Research Article

Improving Precision and Reducing Runtime of Microscopic Traffic Simulators through Stratified Sampling

Khewal Bhupendra Kesur

School of Statistics and Actuarial Science, University of the Witwatersrand, Private Bag 3, Wits 2050, South Africa

Correspondence should be addressed to Khewal Bhupendra Kesur; khewalkesur@gmail.com

This paper examines the application of Latin Hypercube Sampling (LHS) and Antithetic Variables (AVs) to reduce the variance of estimated performance measures from microscopic traffic simulators. LHS and AV allow for a more representative coverage of input probability distributions through stratification, reducing the standard error of simulation outputs. Two methods of implementation are examined, one where stratification is applied to headways and routing decisions of individual vehicles and another where vehicle counts and entry times are more evenly sampled. The proposed methods have wider applicability in general queuing systems. LHS is found to outperform AV, and reductions of up to 71% in the standard error of estimates of traffic network performance relative to independent sampling are obtained. LHS allows for a reduction in the execution time of computationally expensive microscopic traffic simulators as fewer simulations are required to achieve a fixed level of precision with reductions of up to 84% in computing time noted on the test cases considered. The benefits of LHS are amplified for more congested networks and as the required level of precision increases.

1. Introduction

Simulation models have become established as invaluable tools for managing signalized road traffic networks. A large range of planning applications from optimizing traffic signal timings, evaluating changes in lane configurations, and testing intelligent traffic systems have been extensively explored. Although deterministic macroscopic simulation models such as TRANSYT [1], TRANSYT-7F [2], and SYNCHRO [3] were initially used, preference is now given to microscopic simulation models such as PARAMICS [4], CORSIM [5], and VISSIM [6]. Microscopic traffic simulators consider the state of vehicles individually rather than modeling an entire stream of traffic in aggregate as is the case with macroscopic traffic simulators. This more detailed consideration allows for a more comprehensive consideration of the complex effects of vehicle interactions as well as catering for the stochastic nature of vehicle arrivals, vehicle routing, and driver behavior to be considered.

The disadvantage of using a microscopic traffic simulator is the increased computing requirement due to the more detailed modeling. The following studies serve as examples to demonstrate the extensive model run-times which can be experienced.

(i) Rouphail et al. [7] required over 7 hours of computing time to evaluate the effectiveness of 625 alternative traffic signal timing policies in CORSIM as part of an optimization search on a moderately sized network with 9 signalized intersections.

(ii) While optimizing traffic signal settings in Stevanovic et al. [8], the evaluation of 7600 alternative traffic signal timing policies in VISSIM for a network with 12 signalized intersections required 90 hours of computing time even though processing was distributed across 10 computers working in parallel.

(iii) Park and Lee [9] compared several traffic signal optimization strategies on a 16-signal arterial network and found that the length of a single VISSIM evaluation required 84 seconds using a single computer. Distributed processing with 8–12 computers was required

TABLE 1: Maximum estimated coefficient of variation for network average delay.

Study	Traffic simulator	Coefficient of variation	Relative error at 95% confidence
Stevanovic et al. [11]	VISSIM	4.48%	8.79%
Yun and Park [12]	CORSIM	9.58%	18.78%
Park et al. [13]	CORSIM	29.55%	57.92%
Rouphail et al. [7]	CORSIM	20.47%	40.12%
Chang and Kanaan [14]	CORSIM	30.69%	60.15%

to reduce computing times and achieve the study objectives.

(iv) Stevanovic et al. [10] required over a month of computing time to complete traffic signal optimization runs for a 12-signal arterial network using VISSIM as an evaluation model.

In addition, due to the stochastic nature of the modelling, microscopic traffic simulators can only provide estimates of underlying network performance metrics. For the case of network average delay, the maximum estimated coefficient of variation (ratio of the standard deviation to the mean) from a number of studies has been calculated and summarized in Table 1.

Since network average delay is known to be normally distributed [15], the relative error at 95% confidence can be obtained by multiplying the coefficient of variation by 1.96. Thus, if we were to perform a single execution of the traffic simulator, obtaining an estimate of network average delay of x, the precision is such that a 95% confidence interval for the true underlying mean could be as wide as $(0.4x, 1.6x)$. Averaging over multiple independent replications of the traffic simulator is a necessity to increase precision. However, this can be computationally expensive as noted earlier. Furthermore, the standard error of the estimator of the mean is proportional to $1/\sqrt{n}$ where n is the number of independent replications of the simulator. Thus, a doubling of the number of simulation runs will reduce the standard error by only 29%.

The large computational burden of microscopic traffic simulation models can serve as a considerable obstacle in practical applications. This paper considers the application of the variance reduction techniques Latin Hypercube Sampling (LHS) and Antithetic Variables (AVs) to improve the precision of estimators from microscopic traffic simulators. LHS and AV are applied to ensure that the limited simulation sample provides a more representative coverage of the vehicle generation process through stratification, thus reducing the standard error of outputs from the traffic simulator. Increased precision will result in improved decision making for a fixed level of computation resources. Alternatively, the number of multiple model runs can be reduced without losing accuracy, thus reducing model run time. Two alternative methods of implementing stratified sampling will be considered and comparisons with the conventional approach of Simple Random Sampling (SRS) will be made. Results will be based on

an evaluation of SRS, LHS, and AV on a set of test networks using the open-source microscopic traffic simulator MSTRANS [15]. The sensitivity of findings to the impact of the number of simulations performed and the level of congestion will also be examined.

The remainder of this paper is organized as follows. First, a description of LHS, AV, and the related simulation variance reduction technique Common Random Numbers (CRNs) is given. A review of the application of variance reduction techniques in microscopic traffic simulators and a closely related set of models known as queuing systems is then provided. An overview of MSTRANS is provided next, followed by a description of how the stratified sampling methods have been implemented. The test networks are then described. The methodology for evaluating the variance reduction is discussed next. Results from the simulation experiments are then presented and analyzed. Finally, a summary of findings is given in the conclusion.

2. Simulation Variance Reduction Techniques

2.1. Latin Hypercube Sampling. Stochastic simulation is used to empirically estimate a property of a random variable Y. Often, an estimate of the mean $E[Y]$ is required. Y is typically a function of a vector of other random quantities $\underline{X} = (X_1, X_2, \ldots)^T$; that is, $Y = f(\underline{X})$, where the joint distribution of $\underline{X}$ is known. For the case of a microscopic traffic simulation model, Y is some overall measure of performance of the traffic network such as the average delay experienced by all vehicles traversing the network, $\underline{X}$ are the interarrival times of vehicles which are stochastic and based on known headway distributions such as the exponential [16], and $Y = f(\underline{X})$ is evaluated computationally based on the logic and parameters of the traffic simulation model.

In order to estimate $E[Y]$, $n \geq 1$ replications (simulations) are performed. Let $\underline{X}^{(i)} = (X_1^{(i)}, X_2^{(i)}, \ldots)$ denote the realization of the vehicle headways on the ith simulation. Let $Y^{(i)} = f(\underline{X}^{(i)})$ denote the network-wide average vehicle delay on the ith simulation. An estimate of $E[Y]$ can be obtained using $(1/n) \sum Y^{(i)}$.

In most simulation applications, $\{\underline{X}^{(i)} \mid 1 \leq i \leq n\}$ are generated independently and this constitutes SRS. The idea behind LHS is to select observations from $\{\underline{X}^{(i)} \mid 1 \leq i \leq n\}$ so that the resulting sample is more representative of
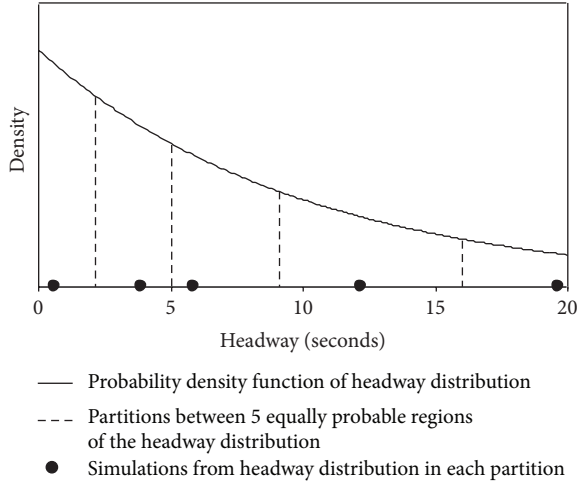
FIGURE 1: Simulated vehicle headways using LHS.

— Probability density function of headway distribution

--- Partitions between 5 equally probable regions
    of the headway distribution

● Simulations from headway distribution in each partition

FIGURE 2: Simulated vehicle headways using AV.

the distribution of $\underline{X}$ [17, 18]. This is achieved by stratified sampling of each marginal random variable in the vector. For example, suppose that X_m denotes the headway between the $(m-1)$th and mth vehicles entering at a particular point of the network. Furthermore, let us suppose that headways are exponentially distributed with a mean of 10 seconds. Assuming that we are performing $n = 5$ simulations, executions of the traffic simulation model with different random number seeds may produce the following observations: $X_m^{(1)} = 8.7$, $X_m^{(2)} = 7.9$, $X_m^{(3)} = 7.5$, $X_m^{(4)} = 6.4$, and $X_m^{(5)} = 9.6$. These observations are a poor representation of the headway distribution as the headway values are all smaller than the mean. A large headway between the $(m-1)$th and mth vehicles will not be covered in any of the five simulations generated. With LHS, the sample is obtained by partitioning the support into five regions of equal probability and selecting a single observation from each region. This is illustrated in Figure 1, and the exam-ple simulated outcomes can be seen to provide a more even coverage of the headway distribution.

This procedure is applied separately to generate the simulated outcomes for each marginal random variable in $\underline{X}$. The orders of simulated observations are randomized before combining to form $\underline{X}^{(i)}$. The approach guarantees an "even" coverage for each univariate outcome. LHS thus produces a "more representative" set of simulated vectors $\{\underline{X}^{(i)} \mid 1 \leq i \leq n\}$. Accordingly, $\{Y^{(i)} \mid 1 \leq i \leq n\}$ will be more representative, and the estimate of $E[Y]$ will have lower variance. Mathematically, LHS produces an unbiased estimator of $E[Y]$ with lower variance than that from SRS [17, 19]. The extent of the variance reduction is problem specific and is evaluated for the case of microscopic traffic simulation in this study.

2.2. Antithetic Variables.
The technique of AV [20] is closely related to LHS. For the case $n = 2$, the value of $X_m^{(1)}$ is simulated at random. Suppose that the simulated outcome is
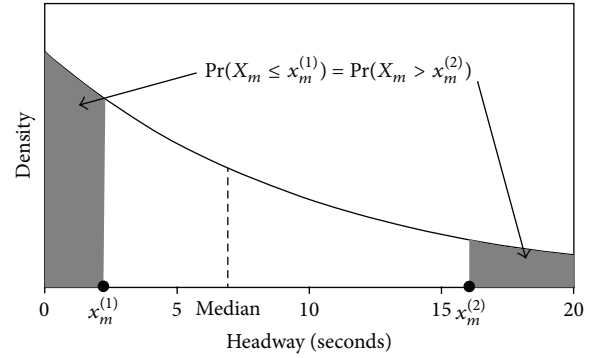
$x_m^{(1)}$. The value of $X_m^{(2)}$ is then set to $x_m^{(2)}$ where $x_m^{(2)}$ satisfies $\Pr(X_m \leq x_m^{(1)}) = \Pr(X_m > x_m^{(2)})$. An example of a possible outcome for the headway distribution covered in the previous section is illustrated in Figure 2.

For $n > 2$ and even, the process is repeated to generate pairs of observations $(X_m^{(2i-1)}, X_m^{(2i)})$ for $1 \leq i \leq n/2$. As with LHS, each marginal random variable in $\underline{X}$ is simulated independently. Simulation using AV will guarantee an equal number of observations on either side of the median. AV is similar to LHS in that the support of each marginal is partitioned into two equiprobable regions with an equal number of observations being simulated in each region. However, LHS can be considered to be more general in that instead of partitioning into two regions, stratification is performed using n equi-probable regions. For n odd, AVs can be applied by generating $(n-1)/2$ paired observations as suggested and generating an additional observation independently.

2.3. Common Random Numbers.
CRN [20] is explained by example with reference to a microscopic traffic simulator. For the case of comparing two alternative signal timing plans, differences in observed performance values are a consequence of the impact of signal timings as well as the fluctuations in experimental conditions (e.g., each signal timing plan may have been subjected to drastically different vehicle arrival patterns). If we apply similar experimental conditions for all system configurations (i.e., identical vehicle arrival patterns), we can then attribute the computed differences in measured performances to the difference in the signal timing policies. CRN can be implemented by utilizing the same random number seed for the microscopic traffic simulator when evaluating the two alternative signal timing plans. LHS and AV are not at odds with CRN as the variance reduction techniques can be combined. For example, assuming that $n > 1$ simulations are performed for each signal timing policy, the arrival patterns for the n simulations can be generated by LHS. These same arrival patterns can then be applied when evaluating both traffic signal timing plans using the microscopic traffic simulator.

3. Previous Applications of Variance Reduction Techniques

3.1. Traffic Simulation. Rathi and Venigalla [21] experimented with AV using an early version of CORSIM [5]. AVs were applied to the generation of all stochastic effects, namely, vehicle characteristics, headways, routing, and driver behavioral decisions on a small three-intersection network. Reductions of up to 36% in the standard deviation of network average travel time were obtained. Rathi [22] found the use of CRN to provide substantial reductions in the variance of the difference between performance measures of alternative traffic signal timing policies. Kesur [23] demonstrated the improvements in the solution quality obtained from state-of-the-art traffic signal optimization algorithms when implementing CRN in the evaluation of alternative signal timing policies. CRN is currently in widespread use and is supported by the proprietary traffic simulators CORSIM [5] and VISSIM [6]. No studies on the possible implementation methods or the effectiveness of applying LHS to the vehicle generation process have been performed as yet.

A similar, though not directly relevant, application is the use of LHS as an experimental design tool in microscopic traffic simulator applications. For example, Park et al. [24] considered the robustness of a signal timing policy for a 9-signal network to changes in mean demand at 8 major source nodes. Three different flow levels were considered at each source node, namely, base, high, and low. With 8 factors and 3 levels for each factor, $3^8 = 6561$ different demand patterns can be tested. LHS was used to reduce the number of demand combinations to 204 by sampling 68 base, high, and low flow scenarios for each source node and then combining these in random order. LHS has also been used for experimental design when identifying significant parameters for the purpose of calibrating microscopic traffic simulators [25].

3.2. Queuing Systems. Traffic networks are special cases of queuing systems. Queuing theory deals with the analysis of systems where customers arrive at a queue according to a deterministic or stochastic process and are served by one or more servers with deterministic or stochastic service times. An isolated road intersection can be modeled as a queuing system with exponentially distributed customer inter-arrival times [26]. For the case of a queuing system with a single server, Page [27] and Mitchell [28] applied AV to the generation of exponentially distributed customer inter-arrival and service times, noting reductions of up to 41% in the standard error of the estimated average customer waiting time when compared to simulation by SRS. Sabuncuoglu et al. [29] experimented with AV and LHS in turn for the same queuing system, observing reductions of up to 16% in the standard error of the estimator of the total time spent in the system relative to SRS. Ross and Lin [30] advocate stratification of the number of customer arrivals in queuing systems as opposed to inter-arrival times and obtained a reduction in the standard deviation of the average customer waiting time of 29%. Streltsov and Vakili [31] have also applied LHS to a range of queuing systems. However, their approach focuses on event based simulation where the state of the system at the occurrence of an event (such as a customer arrival or departure) is considered.

4. MSTRANS Overview

Comparisons between SRS, LHS, and AV are made using Microscopic Stochastic Traffic Network Simulator (MSTRANS) which is a stochastic microscopic traffic simulator developed by the author [15]. This model was chosen for the following reasons.

(i) MSTRANS is open-source with a full code listing available at [32] or directly from the author. Modifications to the vehicle generation algorithms to implement LHS and AV could thus be made directly.

(ii) Several validation exercises have been performed, and MSTRANS has been found to give results comparable to CORSIM [5], in terms of both the mean and standard error of estimators of network performance in aggregate, as well as on individual links [15].

(iii) The model has been extensively used in several applications related to optimization of traffic signal timings [23, 33–35].

(iv) MSTRANS has a rapid execution speed relative to the commercially available microscopic traffic simulators, allowing for the large number of model executions required to achieve the study objectives [15].

Networks in MSTRANS are represented schematically using graphs with nodes representing vehicle entry points or intersections and links representing the road. The model logic and parameters are based on findings from the literature on driver behavior and vehicle characteristics. The model applies a fixed increment time step to advance the simulation. Vehicle status and kinematics are updated each second, along with the traffic signal indications. New vehicles are generated at the network boundary nodes with exponentially distributed headways. Vehicle routing through the network is generated stochastically based on specified expected turning proportions. Linear acceleration and constant deceleration models are applied to lead vehicles. The behavior of following vehicles is based on the Gipps microscopic car-following model [36]. Left-turning vehicles receiving unprotected green utilize a gap-acceptance model where the probability of accepting a gap increases with the gap size according to a logistic function. Stopping decisions on amber signals are also made using a logistic model. Response delays are modeled and lane changes are governed by pragmatic rules. The delay measure produced is control delay which is the difference between actual and uninterrupted travel times. The number of stops is computed using the approach of Rakha et al. [37] which accountings for partial stops and is appropriate in both undersaturated and oversaturated conditions.

An initialization period is completed before results are recorded. Network performance measures are obtained by averaging over all vehicles that complete their trips through the network during the analysis period. However, to account

for the delay and stops to be experienced by residual queues remaining at the end of the analysis period, the run length is extended until all remaining vehicles have cleared the network.

A thorough account of the functional details of the model and a review of the literature justifying the logic and parameters are given in [15].

5. Implementation of Stratified Sampling

Applications of LHS and AV in queuing systems have been covered in the literature review. Of the studies mentioned, [27–29] implement stratification to the customer inter-arrival times, while [30] applies stratification to the number of customer arrivals. Extending these ideas to the realm of microscopic traffic simulators, two methods of implementing stratification are considered. We refer to implementations of LHS or AV using the first method as LHS1 and AV1, respectively. Implementations of LHS or AV under the second method are referred to as LHS2 and AV2, respectively.

5.1. Method 1. Under this implementation, stratification is applied to the generation of vehicle inter-arrival times. In addition, stratification is applied to generate the sequence of turning decisions for each vehicle.

5.1.1. Vehicle Interarrival Times. Let n be number of replications of the traffic simulator, T duration of simulation period (including duration of initialization period) (seconds), μ_k average headway between vehicles entering at boundary node k (seconds), $X_k^{(i)}(j)$ headway between the $(j-1)$th and jth vehicles entering at boundary node k in simulation i (seconds), and $I_k^{(i)}(j)$ indicator as to whether the entry of the $(j-1)$th vehicle at boundary node k in simulation i occurs before the end of the simulation period

$$I_k^{(i)}(j)$$
$$= \begin{cases} 1, & \text{if } j = 1 \text{ or } \left(j > 1 \text{ and } \sum_{l=1}^{j-1} X_k^{(i)}(l) < T \right), \\ 0, & \text{otherwise.} \end{cases} \quad (1)$$

For each boundary node k and vehicle arrival j, $\sum_{i=1}^{n} I_k^{(i)}(j)$ vehicle inter-arrival times are generated by stratification of the exponential distribution with mean μ_k as per the procedures described in Section 2. Note that $\sum_{i=1}^{n} I_k^{(i)}(j) \leq n$ headways are generated as some simulations may have fewer than j vehicles entering at node k during the finite simulation period. For each boundary node k, the previous process begins with $j = 1$ and is repeated for increasing j until $\sum_{i=1}^{n} I_k^{(i)}(j) = 0$.

5.1.2. Vehicle Turning Decisions. For each vehicle generated, a route through the network is assigned. The route is comprised of a sequence of turning decisions at each intersection. Stratification is applied to ensure a more representative coverage of routes through the network. The manner in which this is
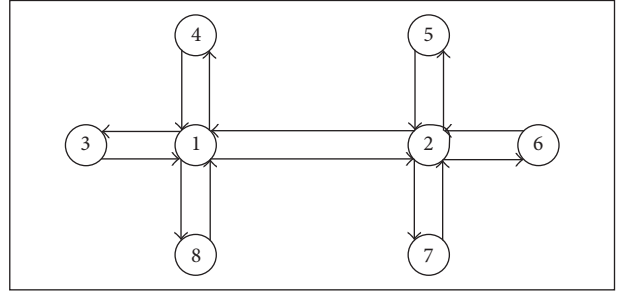

FIGURE 3: Two-intersection network.

achieved is demonstrated by example using the two-intersection network in Figure 3. Let us assume that the expected proportion of left turning, through, and right turning traffic on each approach to an intersection is 1/4, 1/2, and 1/4, respectively. For a vehicle entering at a particular node, the sequence of possible turning decisions is completely determined if the exit node is known. Let D_k denote the exit node for a vehicle arriving at boundary node k.

We will consider routing decisions for vehicles entering at node 3. If we enumerate the set of turning decisions in the order, left, through, and right, starting with the first turning decision and iterating for all remaining turning decisions, the distribution of D_3 will be given by

$$\Pr(D_3 = d) = \begin{cases} \dfrac{1}{4}, & \text{for } d = 4, \\[2mm] \dfrac{1}{2} \times \dfrac{1}{4} = \dfrac{1}{8}, & \text{for } d = 5, \\[2mm] \dfrac{1}{2} \times \dfrac{1}{2} = \dfrac{1}{4}, & \text{for } d = 6, \\[2mm] \dfrac{1}{2} \times \dfrac{1}{4} = \dfrac{1}{8}, & \text{for } d = 7, \\[2mm] \dfrac{1}{4}, & \text{for } d = 8. \end{cases} \quad (2)$$

According to the inverse transform method [20], an observation from D_3 can be obtained by simulating a uniform random number U in the interval $(0, 1)$ and evaluating the function $g_3(U)$ which is graphed in Figure 4.

LHS and AV can be implemented by applying the stratification to simulate U and then evaluating the function $g_3(U)$. For the case of LHS with $n = 8$, we will then have the following trips for the jth vehicle entering at node 3:

  (i) two trips exiting via node 4;

 (ii) one trip exiting via node 5;

(iii) two trips exiting via node 6;

 (iv) one trip exiting via node 7;
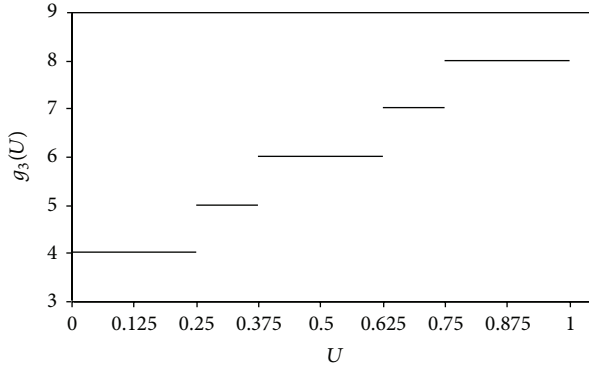
  (v) two trips exiting via node 8.

FIGURE 4: Function used to simulate route of vehicle entering at node 3.

If only $n = 4$ simulations are performed, LHS stratification will ensure that we have

(i) one trip exiting via node 4;

(ii) two trips through to intersection 2, and of these trips, one trip is equally likely to be a left turn and exit via node 5 or a through movement and exit via node 6, while the remaining trip is equally likely to be a through movement and exit via node 6 or a right turn and exit via node 7;

(iii) one trip exiting via node 8.

5.2. Method 2. In this alternative implementation, stratification is applied to vehicle counts as opposed to vehicle headways. In addition to vehicle counts, stratification is applied to the generation of vehicle entry times by considering the distribution of entry times conditional on the simulated count.

5.2.1. Vehicle Counts. Assuming exponentially distributed headways, we have from theory on the Poisson process that the number of vehicles entering at boundary node k during the simulation period will have a Poisson(T/μ_k) distribution [20]. Thus, if we simulate vehicle counts rather than vehicle headways, we can apply stratification to the number of vehicles entering the network. However, rather than applying stratification to the number of vehicles entering at boundary node k, a more detailed stratification is performed. This is explained using the two-intersection network from the previous section as an example. From theory on the Poisson process, we have that the number of vehicles entering at node 3 and exiting at node d during the simulation period will be distributed Poisson$((T/\mu_3) \times \Pr(D_3 = d))$ and independent for each d [20]. Thus, rather than simulating all vehicles entering at node 3 and then assigning routes to individual vehicles, we can rather simulate the count of vehicles entering at source node 3 separately for each possible exit node d using independent Poisson random variables. Stratification of each Poisson random variable can be performed as per the procedures described in Section 2. Thus, rather than merely ensuring a representative coverage of the distribution of the number of vehicle entering at each boundary node,

the procedure described will ensure a representative sample for the distribution of the number of vehicles traversing each possible route through the network.

5.2.2. Vehicle Entry Times. From theory on the Poisson process [20], it can be shown that once the vehicle count has been simulated, the entry times of individual vehicles can be simulated independently from a uniform distribution over the interval $(0, T)$. These entry times need to be sorted into increasing order, to obtain the entry times of successive vehicles. Using this result, stratification is applied to the generation of the vehicle entry times as follows.

Let $\eta_{k,l}^{(i)}$ be number of vehicles entering at node k and exiting at node l in simulation i, $\eta_{k,l}$ maximum number of vehicles entering at node k and exiting at node l over the n simulations $= \max\{\eta_{k,l}^{(i)} \mid 1 \le i \le n\}$, $X_{k,l}^{(i)}(j)$ entry time of the jth vehicle entering at node k and exiting at node l in simulation i, and $I_{k,l}^{(i)}(j)$ indicator as to whether at least j vehicles enter at node k and exiting at node l in simulation i where

$$I_{k,l}^{(i)}(j) = \begin{cases} 1, & \text{if } j \le \eta_{k,l}^{(i)}, \\ 0, & \text{otherwise.} \end{cases} \tag{3}$$

First, the number of vehicles entering at node k and exiting at node l, $\eta_{k,l}^{(i)}$ is simulated for all i by stratification as described in Section 5.2.1. Thereafter, for each vehicle $j \in \{1, 2, \ldots, \eta_{k,l}\}$, $\sum_{i=1}^{n} I_{k,l}^{(i)}(j)$ arrival times are simulated by stratification of the $U(0, T)$ distribution using the procedures described Section 2. The generated arrival times $\{X_{k,l}^{(i)}(j) \mid 1 \le j \le \eta_{k,l}^{(i)}\}$ along the route are then sorted in increasing order for each simulation i.

6. Test Networks

SRS, LHS, and AV are compared based on simulation results for a 9-signal arterial network and a 14-signal grid network. The arterial network is Canal Street in New Orleans, Louisiana [38]. The grid network is based on a data set for downtown Ann Arbor, Michigan [39]. The structure and spacing's for each network are given in Figure 5.

Two through lanes, exclusive turning lanes, and a free flow speed of 37 mph (60 km/h) have been assumed for each link. The average total flows into the arterial and grid networks are 8191 and 6546 vehicles/hour, respectively. A breakdown of the average flows by boundary node and the expected turning proportions at each approach is given in Gartner et al. [38] for the arterial network and Kesur [15] for the grid network. The flow rates as given constitute undersaturated conditions. The average flows are increased by 50% for the arterial network and 60% for the grid network to construct oversaturated conditions. The following four test networks are thus considered:

(i) arterial undersaturated;

(ii) arterial oversaturated;

(iii) grid undersaturated;
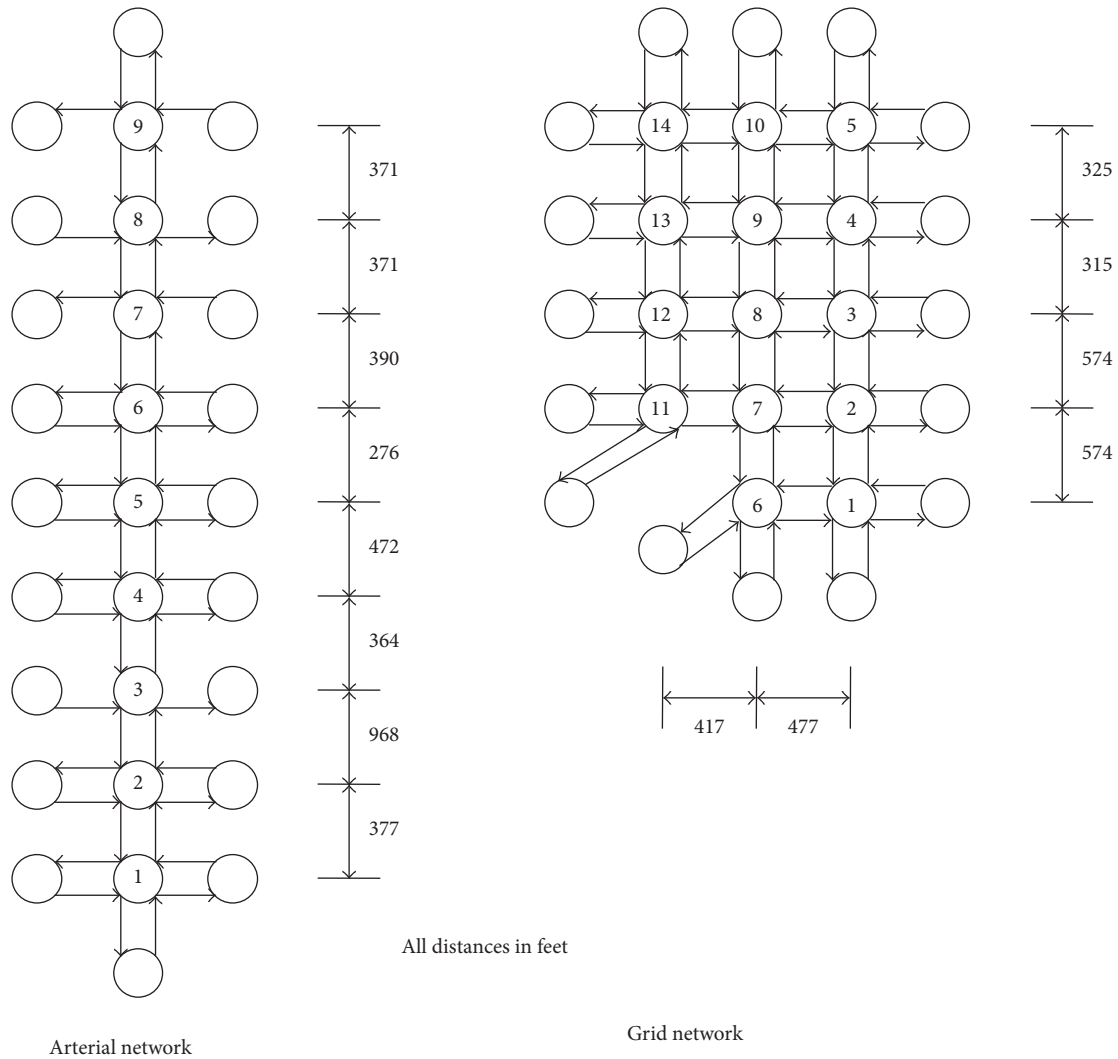
(iv) grid oversaturated.

FIGURE 5: Test networks.

A 15-minute initialization period is utilized followed by a 15-minute analysis period in each case. The implemented fixed-time traffic signal timing scheme for each network was that obtained from the delay minimizing optimization algorithm of Kesur [23]. A model calibration exercise was not performed and the default MSTRANS model parameters [15] which have been set based on findings on the literature have been used. The test networks and model parameters were felt to be generic enough to achieve the study objective which is an evaluation of the effectiveness of the proposed variance reduction techniques.

7. Evaluation of Variance Reduction

The improvements in precision of mean estimates of network averages for delay, number of stops, and speed are quantified by comparing the variance of these quantities under SRS, LHS, and AV. The estimators are compared for $n \in \{2, 4, 10, 20, 50, 100\}$. A range of different values of n are tested as the required number of replications of the microscopic

traffic simulator is problem specific. For example, in traffic signal optimization, $n \in \{1, 2\}$ has been found to be optimal as it allows for a large number of signal timing combinations to be tested [15, 33]. A model calibration exercise would also make use of a small value of n [25]. However, for a planning or evaluation application, it is not uncommon to perform up to 100 replications.

Under SRS, the variance of the estimator of network average delay obtained by averaging results from n independent replications of MSTRANS is given by

$$\operatorname{Var}\left(\widehat{E}[Y]_{\mathrm{SRS}}\right) = \frac{1}{n} \operatorname{Var}(Y), \tag{4}$$

where Y is the network average delay. The quantity $\operatorname{Var}(Y)$ was estimated by performing $N = 1000$ evaluations of each test network in MSTRANS with a different random number seed each time and using the sample variance as an unbiased estimator.

TABLE 2: Coefficient of variation of estimators of network performance under SRS, LHS1 and LHS2, and AV1 and AV2.

Network	n	Coefficient of variation of estimator for network average delay					Relative reduction in coefficient of variation of estimator for network average delay					
		SRS	LHS1	LHS2	AV1	AV2	LHS1 versus AV1	LHS2 versus AV2	LHS2 versus LHS1	AV2 versus AV1	LHS1 versus SRS	LHS2 versus SRS
Arterial undersaturated	2	5.0%	4.2%	4.0%	4.3%	4.0%	1%	1%	6%	6%	16%	21%
	4	3.5%	2.6%	2.6%	2.8%	2.7%	8%	6%	0%	2%	27%	27%
	10	2.2%	1.5%	1.4%	1.9%	1.7%	18%	18%	11%	12%	31%	39%
	20	1.6%	1.0%	0.9%	1.3%	1.2%	24%	20%	5%	10%	38%	41%
	50	1.0%	0.6%	0.6%	0.8%	0.8%	22%	27%	11%	5%	37%	44%
	100	0.7%	0.4%	0.4%	0.6%	0.5%	30%	26%	7%	12%	39%	44%
Arterial oversaturated	2	6.9%	5.0%	4.8%	4.7%	3.4%	−8%	−41%	4%	27%	26%	29%
	4	4.8%	2.7%	2.1%	3.3%	2.4%	16%	11%	22%	27%	43%	56%
	10	3.1%	1.4%	1.1%	2.1%	1.5%	36%	28%	20%	29%	56%	65%
	20	2.2%	0.8%	0.7%	1.5%	1.0%	43%	30%	16%	31%	62%	68%
	50	1.4%	0.5%	0.4%	0.9%	0.7%	46%	37%	15%	28%	65%	70%
	100	1.0%	0.3%	0.3%	0.6%	0.5%	51%	42%	11%	24%	68%	71%
Grid undersaturated	2	2.2%	2.0%	1.9%	2.1%	2.0%	6%	5%	5%	5%	9%	13%
	4	1.6%	1.3%	1.3%	1.3%	1.4%	−1%	4%	1%	−4%	14%	15%
	10	1.0%	0.8%	0.8%	0.9%	0.8%	12%	8%	0%	4%	21%	21%
	20	0.7%	0.6%	0.5%	0.6%	0.6%	3%	9%	7%	0%	16%	22%
	50	0.4%	0.3%	0.3%	0.4%	0.4%	9%	9%	1%	1%	23%	24%
	100	0.3%	0.3%	0.2%	0.3%	0.3%	4%	13%	9%	−1%	20%	27%
Grid oversaturated	2	9.6%	8.2%	7.3%	7.9%	6.2%	−5%	−17%	11%	20%	14%	24%
	4	6.8%	5.0%	4.3%	5.2%	4.5%	4%	6%	15%	14%	26%	37%
	10	4.3%	2.6%	2.1%	3.5%	2.9%	26%	26%	17%	18%	40%	50%
	20	3.0%	1.7%	1.4%	2.4%	1.9%	30%	26%	15%	20%	44%	53%
	50	1.9%	1.0%	0.9%	1.6%	1.3%	36%	29%	12%	20%	48%	54%
	100	1.4%	0.7%	0.6%	1.1%	0.9%	39%	31%	7%	17%	51%	55%

Analytical expressions for the variance of estimators obtained using AV and LHS do not exist. Instead, the replicated sampling technique is used [18]. This involves producing $N = 1000$ independent super samples for each stratified sampling method $k \in \{\text{LHS1, LHS2, AV1, AV2}\}$ and n. Let $Y_k^{(i,j)}(n)$ denote network average delay in the ith simulation out of n on the jth independent super sample using stratified sampling method k. We have that $\overline{Y_k^{(j)}}(n) = (1/n) \sum_{i=1}^{n} Y_k^{(i,j)}(n)$ is the jth independent estimate of network average delay using stratified sampling method k with n simulations. An unbiased estimator for the variance of the estimator of network average delay using stratified sampling method k with n replications of MSTRANS is given by

$$\text{Var}\left(\widehat{E}[Y]_k\right) = \frac{1}{N-1} \sum_{j=1}^{N} \left(\overline{Y_k^{(j)}}(n) - \frac{1}{N} \sum_{j=1}^{N} \overline{Y_k^{(j)}}(n) \right)^2. \quad (5)$$

When quoting results, rather than providing figures for the variance, we provide the coefficient of variation. This allows for easier comparison as this measure is unit-free.

Although the previous discussion only covers network average delay, the variance of estimates for the network average number of stops and speed is obtained in a similar way.

8. Results

8.1. Improvements in Precision. The coefficient of variation for estimates of network average delay is given in Table 2 for each test network, sampling method, and value of n. The relative reductions in the coefficient of variation using different stratified sampling methods are also provided.

From Table 2, we note that the stratified sampling techniques reduce the variability of network average delay relative to SRS in all instances. From columns 8 and 9, we find that LHS offers larger variance reductions relative to AV in all cases expect for when $n = 2$ in oversaturated flow conditions. We thus recommend stratification using LHS instead of AV as it produces larger variance reductions in general. Larger variance reductions under LHS are expected as the method allows for stratification of each univariate outcome to a greater extent. AV has the potential to outperform LHS when $n = 2$ as it imposes a larger negative correlation between the two observations of each univariate outcome.

TABLE 3: Estimated values of p under LHS1 and LHS2.

| Network | Estimated values of p | | | | | |
| | LHS1 | | | LHS2 | | |
	Delay	Number of stops	Speed	Delay	Number of stops	Speed
Arterial undersaturated	0.260	0.374	0.250	0.308	0.419	0.295
Arterial oversaturated	0.565	0.582	0.563	0.662	0.673	0.647
Grid undersaturated	0.128	0.216	0.108	0.156	0.328	0.137
Grid oversaturated	0.348	0.245	0.358	0.426	0.336	0.436

Columns 10 and 11 demonstrate that stratification using Method 2 generally provides a larger variance reduction than that obtained by Method 1. This can be explained by the fact that Method 2 applies stratification to a greater extent than Method 1. While both methods apply stratification to the distribution of vehicle entry times directly, Method 2 applies stratification to the distribution of vehicle counts to a greater degree in that partitioning is applied directly to the count of vehicles traversing each route through the network. With Method 1 this stratification is only applied indirectly through the following effects.

(i) The stratification of routing decisions of individual vehicles ensures a more even coverage of routes through the network in aggregate.

(ii) Vehicle counts and vehicle headways at each boundary node are related; that is, simulated vehicle count = [average of headway of all simulated vehicles]$^{-1}$. The stratification of the headway distribution of individual vehicles implies a more representative sampling of the average headway and as a consequence a more representative coverage of the vehicle count.

However, the scope for application of Method 1 is more general in that it can be applied to headway distributions other than the exponential. Although the exponential headway distribution is the most widely used [16], the Gamma and Lognormal distributions have been recommended for higher levels of traffic flow [40, 41].

From Columns 12 and 13, we find that the improvements in precision of estimators obtained using simulation with LHS as opposed to SRS increase as more simulations are performed. For instance, the relative reduction in the standard error of network average delay using LHS2 for the arterial network in oversaturated conditions starts at 29% for $n = 2$ and increases to 71% for $n = 100$. This can be explained by the fact that LHS applies a more detailed stratification of each marginal random variable for larger n. Furthermore, the benefits of employing LHS are larger in oversaturated flow conditions.

Although the impacts of stratified sampling on the network averages for the number of stops and speed are not provided, the overall findings and conclusions are similar.
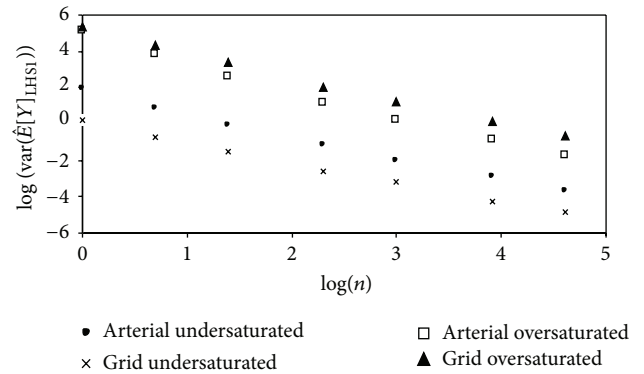

FIGURE 6: Approximately linear relationship.

8.2. Reduction in Model Runtime. The variance of estimators for sampling method $k \in \{\text{LHS1}, \text{LHS2}\}$ has been examined and the following approximate relationship was observed:

$$\text{Var}\left(\widehat{E}[Y]_k\right) \cong \frac{\text{Var}(Y)}{n^{p+1}}. \tag{6}$$

In order to more easily demonstrate this fact and to obtain an estimate of p, we can convert (6) into a linear relationship by taking logs

$$\log\left(\text{Var}\left(\widehat{E}[Y]_k\right)\right) \cong \log\left[\text{Var}(Y)\right] - (p+1)\log(n). \tag{7}$$

The quantities $\log(\text{Var}(\widehat{E}[Y]_k))$ and $\log(n)$ are plotted against each other for the case of network average delay under LHS1 in Figure 6, and the relationship can be observed to be close to linear.

A similar result was observed for LHS2 and when considering performance measures other than network average delay. Least-squares estimates of p are given in Table 3.

From (4) and (6), we have that the number of simulations required under stratified sampling to obtain estimators with the same precision as those obtained using SRS with n simulations is given by $n^{1/(p+1)}$. These are tabulated in Table 4.

From Table 4, we find that for the arterial network in oversaturated conditions, the estimator for network average delay obtained by averaging the results of 100 simulations with SRS has the same precision as that obtained by averaging just 16 simulations if LHS2 was applied instead. Since computing time increases linearly with the number of simulations, LHS provides an 84% runtime reduction in this instance.

TABLE 4: Comparison of sample sizes under SRS and LHS for the same level of precision.

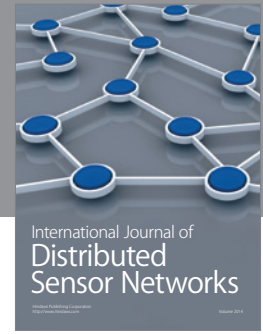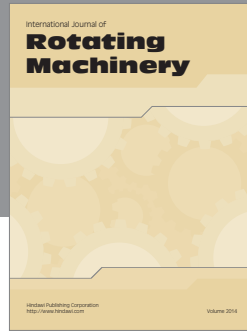| Network	SRS sample size	Equivalent sample size under LHS1 to achieve identical precision to SRS			Equivalent sample size under LHS2 to achieve identical precision to SRS		
		Delay	Number of stops	Speed	Delay	Number of stops	Speed
Arterial undersaturated	2	2	2	2	2	2	2
	4	3	3	3	3	3	3
	10	6	5	6	6	5	6
	20	11	9	11	10	8	10
	50	22	17	23	20	16	20
	100	39	29	40	34	26	35
Arterial oversaturated	2	2	2	2	2	2	2
	4	2	2	2	2	2	2
	10	4	4	4	4	4	4
	20	7	7	7	6	6	6
	50	12	12	12	11	10	11
	100	19	18	19	16	16	16
Grid undersaturated	2	2	2	2	2	2	2
	4	3	3	3	3	3	3
	10	8	7	8	7	6	8
	20	14	12	15	13	10	14
	50	32	25	34	29	19	31
	100	59	44	64	54	32	57
Grid oversaturated	2	2	2	2	2	2	2
	4	3	3	3	3	3	3
	10	6	6	5	5	6	5
	20	9	11	9	8	9	8
	50	18	23	18	16	19	15
	100	30	40	30	25	31	25

9. Conclusion

The application of LHS and AV to reduce the variance of output measures from microscopic traffic simulators has been explored and found to provide considerable benefits. Two methods of implementing stratified sampling are presented, one which offers greater variance reduction and another that has more general applicability. LHS is found to outperform AV in general. For the same level of computational resources, reductions of up to 71% in the standard error of the equivalent estimates of traffic network performance under SRS are obtained using LHS. Alternatively, in the most extreme case, the same level of precision can be maintained by using LHS with only 16% of the original computational resource level. The improvements in precision and reductions in runtime can allow for better decision making and increased scope of use of microscopic traffic simulators. Circumstances which place the largest burden on computing time are requirements for high levels of precision or the evaluation of congested networks. However, it is in these difficult situations where LHS is the most effective. The overhead and complexity of implementing LHS is minimal, and it is recommended for inclusion in current commercial microscopic traffic simulators.

References

[1] D. I. Robertson, "TRANSYT: a traffic network study tool," Road Research Laboratory Report LR253, 1969.

[2] University of Florida, *Traffic Network Study Tool, TRANSYT-7F, United States Version 11*, McTrans Center, University of Florida, 2009.

[3] Trafficware, *SYNCHRO 5.0 User's Guide*, Albany California, 1999.

[4] Quadstone Limited, *Quadstone Paramics V5.0 Manual*, Edinburgh, 2004.

[5] Federal Highway Administration, *Traffic Software Integrated System (TSIS) Version 6.3 User Guide*, Federal Highway Administration Office of Operations Research, Development and Technology, 2011.

[6] Planung Transport Verkehr AG, *VISSIM 5.40 User Manual*, PTV, Karlsruhe, Germany, 2012.

[7] N. M. Rouphail, B. Park, and J. Sacks, "Direct signal timing optimization: strategy development and results," Tech. Rep. 109,

National Institute of Statistical Sciences, 2000, http://www.niss.org/.

[8] A. Stevanovic, P. T. Martin, and J. Stevanovic, "VisSim-based genetic algorithm optimization of signal timings," *Transportation Research Record: Journal of the Transportation Research Board*, no. 2035, pp. 59–68, 2007.

[9] B. Park and J. Lee, "Optimization of coordinated-actuated traffic signal system: stochastic optimization method based on shuffled frog-leaping algorithm," *Transportation Research Record: Journal of the Transportation Research Board*, no. 2128, pp. 76–85, 2009.

[10] A. Stevanovic, J. Stevanovic, and C. Kergaye, "Optimizing signal timings to improve safety of signalized arterials," in *Proceedings of the 3rd International Conference on Road Safety and Simulation*, Indianapolis, Ind, USA, September 2011.

[11] A. Stevanovic, J. Stevanovic, and P. T. Martin, "Optimizing signal timings from the field: VISGAOST and VISSIM—ASC/3 software-in-the-loop simulation," *Transportation Research Record: Journal of the Transportation Research Board*, no. 2128, pp. 114–120, 2009.

[12] I. Yun and B. Park, "Stochastic optimization method for coordinated actuated signal systems," Research Report UVACTS-15-0-102, Center for Transportation Studies, University of Virginia, 2005.

[13] B. Park, N. M. Rouphail, J. P. Hochanadel, and J. Sacks, "Evaluating reliability of transyt-7F optimization schemes," *Journal of Transportation Engineering*, vol. 127, no. 4, pp. 319–326, 2001.

[14] G. L. Chang and A. Kanaan, "Variability assessment for TRAF-NETSIM," *Journal of Transportation Engineering*, vol. 116, no. 5, pp. 636–657, 1990.

[15] K. B. Kesur, *Traffic Signals: Advances in Genetic Algorithm Optimization*, VDM, 2011.

[16] C. J. Messer and J. A. Bonneson, *NCHRP Web Document 12: Capacity Analysis of Interchange Ramp Terminals*, Transportation Research Board, National Research Council, Washington DC, USA, 1997.

[17] M. D. McKay, R. J. Beckman, and W. J. Conover, "A comparison of three methods for selecting values of input variables in the analysis of output from a computer code," *Technometrics*, vol. 21, no. 2, pp. 239–245, 1979.

[18] J. C. Helton and F. J. Davis, "Latin hypercube sampling and the propagation of uncertainty in analyses of complex systems," *Reliability Engineering and System Safety*, vol. 81, no. 1, pp. 23–69, 2003.

[19] M. Stein, "Large-sample properties of simulations using latin hypercube sampling," *Technometrics*, vol. 29, no. 2, pp. 143–151, 1987.

[20] A. M. Law and D. W. Kelton, *Simulation Modeling and Analysis*, McGraw-Hill, New York, NY, USA, 3rd edition, 2000.

[21] A. K. Rathi and M. M. Venigalla, "Variance reduction applied to urban network traffic simulation," *Transportation Research Record: Journal of the Transportation Research Board*, no. 1365, pp. 133–146, 1992.

[22] A. K. Rathi, "The use of common random numbers to reduce the variance in network simulation of traffic," *Transportation Research Part B*, vol. 26, no. 5, pp. 357–363, 1992.

[23] K. B. Kesur, "Advances in genetic algorithm optimization of traffic signals," *Journal of Transportation Engineering*, vol. 135, no. 4, pp. 160–173, 2009.

[24] B. Park, N. M. Rouphail, and J. Sacks, "Assessment of stochastic signal optimization method using microsimulation," *Transportation Research Record: Journal of the Transportation Research Board*, no. 1748, pp. 40–45, 2001.

[25] B. Park and H. Qi, "Development and evaluation of a procedure for the calibration of simulation models," *Transportation Research Record: Journal of the Transportation Research Board*, no. 1934, pp. 208–217, 2005.

[26] F. V. Webster, "Traffic signal settings," Road Research Technical Paper 39, H.M.S.O., London, UK, 1958.

[27] E. S. Page, "On monte carlo methods in congestion problems. II. Simulation of queuing systems," *Operations Research*, vol. 13, pp. 13–35, 1967.

[28] B. Mitchell, "Variance reduction by antithetic variates in GI/G/1 queuing simulations," *Operations Research*, vol. 21, pp. 988–997.

[29] I. Sabuncuoglu, M. M. Fadiloglu, and S. Çelik, "Variance reduction techniques: experimental comparison and analysis for single systems," *IIE Transactions*, vol. 40, no. 5, pp. 538–551, 2008.

[30] S. M. Ross and K. Y. Lin, "Applying variance reduction ideas in queuing simulations," *Probability in the Engineering and Informational Sciences*, vol. 15, no. 4, pp. 481–494, 2001.

[31] S. Streltsov and P. Vakili, "Variance reduction algorithms for parallel replicated simulation of uniformized Markov chains," *Discrete Event Dynamic Systems: Theory and Applications*, vol. 6, no. 2, pp. 159–180, 1996.

[32] K. B. Kesur, *Advances in genetic algorithm optimization of traffic signals [M.S. thesis]*, University of the Witwatersrand, Johannesburg, South Africa, 2007, http://wiredspace.wits.ac.za/bitstream/handle/10539/4900/TrafficSignalOptimization.pdf.

[33] K. B. Kesur, "Generating more equitable traffic signal timing plans," *Transportation Research Record: Journal of the Transportation Research Board*, no. 2192, pp. 108–115, 2010.

[34] K. B. Kesur, "Optimization of mixed cycle length traffic signals," *Journal of Advanced Transportation*, 2013.

[35] K. B. Kesur, "Multi-objective optimization of delay and stops in traffic signal networks," in *Metaheuristic in Water, Geotechnical and Transport Engineering*, A. H. Gandomi, X. Yang, S. Talatahari, and A. Alavi, Eds., Elsevier, 2012.

[36] P. G. Gipps, "A behavioural car-following model for computer simulation," *Transportation Research Part B*, vol. 15, no. 2, pp. 105–111, 1981.

[37] H. Rakha, Y. S. Kang, and F. Dion, "Estimating vehicle stops at undersaturated and oversaturated fixed-time signalized intersections," *Transportation Research Record: Journal of the Transportation Research Board*, no. 1776, pp. 128–137, 2001.

[38] N. H. Gartner, S. G. Assmann, F. Lasaga, and D. L. Hou, "MULTIBAND—a variable-bandwidth arterial progression scheme," *Transportation Research Record: Journal of the Transportation Research Board*, no. 1287, pp. 212–222, 1990.

[39] C. Stamatiadis and N. H. Gartner, "MULTIBAND-96: a program for variable-bandwidth progression optimization of multiarterial traffic networks," *Transportation Research Record: Journal of the Transportation Research Board*, no. 1554, pp. 9–17, 1996.

[40] A. S. Al-Ghamdi, "Analysis of time headways on urban roads: case study from Riyadh," *Journal of Transportation Engineering*, vol. 127, no. 4, pp. 289–294, 2001.

[41] M. Mei and A. G. R. Bullen, "Lognormal distribution for high traffic flows," *Transportation Research Record: Journal of the Transportation Research Board*, no. 1398, pp. 125–128, 1993.

Journal of
Engineering

The Scientific
World Journal

International Journal of
Rotating
Machinery

Journal of
Sensors

International Journal of
Distributed
Sensor Networks

Advances in
Civil Engineering

Journal of
Control Science
and Engineering

Journal of
Robotics

Journal of
Electrical and Computer
Engineering

Hindawi

Submit your manuscripts at
http://www.hindawi.com

Advances in
OptoElectronics

VLSI Design

International Journal of
Navigation and
Observation

Modelling &
Simulation
in Engineering

International Journal of
Aerospace
Engineering

International Journal of
Chemical Engineering

International Journal of
Antennas and
Propagation

Active and Passive
Electronic Components

Shock and Vibration

Advances in
Acoustics and Vibration