Proceedings of the 1998 Winter Simulation Conference
D.J. Medeiros, E.F. Watson, J.S. Carson and M.S. Manivannan, eds.

DEVELOPMENT AND APPLICATION OF A VALIDATION FRAMEWORK FOR TRAFFIC SIMULATION MODELS

Lei Rao
Larry Owen

ITT Systems Corp.
P.O.Box 15012
Colorado Springs, CO 80935-5012, U.S.A.

David Goldsman

School of Industrial and Systems Engineering
Georgia Institute of Technology
Atlanta, Georgia 30332-0205, U.S.A.

ABSTRACT

This paper discusses the concept of validation and proposes a multistage validation framework for traffic simulation models. The framework consists of conceptual validation and operational validation. The operational validation involves two levels of statistical tests: a two-sample *t* test and a two-dimensional two-sample Kolmogorov-Smirnov test. The validation experience employing the proposed framework demonstrates the fact that while a model can be valid for one level of detail, it can be invalid for another. The validation results also illustrate that the proposed multistage validation procedure can account for the complexity of the validation task and its conclusions.

1 INTRODUCTION

Traffic system operation is characterized by the flow of mobile elements (users and vehicles) through facilities (roadways and control devices). The flow of the mobile elements is a complex interactive process that is a function of facility design, user objectives, perceptions and reactions of drivers, and vehicle dynamics. A traffic system simulation is a symbolic software model for conducting experiments on a traffic system. The purpose of the experiments is to design and modify the facilities to optimize safety and efficiency of traffic flow.

Since the emergence of Intelligent Transportation Systems (ITS) in the early 1990s, simulation has become an invaluable tool for evaluating ITS strategies. While considerable research efforts have been devoted to the development of traffic simulation models, validation, which is an integral part of the "model development life cycle," has not received enough attention. In fact, most of the simulation models developed in the traffic engineering community do not have guidelines for validation. For instance, it is usually the users' responsibility to choose a number of parameters to vary in order to study the way the simulation behaves, and to understand the significance of any differences between observed measures from the real world and simulated measures of effectiveness (MOE's). However, no specifications are provided for making those parameter adjustments and interpreting those differences.

In this paper, we review validation methods in the literature and propose a multistage validation framework for traffic simulation models. In the next section, the concept of validation is discussed. Section 3 proposes a validation framework. In Section 4, we present a validation experience, while Section 5 summarizes the paper.

2 VALIDATION

Validation is generally defined as the act of determining whether a simulation model reasonably represents or approximates the real system for its intended use (Fishman and Kiviat 1968, Sargent 1982, Law and Kelton 1991). Validation is a purpose-specific task. Balci and Sargent (1981) argue that a simulation model should be developed for a specific purpose or application and its adequacy or validity should be evaluated only in terms of that purpose with regard to the relevant experimental frame(s). Moreover, since increasing the validity of a model beyond a certain level may be quite expensive (e.g., more data collection may be required), it is more cost-effective for a simulation model to be validated relative to those MOE's that will actually be used for decision making (Law and Kelton 1991). Thus, the purpose of the model determines what aspects of the model to validate and their levels of detail.

Determining the validity of a simulation model is not a binary decision in which the model is simply deemed valid or invalid; rather, validity should be considered one of degree depending on the model's purpose. Shannon (1975) suggests that since no model is absolutely correct in the sense of a one-to-one correspondence between itself and real life, simulation modeling is probably not a search for absolute truth or correctness but rather a succession of theories that will progressively approach the truth. In fact, Schlesinger *et al.* (1979) go a step further to define validation as a substantiation that a computerized model within its domain of applicability possesses a satisfactory range of accuracy consistent with the intended application of the model.

3 MULTISTAGE VALIDATION FRAMEWORK

Since no single test can ever demonstrate the sufficiency of a simulation to reflect real-world behavior, the approach to transferring model confidence involves many layers and tiers with several tests being performed at different stages and levels (Figure 1).

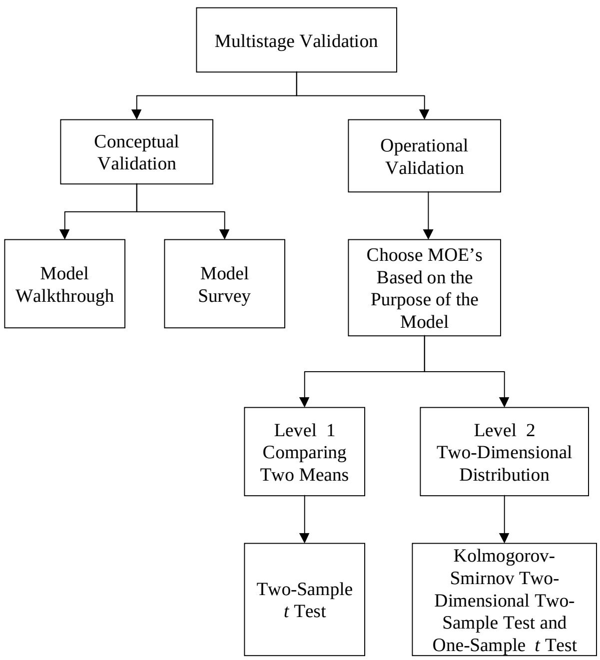

Figure 1: Multistage Validation Framework

In fact, the validation process is often divided into two stages: conceptual and operational. Conceptual validation assesses simulation models against sound and accepted theoretical foundations. The operational validation process involves comparisons between model predictions and measured real-world system behavior. While they are two distinct processes, conceptual validation is not necessarily a precursor to operational validation. Rather, conceptual validation is a concurrent and reoccurring process that takes place in conjunction with operational validation. Conceptual validation may be reexamined to explain anomalous or inconsistent behavior detected during operational validation.

3.1 Conceptual Validation

Conceptual validation consists of identification and evaluation of the model's underlying theory (usually described in the model's documentation and supporting academic literature) and comparisons of the methodology of the model with that of alternative approaches. It results in a qualitative assessment of a model's theoretical underpinnings, as well as its implementation, evaluated in the light of sound and accepted theoretical methods. For traffic simulation models, the focus of conceptual validation is on the underlying traffic flow theory. The primary methods employed during conceptual validation are model survey and model walkthrough.

3.1.1 Model Survey

The purpose of validation in this paper is to transfer confidence in model performance to the traffic engineering community. Thus, an important step is to engage this community, which is the end user of the traffic simulation models, in a continuing dialog regarding the methodology, data requirements and results of traffic model validation. The community brings three important perspectives to traffic model validation:

- Researchers supply the information regarding the development of new and improved theoretical models.
- Developers provide the implementation of the concepts into useable software and continue to maintain and modify the traffic models.
- Practitioners apply the models to real-world problems and provide feedback regarding their modeling experience.

Without the experience and opinions of the members of this community, the validation effort will not be able to address the concerns of the end users of the models. In practice, a questionnaire may be sent to these members to conduct this survey.

3.1.2 Model Walkthrough

A model walkthrough involves a small group of qualified individuals or "experts," who carefully review and revisit the model's logic and documentation. This group may also contrast existing logic with alternative methods as well as review the basic structure of the model.

3.2 Operational Validation

The procedures associated with operational validation are designed to present quantitative measurements of the consistency between model prediction and operational measurements from real-world systems. The most definitive test of a simulation model's operational validity is establishing that its output data closely resemble the output data that would be expected from the actual (or proposed) system using identical inputs (Law and Kelton 1991, Carson 1986). In particular, Law and Kelton (1991) point out that if the two sets of output data compare "favorably," then the model of the existing system is considered "valid." (The accuracy required from the model will depend on its intended use and the utility function of the manager.)

Validating a simulation model is essentially validating a set of MOE's, which are chosen by users to reflect model performance based on the intended application of the model. Users believe that if these MOE's are valid, then the model tends to be valid for its purpose. In fact, operational validation comprises two phases. First, users identify the MOE's. This can be accomplished by discussion with "system experts." Second, operational tests for the chosen MOE's are conducted. While it is often ignored in the validation discussion, the first phase is very important, because it establishes a set of reference points for the second phase. If the reference points are not chosen correctly, no matter what operational tests one conducts in the second phase, conclusions regarding the validity of the model will be flawed.

There is a wide array of methods available for the second phase of operational validation. These methods include sensitivity analysis (Law and Kelton 1991), analysis of variance (Garratt 1974, Van Horn 1971), chi-square tests (Gerlough and Huber 1975), regression analysis (Van Horn 1971, Taylor 1979), Wilcoxon signed-rank test (Emshoff and Sisson 1970), Theil's inequality coefficient (Theil 1961), spectral analysis (Fishman and Kiviat 1968, Emshoff and Sisson 1970), and the standardized time series technique (Chen and Sargent 1987, Schruben 1983). However, there is no completely definitive approach. The decision of choosing a specific approach is often based on the characteristics of the model (or the system) and its intended use.

In this section, we propose a two-level statistical procedure for traffic simulation models. The levels identify the progression of the validation process and different applicabilities of the model, and impose different requirements on the extent of the data collection.

3.2.1 First Level: Comparing Two Means

The first validation level involves comparison of averages of standard traffic flow characteristics (e.g., throughput, speed, density, and volume) between the simulation and the real world. The procedure is illustrated as follows:

Step 1: Users choose validation MOE's (e.g., speed and headway), which must be representative of the model performance for the intended application of the model.

Step 2: For each MOE defined in *Step 1* and each real-world data set, repeat *Steps 2.1-2.4*:

Step 2.1: Users choose the level of significance (LOS) of the test based on the purpose of the model for the specific MOE.

Step 2.2: To account for the variability of simulation results, multiple (independent) simulation runs are conducted based on different random number seeds. Then the average measures of these simulation results are computed.

Step 2.3: A two-sample t test is conducted. The null hypothesis in this test is that the MOE measured from the system and the MOE computed from the simulation model are the same, and the alternative hypothesis is that they are different. The p-value is then computed.

Step 2.4: If the p-value is less than the predefined LOS, then the null hypothesis is rejected and the alternative hypothesis is accepted, indicating that the model is invalid; otherwise, the model will be regarded as valid for the chosen MOE at the LOS.

Step 3: If the model is valid for all the chosen MOE's for all the real-world data sets, the model is considered valid for its intended purpose. Otherwise, the model is invalid for that purpose.

3.2.2 Second Level: Two-Dimensional Distribution

In the first level, comparison is conducted with respect to a single MOE. However, the chosen MOE's are sometimes correlated. For instance, when other conditions are the same, higher speed tends to be correlated with lower headway. Moreover, while it may be sufficient for macroscopic traffic models, the first-level test may be insufficient for microscopic traffic models, which model individual vehicles at a higher level of detail, i.e., in more detail. Finally, the assumption that the data are distributed normally may not be applicable for some data sets.

To overcome the above difficulties, we introduce the two-dimensional (e.g., speed vs. headway) two-sample Kolmogorov-Smirnov (K-S) test as our second-level test. The K-S test is a nonparametric test, which does not require explicit distributional assumptions about the underlying processes. It can be employed to test whether the values from the real world and those from the simulation are from the same distribution.

The K-S test is not well-defined in more than one dimension. The version of the K-S test for comparison between two two-dimensional distributions is due to Fasano and Franceschini (1987). The significance level of an observed value d of the test statistic D (as a disproof of the null hypothesis that the distributions are the same) is given approximately by the formula (Press *et al*. 1992):

$$\text{Probability}(D > d) = Q_{KS}\left(\frac{\sqrt{N}d}{1+\sqrt{1-r^2}(0.25-0.75/\sqrt{N})}\right)$$

where

Q_{KS} = a complex monotonic function (see Press *et al*. 1992 for details);

D = K-S statistic, defined as the maximum value of the absolute difference between the two two-dimensional cumulative distribution functions;

$N = \frac{N_1 N_2}{N_1 + N_2}$, where N_1 and N_2 are sample sizes;

$r^2 = \frac{r_1^2 + r_2^2}{2}$, where r_1 is the sample coefficient of correlation between the two variables for the first distribution, and r_2 is the sample coefficient of correlation between the two variables for the second distribution.

The two-dimensional K-S statistic provides an indication of the consistency between the real world and the simulation. The $\text{Probability}(D > d)$ has a value between zero and one, with zero indicating different distributions and unity indicating perfect correlation. Press *et al.* (1992) point out that “when the indicated probability is greater than 0.20, …the implication that the data and model (or two data sets) are not significantly different is certainly correct.” Thus, this probability can be used to test the agreement between the real-world and the simulation data. The procedure in this level of validation is summarized as follows:

Step 1: Users choose two-dimensional MOE pairs (e.g., headway vs. speed), which must be representative of the model performance for the intended application of the model.

Step 2: For each MOE pair, repeat *Steps 2.1-2.4*:

Step 2.1: Users choose the LOS of the test based on the purpose of the model for the specified MOE pair.

Step 2.2: For all the real-world data sets and all the different simulation runs based on each real-world data set, conduct the two-dimensional two-sample K-S test between the real world and the simulation, and generate a K-S statistic matrix.

Step 2.3: Conduct a one-sample *t* test for the mean of the K-S statistic matrix (values in the matrix are assumed to be independent and identically distributed). Informally, the hypothesis to be tested in this paper is as follows:

H_0: $\text{Probability}(D > d)$ = 0.2

H_1: $\text{Probability}(D > d)$ < 0.2 (i.e., the model is invalid)

Then compute the one-sample *t* statistic and the associated p-value.

Step 2.4: Compare the p-value with the predefined LOS. If the p-value is less than the LOS, we reject the null hypothesis, indicating that the model is statistically significant from the real world; otherwise, the model tends to be valid at the LOS for the specified MOE pair.

Step 3: If the model is valid for all the chosen MOE pairs, the model is considered valid for its intended purpose. Otherwise, the model is invalid.

3.3 Data Collection

Collecting valid data is essential for the success of operational validation. The approach for traffic data collection is to take detailed measurements of traffic flow during a given time period. Data collection plans are prepared based on the desired level of detail and the condition that the data collection is not intrusive to normal traffic operations. Subject to resource limitations, day-to-day measurements, with approximately the same traffic demand volume and conditions, are required to incorporate the variability inherent in real-world behavior. For instance, to study morning rush-hour behavior, one can take measurements from Tuesday to Thursday during the same time period (e.g., 7:00 am - 8:00 am) for two weeks.

4 VALIDATION EXPERIENCE

In this section, we present an effort to validate CORSIM (CORridor SIMulation, version 1.02 beta) employing the proposed framework. CORSIM is currently the most extensive and widely used microscopic traffic simulation model. In urban street networks, one of the main objectives of traffic study is to synchronize traffic signals so that a platoon of cars being released from one signal arrives at the next one without interruption. To achieve the above objective and account for traffic instability, a traffic simulation model must mimic the behavior of platoon dispersion accurately. In this paper, we attempt to validate this behavior in a real-world traffic setting. This is achieved by observing the progression of platoon dispersion from an upstream node (or intersection) to a downstream node, and comparing the observation to that from CORSIM.

4.1 Conceptual Validation

4.1.1 Model Survey

The model survey was sent to various members of the traffic engineering community. Some of the key comments are summarized in Table 1. These comments helped us understand and address the salient issues in our validation effort.

Table 1: Key Comments from the Model Survey

It should be demonstrated that various sub-systems (e.g., car-following and lane-changing logic) of the model work reasonably well. An understanding of the difference between actual and predicted MOE's is necessary. The inputs having the greatest influence on output and their uncertainty should be identified.
In the validation process, all parameters and their ranges for various scenarios should be considered. A hierarchy of the critical parameters may be useful.
In data collection, the level of detail and the variables to be collected are very important. Increased use of probe cars, video technology, differential GPS will have an impact on the data collection.

4.1.2 Model Walkthrough

We conducted an investigation of the theory and assumptions underlying the CORSIM model and its submodels. In particular, the mathematical, logical and causal relationships used in the model were examined. Moreover, a literature review of recent traffic modeling theories was performed.

Because the validity of the model is dependent on its specific uses, our investigation found that CORSIM is conceptually "valid" for most practical purposes. However, the model still needs enhancement. For example, we discovered that the GM "car-following" logic is more robust than the Pitt car-following logic, which is currently used in the model, and therefore recommended employing the GM logic in future modifications of CORSIM.

4.2 Operational Validation

4.2.1 Test Scenario and Data Collection

The test site was selected at a link on the Union Boulevard between Austin Bluff and Academy Boulevard, Colorado Springs, Colorado. Union Boulevard is an arterial that has three lanes, and the specified link has a length of about 1.3 miles. The test scenario and the data collection points are illustrated in Figure 2. Three video cameras were set up to observe and record the progression of platoons. Microscopic data (e.g., speed and headway) associated with individual vehicles can be collected at several points within the section of interest.

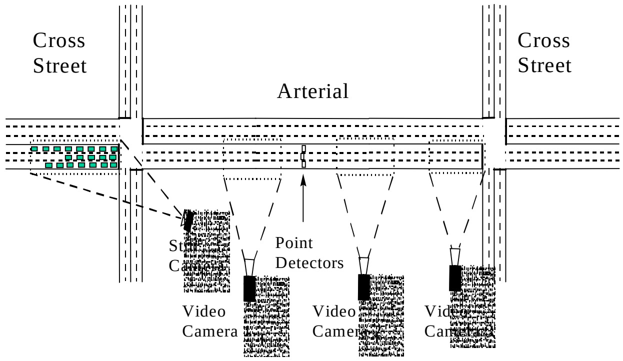

Figure 2: Test Scenario and Data Collection

Five platoons were identified for comparison after data collection and reduction. In CORSIM, each platoon composition and dispersion scenario is coded based on real-world information. For each platoon, ten independent simulation runs were conducted.

4.2.2 First Level: Comparing Two Means

We chose speed and headway as validation MOE's, and LOS at 0.1 for the purpose of model application. The average MOE's of the ten simulation runs for each individual vehicle were computed. We repeated the above process for all the platoons and reported the results along with the corresponding real-world data. This generates Table 2 (speed) and Table 3 (headway). In these two tables, we performed two-sample *t* tests between the simulation and the real-world data for each platoon, and computed the corresponding p-values.

Table 2: Comparison of Mean Speed (feet/second) for Different Platoons

	Platoon 1		Platoon 2		Platoon 3		Platoon 4		Platoon 5	
Vehicle ID	Field	Model	Field	Model	Field	Model	Field	Model	Field	Model
1	53.33	47.40	59.26	51.70	59.26	48.40	64.00	53.40	61.54	51.80
2	53.33	49.80	61.54	54.10	66.67	65.00	59.26	53.10	59.26	59.50
3	53.33	51.90	80.00	58.20	64.00	68.30	66.67	55.90	57.14	58.60
4	55.17	52.20	55.17	55.40	55.17	62.40	48.48	60.80	53.33	59.80
5	55.17	54.50	55.17	51.10	55.17	58.30	50.00	57.70	55.17	58.50
6	55.17	55.70	53.33	52.90	57.14	64.90	50.00	66.10	61.54	63.50
7	57.14	56.70	53.33	56.10	53.33	63.00	72.73	55.30	66.67	60.30
8	57.14	58.20	53.33	64.10	53.33	58.80	66.67	54.30	66.67	63.80
9	57.14	59.00	55.17	62.20	64.00	63.10	59.26	56.30	66.67	64.80
10	61.54	60.10	61.54	60.30	59.26	62.80	64.00	61.00	66.67	61.20
11	61.54	63.40	61.54	58.20	64.00	59.30	64.00	64.50	66.67	64.30
12	61.54	65.80	64.00	54.80	61.54	57.90	76.19	60.00	61.54	58.50
13	69.57	71.30	55.17	54.40	57.14	62.30	59.26	60.70	61.54	59.20
14	69.57	75.40	53.33	55.30	61.54	58.20	59.26	56.20	59.26	55.30
15			53.33	60.30	66.67	64.20	64.00	58.30		
16			51.61	55.40	57.14	56.40				
17			66.67	59.30	53.33	57.30				
18			64.00	57.50						
19			61.54	54.80						
20			64.00	55.30						
21			57.14	54.30						
22			57.14	55.90						
23			55.17	54.20						
p-Value	0.98		0.11		0.42		0.16		0.27	

Table 3: Comparison of Mean Headway (seconds) for Different Platoons

	Platoon 1		Platoon 2		Platoon 3		Platoon 4		Platoon 5	
Vehicle ID	Field	Model	Field	Model	Field	Model	Field	Model	Field	Model
1	0.80	0.68	1.10	3.83	1.60	7.38	5.00	3.98	4.00	4.65
2	0.80	0.94	1.20	1.33	1.50	3.02	1.30	1.08	2.60	1.64
3	0.90	1.04	2.10	1.71	0.80	2.21	3.40	1.42	3.30	1.55
4	0.90	1.17	3.70	2.24	3.20	3.64	2.20	2.75	2.50	2.80
5	1.00	1.24	0.80	2.10	3.00	4.80	1.60	3.88	1.50	4.22
6	1.40	1.40	1.00	1.89	2.00	2.52	1.50	1.71	2.90	2.06
7	1.40	1.54	1.10	2.78	1.50	2.95	1.80	2.48	3.40	1.75
8	1.60	1.78	0.60	2.22	0.90	3.11	1.00	1.85	2.10	1.18
9	1.70	2.09	2.40	1.58	4.10	1.40	1.40	2.45	2.30	2.06
10	1.80	2.24	4.90	1.63	3.90	2.77	1.80	2.83	1.50	2.25
11	2.00	2.80	2.60	2.90	0.80	2.09	2.00	1.82	2.30	2.34
12	2.10	3.59	0.70	1.17	0.70	2.27	2.20	2.59	1.80	2.57
13	2.90	5.22	2.90	1.77	1.60	1.61	2.00	1.58	1.40	1.56
14	5.80	6.62	1.20	1.16	2.00	2.68	0.80	2.53	2.20	2.50
15			1.10	1.74	4.30	1.60	4.70	1.54		
16			1.70	3.04	3.50	2.81				
17			2.10	1.53	2.30	1.73				
18			2.60	1.53						
19			1.00	1.65						
20			2.20	1.45						
21			0.70	2.23						
22			3.30	1.87						
23			1.90	2.48						
p-Value	0.38		0.64		0.17		0.76		0.89	

For headway comparisons, the p-values range from 0.17 to 0.89 for five platoons; thus, the null hypothesis can not be rejected, so the simulation and the real world are not significantly different at level 0.1 with respect to headway. The same conclusion can be reached for speeds (the p-values range from 0.11 to 0.98). Since CORSIM is valid at LOS = 0.1 for the chosen MOE's (headway and speed) for all the real-world data sets, we believe that the model is valid for its intended purpose.

4.2.3 Second Level: Two-Dimensional Distribution

Here we chose speed and headway as a validation MOE pair, and LOS at 0.1 for the purpose of model application. Further, we conducted the K-S two-dimensional two-sample test in a C++ environment to post-process the data from both the real world and CORSIM. For each of the five platoons and its corresponding simulation runs, a K-S statistic for headway-speed distribution was computed, and a K-S statistic matrix was generated (Table 4 and Figure 3).

We then conducted a one-sample *t* test for the mean of the K-S statistic matrix: average = 0.133, stdev = 0.165, degrees of freedom = 49, t = -2.87, p-value = 0.003. Since the p-value (0.003) is less than the LOS (0.1), we reject the null hypothesis that the mean of the K-S matrix is at least 0.2, and unfortunately conclude that the model is invalid at the specified LOS for the specified MOE pair. Thus, the model tends to be invalid for the purpose specified in this test. It should be noted, however, that the CORSIM model used in this test is version 1.02 beta, and a more-recent enhanced version may perform better than the one used in this paper.

The computational results for this two-level procedure illustrate the fact that while a model can be valid for one level of detail, it can be invalid for another. In this test case, CORSIM (1.02 beta) is valid for a lower level of detail test (one-dimensional comparison), but invalid for a higher level of detail test (two-dimensional distribution). Since the requirements of the level of detail are determined by the purpose of the model, we conclude that the model is valid for some purposes (e.g., simulate macroscopic traffic behavior), while invalid for others (e.g., simulate detailed microscopic vehicle behavior). Thus, the results demonstrate that the proposed multistage validation procedure can account for the complexity of validation task and its conclusions.

5 SUMMARY

In this paper, we reviewed the concept of validation for simulation models, and proposed a validation framework for traffic simulation models. The framework consists of conceptual and operational validation processes. The operational validation involves two levels where statistical tests are performed in a systematic manner. In particular, a *t* test and a K-S test are conducted at different levels of detail. These levels correspond to the desirable degree of accuracy of the model and its applicability. A test case was selected to validate CORSIM's prediction of platoon dispersion. The results demonstrate that validation is not a binary decision, but rather a decision based on the model's intended use. The validation experience further illustrates the necessity and advantages of the proposed multistage validation procedure.

Table 4: p-values for K-S Two-Dimensional Two-Sample Tests

	Platoon 1	Platoon 2	Platoon 3	Platoon 4	Platoon 5
RUN1	2.00E-01	4.85E-02	6.98E-02	3.48E-01	2.80E-02
RUN2	2.98E-01	2.70E-05	1.08E-01	2.41E-01	1.30E-01
RUN3	2.01E-01	2.12E-02	1.61E-01	9.12E-01	1.30E-01
RUN4	2.00E-01	3.25E-02	6.98E-02	3.47E-01	8.10E-02
RUN5	2.00E-01	3.26E-02	4.40E-02	2.41E-01	8.08E-02
RUN6	2.08E-03	1.02E-01	4.40E-02	1.62E-01	2.79E-02
RUN7	4.56E-04	1.10E-03	1.47E-03	1.29E-02	1.02E-03
RUN8	4.20E-03	4.86E-02	6.99E-02	1.62E-01	4.85E-02
RUN9	4.55E-04	3.26E-02	1.08E-01	4.79E-01	2.07E-04
RUN10	1.30E-01	3.50E-01	4.45E-01	1.62E-01	2.79E-02

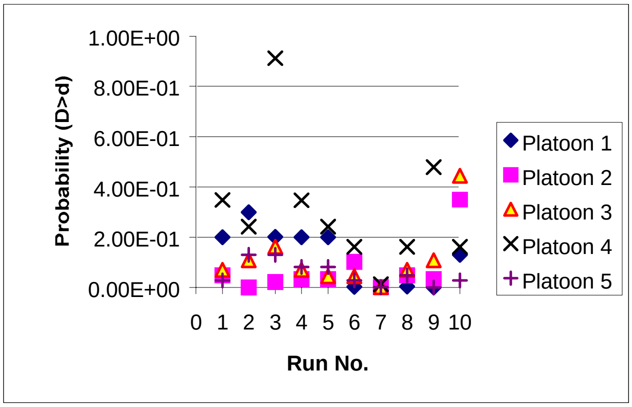

Figure 3: Histogram from K-S Two-Dimensional Two-Sample Tests

ACKNOWLEDGMENTS

This work was performed for the Federal Highway Administration, U.S. Department of Transportation, as part of contract DTFH61-95-C-00125. The third author acknowledges support from National Science Foundation grant DMI-9622269. We also thank numerous personnel at ITT Systems Corp., various members in the traffic engineering community, and Gamze Tokol for their constructive suggestions.

REFERENCES

Balci, O., and R.G. Sargent. A Methodology for Cost-Risk Analysis in the Statistical Validation of Simulation Models. *Communications of Assoc. Comput. Mach.*, 24, 1981, pp. 190-197.

Carson, J.S. Convincing Users of Model's Validity is a Challenging Aspect of Modeler's Job. *Industrial Engineering*, 18, June 1986, pp. 74-85.

Chen, B.C., and R.G. Sargent. Using Standardized Time Series to Estimate the Difference Between Two Stationary Stochastic Processes. *Operations Research*, 35, 1987, pp. 428-436.

Emshoff J.R., and R.L. Sisson. Design and Use of Computer Simulation Models. MacMillan Company, New York, NY, 1970.

Fasano, G. and A. Franceschini. Monthly Notices of the Royal Astronomical Society, Vol. 225, pp. 155-170, 1987.

Fishman, G.S., and P.J. Kiviat. The Statistics of Discrete-Event Simulation. *Simulation*, 10, 1968, pp.185-195.

Garratt, M. Statistical Validation of Simulation Models. In *Proceedings of the 1974 Summer Computer Simulation Conference*. Simulation Councils, La Jolla, CA, pp. 915-926.

Gerlough, D.L., and M.J. Huber. *Traffic Flow Theory, A Monograph.* Transportation Research Board, Special Report 165, 1975.

Law, A.M., and W.D. Kelton. *Simulation Modeling & Analysis*, 2nd edition, McGraw-Hill, Inc., 1991.

Press, W.H., S.A. Teukolsky, W.T. Vetterling, and B.P. Flannery. *Numerical Recipes in C*, 2nd Edition, Cambridge University Press, 1992.

Sargent, R.G. Verification and Validation of Simulation Models, In *Progress in Modeling & Simulation (*ed. F. E. Cellier), Academic Press, 1982.

Schlesinger, S., R.E. Crosbie, G.S. Innis, C.S. Lalwani, J. Loch, R.J. Sylvester, R.O. Wright, N. Kheir, and D. Bartos. Terminology for Model Credibility. *Simulation*, 32, 1979, pp. 103-104.

Schruben, L.W. Confidence Interval Estimation Using Standardized Time Series. *Operations Research*, 31, No.6, 1983, pp.1090-1108.

Shannon, R.E. *Systems Simulation: The Art and Science*. Prentice-Hall, Englewood Cliffs, NJ, 1975.

Taylor, M.A.P. Evaluating the Performance of a Simulation Model. *Transportation Research*, Vol. 13A, 1979, pp. 159-173.

Theil, H. *Economic Forecast and Policy*. North-Holland Publishing Company, Amsterdam, Netherlands, 1961.

Van Horn, R.L. Validation of Simulation Results. *Management Science*, Vol.17, No. 5, 1971, pp. 247-258.

AUTHOR BIOGRAPHIES

LEI RAO is a transportation engineer at ITT Systems Corp. He holds an M.S. in Civil Engineering from the Ohio State University. His research interests include traffic simulation, model validation and calibration, traffic flow theory and dynamic traffic assignment.

LARRY OWEN is the Principal Analyst for the ITT Systems' Transportation Group. For the past several years he has been doing research in adaptive traffic signal control, traffic flow modeling and validation under contract to the Federal Highway Administration. Mr. Owen has over 25 years experience in simulation development and application, systems analysis, and operations research.

DAVID GOLDSMAN is an associate professor in the School of Industrial and Systems Engineering at Georgia Tech. His research interests include simulation output analysis and ranking and selection procedures. He is an active participant in the Winter Simulation Conference, having been the Program Chair in 1995.