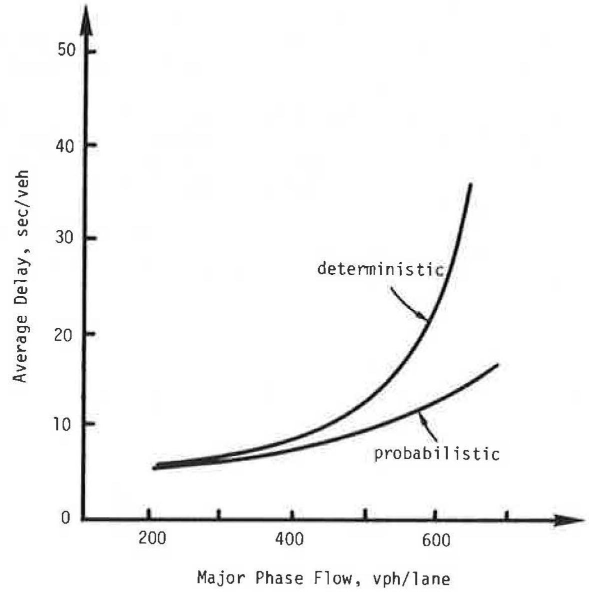

FIGURE 3 Effects of using average headways (simulation period = 1 hr).

interval is 0 sec and each phase has a clearance interval of 4 sec. The signal is allowed to rest in red if no actuations of the detectors take place. To extend the green continuously under these conditions, at least one vehicle should always be in one of the detection areas associated with a given signal phase. And, when the traffic flows are light, the delay of a vehicle is primarily attributable to the need to decelerate. The curve labeled probabilistic is obtained from the actual probabilistic distributions of the arrival and the departure times. The other curve labeled deterministic is based on the averages given in Table 3.

When the flows are light ($Q \leq 400$ vph), queue lengths are short and the opportunity for premature termination of the green rarely exists. As a result, the use of averages brings about only slightly higher estimates (up to 1.5 sec/vehicle). When the flows exceed 500 vph, the averages given in Table 3 overestimate the probability of premature termination of the green. The resulting estimates of average delays deviate significantly from the probabilistic estimates.

CONCLUSIONS

The stability characteristics of simulation results can be assessed in terms of (a) their dependence on or independence of the length of simulation period, (b) within-run variation, and (c) between-run variation. These characteristics are inherent in the operation of a signal control, but they complicate the application of simulation models.

It is convenient to infer from simulation results that one alternative improves the efficiency of a control by a certain percentage. But it is more difficult to determine whether the alleged improvement is real. For proper interpretation of simulation results, a user first has to know whether the results are time dependent. Furthermore, when uniform flow rates are used as inputs and the operation of a control is not in an unstable state, a user has to ensure that the results from a simulation run represent stabilized estimates of the performance of a control.

A simple algorithm may be incorporated in a model to detect the existence of an unstable-state operation. The algorithm should also be capable of determining whether estimates of measures of performance have reached stabilized values. Such an algorithm should at least be implemented for system-wide estimates and, at the option of a user, for estimates related to individual components of a system. To assist users in the interpretation and comparison of simulation results, a model should provide the following outputs for key measures of performance: estimated mean, standard deviation of the mean, and number of observations. Estimated means recorded as a function of elapsed real time are also useful.

Between-run variations always exist in microscopic simulation of signal control. Because of this, replicated runs are necessary unless the accuracy of simulation results is not a concern. Introducing between-run variations into a single run can eliminate the need for replicated runs. But no established mechanisms are available for such an application.

When replicated runs are performed, it is advisable to obtain an equal number of observations for each run. This will simplify the estimation of the mean of a measure of performance and its standard deviation. Negatively correlated simulation runs are preferred to independent runs. They can reduce the number of runs needed to achieve a specific level of estimation accuracy.

The use of a single string of random numbers would make comparative analysis based on identical sequences of events impossible. One should realize, however, that in a simulation analysis only arriving headways can really be rigidly controlled for comparisons of alternatives. Other events resulting from the interaction of flows and signals will vary from one alternative to another. Only the probability distributions of such events can be controlled through the use of multiple strings of random numbers. In generating the distribution of an event, the use of multiple strings does not appear to have a real advantage over the use of a single string. In fact, both approaches may fail to generate a desired distribution. Therefore, it may be more important to ensure that the use of a seed number will generate a distribution correctly.

Aggregation of data to provide inputs runs the risk of incurring biased estimates. The bias may result from reduced variations in representing the real world. It may also result from incorrect representation of the operation of a control. Whenever possible, it is wise to avoid the use of aggregated data, particularly when heavy flows are involved.

The discussion presented herein perhaps overemphasizes the need to determine the level of accuracy of estimated measures of performance. The profound problems associated with the application of simulation models may be understated. The accuracy check of simulation results is meaningless unless a model is capable of generating reliable information. Furthermore, the level of accuracy of estimates has only to be compatible with the purpose of a simulation study. In any case, interpretation of simulation results should be done very cautiously if their accuracy is uncertain.

REFERENCES

1. Peat, Marwick, Mitchell, and Co. Network Flow Simulation for Urban Traffic Control System--Phase II: Vol. 4, User's Manual for UTCS-1 Network Simulation Model. FHWA-RD-73-86. FHWA, U.S. Department of Transportation, May 1973.

2. Traffic Network Analysis with NETSIM, A User Guide. FHWA-IP-80-3. FHWA, U.S. Department of Transportation, Jan. 1980.
3. Application of Traffic Simulation Models. TRB Special Report 194. TRB, National Research Council, Washington, D.C., 1981, 114 pp.
4. F.V. Webster. Traffic Signal Settings. Road Research Technical Paper No. 39. Great Britain Road Research Laboratory, London, England, 1958.
5. K.A. Brownlee. Statistical Theory and Methodology in Science and Engineering. Wiley, New York, 1965.
6. D.R. Drew. Traffic Flow Theory and Control. McGraw-Hill, New York, 1968.
7. J.R. Emshoff and R.L. Sisson. Design and Use of Computer Simulation Models. MacMillan, New York, 1970.
8. S. Yager and E.R. Case. Summary Evaluation of UTCS-1/NETSIM in Toronto. In TRB Special Report 194: Application of Traffic Simulation Models, TRB, National Research Council, Washington D.C., 1981, pp. 95-99.
9. S. Yager and E.R. Case. Using TRANSYT for Evaluation. In TRB Special Report 194: Application of Traffic Simulation Models, TRB, National Research Council, Washington, D.C., 1981, pp. 100-103.
10. D.I. Robertson. TRANSYT: A Traffic Network Study Tool. RRL Report LR253. Road Research Laboratory, London, England, 1969.

Publication of this paper sponsored by Committee on Traffic Control Devices.

Evaluation of Engineering Factors Affecting Traffic Signal Change Interval

MYUNG-SOON CHANG, CARROLL J. MESSER, and ALBERTO SANTIAGO

ABSTRACT

Engineering factors affecting traffic signal change interval (yellow and all-red) are reviewed, particularly in terms of drivers' perception and brake reaction time (t), and deceleration rate (d). Using driver behavior data collected from time-lapse cameras, two hypotheses were tested: (a) t and d are dependent on speed and (b) there is an interactive effect of t (prebraking) and d (postbraking) on drivers' braking distance. All hypotheses from statistical analysis results were accepted. It is concluded that joint consideration instead of independent consideration should be given to t and d when selecting their values. Furthermore, it is recommended that different t's and d's for different approach speeds should be used rather than a single value (as used in current practice) for all approach speeds in signal change interval design.

To stop or not to stop? That is the question asked when drivers see a green light ahead change to yellow and then to red. If the driver fails to respond safely, a major right-angle collision at the intersection is possible. On the other hand, if the driver overreacts, a hazardous rear-end collision is likely. Because of the complexity of the driver-vehicle-traffic control system involved and the potential severe consequences of system failure (a fatal accident), the design of the traffic signal yellow time and any following all-red interval should be optimized based on the best understanding of the engineering factors involved. The magnitude of the problem requires that traffic engineers do no less.

PERCEPTION AND BRAKE REACTION TIME (t)

AASHO (1) recommends a total of 2.5 sec for perception and brake reaction time. The Institute of Traffic Engineers (ITE) Handbook (2,3) assumed a perception-brake reaction time for a yellow signal of 1 sec. Actual drivers' stopping distance data reported by Williams (4) and Sheffi (5) are analyzed using different deceleration rates from 8 to 15 feet per second per second (fps²). The results, which indicate three categories of t under normal driving behavior, are as follows:

1. Forced stopping: When more than 85 percent of the drivers go through the intersection, those 15 percent or less of the drivers who decide to stop take less than 1 sec of perception and brake reaction time.
2. Indecision stopping: When half of the drivers decide to stop, the perception and brake reaction time is 1 to 1.5 sec.
3. Comfortable stopping: When the majority of drivers decides to stop, their perception and brake reaction time is 1.5 to 3.0 sec.

An analysis of Williams' (4) and Sheffi's (5) results also indicates that perception and brake