Validation of a Traffic Model

WILHELM B. CRONJE

ABSTRACT

Simulation is very useful for generating traffic data. However, for validating a traffic model, field data are essential. Because of the random variation of the traffic arriving at signalized intersections, a large number of observations are necessary at a particular intersection. To obviate the collection of a large amount of data, the Bootstrap technique was applied to a limited amount of field data that were collected at three fixed-time signalized intersections. In addition, the data were supplemented by simulation to cover a wide range of cycle lengths, types of flow, degrees of saturation, and ratios of the variation in the mean of the number of arrivals per cycle. A recently developed traffic model based on a Markov process and a geometric probability distribution, the Modified Geometric Model (M Geom Model), was used to estimate the measures of effectiveness commonly used for optimization purposes, namely, delay and stops. Satisfactory results were obtained, indicating that a limited amount of field data is required for validating a traffic model.

Traffic arrival at a signalized intersection is a stochastic process. If a traffic count is taken on a specific day for a particular period at an intersection and the delay and stops are calculated, when the count is repeated for the same time period the next day, it is possible that the values obtained for delay and stops may differ appreciably. Consider Set A of generated arrivals obtained over 15 cycles with the following conditions: cycle length (c) is 50 sec; degree of saturation (x) is 1.05; flow (Q) is 500 vehicles per hour; and ratio of variance in the mean of arrivals per cycle (I) is 0.5.

Set A: 8, 8, 6, 6, 8, 8, 7, 8, 8, 5, 9, 8, 9, 8, 10: Σ 116. Initial queue length is 2 vehicles; average departures per cycle are 7.4 vehicles.

The equation for obtaining the overflow at the end of a cycle is as follows:

$$K = B + Z - V \tag{1}$$

where

- K = overflow at the end of the cycle,
- B = overflow at the beginning of the cycle,
- Z = arrivals per cycle, and
- V = departures per cycle.

If Equation 1 is applied to the generated arrivals, the following values for K are obtained at the end of 15 cycles: 2.6, 3.2, 1.8, 0.4, 1.0, 1.6, 1.2, 1.8, 2.4, 0.0, 1.6, 2.2, 3.8, 4.4, 7.0.

Suppose that the order of arrivals is changed to Set B:

Set B: 9, 8, 7, 8, 6, 6, 9, 10, 8, 8, 8, 8, 8, 5, 8: Σ 116.

The following values for K are now obtained: 3.6, 4.2, 3.8, 4.4, 3.0, 1.6, 3.2, 5.8, 6.4, 7.0, 7.6, 8.2, 8.8, 6.4, 7.0.

Department of Civil Engineering, University of Stellenbosch, Stellenbosch 7600, Republic of South Africa.

The two sets of overflow are plotted in Figure 1. Although the area under the curves is not the delay, it is directly related to delay and obviously the delay differs appreciably for the two sets.

Calculating delay as the area under the queue-length curve and stops as the number of arrivals while there is a queue plus the overflow from the previous cycle, the values for the two sets of arrivals obtained are indicated in Table 1 alongside those obtained by the Modified Geometric Model (M Geom Model), a model recently developed by Cronjé (1) based on a Markov process and the geometric probability distribution for estimating delay and stops.

From the values in Table 1 it is abundantly clear that the order of the arrivals affects delay and stops appreciably. If actual counts have to be used, a fairly large sample will be required to properly estimate the population measures of effectiveness. On the other hand, the larger the sample, the longer is the calander period over which the counts are taken, and during this period changes in the traffic flow conditions may occur that bias the sample. This problem is eliminated with computer simulation, which will therefore be applied.

PROCEDURE

In order to ascertain how variable traffic arrivals are at intersections, data were collected at fixed-time Intersection A. Observations were carried out on the same through approach lane on 6 days during the same peak hour. Data were collected on Monday through Thursday of the first week followed by Monday and Tuesday of the following week. The data were analyzed for variability. First the standard statistics--namely, the mean, the standard deviation, the coefficient of variation, and the coefficient of skewness--were computed. Second the Bootstrap technique developed by Efron (2) was applied to each set of data. The Bootstrap technique consists essentially of simulating a set of data by utilizing the generation of random numbers. It has the useful property of being independent of a particular distribution. Being a relatively recently developed technique, it will be illustrated by means of an example.

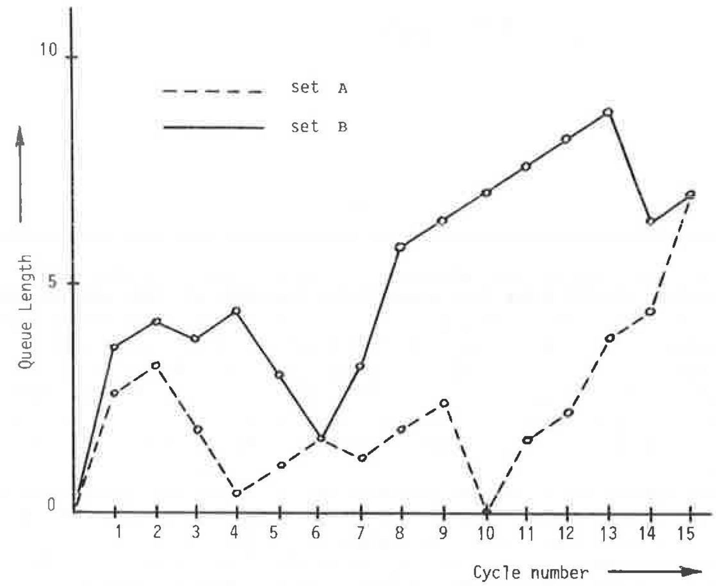

FIGURE 1 Queue length at end of each of 15 cycles.

TABLE 1 Delay and Stops for Two Sets of Generated Data

	Delay		Stops			
Set	Actual	M Geom	Actual	M Geom	ΔD (%)	ΔN (%)
A	3,770	6,797	150	210	80.3	40.0
B	6,053	6,797	192	210	12.3	9.4

Note: ΔD = difference in delay; ΔN = difference in stops.

Suppose that the following set of arrivals per cycle was collected at a traffic light over five cycles: 12, 8, 17, 14, 11. Since this is a random process, these arrivals could include many different patterns. Suppose that the first random number generated is 8,059. Converting to a random fraction gives 0.8059. This random fraction multiplied by 5, the number of observations, gives 4.0295. Adding 1 and dropping the fractional part gives 5. The number of arrivals for the first cycle of a new set is then the fifth observation of the collected set, namely, 11. This procedure is then repeated until five random numbers have been generated. This procedure can be repeated to give as many sets as required with a maximum of n^n, but always by manipulating the observed set. Please note that more than one cycle may have the same number of arrivals, implying that not all observed values are represented in each simulated set.

Bootstrap samples of 320 were drawn from the data points collected for each day, with replacement (Efron uses a number of Bootstrap replications as low as 100), and each simulated set was compared with the M Geom Model on a cost basis. Cost was obtained by applying monetary rates to the measures of effectiveness, namely, delay and stops. The reason for the sample size of 320 is that the generation of additional data later in the paper consists of 8 cycle lengths, 4 flows, and 10 values of I, giving $8 \cdot 4 \cdot 10 = 320$ combinations.

The analysis of the variability gave satisfactory results. Consequently the Bootstrap technique was also applied to data collected at two other fixed-time intersections, B and C. Data for Intersections A, B, and C were collected over 17, 50, and 60 cycles, respectively.

The determination of the required sample size in order to obtain meaningful Bootstrap results is a difficult problem. There is no general theory available and each application has to be investigated individually. In this application, the sample size required will depend on a number of factors, for example, the I-value and the distribution of arrivals. However, Efron has obtained meaningful results with samples as small as 15. From this experience it would appear that sample sizes of 17, 50, and 60 would be adequate.

The problem with small samples is that the tails of the true distribution may be suppressed. In this case, data were collected during six peak hours. Had data been collected over more peak hours it is possible that extreme values might have been observed. These values would undoubtedly affect delay and stops. This problem is not easy to solve. On the other hand, if data are collected during heavy peak-hour flows, extreme values are not so likely to occur.

In order to supplement the data to cover a wide range of cycle length, flow, degree of saturation, and ratio of the variation in the mean of the arrivals per cycle, data were computer generated. Satisfactory results were obtained in this case also.

In the development of the M Geom Model the following arrival and departure patterns as described by Miller (3) are assumed:

> In practice, vehicles start to slow down some distance away from a red signal. When the signal changes to green, vehicles gradually accelerate to a final speed. During both the deceleration and acceleration they are being delayed. Let delay be defined as the difference between delayed and undelayed journey times over a distance which includes the deceleration and acceleration distances on each side of a junction. The delay then is unaltered if, instead of slowing down gradually at the back of the queue, each vehicle maintains steady speed until it reaches the stop-line where it decelerates instantaneously, and then when the signals are green accelerates instantaneously to its final speed when its turn comes. This is the kind of behaviour which will be assumed so that delays can be derived mathematically. To ensure that delays are the same, the first vehicle must stay at the stop-line until a

little while after the start of the actual green phase to allow for its acceleration delay in practice. This intrusion upon the green phase is called lost time.

Because of the model assumptions, all the observed values at the back of the queue had to be adjusted to convert to arrivals at the stop line. This was done as outlined by Branston (4). The application of the equations will now be shown by means of an example.

The adjustment equations developed by Branston (4) are as follows:

$$Q_j = [Y_j + (Q_{j-1} \cdot d_{j-1})/v]/[1 + (d_j/v)] \quad (2)$$

$$k_j = Q_j/v \quad (3)$$

$$S_j = R_j - k_j \cdot d_j \quad (4)$$

where

- Q_j = count that would be observed at the stop line under the zero length assumption;
- Y_j = count by the upstream observer of the number of arriving vehicles per cycle;
- Q_{j-1} = count similar to Q_j, but in the preceding interval;
- d_j = distance between the observer at the stop line and the upstream observer at the end of the interval;
- d_{j-1} = distance similar to d_j, but at the end of the preceding interval;
- v = average running speed, assumed the same for all intervals;
- k_j = average concentration of vehicles between the observers under the zero length assumption;
- R_j = queue length measured by the upstream observer at the end of the interval; and
- S_j = queue length corresponding to R_j that would be measured under the zero length assumption.

Q_j and S_j are therefore the adjusted values that should be used for comparison purposes. The application of Equations 2 through 4 will now be given with data from Table 2.

TABLE 2 Traffic Count for Intersection B

Cycle	B	Z	V	Cycle	B	Z	V
1	2	15	13	26	18	13	12
2	4	13	13	27	19	12	14
3	4	11	12	28	17	10	13
4	3	13	11	29	14	12	12
5	5	11	11	30	14	14	15
6	5	12	12	31	13	17	11
7	5	9	9	32	19	15	14
8	5	6	5	33	20	13	10
9	6	12	12	34	23	13	11
10	6	13	11	35	25	14	13
11	8	14	13	36	26	10	12
12	9	11	12	37	24	16	13
13	8	12	12	38	27	11	15
14	8	13	13	39	23	15	12
15	8	14	12	40	26	11	14
16	10	9	11	41	23	12	13
17	8	12	12	42	22	15	12
18	8	15	14	43	25	11	15
19	9	14	12	44	21	17	12
20	11	21	13	45	26	11	11
21	19	14	12	46	26	12	13
22	21	12	14	47	25	11	12
23	19	16	11	48	24	9	10
24	24	9	14	49	23	14	13
25	19	12	13	50	24	10	13

The details for Intersection B are as follows: c is 80 sec, effective green time (g) is 22 sec, average running speed (v) is 45 km/hr, and average vehicle length in queue (q_1) is 5.91 m.

Using data from the first cycle in Table 2, the following values are obtained: $Y_1 = 15$; $d_1 = 4 \cdot 5.91 = 23.64$; and $d_{j-1} = d_0 = 0$. (Branston states that $d_0 \approx 0$ if the initial queue is less than 20 m. In this case, B = 2, giving a queue length = $2 \cdot 5.91$ = 11.82 < 20 m.) Substituting these values into Equation 2 gives

$$Q_1 = (15 + 0)/[1 + (4 \cdot 5.91 \cdot 3.6)/(80 \cdot 45)]$$
$$= 14.654$$

Substituting $Q_1 = 14.654$ into Equation 3 gives

$$k_1 = (14.654 \cdot 3.6)/(80 \cdot 45) = 0.015$$

Substituting $k_1 = 0.015$, $d_1 = 23.64$, and $R_1 = 4$ into Equation 4 gives

$$S_1 = 4 - 0.015 \cdot 23.64 = 3.645$$

Similarly

$$Q_2 = [13 + (4 \cdot 14.654 \cdot 5.91 \cdot 3.6)/(80 \cdot 45)]$$
$$\div [1 + (4 \cdot 5.91 \cdot 3.6)/(80 \cdot 45)] = 13.038$$

$$k_2 = (13.038 \cdot 3.6)/(80 \cdot 45) = 0.013$$

$$S_2 = 4 - 0.013 \cdot 4 \cdot 5.91 = 3.693$$

The values obtained are shown in Figure 2.

ANALYSIS

To ascertain how delay and stops can vary in practice, data were collected at signalized Intersection A on the same lane for the same peak period over 17

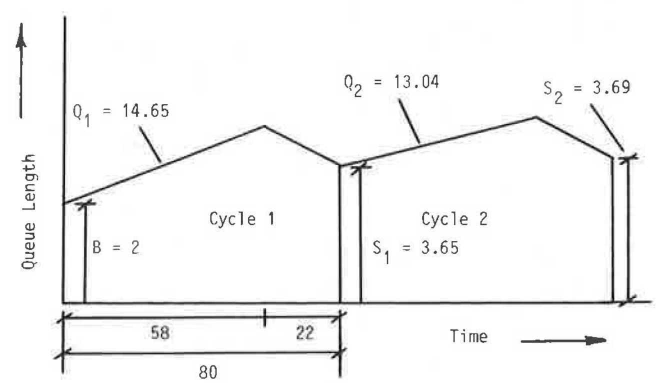

FIGURE 2 Indicating adjusted data for first two cycles of Table 2.

cycles. The data collected are given in Table 3. The data were adjusted as previously described and delay and stops were obtained. The intersection data and the mean of the flows were also applied to the M Geom Model; the results are indicated in Table 4.

From an inspection of Table 4 it is clear that there is variation in the total flow and also in the value of I, which partially explains the variation in delay and stops. The percentage differences indicate the degree to which the M Geom Model differs from the actual delay and stops for a particular day. What is revealing, however, is that for the same flow on the first and fourth days, there is appreciable difference in the delay and stops (these

TABLE 3 Traffic Count for Intersection A

Cycle	B	Z	V	Cycle	B	Z	V
Monday, Oct. 22, 1984							
1	0	16	15	10	3	21	21
2	1	30	28	11	3	24	24
3	3	18	21	12	3	20	19
4	0	32	25	13	4	28	27
5	7	19	26	14	5	29	26
6	0	29	26	15	8	19	24
7	3	18	20	16	3	26	25
8	1	21	21	17	4	17	19
9	1	25	23				
Tuesday, Oct. 23, 1984							
1	0	27	26	10	2	16	17
2	1	28	27	11	1	12	11
3	2	22	22	12	2	19	18
4	2	12	13	13	3	23	24
5	1	30	28	14	2	27	25
6	3	24	26	15	4	20	21
7	1	18	19	16	3	23	23
8	0	10	9	17	3	18	20
9	1	22	21				
Wednesday, Oct. 24, 1984							
1	0	27	25	10	0	14	14
2	2	19	16	11	0	25	25
3	5	22	21	12	0	16	15
4	0	29	26	13	1	21	21
5	3	27	26	14	0	27	26
6	4	32	26	15	1	18	19
7	10	23	29	16	0	25	24
8	4	14	18	17	1	31	27
9	0	19	19				
Thursday, Oct. 25, 1984							
1	1	16	17	10	0	28	27
2	0	18	18	11	1	27	26
3	0	16	16	12	2	26	27
4	0	31	27	13	1	18	19
5	4	28	27	14	0	27	27
6	5	23	25	15	0	22	22
7	3	26	27	16	0	30	27
8	2	14	16	17	3	21	23
9	0	21	21				
Monday, Oct. 29, 1984							
1	3	21	24	10	0	12	12
2	0	25	23	11	0	25	24
3	2	26	21	12	1	16	17
4	1	18	19	13	0	25	23
5	0	22	22	14	2	31	31
6	0	14	14	15	2	22	23
7	0	18	18	16	1	23	23
8	0	19	19	17	1	26	26
9	0	18	18				
Tuesday, Oct. 30, 1984							
1	0	28	25	10	0	23	22
2	3	25	27	11	1	22	23
3	1	31	28	12	0	29	27
4	4	30	26	13	2	29	26
5	8	21	26	14	5	23	25
6	3	15	18	15	3	17	18
7	0	15	15	16	2	20	21
8	0	18	17	17	1	29	28
9	1	16	17				

values are underscored). The reason for this difference is the order of the arrivals, as stated previously. In Table 4 the small differences between the statistics indicate that the samples are representative. The coefficient of kurtosis was not calculated because Benjamin and Cornell (5) state that for small samples its use is not recommended. Therefore the Bootstrap technique developed by Efron (2) was applied to the data of Intersection A to obtain a distribution of the cost difference between the delay and stops obtained from the simulated arrivals and the M Geom Model. The monetary rates used in the Republic of South Africa (6) [0.0174 cent/vehicle-sec of delay and 3.1 cents/stop (1 Rand = 100 cents)] were applied to delay and stops. This distribution indicates how variation of traffic flow in practice can affect the cost. The arrivals were rearranged 320 times, giving 320 sets of data. According to Benjamin et al. (5), if the number of data values is m, the number of intervals between the minimum and maximum values in the sample (k) should be given approximately by

$$\bar{k} = 1 + 3.3 \log m \tag{5}$$

For the sample in which m is 320, this equation gives

$$\bar{k} = 1 + 3.3 \cdot 2.505 = 9.27 \approx 10$$

The groups and frequencies for the 320 sets of data for each day were combined and are indicated in Table 5 together with the mean and the standard deviation. The histograms for the 320 sets of data for each day were combined to give Figure 3. The χ^2-test was applied to the data and the results for each day's data are indicated in Table 6.

From an inspection of the results obtained for Intersection A as indicated in Tables 5 and 6 and Figure 3, it is clear that there is close agreement between the distributions as indicated by the following: there is not much between-group variation for the data because the individual standard deviations are 8.83, 9.36, 9.76, 8.75, 10.02, and 5.94 and the combined standard deviation is 9.24. The observed distributions are all approximately normal in shape (in fact, for four out of six one cannot reject the hypothesis at the 5 percent level in spite of the relatively large sample size of 320, and the remaining two are rejected by a very narrow margin). These results indicate that in order to obtain a distribution for a particular intersection only one set of representative data needs to be simulated by the Bootstrap technique. This technique was therefore applied to the data of Intersections B and C indicated in Table 7 and also in Tables 2 and 8.

The groups, frequencies, means, and standard deviations for the 320 sets of data are indicated in Table 9 for Intersections B and C. The histograms of the data are shown in Figures 4 and 5, respectively. The M Geom Model was also applied to the data indicated in Table 7 and the cost was obtained for delay and stops. The difference between this cost and the cost obtained for delay and stops for the single sets of data given in Tables 2 and 8 is indicated by dashed arrows in Figures 4 and 5 for Intersection B (-3.32 percent) and Intersection C (21.84 percent), respectively.

The percentage of cost difference for the Intersection B data falls very close to the zero value, whereas that for the Intersection C data is much further away. However, the distributions indicate how the cost difference can vary because of the variation in the order of arrivals.

VALIDATION BY GENERATED DATA

To supplement the limited amount of observed data, arrivals at a traffic-signalized intersection were generated on the computer for various values of the cycle length (c), degree of saturation (x), flow (Q), and the ratio of the variance in the mean of arrivals per cycle (I). To introduce a large degree of variation into the generated data, I was assigned

TABLE 4 Results for Intersection A

Day	I	Q	D	N	ΔD (%)	ΔN (%)	$\bar{x}$	$\bar{s}$	v	$\bar{g}$
Monday	1.15	384.3	42,532	438	34.4	18.4	23.06	5.15	0.22	0.26
Tuesday	1.66	344.1	35,510	380	12.2	2.7	20.65	5.86	0.28	-0.25
Wednesday	1.39	381.4	38,664	418	22.2	13.0	22.88	5.64	0.25	-0.07
Thursday	1.22	384.3	36,995	413	16.9	11.6	23.06	5.31	0.23	-0.19
Monday	1.14	353.9	32,587	373	3.0	0.8	21.24	4.93	0.23	-0.06
Tuesday	1.36	383.3	39,387	423	24.5	14.3	23.00	5.60	0.24	-0.04

Note: c = cycle length (sec); g = effective green time (sec); s = saturation flow (vehicles/sec); I = ratio of the variance in the mean of the number of arrivals per cycle; Q = arrival flow (vehicles/hr); D = delay over 17 cycles (vehicle-sec); N = stops over 17 cycles; ΔD% = percentage difference in delay between the M Geom Model and the delay for the particular day; ΔN% = percentage difference in stops between the M Geom Model and the stops for the particular day; $\bar{x}$ = sample mean; $\bar{s}$ = sample standard deviation; v = coefficient of variation; and $\bar{g}$ = coefficient of skewness.

For Intersection A, c = 216 sec, g = 50 sec, and s = 0.55 vehicle/sec. Results for the M Geom Model: Q = 371.9, D = 31,645, N = 370.

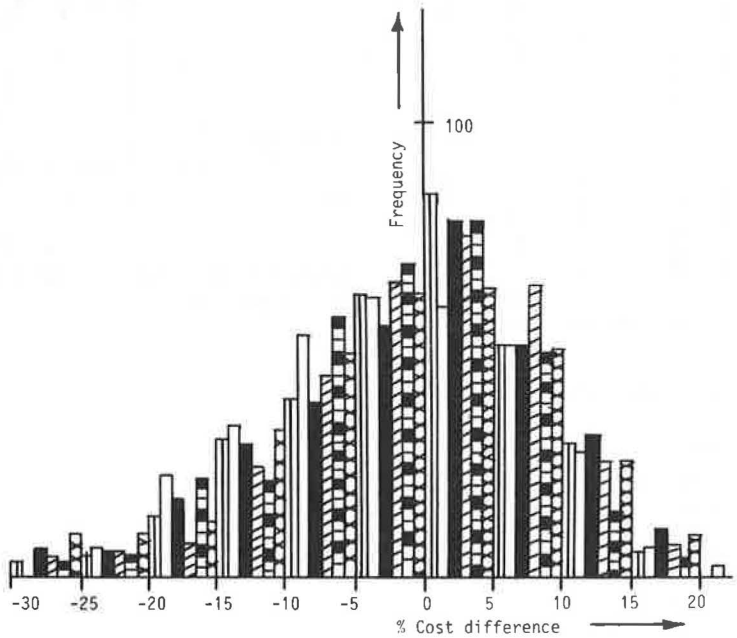

FIGURE 3 Combination of distributions of data for Intersection A.

TABLE 5 Frequencies for Intersection A Data

Cost Interval (%)	Generated Frequency
-30 to -25	25
-25 to -20	33
-20 to -15	92
-15 to -10	169
-10 to -5	280
-5 to 0	374
0 to 5	437
5 to 10	316
10 to 15	151
15 to 20	41
20 to 25	2
	1,920 (6·320)

Note: $\bar{x}$ = -1.081 percent and $\bar{s}$ = 9.236 percent.

TABLE 6 Results of χ^2-Test for Each Day's Data

Day	$\bar{x}$	$\bar{s}$	χ^2	f	$\chi^2_{0.05}$
Monday	-0.56	8.83	13.06	6	12.59
Tuesday	-1.45	9.36	11.91	6	12.59
Wednesday	-0.69	9.76	12.98	6	12.59
Thursday	-0.11	8.75	11.15	6	12.59
Monday	-1.81	10.02	10.18	7	14.07
Tuesday	1.69	5.94	9.25	6	9.47

Note: f = degrees of freedom, χ^2 = calculated value, $\chi^2_{0.05}$ = tabulated value.

TABLE 7 Data for Intersections B and C

Intersection and Cycle	c	g	$\bar{s}$	$\bar{x}$	I	Q
B						
Cycles 1-50	80	22	0.56	1.02	0.38	566.7
C						
Cycles 1-60	70	30	0.49	0.97	1.10	731.2

the following values (data collected in previous studies by the author indicate that these are practical values): 0.5, 0.6, 0.7, 0.8, 0.9, 1.1, 1.2, 1.3, 1.4, and 1.5. The cycle length was assigned the following values: 50, 60, 70, 80, 90, 100, 110, and 120. The degree of saturation and the flow were assigned the following paired values:

x	Q (vehicles/hr)
1.05	500
1.10	650
1.15	800
1.20	900

Arrivals were generated over 15 cycles and converted to flow. The cycle length and degree of saturation were kept the same and the corresponding effective green time g (in seconds) was calculated. The input and output data obtained for c = 50 sec, x = 1.05,

TABLE 8 Traffic Count for Intersection C

Cycle	B	Z	V	Cycle	B	Z	V
1	2	8	10	31	5	16	15
2	0	13	13	32	6	11	16
3	0	16	12	33	1	9	10
4	4	23	16	34	0	17	16
5	11	17	15	35	1	21	16
6	13	11	15	36	6	17	18
7	9	13	16	37	5	12	16
8	6	15	15	38	1	14	14
9	6	20	17	39	1	14	15
10	9	13	17	40	0	18	18
11	5	15	16	41	0	15	9
12	4	21	16	42	6	7	13
13	9	9	17	43	0	15	13
14	1	20	14	44	2	13	15
15	7	11	17	45	0	14	14
16	1	14	14	46	0	10	10
17	1	17	18	47	0	11	11
18	0	16	16	48	0	14	14
19	0	19	18	49	0	9	8
20	1	19	16	50	1	7	7
21	4	14	17	51	1	17	14
22	1	28	18	52	4	7	10
23	11	9	15	53	1	5	6
24	5	13	17	54	0	11	10
25	1	16	16	55	1	12	13
26	1	16	17	56	0	13	12
27	0	19	15	57	1	17	16
28	4	11	15	58	2	11	13
29	0	17	15	59	0	15	13
30	2	18	15	60	2	10	12

TABLE 9 Frequencies for Intersection B and C Data

Cost Interval (%)	Generated Frequency
Intersection B	
-75 to -60	6
-60 to -45	7
-45 to -30	22
-30 to -15	36
-15 to 0	45
0 to 15	50
15 to 30	74
30 to 45	55
45 to 60	25
	320
Intersection C	
-140 to -120	5
-120 to -100	9
-100 to -80	10
-80 to -60	19
-60 to -40	27
-40 to -20	37
-20 to 0	60
0 to 20	95
20 to 40	58
	320

Note: For Intersection B, $\bar{x}$ = 8.34 and $\bar{s}$ = 28.69. For Intersection C, $\bar{x}$ = -13.25 and $\bar{s}$ = 38.78.

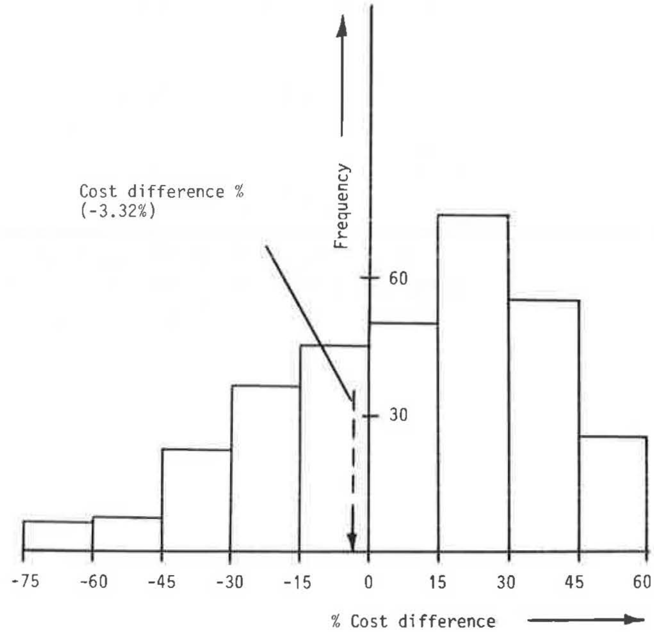

FIGURE 4 Histogram of sample from simulated arrivals for Intersection B data.

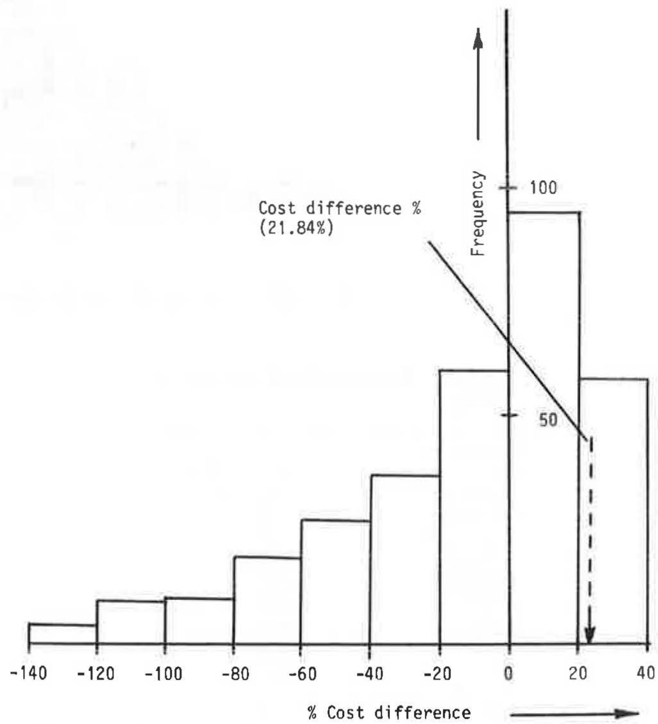

FIGURE 5 Histogram of sample from simulated arrivals for Intersection C data.

and Q = 500 are given in Table 10. A total of 320 sets of data were obtained.

Delay and stops for the generated arrivals were calculated as previously described. The output data indicated in Table 10 were applied to the M Geom Model and the delay and stops were obtained. A histogram of the data is shown in Figure 6 and the frequencies are given in Table 11.

DISCUSSION OF RESULTS

In the case of the wide range of generated data, the M Geom Model estimates cost on the average within -1.906 percent. In the case of the simulated field data the means are as follows: Intersection A, -1.081 percent; Intersection B, 8.34 percent; Intersection C, 13.25 percent. The percentage of data continued between $\bar{x} + \bar{s}$ and $\bar{x} - \bar{s}$ for the generated case and for Intersections A, B, and C is 67.5, 67.8, 64.9, and 70.1 percent, respectively. This indicates not only that the means are relatively low but that there is also appreciable cluster around the mean. In the light of these results it can be concluded that validation of the M Geom Model has been established.

TABLE 10 Data for Generating Arrivals

Input				Output			
c	x	Q	I	c	x	Q	g
50	1.05	500	0.5	50	1.05	542.4	14.4
			0.6			547.2	14.5
			0.7			556.8	14.7
			0.8			499.2	13.2
			0.9			552.0	14.6
			1.1			499.2	13.2
			1.2			537.6	14.2
			1.3			432.0	11.4
			1.4			480.0	12.7
			1.5			465.6	12.3

TABLE 11 Frequencies for Generated Data

Cost Interval (%)	Generated Frequency
-40	6
-40 to -30	6
-30 to -20	23
-20 to -10	54
-10 to 0	81
0 to 10	80
10 to 20	51
20 to 30	17
30 to 40	2
	320

Note: $\bar{x} = -1.906$ percent and $\bar{s} = 15.173$ percent.

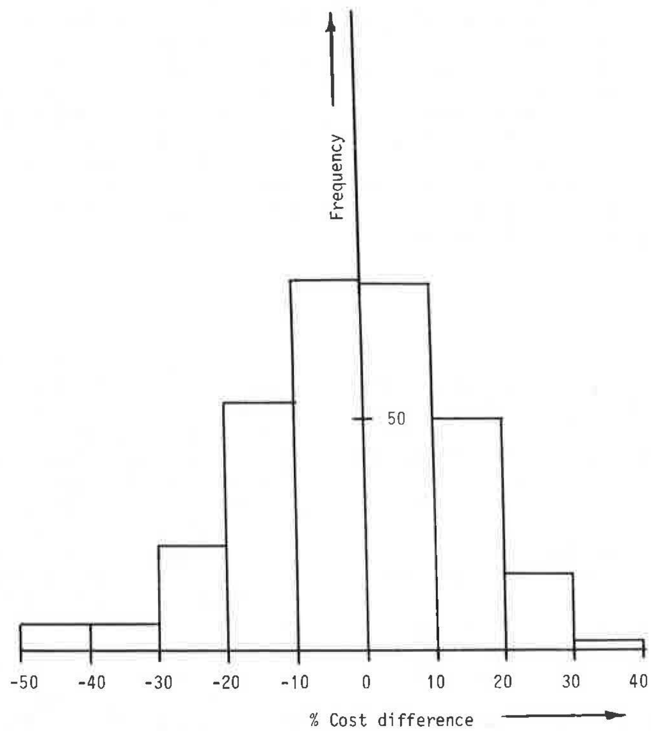

FIGURE 6 Histogram of sample from generated arrivals for the M Geom Model.

The results further indicate that because of the variation in the arrival of traffic at signalized intersections, the Bootstrap technique must be applied to simulate the field data. Unless this is done, the results can be very misleading.

In conclusion, for the validation of a traffic model a limited amount of field data is required. Some can be simulated by the Bootstrap technique and supplemented by generated data to cover a wide range of cycle lengths, degrees of saturation, types of flow, and ratios of the variance in the mean of the arrivals per cycle.

REFERENCES

1. W.B. Cronjé. Optimization Model for Isolated Signalized Traffic Intersections. In Transportation Research Record 905, TRB, National Research Council, Washington, D.C., 1983, pp. 80-83.
2. B. Efron. The Jacknife, The Bootstrap and Other Resampling Plans. Society for Industrial and Applied Mathematics, Philadelphia, Pa., 1982.
3. A.J. Miller. Settings for Fixed-Cycle Traffic Signals. Operational Research Quarterly, Vol. 14, No. 4, 1963.
4. D. Branston. A Comparison of Observed and Estimated Queue Lengths at Oversaturated Traffic Signals. Traffic Engineering and Control, Vol. 19, July 1978.
5. J.R. Benjamin and C.A. Cornell. Probability, Statistics and Decision for Civil Engineers. McGraw-Hill Book Company, New York, 1970.
6. I.C. Schutte. Manual on the Economic Evaluation of Transportation Projects, Part 2: National Cost Data. CSIR Draft Manual K64. Council for Scientific and Industrial Research, Pretoria, South Africa, 1983.

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