Procedure for Validation of Microscopic Traffic Flow Simulation Models

RAHIM F. BENEKOHAL

Model verification and validation are two important tasks in developing a traffic simulation model. Traffic simulation models have unique characteristics because of the interaction among the drivers, vehicles, and roadway. The effects of the interaction on traffic flow should be considered in verification and validation of the models. If these two tasks are not properly performed, a traffic simulation model may not provide accurate results. A procedure for verification and validation of microscopic traffic simulation models is developed, and its application to a car-following simulation model, CARSIM, is demonstrated. The validation part of the procedure is emphasized. The validation efforts are performed at the microscopic and macroscopic levels. For validation at the microscopic level, the speed change patterns and trajectory plots obtained from simulation models are compared with those from a field data. For validation at the macroscopic level, the average speed, density, and volume for simulated platoons are compared with those of field data. Also, variation of these parameters when the platoons go through a disturbance and interrelationships between these variables computed from the simulation models and the field data are examined. Regression analysis and analysis of variance of the simulation results versus the field data are discussed. The procedure may be considered as a step toward development of a comprehensive systematic approach for verification and validation of traffic simulation models.

Two important tasks in developing a traffic simulation model are verification and validation of the model. Verification is to check if the model behaves as the experimenter assumes it does, and validation is to test whether the simulation model reasonably approximates a real system (*1*). Traffic simulation models have unique characteristics because of the interaction among the drivers, vehicles, and roadway environment. The effects of the interaction on traffic flow should be considered in verification and validation of the models. A traffic simulation model may not provide accurate results if these two tasks are not properly performed.

Often the users of traffic simulation models do not examine how well the verification and validation steps are carried out, or do not have access to such information. Consequently, the users rely on the model performance assuming that enough verification and validation have been done. Even when the information is available, it is difficult for most users to compare validation of one model to another model because they are validated differently. Thus, there is a need for developing a systematic approach for verification and validation of traffic simulation models to provide some degree of consistency and to increase the reliability of the models.

A procedure for verification and validation of traffic flow simulation models is suggested. The procedure is discussed and its application for verification and validation of a car-following simulation model, CARSIM (*2*), is demonstrated. Following this approach in validation of a model most likely will increase the reliability of the simulation results. However, using the suggested approach does not guarantee that this model will simulate the real-world conditions better than another model. The procedure may be considered as a step toward development of a comprehensive systematic approach for verification and validation of traffic simulation models.

Department of Civil Engineering, University of Illinois, 205 N. Mathews Avenue, Urbana, Ill. 61801-2397.

BACKGROUND

Validation

A model should be validated under different experimental conditions to obtain a high model confidence. Validation is to see whether there is an adequate agreement between the model and the system being modeled. Annino and Russell (*3*) stated that using unverified models and lack of understanding the system were among most frequent causes of simulation failure. Sargent (*4*) suggested that model validation should consist of conceptual validation, computerized validation, operational validation, and use of adequate and correct data.

For the conceptual validation, the theories, the assumptions, and the relationships used are checked to ensure they are correct and proper for each submodel and for the overall model. Sargent (*4*) suggested using the tracing and face validation techniques. In tracing, the behavior of different entities (e.g., vehicles) is traced through each submodel and overall model to determine if the model's logic is correct and whether the necessary accuracy is obtained. In face validation, the experts in the subjects are asked to evaluate the logic of the submodel and the model and the input-output relationships.

In computerized model validation, it is checked to ensure that the conceptual model is implemented, and the computer program runs properly (error free). Each submodel is tested to see if it works properly and the overall model is executed under different conditions to investigate input and output relations. Operational validity is ensuring that the simulation model is a reasonable and accurate representation of the real system with certain levels of confidence. Here the validation can be done subjectively, such as graphical representation and examination of it, or can be done objectively such as using statistical techniques.

Various statistical techniques have been used for validation of simulation models. Torres et al. (*5*) developed a statistical guideline based on differences between the real world and traffic simulation results. Gafarian and Walsh (*6*) used travel time and velocity of vehicles as the measure of effectiveness of simulation model, and compared the simulation results with observed values using the Wilcoxon signed-rank test. Mihram (*7*) discussed the five stages in model building and the procedures for comparison of the results from several independent replications of a simulation model with that of one or more observed data.

Kleijnen (*8*) discussed techniques for validation of simulation models using chi squared, factor analysis, spectral analysis, and regression analysis between the actual and the simulation outputs. Naylor et al. (*9*) provided three alternative forms of analysis of variance for analysis of output from computer simulation experiments and for making a decision about their differences and ranking. Kleijnen (*10*) discussed how regression analysis is used to obtain a metamodel (a model explaining the simulation model) and how the effect of qualitative or quantitative factors can be investigated by using the weighted least squares technique. The four steps suggested by Sargent (*4*) were used to develop a procedure that will be discussed after review of the truncation and replication policies.

Truncation and Replication

The results of simulation studies should be collected after the system reaches a steady state condition. The data before the steady state conditions (initial transient state) may be eliminated to minimize the bias on the mean value of the response variable. There is not a definite rule on how the bias should be eliminated and how much of the simulation run should be truncated for this purpose. One way of finding out how much of the data should be truncated is by plotting the response variable versus time and locating the beginning of the region of steady state condition.

Start-up policies in simulation and reducing the effect of initial transient state were surveyed by Wilson and Pritsker (*11,12*). They found that deleting data from the beginning of the simulation output to reduce initial transient effect (bias reduction) causes loss of information and an increase in the variance. The net effect of deleting the initial observations was to increase the mean square error of the sample mean. Starting from empty and idle condition (no truncation of initial condition) provided a lower estimate of mean square error. Considering bias, variance reduction, and mean square error, they suggested starting as close to the steady state mode as possible and keeping all data. Furthermore, they stated, "The judicious selection of an initial condition appears to be more effective than truncation in improving the performance of the sample mean as an estimator of the steady-state mean." Also, the research has indicated that for small and well-behaved models truncation should not be performed, but for large models this may not be the case (*13*). Even after truncating the transient state, the simulation results will continue to fluctuate because of the stochastic variable used in the model.

Kleijnen (*14*) discussed that replication of simulation runs is the only alternative for gathering statistics about terminating systems. In a terminating system, the simulation run ends if a specific event occurs. In order to obtain independent replications, different random number seeds ought to be used. Initialization for replicated runs is performed by throwing away some observations from the beginning of each run. The rule-of-thumb is that the observations may be discarded as long as the response variable continues to increase or decrease. From several replications, the response can be obtained for each run and then the mean and the variance of the responses can be found. For independent results, a *t*-test can be run and the confidence level for the mean of observations can be found. However, in simulation the results do not always meet the requirements of a traditional *t*- or *f*-test. Kleijnen (*15*) provided techniques to compare means and variances of two simulations in which the outcomes from replications are pairwise correlated. To compare the means of autocorrelated observations of two simulation experiments, Fishman (*16*) suggested that the difference of the sample means be treated as a normal variate for a sufficiently long sample record. This procedure is not as good as comparing autocorrelation structure of the two experiments, but it suffices as an initial step for comparison (*15,16*).

When two operating conditions are to be compared, one can introduce a negative correlation between replications of runs under one operating condition to reduce the variance for within runs; and then introduce a positive correlation between runs under different operating conditions to reduce the variance for the difference between runs. This procedure is suggested as an efficient experiment design if there is not an initial transient phase (*17*). This variance reduction method uses antithetic variate and common random numbers jointly. Kleijnen (*18*) discusses possible undesirable effects of such a combination and indicates that, because of cross correlations, the results may even be worse than using each method alone. He compared the three methods for different conditions but could not determine which method was best for all systems.

SUGGESTED APPROACH

The suggested approach has two major tasks: verification and validation. Each task is performed at microscopic and macroscopic levels. At each level, several steps are suggested and application of some of them is discussed. The verification task includes efforts similar to those suggested by Sargent (*4*) for the conceptual and computerized validation, and the validation task is comparable to the operational validation discussed by Sargent (*4*).

Some forms of the steps suggested have been used in validation of other traffic simulation models. The author has greatly benefited from numerous studies in developing this approach. For instance, trajectory comparison and speed fluctuation were used in validation of INTRAS (*19,20*). In validation of other simulation models, comparable efforts have been made to ensure validity of the simulation results. The steps suggested here should be considered as the starting point for the efforts needed for verification and validation of a microscopic traffic simulation model, in general. For a particular simulation model, the developer should decide how suitable the suggested approach is, and take additional appropriate steps needed to obtain reliable results.

In the following sections, the verification task will be outlined, because of space limitation, and the validation task will be discussed in detail. Benekohal (*2,21*) provided further information on the verification and validation steps. The field data needed and the data used in validation of CARSIM are briefly described in the following sections.

Field Data

The data used for verification and validation should be accurate and appropriate for the model. A data set different than the one used for verification should be used for validation of the model. The data sets used for validation of CARSIM are the Ohio State University trajectory data collected using aerial photogrammetric techniques (*22*). The aerial photographs were taken in 1-sec time intervals from an elevation of about 3,000 ft. The location of a vehicle was determined with an accuracy of ± 0.50 ft and the speed was determined with an accuracy of ± 1.0 mph.

The field data provided a complete record of spacings, headways, longitudinal positions, and velocities for individual vehicles. The data were from median lane of I-71, near Columbus, Ohio, with normal traffic. Four platoons were used in validation of CARSIM. The platoon selected for presentation was Platoon 123, which went through a severe disturbance. The platoon had 15 vehicles with no vehicle entering or leaving the platoon.

Model Verification

Verification ensures that the model behaves as intended. The program should be debugged first to eliminate any coding errors and programming problems. Then, the logic of different components of the model, such as car following, lane changing, merging, or diverging, should be carefully reviewed. Also, the acceleration and deceleration patterns, velocity change patterns, trajectory plots, and headways obtained from the simulation model should be examined. Sensitivity of these parameters to changes in the input variables should be studied. Model calibration, if needed, should be performed using field data for fine tuning of some of the variables in the model.

For verification of CARSIM, the following parameters were systematically varied and the sensitivity of the model outputs to the changes were carefully examined: maximum deceleration rate of a vehicle, compliance level of drivers, start-up delay of stopped vehicles, reaction time of drivers, buffer space between vehicles, and traffic mix. Then, different disturbances in traffic flow were induced to a platoon of 15 vehicles and their effects at microscopic and macroscopic levels were examined. A disturbance is induced when the leader of a platoon is required to decelerate, stop, and accelerate to a specified speed.

For the microscopic level verification, the effects of a regular disturbance, an emergency stop, and a stop-and-go operation on individual vehicle's trajectory, speed, and acceleration or deceleration were analyzed at high- and low-volume levels. For the macroscopic level verification, the effects of the regular disturbance and an emergency deceleration on average speed, density, volume, and average headway were examined at high- and low-volume levels. For example, the acceleration and deceleration patterns for a 15-car platoon in a stop-and-go condition are shown in Figure 1. The patterns are shown for the 1st, 4th, 10th, and last car in the platoon. Benekohal (*21*) provided further information on verification.

Model Validation

Two levels are proposed for validation of a microscopic traffic simulation model: (a) microscopic level, and (b) macroscopic level. At the microscopic level, the attributes of individual vehicles such as location, time, headway, and speed computed from the simulation model are compared with those obtained from the field data. At the macroscopic level, the aggregate parameters such as the average speed, density, and volume of a platoon of vehicles computed from the simulation model are compared with the results from field data. The steps for validation of CARSIM are shown in Figure 2.

For validation of CARSIM, four platoons covering a wide range of traffic conditions were used. The results for one of the platoons (Platoon 123) are presented here. The results for the other platoons are given elsewhere (*2,21*). Five independent replications were made for each traffic condition using different random number seeds. In each one of these runs, the attributes of all vehicles were generated randomly from the respective distributions. However, the location and speed of the leader of the simulated platoon were set equal to that of the leader of the field data. Thus, the leaders of the platoons had the same location and speed, but the followers may or may not.

It is recommended to use different platoons and for each platoon to make independent replications. The number of platoons and replications would depend on the system to be simulated and the range of variation of the response variable. The platoons should represent the real-world traffic conditions that the model is likely to simulate. When the range of the responses from different replications is narrow, a few runs might be enough. However, when a large variation is observed, more replications are needed. This topic will be discussed more in the macroscopic validation section.

Validation at Microscopic Level

The microscopic level validation is, perhaps, the most difficult task in validating a traffic simulation model. For microscopic validation, the variation of some or all of the following parameters should be examined: speed profile of an individual vehicle, location of the vehicle on the road, time headways between vehicles, and spacing between successive vehicles. For microscopic level validation of CARSIM, the speed change patterns and trajectory plots generated by CARSIM were compared with those obtained from the field data. The speed and location of every vehicle were determined at 1-sec time intervals in the simulation and field data. In the following sections, the changes on these variables will be examined.

Speed Profile Comparison

The speed of an individual vehicle was computed at 1-sec time intervals and a speed profile for each vehicle was generated by plotting speed versus time. The average speed of a vehicle

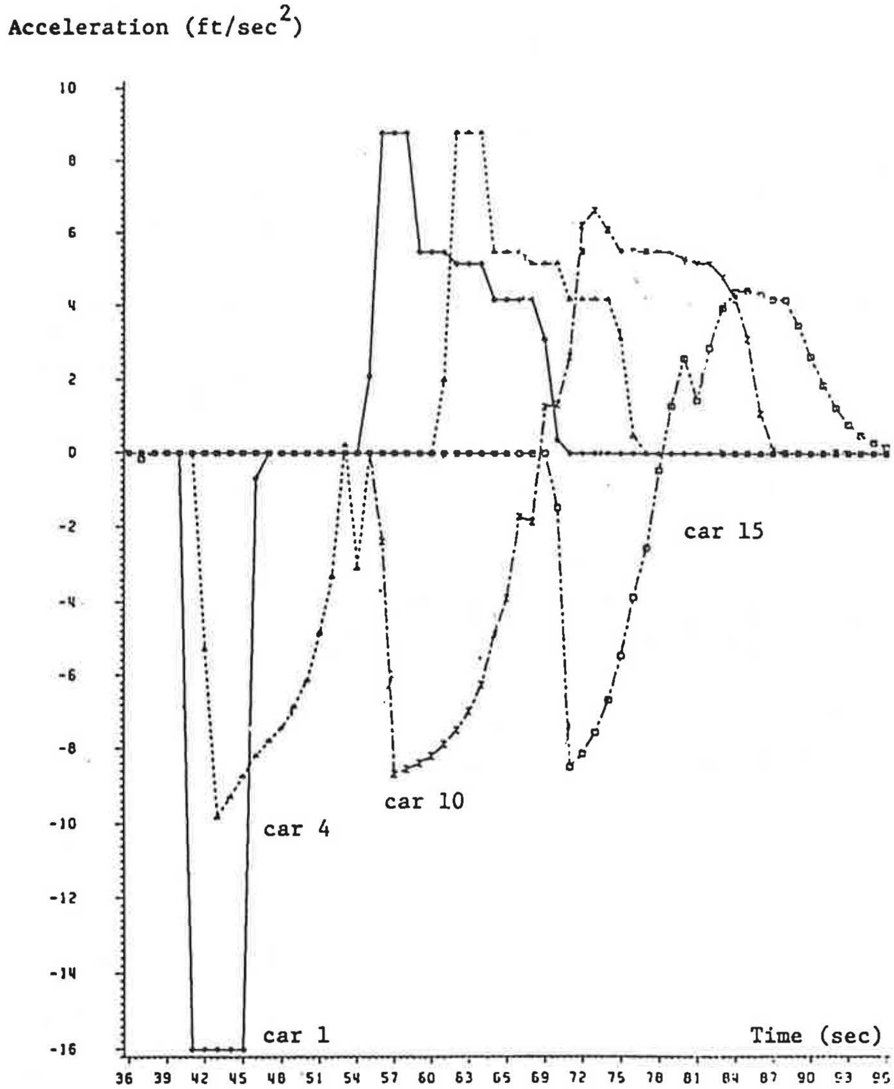

FIGURE 1 Acceleration-deceleration patterns for a platoon of vehicles in a stop-and-go condition. The leader of the platoon decelerates at 16 ft/sec^2, stops for 9 sec, and accelerates to a desired speed.

at a given time interval was computed as the mean of five speeds from the replications. Then, the average speed of an individual vehicle at a given time was compared to the observed speed from field data. Comparison of the speeds from simulation and field data are shown in Figure 3, which shows the speeds for the lead car, the 4th car, and the last (15th) car of Platoon 123. The other cars could not be shown because the plot became cluttered.

The simulation model generated speed change patterns similar to those from the field data. All vehicles came to a complete stop and then accelerated to reach their desired speed. The similarity of the speed change patterns indicates that the simulation model replicates the real-world traffic disturbance with an acceptable accuracy. An acceptable level of accuracy would depend on the model and the purpose the model is used for. Here, the similarities of speed change patterns were considered the important criteria in accepting accuracy of the model. Speed difference at each time interval may also be considered as the criteria; however, the difference may exhibit a large fluctuation because the speeds are updated at short time intervals (e.g., 1 sec). In the simulated platoon, the vehicles exhibited less speed fluctuation than the vehicles in the field data, because of the fact that the model was programmed to mimic only limited characteristics of real drivers.

From the comparison of speed change patterns, one may conclude how well the model duplicates the real-world speed change patterns in various traffic conditions. The criteria for the operational validation may be comparison of the shape of the speed profile for a vehicle generated by the model and the field data. The shift between the two profiles for a vehicle is not as important as the similarity of the shape of the profiles, because the shift would depend on the drivers' characteristics, but the profile would reflect the model's capability to simulate actual speed profiles. A driver with a longer reaction time would cause a larger shift than a driver with a shorter reaction time, but the profile for both drivers may look similar.

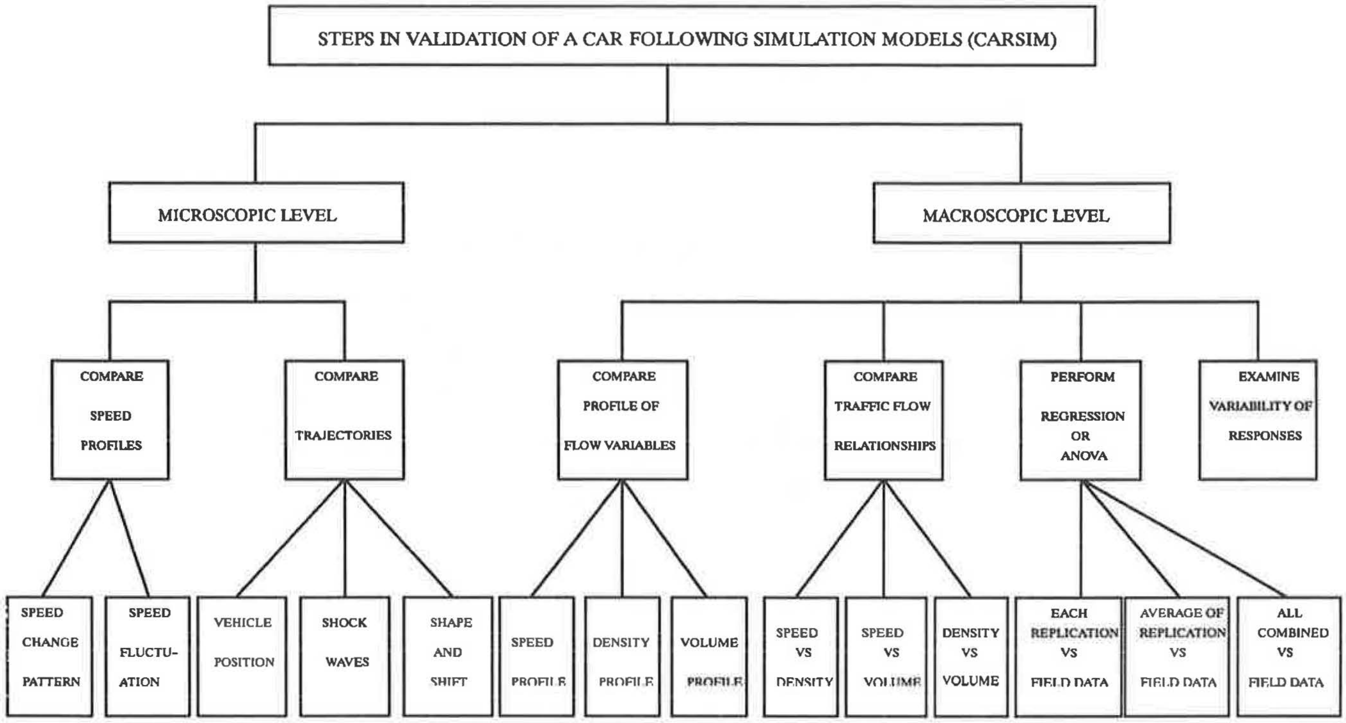

FIGURE 2 Steps in validation of the car-following simulation model (CARSIM).

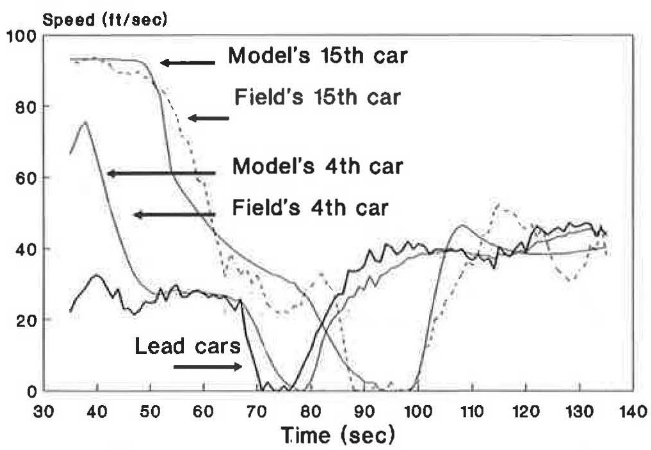

FIGURE 3 Comparison of speed change patterns from the simulation model versus field data for the first, fourth, and last cars in Platoon 123.

Trajectory Comparison

A vehicle trajectory was obtained by plotting the position of the vehicle versus the time in 1-sec time intervals. The average of five numbers obtained from the replications was used as the location of the vehicle at a given time. For Platoon 123, the trajectory plots for every third vehicle including the last vehicle are shown in Figure 4.

When there is a severe disturbance in the traffic flow, it is challenging for a simulation model to generate trajectories that are close enough to the actual trajectories. However, in normal traffic conditions it is not difficult to obtain trajectory plots that are close to the trajectory plots from field data. Thus, models validated only in free flow traffic conditions may not accurately simulate high-density traffic conditions. Also, models validated using data only either from the acceleration or from the deceleration phase of a traffic disturbance may not accurately simulate traffic flow in stop-and-go conditions.

The criteria for evaluating the similarity between the model and field data may include vehicle location, difference in location between the model and real platoon, shape of the plot for individual vehicles, general shift up or down, shift either before or after the disturbance, and location where a vehicle slows down, stops, starts, and recovers. It is important to use short time intervals (a few seconds) in generating trajectory plots, if the model will be used for detailed studies. Otherwise, the model may not accurately show the behavior of traffic within that interval. For instance, Figure 4 shows that the vehicles in the platoon stopped and moved in less than 20 sec. If the data had been collected every 30 sec, the stop-and-go behavior of the platoon would have been missed.

The microscopic validation may be conducted subjectively or objectively, as suggested by Sargent (*4*). The graphical comparison of (subjective validation) the speed change patterns and trajectory plots was found to be sufficient at this level. The objective validation (statistical technique) was not used because of a strong correlation between successive points on the speed profile or the trajectory plots for a vehicle. Once satisfactory results were obtained from the microscopic level validation, the macroscopic validation was started.

Validation at Macroscopic Level

For macroscopic validation, the overall performance of a platoon of vehicles should be evaluated rather than the perfor-

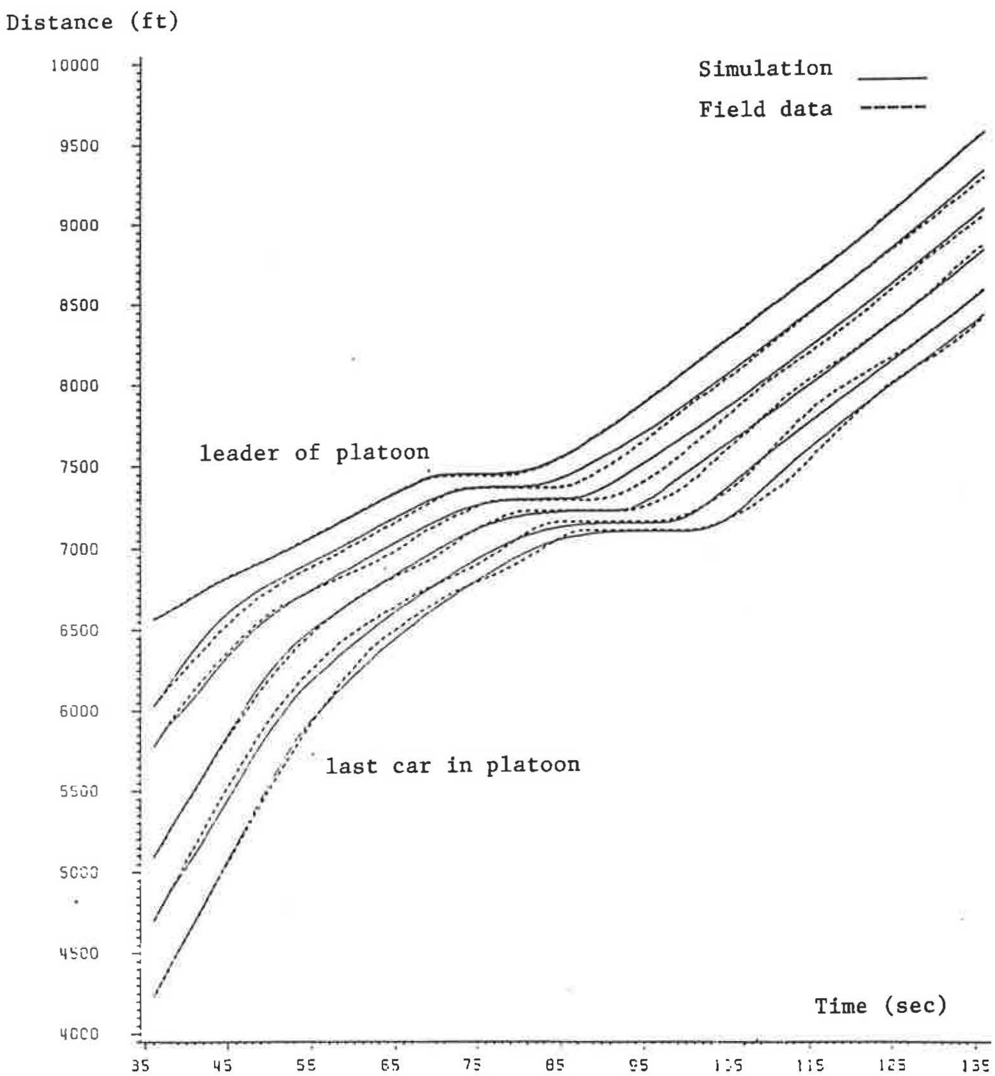

FIGURE 4 Comparison of trajectories of vehicles from the simulation model versus field data for Platoon 123. The trajectories are shown for every third car in a platoon of 15 cars.

mance of an individual vehicle. The macroscopic level validation may not reveal as much detailed information about the model capabilities as the microscopic level, because the variables are the average values for all vehicles in the platoon. For instance, the average speeds for a simulated and an actual platoon might be close, but the speed of individual vehicles may still be different. Likewise, the density of a simulated platoon might be close to that of a field platoon, but spacing between successive vehicles may be considerably different.

For the macroscopic level validation, the following comparisons are suggested:

1. Comparison of profile of traffic flow variables,
2. Comparison of fundamental relations of traffic flow, and
3. Comparison of simulation results versus field data.

In addition to the comparisons, the range of variation of the response variables should also be examined. Application of these concepts in validation of CARSIM is discussed in the following sections.

Comparison of Profile of Traffic Flow Variables

Traffic flow variables used for the comparison were the speed, density, and volume. These variables were computed at 1-sec time intervals and their variations over time (profile) were compared to the field data. The plot of the average speed from the simulation runs versus the speed from the field data for Platoon 123 is shown in Figure 5. The platoon suffered from a severe kinematic disturbance and recovered immediately. The platoon traveling at a speed of more than 80 ft/sec reached a speed of near zero in less than 1 min. The simulation results are close to the actual traffic speeds. The simulation curve exhibits less local fluctuation than the curve for the field data, as expected.

The plots of density versus time for Platoon 123 and the simulation counterparts are shown in Figure 6. The graphs show the same patterns and fluctuations for both simulated and actual platoons. The time a simulated platoon reaches the jam density is close to that of the actual platoon. The density of a platoon is computed from the distance between

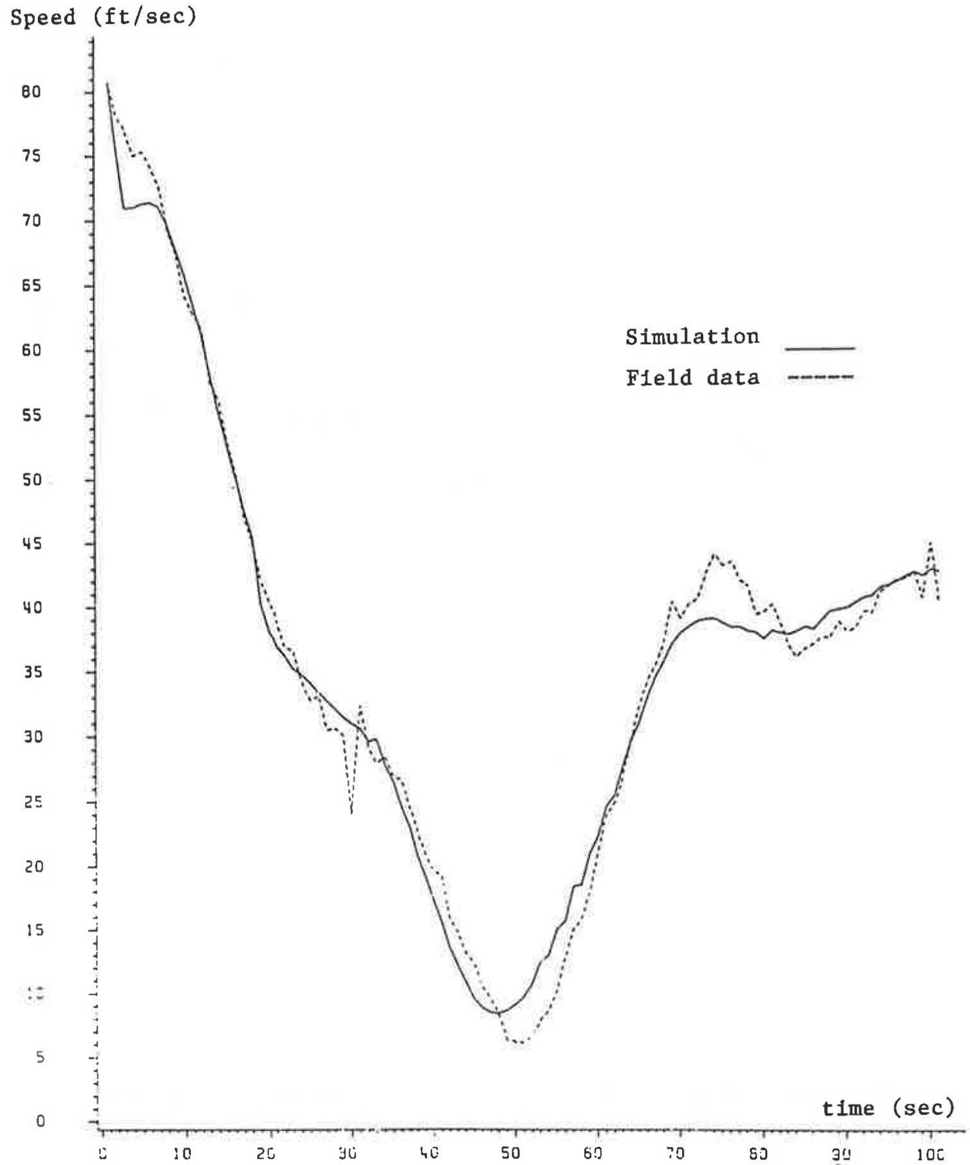

FIGURE 5 Comparison of the average speed from the simulation model versus field data for Platoon 123.

the first and the last car in the platoon. The distance is dependent on the spacing between these two cars. Therefore, one should be careful in using the density of a platoon for comparison of actual and simulated results.

Comparison of volume profile for the simulated and actual platoon is shown in Figure 7. The volume is computed as product of speed and density. The difference between the simulated and actual volume may be large because of the multiplication. Another reason for the large difference might be that the volume may not be equal to the product of speed and density when traffic flow breaks down (critical density). Thus, the volume comparison is not recommended when traffic density reaches its critical range.

Fundamental Relations of Traffic Flow

At this level, the fundamental relationships between traffic flow parameters (speed, density, and volume) obtained from the actual and simulated platoons should be examined. The speed-density, speed-volume, and density-volume relationships from the field data should be compared with those from the simulation model. The speed-density relationships for simulated and actual platoons are shown in Figure 8. The actual data exhibited a nonlinear relationship between speed and density, and a loop representing the hysteresis phenomenon (*19*) when the traffic flow breakdown occurs. The simulation results exhibit a similar relationship and the same phenom-

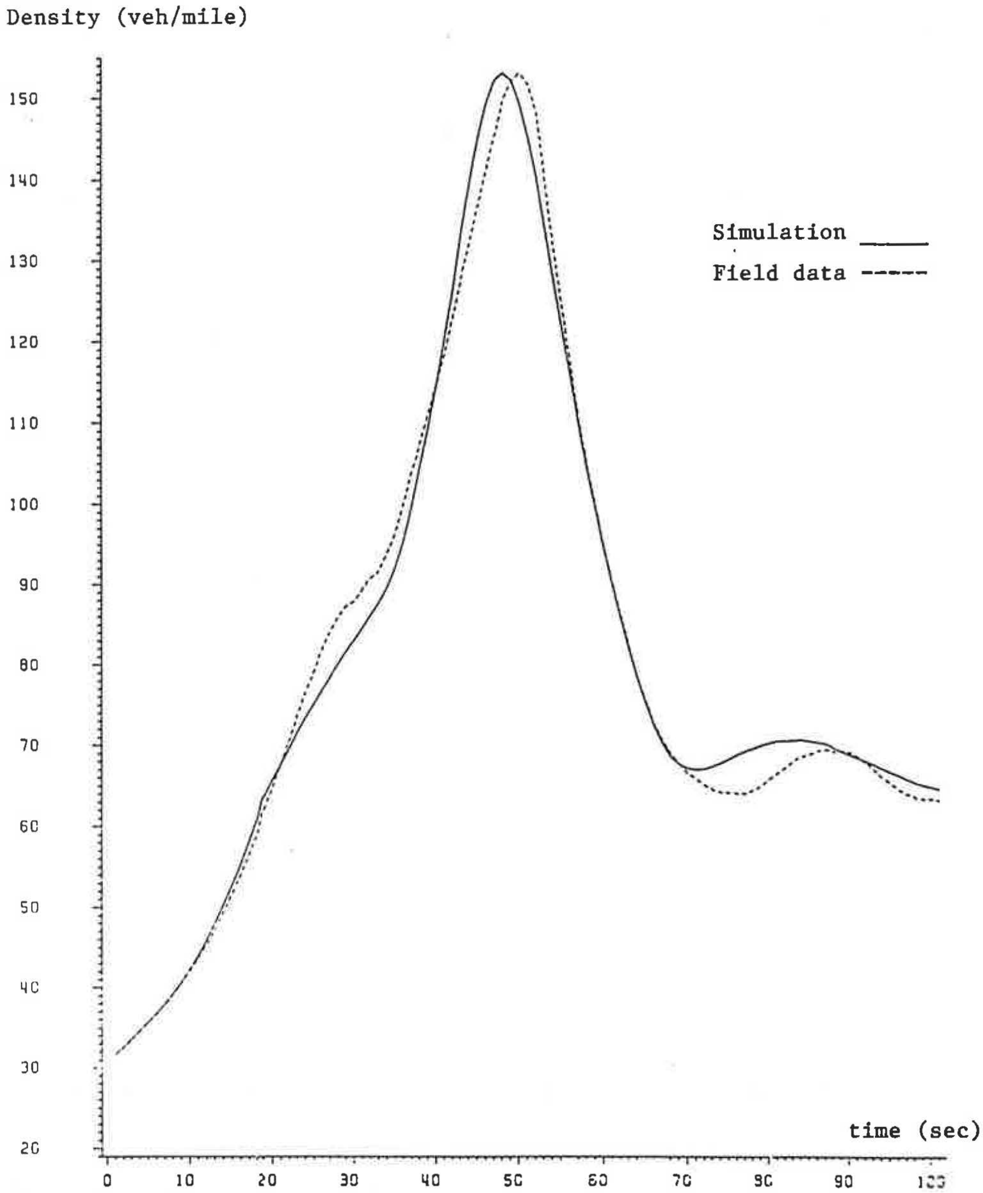

FIGURE 6 Comparison of density from the simulation model versus field data for Platoon 123.

enon. The loop from the simulation model is less distinct than that of the field data because the simulation results are the average of five replications.

Similar comparisons should be made for speed-volume and volume-density relationships. In addition to the graphical presentation of the results, the statistical analysis of the simulation results versus the actual data is also carried out. The statistical analyses will be discussed in the following section.

Comparison of Simulation Results Versus Field Data

Traffic parameters computed from the simulation model should also objectively (e.g., statistical analysis) be compared to the values from field data. Statistical techniques such as regression analysis, analysis of variance, or time series analysis may be used for comparison of the results. The appropriate method should be selected on the basis of factors such as type of data available, relationship with the other parameters, dependency to the other variables, etc. Application of regression analysis and analysis of variance (ANOVA) for model performance evaluation are discussed here. For comparison of the model performance, regression analysis of speed, density, and volume computed from the simulation model versus those from the field data is carried out. For each time interval, the average speed, density, and volume were computed from simulation and field data. The general form of the regression lines is

$$P_{\text{model}} = \text{b0} + \text{b1} * (P_{\text{field}})$$

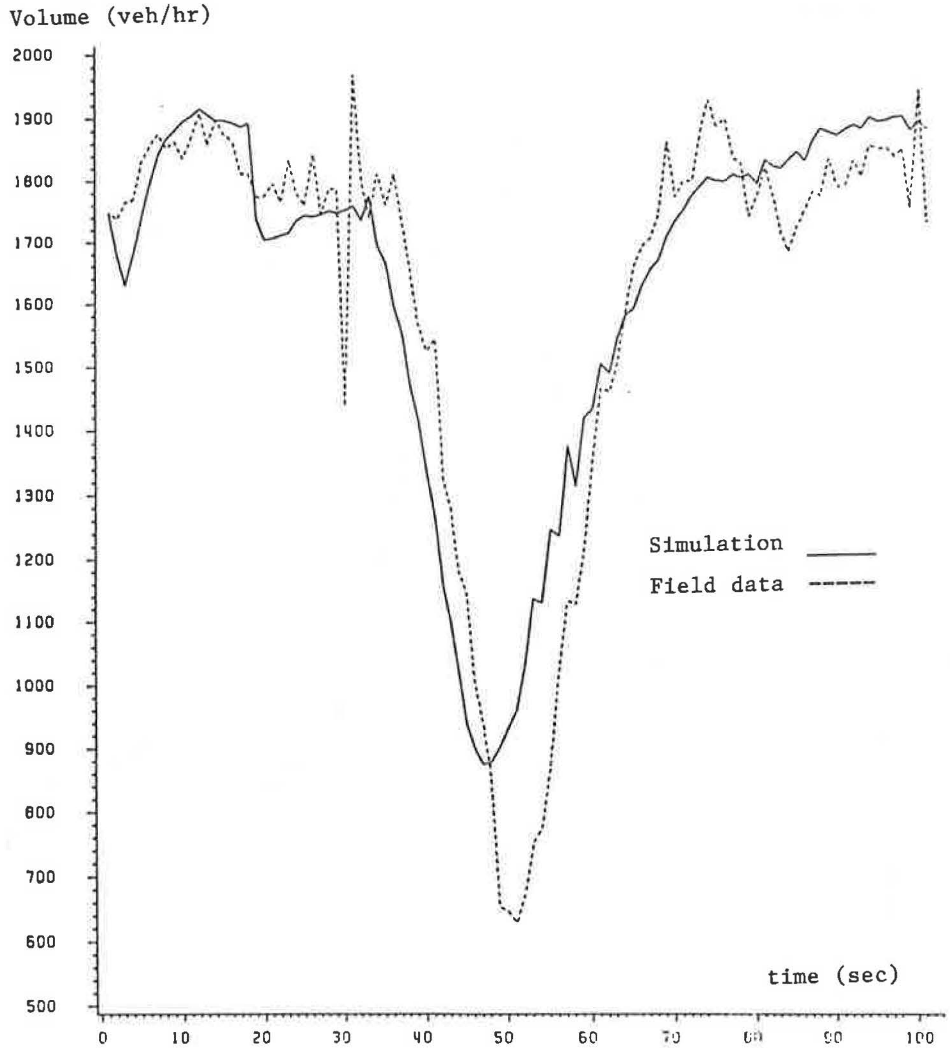

FIGURE 7 Comparison of volume from the simulation model versus field data for Platoon 123.

where

P_{model} = Speed (density or volume) from simulation model,
P_{field} = Speed (density or volume) from field data,
b0 = Y-intercept of the regression line, and
b1 = Slope of the regression line.

Independent replications must be made for any simulation model with stochastic parameter to account for variability of the parameters. However, replication of simulation runs creates more than one set of data for a given condition. One must be careful in interpreting the differences among the replications. The following comparisons may be considered among independent replications:

1. Comparison of an individual replication versus field data,
2. Comparison of average of replications versus field data,
3. Comparison of all replications combined versus field data.

For comparison of individual replication versus field data, speed, density, and volume computed from each simulation run are regressed over the values from field data. The coefficients of the regression lines, variance of the coefficients, and R^2 values are presented in Table 1. Note that, s(b0) and s(b1) are the variances of b0 and b1. The R^2 values and the coefficients of the regression lines indicate that there is a strong agreement between the simulation results and the field data. The slope and y-intercept of the regression lines as well as the R^2 values for a given parameter do not exhibit a large variation among the replications. When the variation is large, more replications should be made.

The advantage of using regression of individual runs over using the average of the five replications is that, one would get additional information about the variation of the response variables among the replications. When the average values are used, this information is no longer available; however, the results are easier to interpret.

For comparing the average of replications versus field data, the speed, density, or volume at a given time is computed as the average values from five replications. Then, the average of five replications is compared to field data. Table 2 presents some of the parameters obtained from the regression analysis. The slopes of the regression lines are close to 1 and the y-

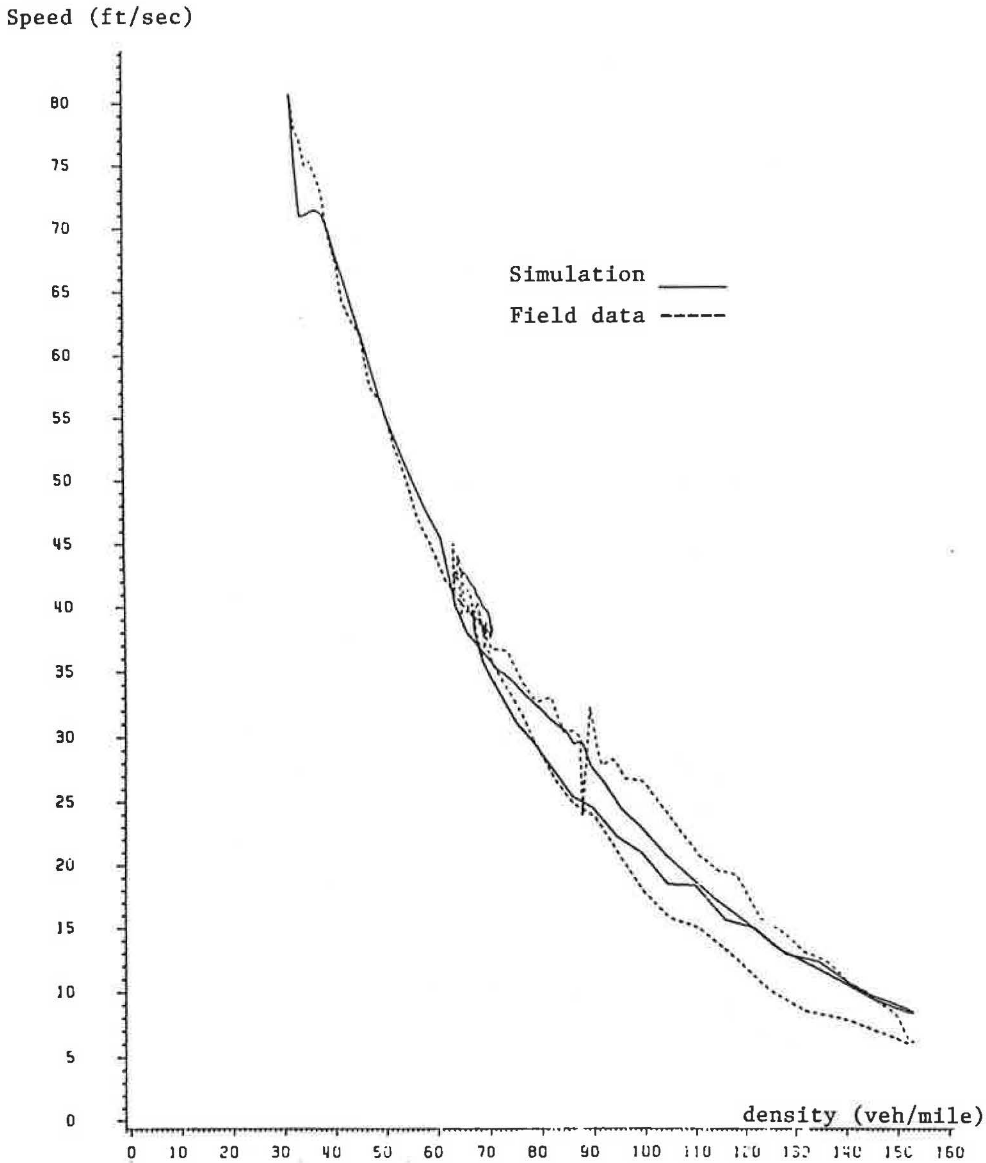

FIGURE 8 Comparison of speed-density relationships from the simulation model versus field data for Platoon 123.

intercepts are close to 0. The R^2 values for speed and density are 0.98 or higher, and for volume it is 0.80 or higher. The results from Table 2 indicate that there is a strong agreement between the speeds or densities computed from the simulation and the field data. As discussed before, the agreement between the traffic volumes is not expected to be as strong as that of speed or density.

When the average values of several replications were compared to field data, the parameters indicated less variation than the same parameter in the individual runs. Although using the average values of the replications makes the comparisons less complicated, finding average of several replications may conceal useful information about the sensitivity of a parameter to an input variable. For instance, the regression lines for speed in Table 1 indicate the range of variation of slope and intercept, but the regression line for speed in Table 2 does not have any variation. Because there are five values from the simulation model for each value from the field data, one might treat them as repeated observations at a given point. The consequence of this assumption is discussed in the following section.

For comparison of all replications combined versus field data, the values obtained from the individual simulation runs are combined to create one data set. Regression analysis of the combined data set versus field data yielded, as it was expected, slopes and y-intercepts equal to that of the average of five replications, see Table 2. However, the variance of slope and y-intercept decreased almost to one-half of that of the average of five replications. There was also a slight decrease on the R^2 values.

The assumption about the repeated data was proven to be incorrect. In regression analysis when there are repeated ob-

TABLE 1 PARAMETERS OF REGRESSION LINES FOR INDIVIDUAL SIMULATION REPLICATIONS VERSUS THE VALUES FROM FIELD DATA FOR PLATOON 123

Repli-cations	b0	b1	R^2	S(b0)	S(b1)
			SPEED		
1	1.09696	0.96812	0.98129	0.54144	0.01344
2	1.96750	0.94091	0.98012	0.54266	0.01348
3	1.91585	0.93815	0.98287	0.50152	0.01245
4	2.44674	0.91579	0.98478	0.46103	0.01144
5	1.10296	0.97226	0.97963	0.56778	0.01409
			DENSITY		
1	1.90606	0.98211	0.98744	0.95538	0.01113
2	2.05298	0.95981	0.98826	0.90241	0.01051
3	1.40735	0.97494	0.99118	0.79343	0.00924
4	2.32060	0.94958	0.99130	0.76726	0.00894
5	-0.49143	1.05514	0.97827	1.35641	0.01580
			VOLUME		
1	353.88146	0.79593	0.81033	64.38405	0.03870
2	277.90502	0.82216	0.82465	63.38825	0.03810
3	415.23342	0.74210	0.82827	56.49940	0.03396
4	383.34560	0.74492	0.88891	44.03049	0.02647
5	316.72237	0.85828	0.79982	71.79334	0.04315

TABLE 2 PARAMETERS OF REGRESSION LINES FOR THE AVERAGE OF FIVE REPLICATIONS AND ALL FIVE REPLICATIONS COMBINED VERSUS THE VALUES FROM FIELD DATA FOR PLATOON 123

Type of Analysis	Variable	b0	b1	R^2	s(b0)	s(b1)
Average of 5 Runs	Speed	1.70554	0.94705	0.98383	0.49177	0.01220
	Density	1.43827	0.98432	0.98922	0.88620	0.01033
	Volume	349.433	0.79265	0.84406	56.9654	0.03424
All 5 Runs Combined	Speed	1.70655	0.94705	0.98100	0.23682	0.00588
	Density	1.44075	0.98432	0.98123	0.52091	0.00607
	Volume	349.472	0.79267	0.80913	28.5566	0.01717

servations at a given point, the lack of fit of the regression line should be examined (*23*). The examination of ANOVA tables falsely indicated the lack of fit for the linear model.

The possibility of fitting a higher-order model was explored, and the residuals were plotted versus the time and predicted values. The plots did not exhibit any definite trend. The residuals were clustered around the line $y = 0$, and approximately made a horizontal band. The lack of the trend, presence of the band, and an R^2 value close to 1 indicated the adequacy of the model. Thus, there was no lack of fit.

The lack-of-fit test is not appropriate for this situation because there is a strong dependency between successive data points. For instance, the density at the current time interval depends on the density at the previous time interval. The false detection of lack of fit was not caused by the inadequacy of the linear regression model, but by using the lack-of-fit test when the data points were strongly correlated.

Variation of the Response Variables

The result obtained from a single simulation run may not be reliable when stochastic variables in the model affect the model's outcome. One method to increase reliability and confidence on the simulation responses is to make several independent replications. Then, the change in the response variable among the replications should be examined. Plots indicating the range of variation of speed, density, and volume among different replications were prepared. The range of variation of speed is shown in Figure 9. Figure 9 indicates that there was small variation from one simulation run to another. When a wider variation is observed, more runs should be made. The number of replications depends on the model and the range of the output parameter.

CONCLUSIONS AND RECOMMENDATIONS

A model verification and validation procedure is suggested and its application to CARSIM is presented. The procedure divides the tasks into microscopic and macroscopic levels. This paper concentrated on the validation part and briefly discussed the verification efforts.

For validation at the microscopic level, the speed change patterns and trajectories of vehicles obtained from the simulation models were compared with those from field data. For validation at the macroscopic level, the average speed, density, and volume computed from the simulation were compared to the field data. The variation of these parameters over time and the relationships between these parameters were examined. Furthermore, the regression of the simulation results versus the field data was discussed. Comparisons of the results indicated that using this procedure enabled CARSIM to provide results close to those from field data.

Some of the steps suggested here have been used in validation of other traffic simulation models. The steps suggested here should be considered as the starting point for the efforts needed for verification and validation of a microscopic traffic simulation model. For a particular simulation model, the developer should decide on appropriateness of the suggested approach and take additional steps needed to obtain reliable

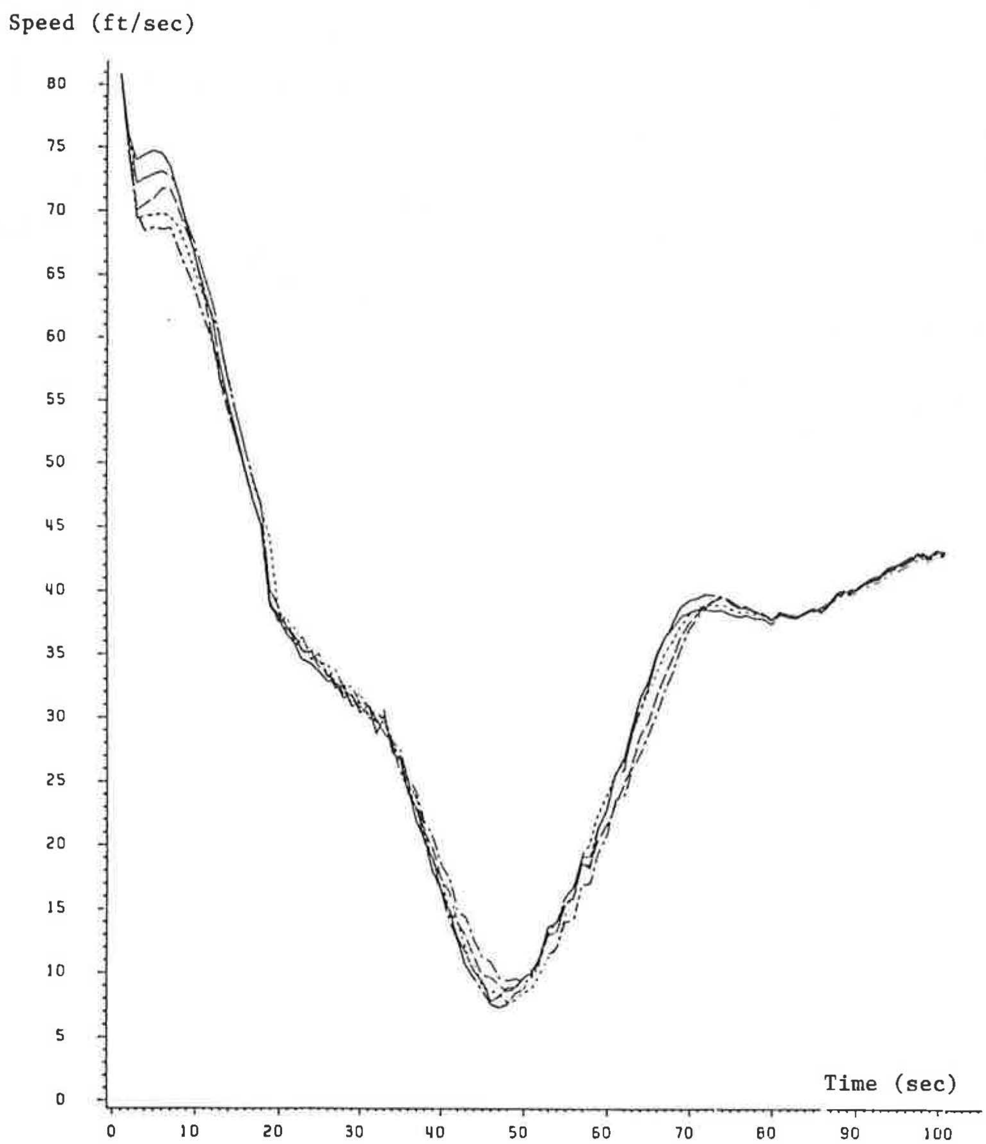

FIGURE 9 Speed variation among five independent replications of a simulation model representing Platoon 123.

results. Questions about number of replications, truncation policy, and traffic conditions to be covered were not discussed. The procedure is intended to be used for validation of microscopic models, although some parts of it may be used for validation of macroscopic models, as well. The approach should be considered as a step toward the development of a comprehensive guideline for validation of traffic simulation models.

REFERENCES

1. G. S. Fishman and P. J. Kiviat. *The Statistics of Discretion-Even Simulation*. SCI Simulation 10, 1968.
2. R. F. Benekohal and J. Treiterer. *A Car Following Model For Simulation of Traffic Flow in Normal and Stop-And-Go Conditions*. In *Transportation Research Record 1194*, TRB, National Research Council, Washington, D.C., 1989.
3. J. S. Annino and E. C. Russell. Ten Most Frequent Causes of Simulation Analysis Failure-and How to Avoid Them. *Simulation*, Vol. 32, No. 6, 1979.
4. R. S. Sargent. Verification and Validation of Simulation Models. *Progress in Modeling and Simulation*, F. E. Cellier (ed.), Academic Press, New York, 1982.
5. J. F. Torres, A. Halati, and A. Gafarian. *Statistical Guidelines for Simulation Experiments*. Executive Summary, JFT Associates, Culver City, 1983.
6. A. V. Gafarian and J. D. Walsh. Methods for Statistical Validation of a Simulation Model for Freeway Traffic Near an On-Ramp. *Transportation Research*, Vol. 4, 1970, pp. 379–324.
7. G. A. Mihram. *Simulation Statistical Foundations and Methodology*. Academic Press, New York, 1972.
8. J. P. C. Kleijnen. Statistical Design and Analysis of Simulation Experiments. *Informatie*, Vol. 17, No. 10, Netherlands, Oct. 1975, pp. 531–535.
9. T. H. Naylor, K. Wertz, and T. H. Wonnacott. Methods for Analyzing Data From Computer Simulation Experiments. *Communications of the ACM*, Vol. 10, No. 11, 1967.
10. J. P. C. Kleijnen. Experimentation with Models: Statistical Design and Analysis Techniques. In *Progress in Modeling and Simulation*, F. E. Cellier (ed.), Academic Press, New York, 1982.
11. J. R. Wilson and A. B. Pritsker. *A Survey of Research on the Simulation Start Up Problem. Simulation*, Vol. 31, No. 2, Aug. 1978.
12. J. R. Wilson and A. B. Pritsker. *Evaluation of Start Up Policies in Simulation Experiments. Simulation*, Vol. 31, No. 3, Sept. 1978.
13. A. A. B. Pritsker and C. D. Pegden. Introduction to Simulation and SLAM. John Wiley, New York, 1979.
14. J. P. C. Kleijnen. *Design and Analysis of Simulations: Practical Statistical Techniques. SCI Simulation*, Vol. 28, No. 3, March 1977.
15. J. P. C. Kleijnen. *Comparing Means and Variances of Two Simulations. SCI Simulation*, Vol. 26, No. 3, 1976.
16. G. S. Fishman. *Problems in the Statistical Analysis of Simulation Experiments: The Comparison of Means and the Length of Sample Records*. Communications of the ACM. Vol. 10, No. 2, 1967, pp. 94–99.
17. J. R. Emeshoff and R. L. Sisson. *Design and Use of Computer Simulation Models*. McMillan Co., New York, 1970.
18. J. P. C. Kleijnen. *Antithetic Variates, Common Random Numbers and Optimal Computer Time Allocation in Simulation. Management Science*, Vol. 21, No. 10, 1975.
19. D. Wicks et al. *Development and Testing of INTRAS, a Microscopic Freeway Simulation Model, Volume 1 and 2, Final Report*. FHWA/RD-80/106-107. FHWA, U.S. Department of Transportation, 1980.
20. R. B. Goldblatt. *Development and Testing of INTRAS, a Microscopic Freeway Simulation Model, Volume 3, Validation and Application, Final Report*. FHWA/RD-80/108. FHWA, U.S. Department of Transportation, 1980.
21. R. F. Benekohal. *Development and Validation of a Car Following Model for Simulation of Traffic Flow and Traffic Waves Studies*. Ph.D. dissertation, Ohio State University, Columbus, 1986.
22. J. Treiterer. *Investigation of Traffic Dynamics by Aerial Photogrammetry Techniques*. Final Report EES 278. Transportation Research Center Department of Civil Engineering, The Ohio State University, Feb. 1975.
23. N. Draper and H. Smith. *Applied Regression Analysis, 2nd ed.* John Wiley, New York, 1981.

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