Variability Analysis of Traffic Simulation Outputs: Practical Approach for TRAF-NETSIM

RAHIM F. BENEKOHAL AND GHASSAN ABU-LEBDEH

Stochastic traffic simulation models, such as TRAF-NETSIM, use random number seeds to generate variables to describe driver, roadway, and traffic characteristics. In analyzing outputs from these models, one should consider the variability of the responses. The variability of NETSIM's output using the methods of replication and batch means was explored. For the batch means method, it is proposed to compute the measures of effectiveness (MOEs) for intermediate time intervals using a proposed interval calculation (PIC) procedure. The MOEs were evaluated at the network, intersection, and link levels of aggregation. Depending on the MOE and level of aggregation, the two methods yielded significantly different results. Hence, depending on the study objective, outputs may need to be examined at different levels of aggregation to obtain meaningful results. The practical implications of the variability are also discussed, and statistical approaches are proposed to deal with output variability. Auto- and cross-correlations must be examined explicitly, particularly when dealing with link MOEs resulting from very short simulation time. Ignoring positive cross-correlation is not detrimental but leads to more conservative confidence intervals. Either the batch means with PIC method or replication method must be used to build confidence intervals. NETSIM's direct output for intermediate time intervals should not be used to build a confidence interval unless an autocorrelation analysis is done. Not using proper statistical procedures can lead to erroneous and misleading conclusions.

Computer simulation models have been used, as a decision tool, to evaluate the effects of alternative traffic control strategies. Simulation results, however, will vary when either the sequence of the random numbers used (internal variables) or the input variables (external variables) are changed. Stochastic traffic simulation models, such as TRAF-NETSIM (henceforth NETSIM) (*1*), use random number seeds (RNSs) to generate random variates to describe roadway, traffic, and driver characteristics. Using a different RNS changes the outcome of the simulation model.

Misleading and erroneous conclusions may be obtained if the variability in NETSIM's output is not considered, especially when alternative measures are being evaluated or the effects of certain traffic-related changes are being quantified. There is, however, no clear guideline on ways to handle the output variability, the length and number of runs needed, and the magnitude of the effects of internal variables on the simulation results. This paper presents a practical approach to deal with these issues. It also examines autocorrelation and cross-correlation issues in NETSIM.

R. F. Benekohal, Civil Engineering Department, University of Illinois at Urbana-Champaign, 205 North Mathews Avenue, Urbana, Ill. 61801. G. Abu-Lebdeh, Champaign-Urbana Urbanized Transportation Study, 1303 North Cunningham Avenue, Urbana, Ill. 61801.

PROBLEM STATEMENT

Consider a situation in which one wants to compare two conditions, such as assessing the traffic impact of a new office park on an existing network, using NETSIM. When running a stochastic simulation model such as NETSIM, a few options are available. The easiest and the most widely used option is running NETSIM for conditions "with" and "without" the office park traffic and comparing the results. However, one may get misleading results using this approach, as illustrated by the following example. Assume that the measure of effectiveness (MOE) used is the average delay per vehicle. Assume that the delay for the base condition is X and that when the office park traffic is added to the network the delay is Y. Does $(Y - X)$ provide enough information to assess the impact of office park traffic? If it does, how large should $(Y - X)$ be to be considered a significant impact?

These questions appear to apply when other software, such as Highway Capacity Software (HCS) (*2,3*), is used for impact assessment. However, a major difference between NETSIM and models such as HCS is that the former is a stochastic model and the latter is a deterministic model. Thus, for a given traffic and roadway condition, HCS results do not vary when the sequence of vehicle arrivals is changed. However, NETSIM's results do vary when the sequence of random events (e.g., sequence of arrival of vehicle) is changed even though traffic and roadway conditions remain unchanged.

In NETSIM the characteristics of a vehicle-driver unit are assigned randomly upon arrival of that vehicle into the system. Assume that using an RNS, the sequence of arrival of vehicles on a given approach is red car, blue car, and white car. Using a different RNS (while keeping traffic and roadway conditions the same) may result in the sequence white car, red car, and blue car. Running NETSIM with these two sequences of arrivals yields different results. It should be noted that in both runs, all input data remain the same except the cars' order of arrival.

Considering that NETSIM's outputs vary due to changes in the internal variables, one should not rely on the difference between Y and X without knowing the variability caused by the change in the sequence of random events. One cannot assess the impact of the office park correctly by getting two delay values from two long runs of NETSIM. Another important question is whether the impact should be assessed at the network level, intersection level, or approach level. If link and intersection data are used, which link or intersection should be used to represent the impact of the office park traffic? If network data are used, how large should the network be?

There are several approaches to dealing with output variability (*4*); the two most widely used are replication and batch means. For several well-known queueing and inventory systems, Law compared the batch means and replication methods and concluded that batch means was superior (*5*). However, Law's findings are not directly applicable to traffic simulation models because traffic flow in urban networks does not necessarily resemble queue behavior.

Gafarian and Halati indicated that developing a confidence interval on the basis of a single run of NETSIM (batch means method) is extremely complex because it involves estimating auto- and cross-correlations of numerator and denominator variables (*6*). This is so because certain NETSIM MOEs, such as average speed, are estimated on the basis of ratios of sample means of observations that are auto- and cross-correlated. They analyzed the output directly produced by NETSIM (henceforth BM/direct) for a single intersection and recommended not using the batch means method. They suggested using the replication method and considering the covariance of the numerator and denominator variables. (It should be noted that the direct NETSIM outputs are cumulative statistics that are inherently correlated.)

Chang and Kanaan applied NETSIM to a congested isolated intersection to assess the variability of NETSIM's output (*7*). They used replication and BM/direct approaches and suggested using the approach given by Fishman (*8*) to find the appropriate batch size. The procedure is complex because it needs an autoregressive analysis. The use of the direct method could explain, at least partly, the correlation that existed among the data that Chang and Kanaan used. Thus, it is not clear when and how the batch means or replication method may be used.

BACKGROUND

The applications of NETSIM have been numerous and diverse, ranging from simulation of complex conditions (*9*) to simulation of simple systems such as alternative control strategies at single intersections (*10*). NETSIM proved to be a very powerful and flexible tool for the entire range of applications. Several case studies involving NETSIM have demonstrated clearly that the program can be used effectively to simulate unconventional settings in which traffic other than automobiles, buses, or trucks is involved (*11–13*).

The NETSIM applications can be grouped into three general categories:

1. Studies evaluating traffic control and geometric alternatives.
2. Studies assessing alternatives as well as NETSIM itself.
3. Studies focusing on NETSIM. Studies in this category can be divided into three general subcategories: those aimed at
   - –Addressing the degree of accuracy of NETSIM by either comparing its results with results from other software or validating them in the field,
   - –Dealing with the issue of variability of NETSIM's output, and
   - –Exploring NETSIM's potential and flexibility as well as its strengths and weaknesses.

This classification is subjective and some studies may fit into more than one group.

Category 1 includes studies that used NETSIM to evaluate different geometric alternatives (*14*) and different signal control strategies and timing plans to optimize travel time and fuel consumption (*11,15–19*). Yauch et al. (*12*) used NETSIM to assess the impact of drawbridges, and Luedtke (*13*) used it to simulate the impact of light rail transit on signal operations. Others used it to compare the results from different software with those of NETSIM (*20,21*). Rathi and Lieberman used NETSIM to evaluate the potential effects of restricting traffic flow on approaches to a congested urban street network (*9*). Papacostas and Willey used NETSIM for analyzing the traffic impact of a real estate development (*22*). Only in limited cases were field observations taken to verify the results. For the most part, however, little or nothing was done to address or account for the issue of variability of the program's output.

Category 2 includes studies that had focuses similar to those of Category 1, except NETSIM itself was of equal importance as some efforts were made to verify or validate the simulation results. Hurley and Radwan used NETSIM to study different aspects of traffic flow and to estimate the impact of various traffic control settings on fuel consumption and delay (*23*). Wong compared field observations with capacity and level of service estimates from NETSIM and HCS (*24*). Radwan and Hatton used NETSIM to simulate traffic operations at conventional and single-point diamond interchanges (*25*). Ten replications were used to minimize the effect of variation. Kim and Messer used NETSIM to evaluate different control strategies for saturated signalized diamond interchanges (*26*). To account for output variability, the same RNSs were used for the paired simulation trials. Torres et al. (*27*) used NETSIM to estimate the impact of lane obstruction on arterial streets. To account for output variability, three replications were used to quantify the significance of any particular combination of factors.

The first subcategory includes those studies that aimed at evaluating NETSIM by comparing its results with those of other programs or with field data. Yagar and Case addressed the issue of sensitivity of NETSIM/UTCS-1 to aggregation of traffic flows and to the RNS (*28*). Davis and Ryan used NETSIM to estimate delay and queue length at an isolated intersection (*10*).

The second subcategory includes those studies that dealt specifically with variability of the NETSIM's output. This issue has come up in several of the studies that used the model, but for the most part the users did not address it. Those who did mostly used multiple runs (*19,21,25–27*). However, the degree of success in neutralizing the effect of variability was hardly addressed. Recent acceptance and widespread use of NETSIM, along with improved features of the program gave rise to new efforts aimed at reducing or otherwise dealing with such variability through variance reduction techniques.

Variance reduction techniques were applied to NETSIM to assess their effectiveness. Rathi and Santiago used the common RNS (*29*), and Rathi and Venigalla applied antithetic variates techniques (*30*). Regardless of the variance reduction issue, the authors stressed that statistical analysis must be used to interpret the simulation output data properly. It should be noted that in applying the variance reduction techniques to TRAF-NETSIM, they assumed that the desired correlations (e.g., synchronization) are attainable with the way TRAF-NETSIM generates random variables.

The last subcategory includes studies that aimed at exploring NETSIM's potential and flexibility as well as drawing attention to its strengths and weaknesses. Wong used the detailed simulation capabilities and graphics of NETSIM to estimate capacity and level of service (*31*). Very few guidelines, however, were given on ways to run TRAF-NETSIM and to deal with the output variabilities.

STUDY APPROACH

Analysis Approach

Batch means and replication methods were used to assess the variability in NETSIM's MOEs due to changes in the internal variables.

These methods were applied to two cases: the base condition (Case 1) and the base condition with volume added (Case 2). For each case, the MOEs for the network, a typical link, and a typical intersection were analyzed using the two methods.

The results for the replication method were obtained by running NETSIM 24 times each for 10 min. The independent replications were achieved by using 24 RNSs from a random number table (*32*). The selected seeds satisfied the NETSIM's requirements for RNS. The results from the batch means were obtained by running NETSIM once for 4 hr, with intermediate results computed every 10 min (batch size = 10 min) using the proposed interval calculation (PIC) procedure.

Proposed Interval Calculation Method

If the true variability of the different MOEs among the intermediate intervals of a long run is to be determined, the statistics must be calculated for the individual batches. The PIC method computes such MOEs. In the PIC method, vehicle trips and phase failures are computed by finding the differences between successive batches. Delays and speeds are computed as described in the following.

To compute the average delay per vehicle for each batch, the total time—which is the sum of the link travel times for all vehicles in that batch—is divided by the number of vehicle trips in that batch. To find the number of vehicle trips or the total travel time for a specific batch, the previous intermediate output values are subtracted from the current output values.

To find the average speed of vehicles during each batch, the total number of miles driven within that batch is calculated and then divided by the total time the vehicles spent during that batch to complete those miles. For illustration, batch statistics for the average delay time and speed are calculated here using direct NETSIM statistics given in Table 1.

PIC delay = (44.54 − 22.07) * 60/(2,227 − 1,110) = 1.207 min/veh-trip
PIC speed = (1,053.03 − 529.36)/(79.64 − 39.71) = 13.11 mph

NETSIM's direct results (BM/direct) were not used. The BM/direct method gives the mean MOEs for the entire simulation run up to that time, not the mean MOEs for each batch. Since confidence intervals and statistical tests cannot be constructed for the BM/direct method, it is dropped out of any further discussion. In some graphs the results from the BM/direct method are shown for comparison.

Description of Network

The batch means and replication methods were implemented on a nine-intersection network in downtown Champaign, Illinois (Figure 1). Dummy nodes were introduced to collect statistics at entry links. Traffic signals had cycle lengths of 60 sec, and the overall network was not congested.

MOEs Used

Three MOEs will be examined: average delay, average speed, and vehicle trips. In NETSIM, delay time is defined as the difference between the actual time that a vehicle spends in the system and the ideal amount of time based on free-flow speed. Speed is defined as the ratio of the total distance traveled by all vehicles in the system to their total travel time. A vehicle completes a trip on a link when it passes the stopline.

TABLE 1 Direct NETSIM Statistics Used To Calculate Average Delay Time and Speed

Elapsed Time	Vehicle-mi	Trips	Delay Time (veh-hr)	Total Time (veh-hr)
0:10:00	529.36	1,110	22.07	39.71
0:20:00	1,053.03	2,227	44.54	79.64

Statistical Methods

Batch Means Method

The batch means method is performed by running the simulation model for one long run and then dividing it into smaller time intervals (batches). For each batch, statistics are collected and variability among batches is used to build a confidence interval on the simulation output. If the batches are long enough, the means from the batches may be uncorrelated. Increasing the length of the batches may reduce their autocorrelation. The advantage of batch means over the replication method is in having only one initialization period. However, the correlation problem between batches, the length of each batch, and the number of batches must be determined carefully.

Replication Method

The replication method is performed by running the simulation for a number of independent runs. The independent simulation runs are made for the same roadway and traffic conditions. Each run will have an initialization time until the system reaches equilibrium condition. After the warm-up time, statistics on system performance are collected. The advantage of this method is that the autocorrelation is eliminated. However, one must know the length of each run and the number of replications. These will depend on the range of variability of the responses. One obvious disadvantage of this method is that each run needs its own initialization period. However, with the speed of today's computers, the initialization for each run may not be prohibitive.

Correlation Among Batches

Two methods are proposed for dealing with the correlation issues among batches. One is using time series analysis, and the other is checking the correlation coefficient. When a long run of NETSIM is divided into batches, statistics for these batches may be treated as stationary time series data. A time series is considered stationary when statistical properties (e.g., mean and variance) of the time series are essentially constant over time. Plot of the mean value for each batch against time will help to determine visually whether the time series is stationary. Analytical techniques can also be used to determine whether a time series is stationary.

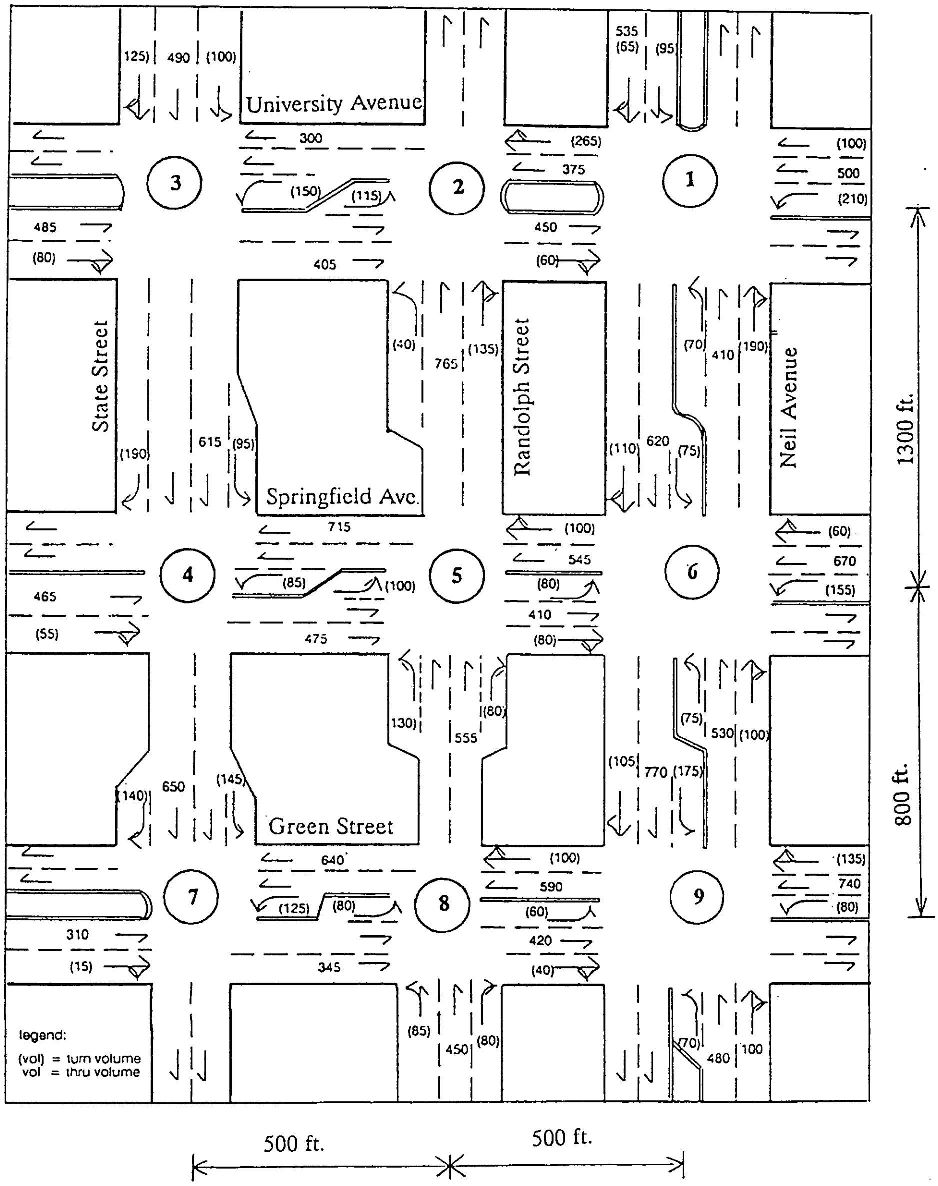

FIGURE 1 Network geometry and traffic volumes.

For a stationary time series of $Z_b, Z_{b+1}, \ldots, Z_n$, the sample autocorrelation at lag k, denoted by r_k, is computed from the following equation (*32*):

$$r_k = \frac{\sum_{t=b}^{n-k} (z_t - \bar{z})(z_{t+k} - \bar{z})}{\sum_{t=b}^{n} (z_t - \bar{z})^2}$$

where

$$\bar{z} = \frac{\sum_{t=b}^{n} z_t}{(n - b + 1)}$$

r_k measures the linear relationship between time series observations separated by a lag of k time units. The value of r_k will be between -1 and $+1$. When r_k is close to -1 or $+1$, the observations separated by a lag of k time units have a strong tendency to move together in a linear fashion (*33*).

The standard error of r_k is

$$S_{rk} = \frac{\left[1 + 2\sum_{j=1}^{k-1} (r_j)^2\right]^{1/2}}{(n - b + 1)^{1/2}}$$

The t_{rk} statistic is

$$t_{rk} = \frac{r_k}{S_{rk}}$$

TABLE 2 Autocorrelations for Average Delay for Lags 1–12

Network

Lag	r_k	$S_{r\,k}$	Lag	r_k	$S_{r\,k}$
1	.00939	.20412	2	-.08439	.20414
3	.12182	.20559	4	.06076	.20858
5	-.15485	.20931	6	-.14545	.21403
7	-.00879	.21811	8	-.01030	.21813
9	-.26227	.21815	10	-.05152	.23091
11	.17652	.23139	12	-.14364	.23694

Intersection 6

Lag	r_k	$S_{r\,k}$	Lag	r_k	$S_{r\,k}$
1	-.14666	.20412	2	-.29162	.20847
3	.25560	.22482	4	-.05741	.23662
5	-.03321	.23720	6	-.07393	.23740
7	-.21173	.23835	8	-.00162	.24607
9	.08054	.24607	10	-.05383	.24716
11	.02738	.24765	12	-.07025	.24778

Link 9-6

Lag	r_k	$S_{r\,k}$	Lag	r_k	$S_{r\,k}$
1	.15073	.20412	2	-.20663	.20871
3	-.18384	.21707	4	.14179	.22346
5	-.05707	.22718	6	-.36020	.22777
7	-.18585	.25039	8	.19916	.25607
9	.30415	.26244	10	-.01755	.27674
11	-.24191	.27679	12	-.05680	.28546

For the network and link delay r_k and s_{rk} values are given in Table 2. The values indicate that there was not a strong correlation at any lag. Lag 1 particularly is of interest to us, because it would indicate how strongly the adjacent batches are correlated. Furthermore, the t_{rk} values are smaller than 1.6. The t_{rk} values greater than 1.6 are considered to be statistically large for lags of 1, 2, and perhaps 3 (*33*). This indicates that there are no spikes at any lags in the data.

Another way of checking correlation between adjacent batches is looking at the correlation coefficient. A procedure that is much simpler than the time series analysis is suggested to examine the correlations among batch means. Correlation coefficient and Lag 1 autocorrelation would provide the same results. Thus, one may compute correlation coefficient (r) if dependency between adjacent batches are considered. The r-values will be between -1 and $+1$. Finding r and interpreting it is much easier than autocorrelation analysis.

DISCUSSION OF RESULTS

Table 3 contains a summary of the differences between the two methods or cases at the three levels of aggregation. The Yes or No entries indicate whether or not the differences between the values shown were statistically significant. Figures 2 and 3 depict the results graphically for the network and Link 9-6, respectively. Only discussion is provided for Intersection 6.

Case 1: Base Network

Average Delay

The average delays from the BM/PIC method were significantly higher than those from the replication method. Delay from the BM/direct method appears to converge to a constant value as the duration of the simulation increases. Some authors incorrectly considered this to be a sign of the stability of NETSIM's output (*30,34*). The convergence does not indicate that a "stable" condition is reached. This convergence is not a real phenomenon in the simulation model or real-world traffic; it occurs because the delay is computed from cumulative statistics. In fact, it is more realistic to have fluctuation in delay than the convergence. The delays computed by the BM/PIC method clearly show that NETSIM does not converge to a value. The cumulative statistics used in the direct method conceal this fluctuation.

Average Speed

The mean speeds estimated by the PIC method exhibited wider range and larger variance. The average speeds estimated by the PIC method were significantly lower than those from the replication method. Similar to delay, speeds from the direct method do not show the true speed fluctuation among batches. It is incorrect to assume that speed reaches a constant value when the duration of simulation is long. The relative speed differential between the two methods is much more pronounced at the link level than at the network level.

Vehicle Trips

The vehicle trips data are important because they are used to compute many of the NETSIM MOEs. At the network level, vehicle trips estimated by the replication method were significantly lower than those estimated by the BM/PIC method. At the link level, the estimated vehicle trips from the BM/PIC and replication methods did not show a statistically significant difference. At the intersection level, the mean vehicle trips estimated by the BM/PIC method are slightly higher than the replication method, but the difference is not statistically significant.

Case 2: Base Network with 120 Through Vehicles Added

The results for Case 1 indicated that the two methods, for the most part, give different results, even though the differences may seem small for practical purposes. The absolute difference between the two methods is not as important as the relative difference, which shows how much more traffic should be added to the network to increase the delay by the amount equal to the difference between the two methods. In fact, one needs to find out how much additional traffic would cause changes similar to those noted between the two methods, and how much added traffic can be "handled" within the internal variability of NETSIM.

In practical terms, one can ask whether NETSIM is sensitive enough to be used for traffic impact studies. Of particular importance is whether the impact assessment should be measured at the

TABLE 3 MOEs for Three Levels of Aggregation and Results of Comparisons Between Different Methods or Cases

Method/Case	Delay			Speed			Vehicle Trips		
	Network (sec./v.trip)	Link 9-6 (sec/v.trip)	Intersection 6 (sec/v.trip)	Network (mph)	Link 9-6 (mph)	Intersection 6 (mph)	Network (v.trips)	Link 9-6 (v.trips)	Intersection 6 (v.trips)
Replication	73.2	29.66	37.33	12.9	11.5	9.9	1110	473	108
Replication Compared to BM/PIC	Yes[1]	Yes	Yes	Yes	Yes	Yes	Yes	No[2]	No
BM/PIC	75.0	32.09	42.82	12.7	11.1	9.1	1123	475	109
BM/PIC Compared to BM/PIC- Added	Yes	Yes	No	Yes	Yes	No	Yes	Yes	Yes
BM/PIC-Added	83.4	35.83	43.06	12.0	10.7	9.0	1143	496	129

Yes[1]: Difference is statistically significant.
No[2]: Difference is not statistically significant.

network, link, or intersection level. To answer these questions, 120 through vehicles were added to all northbound links composing Neil Street. The number was selected such that all of the added vehicles will go through the links without oversaturating them. Output for network, Link 9-6, and Intersection 6 with the 120 vehicles added (henceforth BM/PIC) were used for comparison purposes.

At the network level, adding 120 vehicles greatly affects average delay, average speed, and number of vehicle trips. For Link 9-6, however, contrary to expectations, the changes in delay and speed are not significant, whereas changes in vehicle trips and phase failures are. At the intersection level, adding 120 vehicles has a significant impact on the three MOEs. Further examination of the other links indicated that the added volume affects the MOEs at links that did not receive additional traffic. This unrealistic effect is due mainly to the effects of external variables on the internal variables in NETSIM.

It should be noted that for any pair of comparisons, the MOEs either decreased or increased at all three levels. However, the amount of increase or decrease was markedly different. Although the changes in delay and speed for Link 9–6 due to the 120 added vehicles were not significant, the same changes attributed to the change of procedure (replication versus BM/PIC) were significant. It is conceivable that the reserved capacity of Link 9-6 may have partly concealed the impact of the 120 vehicles.

Practical Implications and Proposed Approach

The results presented in this paper have significant implications to the users of NETSIM. Such issues as the method to use (replication, BM/direct, or BM/PIC), the length and number of runs, and the amount of error to be tolerated were shown to be relevant. Given the stochastic nature of NETSIM, it may not be possible, or even necessary, to provide ready-to-use answers to all of these questions, but there should at least be some guidelines. The following section provides one possible approach to dealing with such issues.

The first question is that of which method to use: batch means (direct and PIC) or replication. Since statistical tests and confidence intervals cannot be performed easily with the BM/direct method, the question becomes whether to use the replication or BM/PIC method.

Other factors to be considered in selecting a method include human resources, computer time (both initialization and simulation), size of system, and previous experience with NETSIM. For large networks in which the initialization period is likely to be longer, it would be inconvenient and time-consuming to use the replication method. Replication will require more runs and may need more human resources, although this can be overcome with some programming. Currently, both methods require considerable time to perform statistical analyses. For the typical uncongested traffic system, neither method offers a clear advantage. The number of technical considerations could be the deciding factor.

Replication Method

The replication method has the advantage in that autocorrelation is eliminated as independence of observations is ensured through the use of different RNSs. The questions, then, are on the number and length of the runs—which are related characteristics. Answering the first question will automatically answer the second.

To start with, users can make *X* (say 10) runs of reasonable length. The length of each replication run would depend mainly on the size of the network and traffic conditions. In general, it should

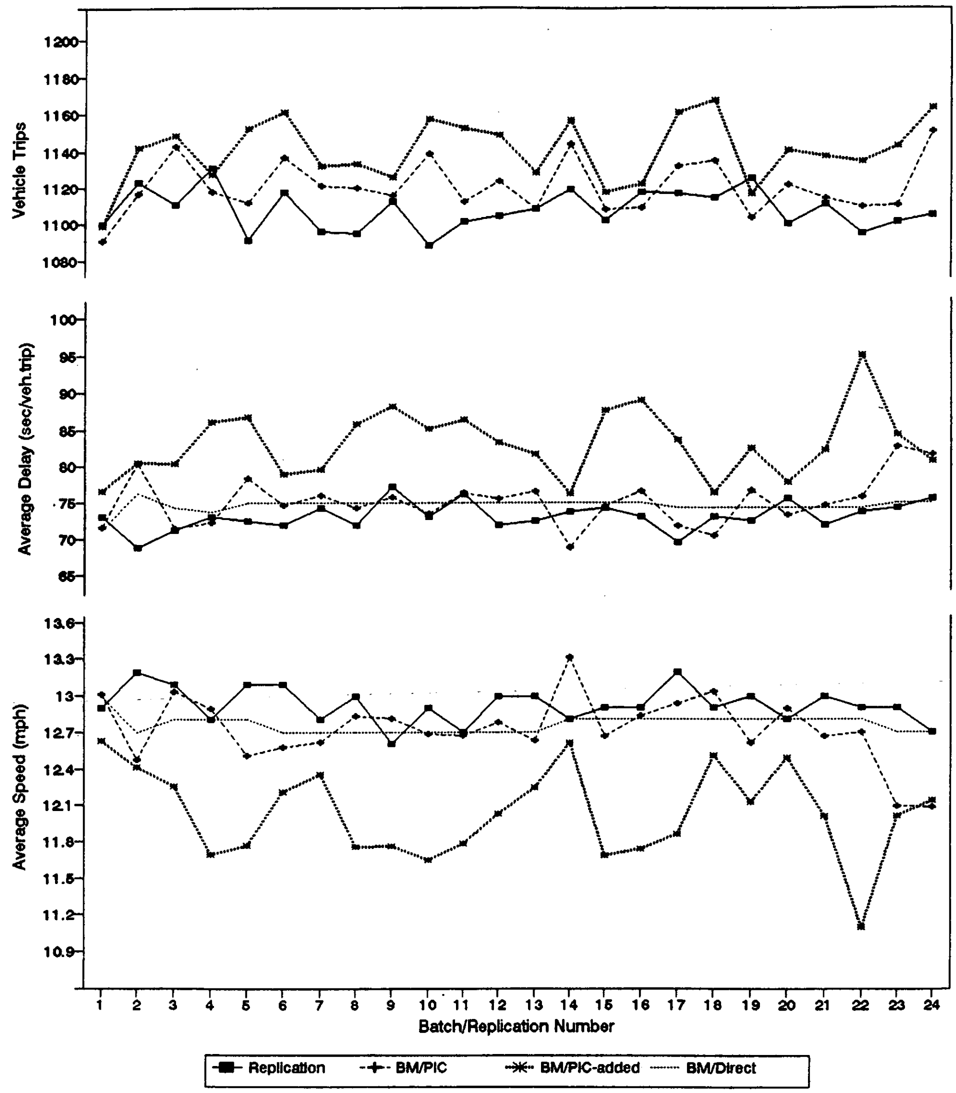

FIGURE 2 Comparison of delay, speed, and vehicle trips at network level.

be long enough that the user believes that it reflects real-world conditions. It is proposed to use a simulation time that is at least 10 signal cycle lengths for small networks. Another proposed way of estimating duration is to make it at least as long as the initialization time given by NETSIM.

Next, find the mean and variance for those observations (runs). Select a confidence level (usually 90 or 95 percent), then calculate the number of observations needed (n), which is a function of the standard deviation and the tolerable error, using Equation 1.

$$n = (ts)^2/e^2 \quad (1)$$

where

s = standard deviation of observations,
e = tolerable error, and
t = critical t-value from a t-distribution table for $(n - 1)$ degrees of freedom and selected confidence level.

If n is greater than X, make a few more runs and repeat the procedure. Once the computed n is less than X, stop. Construct confidence intervals using the results from X runs.

The tolerable error depends on the accuracy that the user desires, which depends on the distribution of data around the mean. For widely dispersed data, the tolerance can be high, but if the data are concentrated around the mean, the tolerance will be low. Thus, it should be a function of the range and dispersion of the data. It is suggested to use 5 to 15 percent of the range when a data set does not have extreme values. When a data set contains extreme values, discard them first and then find the range.

BM/PIC Method

If a user decides to use the BM/PIC method, the length of the simulation run, T, must be determined. T should be divided into X intervals (e.g., start with 10 intervals), each being T/X. The value of T/X

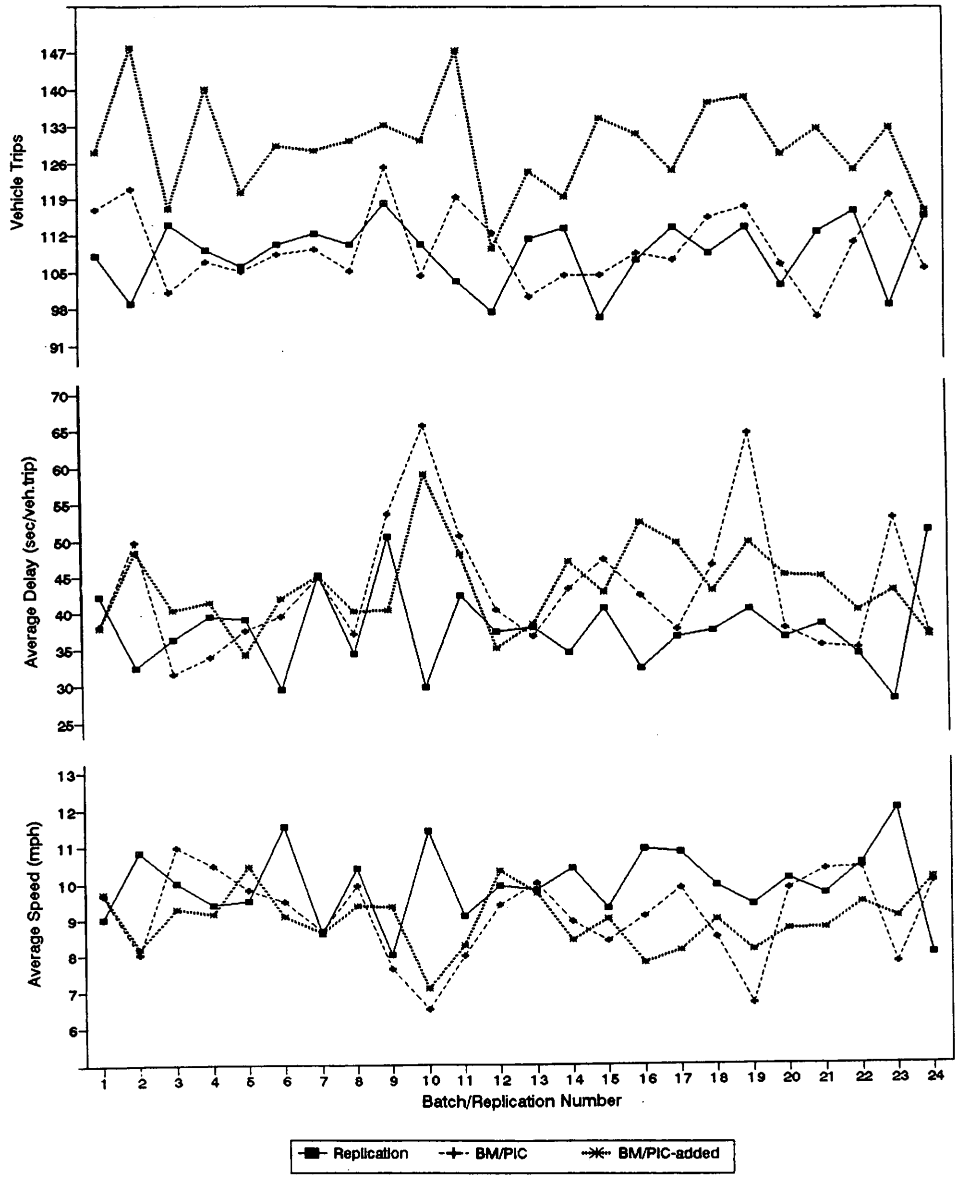

FIGURE 3 Comparison of delay, speed, and vehicle trips at link level.

depends on the size of the network, anticipated correlation among batches, traffic conditions, and other factors. It is suggested that the starting value for *T/X* be equal to an integer value of the cycle length, say, 10 average cycle lengths (if cycle length is 60 sec, starting *T/X* would be 10 min) or at least to the initialization time. Knowing the starting value of *T/X* and *X*, one can determine *T*. Run the simulation and compute MOEs for the *X* batches. Then compute the mean and variance for these batches.

Now, one needs to examine the correlation among batches using the procedures outlined before. If correlation is not significant, one may use this batch size to find the number of batches (shortening the batch size until the correlation becomes significant is optional at this point). To determine the number of batches, use the procedure described in the replication section. If the correlation is a problem, the batch length should be increased. Fishman suggested doubling the batch size to expeditiously arrive at an appropriate batch length (*8*). In the context of this study, it corresponds to doubling the batch length. However, one may consider increasing batch size at a slower rate, perhaps by 50 percent at a time.

Using the new batch size, run the simulation and examine the correlation again. If the correlation is no longer significant, then use the procedure described before to find the number of batches. Knowing the batch size and number of batch, one can easily find the length of the simulation run. However, if the correlation problem persists and the batch length becomes unreasonably long, consider using the replication method.

Choosing a Method

If tests did not reveal any serious correlation problems (with the BM/PIC results), the user is basically free to use either method.

However, users need to be aware of the following characteristics exhibited by output from each method:

- *The vehicle trips estimated by the replication method are consistently lower than the batch means.* The delays estimated by the replication method are consistently lower than the batch means (which is partly due to the lower estimate of vehicle trips).
- *The variability of the BM/PIC method output tends to be higher than that of the replication.* As indicated earlier, statistically there is no clear-cut advantage of using one method over the other for traffic simulation.

CONCLUSIONS AND RECOMMENDATIONS

This study explored the variability of NETSIM's output when batch means and replication methods are used. For purposes of statistical analysis, it is proposed to compute the MOEs for intermediate time intervals using the PIC procedure. The proposed PIC procedure is used with the batch means method.

Three MOEs were evaluated at three levels of aggregation; network, intersection, and link. The BM/PIC method resulted in average delays that are significantly higher than those of the replication method at network, intersection, and link levels. The opposite trend was true for the average speeds. Depending on which MOE is being examined and at what level of aggregation, it was apparent that the two methods can issue very different results.

The practical implications of variability were discussed, and approaches were proposed to help users utilize NETSIM and properly account for its variability. Suggestions also were made with regard to batch length, batch size, and number of runs.

The batch means (when the MOEs are computed by the PIC method) or replication methods should be used to compute confidence intervals for the MOEs. Depending on the objective of the study, the responses may need to be viewed at different levels of aggregation in order to properly assess the magnitude of the phenomena being studied. Not using a proper statistical procedure to make inferences about the results can lead to erroneous results. Output obtained directly from NETSIM for intermediate time intervals should not be used to build confidence intervals because the responses are autocorrelated, and as such complicated statistical procedures are necessary to construct proper confidence intervals.

Auto- and cross-correlation may exist for MOEs collected for a link for a very short period (e.g., 1 min). However, the correlations are likely to become weaker when data are collected at the intersection or network level and for a longer period. Cross-correlation and its magnitude should be examined, particularly when a significant negative correlation exists. Ignoring a positive correlation results in more conservative confidence interval.

Finally, Gafarian and Halati ruled out the use of a single long run (batch means method) to build confidence intervals on the MOEs because the direct output from NETSIM is autocorrelated and may be cross-correlated (*6*). They recommended using the replication method and including covariance of the variables to construct confidence intervals for the MOEs that are cross-correlated. When significant negative cross-correlation exists, the authors concur and recommend including a covariance term regardless of which method is used. However, when cross-correlation is not significant or is positive (which it is in most cases), not including the covariance term would result in a more conservative estimation of confidence intervals, which is desirable. Note that cross-correlation exists only when the MOE of interest is actually a ratio of the means. The BM/PIC method should be built into NETSIM's output structure. A comprehensive study dealing with the sensitivity and output variability of TRAF-NETSIM is recommended.

REFERENCES

1. *TRAF-NETSIM User's Manual.* FHWA, U.S. Department of Transportation, 1992.
2. *Highway Capacity Software.* FHWA, U.S. Department of Transportation, 1987.
3. *Special Report 209: Highway Capacity Manual.* TRB, National Research Council, Washington, D.C., 1985.
4. Law, A. M., and W. D. Kelton. *Simulation Modeling and Analysis*, 2nd ed. McGraw-Hill, Inc., 1991.
5. Law, A. M. Confidence Intervals in Discrete Event Simulation: A Comparison of Replication and Batch Means. *Naval Research Logistics Quarterly*, Vol. 24, No. 4, Dec. 1977, pp. 667–678.
6. Gafarian, A. V., and A. Halati. Statistical Analysis of Output Ratios in Traffic Simulation. In *Transportation Research Record 1091*, TRB, National Research Council, Washington, D.C., 1986.
7. Chang, G. L., and A. Kanaan. Variability Assessment for TRAF-NETSIM. *Journal of Transportation Engineering*, Vol. 116, No. 5, Sept.–Oct. 1990, pp. 636–657.
8. Fishman, G. S. *Principles of Discrete Event Simulation.* Wiley and Sons, New York, 1978.
9. Rathi, A. K., and E. B. Lieberman. Effectiveness of Traffic Restraint for a Congested Urban Network: A Simulation Study. In *Transportation Research Record 1232*, TRB, National Research Council, Washington, D.C., 1989, pp. 95–102.
10. Davis, C. F., and T. A. Ryan. Comparison of NETSIM Results with Field Observations and Webster Predictions for Isolated Intersections. In *Special Report 194: Application of Traffic Simulation Models*, TRB, National Research Council, Washington, D.C., 1981, pp. 91–95.
11. Labrum, W. D. Application of NETSIM Computer Simulation Model to Traffic Control Problems. In *Special Report 194: Application of Traffic Simulation Models*, TRB, National Research Council, Washington, D.C., 1981, pp. 42–50.
12. Yauch, P. J., J. C. Gray, and W. A. Lewis. Using NETSIM To Evaluate the Effects of Drawbridge Operations on Adjacent Signalized Intersections. *ITE Journal*, Vol. 58, No. 5, May 1988, pp. 35–39.
13. Luedtke, P., S. Smith, H. Lieu, and A. Kanaan. Simulating DART's North Central Light Rail Line Using TRAF-NETSIM. Presented at ITE Meeting, the Hague, the Netherlands, Sept. 1993.
14. Hagerty, B. R., and T. L. Maleck. NETSIM: A User Perspective. In *Special Report 194: Application of Traffic Simulation Models*, TRB, National Research Council, Washington, D.C., 1981, pp. 40–42.
15. Maki, R. E., and D. R. Branch. Signal Timing Optimization and Evaluation: Route M-53, Macomb County. In *Special Report 194: Application of Traffic Simulation Models*, TRB, National Research Council, Washington, D.C., 1981, pp. 80–83.
16. Maki, R. E., and J. J. Saller. System Timing Optimization and Evaluation of US-12, Detroit. In *Special Report 194: Application of Traffic Simulation Models*, TRB, National Research Council, Washington, D.C., 1981, pp. 83–86.
17. Schafer, B. F. Comparison of Alternative Traffic Control Strategies at a T-Intersection. In *Special Report 194: Application of Traffic Simulation Models*, TRB, National Research Council, Washington, D.C., 1981, pp. 87–89.
18. Slee, K. L. Signal System Modernization and Time Optimization: Ludington Street, Escanaba. In *Special Report 194: Application of Traffic Simulation Models*, TRB, National Research Council, Washington, D.C., 1981, p. 86.
19. Bruce, E. L., and J. E. Hummer. Delay Alleviated by Left-Turn Bypass Lanes. In *Transportation Research Record 1299*, TRB, National Research Council, Washington, D.C., 1991, pp. 1–8.
20. Nemeth, Z. A., and J. R. Mekemson. Comparison of SOAP and NETSIM: Pretimed and Actuated Signal Control. In *Transportation Research Record 904*, TRB, National Research Council, Washington, D.C., 1983, pp. 84–89.

21. Sadegh, A., and A. E. Radwan. Comparative Assessment of 1985 HCM Delay Model. *Journal of Transportation Engineering*, Vol. 114, No. 2, March 1988, pp. 194–208.
22. Papacostas, C. S., and M. Willey. Use of TRAF-NETSIM to Estimate the Traffic Impacts of an Urban-Resort Area Development. In *Microcomputers in Transportation* (J. Chow, ed.), ASCE, New York, 1992.
23. Hurley, J. W., and A. E. Radwan. Traffic Flow Simulation: User Experience in Research. In *Special Report 194: Application of Traffic Simulation Models*, TRB, National Research Council, Washington, D.C., 1981, pp. 50–54.
24. Wong, J. Y. Comparing Capacities and Delays Estimated by Highway Capacity Software and TRAF-NETSIM to Field Results. *Compendium of Technical Papers*, ITE, 1990, pp. 224–227.
25. Radwan, A. E., and R. L. Hatton. Evaluation Tool of Urban Interchange Design and Operation. In *Transportation Research Record 1280*, TRB, National Research Council, Washington, D.C., 1990, pp. 148–155.
26. Kim, Y., and C. J. Messer. *Traffic Signal Timing Models for Oversaturated Signalized Interchanges*. Research Report 1148-2. Texas Transportation Institute, Texas A&M University, College Station, Jan. 1992.
27. Torres, J. F., A. Halati, and M. Danesh. *Impact of Lane Obstruction*, Vol. 2. Research Report. FHWA Contract DTFH61-84-C-00064. JFT Associates; FHWA, U.S. Department of Transportation, Feb. 1986.
28. Yagar, S., and E. R. Case. Summary Evaluation of UTCS-1/NETSIM in Toronto. In *Special Report 194: Application of Traffic Simulation Models*, TRB, National Research Council, Washington, D.C., 1981, pp. 95–99.
29. Rathi, A. K., and A. J. Santiago. Identical Traffic Stream in the TRAF-NETSIM Simulation Program. *Traffic Engineering and Control*, Vol. 31, No. 6, June 1990, pp. 351–355.
30. Rathi, A. K., and M. M. Venigalla. Variance Reduction Applied to Urban Network Traffic Simulation. In *Transportation Research Record 1365*, TRB, National Research Council, Washington, D.C., 1992, pp. 133–143.
31. Wong, S. Y. Capacity and Level of Service by Simulation: A Case Study of TRAF-NETSIM. *Proc., International Symposium on Highway Capacity*, Karlsruhe, Germany, July 1991, pp. 467–483.
32. Ott, L. *An Introduction to Statistical Methods and Data Analysis*, 3rd ed. PWS-Kent Publishing Company, Boston, Mass., 1988.
33. Bowerman, B., and L. O'Connell. *Time Series Forecasting Unified Concepts and Computer Implementation*. Duxbury Press, 1987.
34. Wong, S. Y. Discussion: Variance Reduction Applied to Urban Network Traffic Simulation. In *Transportation Research Record 1365*, TRB, National Research Council, Washington, D.C., 1992, pp. 143–146.

Publication of this paper sponsored by Committee on Traffic Flow Theory and Characteristics.