

THEMIS SIGNAL ANALYSIS STATISTICS RESEARCH PROGRAM

TRANSPORTATION EXAMPLE OF STATISTICAL VALIDATION

FOR SIMULATION MODELS

by

A. V. Gafarian and John E. Walsh

Technical Report No. ~~68~~ 65
Department of Statistics THEMIS Contract

April 30, 1970

Research sponsored by the Office of Naval Research
Contract N00014-68-A-0515
Project NR 042-260

Reproduction in whole or in part is permitted
for any purpose of the United States Government.

This document has been approved for public release
and sale; its distribution is unlimited.

DEPARTMENT OF STATISTICS
Southern Methodist University

TRANSPORTATION EXAMPLE OF STATISTICAL VALIDATION

FOR SIMULATION MODELS*

A. V. Gafarian and John E. Walsh
System Development Corporation Southern Methodist University

ABSTRACT

Statistical validation of simulation models has much interest in military applications. In particular, there is interest in transportation problems during emergency and standard situations. However, collection of suitable data is difficult and almost necessarily is limited to standard situations. Such a situation is examined in this paper. Vehicle traffic on a section of a freeway (expressway) near an on-ramp is simulated by a Monte Carlo model. Ability of the model to represent actual traffic flow is investigated by observing operational systems and making comparisons with results from corresponding simulations. Measures of effectiveness are used that compare statistics for operational systems with values of these properties for corresponding simulations (to see if the difference is significant). Separate comparisons are made for three kinds of statistics: 1) Average travel times for vehicles entering on the freeway; 2) Average travel times for vehicles entering on-ramp; 3) Some averages of velocities. Two phases occur, depending on the data for operational systems. Data for the first phase are crude and incomplete, so that correspondence of the model to an operational system is rough (vehicles enter observation on a random basis, etc.). Then, the for the second phase are to be complete and rather accurate. Then, the vehicles can enter observation at the same times, locations, and velocities as for the operational system. Also, a more refined measure of effectiveness can be used. For the first phase, the basic comparison is between expected values of the statistics. Expected values and variances can be compared for the second phase.

*Research partially supported by System Development Corporation's Independent Research Program and Mobil Research and Development Corporation. Also associated with ONR Contract N00014-68-A-0515.

INTRODUCTION

An approach to statistical validation of Monte Carlo simulation models by comparison with operational systems is given in ref. 1. Here that approach is followed in the development of methods for validating a model of traffic flow. The traffic flow situation is described first. Then, choice of measures of effectiveness, involving observational statistics, and the two phases of data collection (from operational systems) are considered. The measure developed for the first phase is also usable for the second phase, but a more desirable method is given for the second phase.

The traffic flow situation that is being considered is that of merging of on-ramp vehicles into the main stream of a freeway (expressway). Fig. 1 is a schematic diagram of the specific geometry, which encompasses a section of the freeway. It shows four lanes of freeway widening to five lanes, when one lane

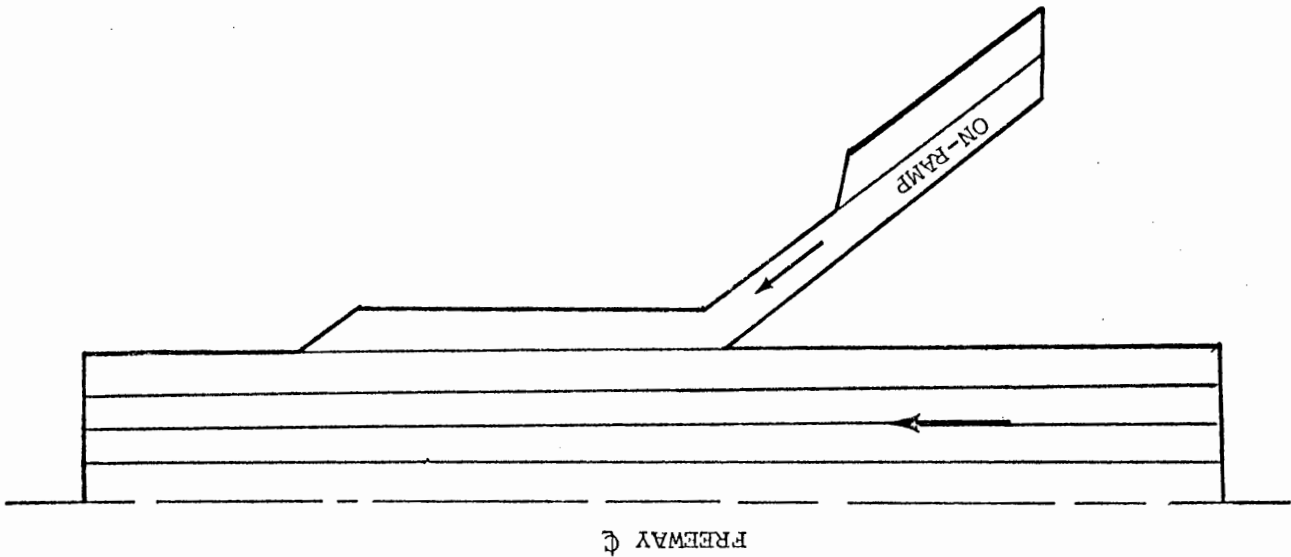


FIGURE 1

is added for acceleration in the merging section, and then subsequently narrowing to four lanes. The on-ramp has two entrance lanes for the arterial traffic. It then narrows to the one lane that ultimately becomes the acceleration lane. The merging model is the first of a sequence of models that will terminate when a complete diamond interchange for vehicles between a freeway and an arterial street has been designed, Fig. 2. The second model of the sequence is

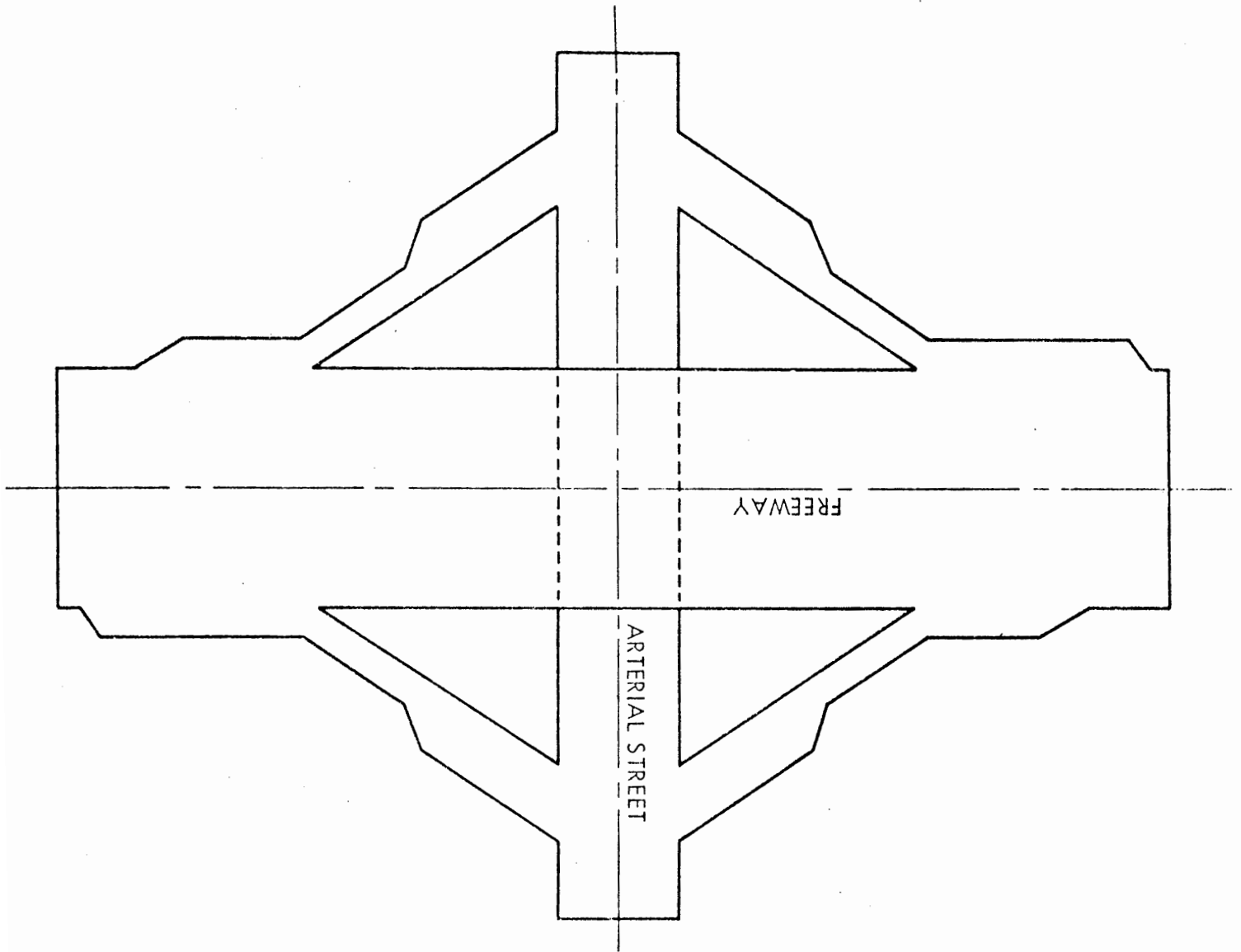


FIGURE 2

a combination of the first model plus an up-stream off-ramp. The third model is the arterial street and its intersections with the ramps. When all of these partial models are complete and validated, the second and third models will be combined appropriately to form the total interchange. It is expected that this procedure will produce a valid total model much sooner and with fewer major difficulties than would be experienced if a total simulation were attempted from the beginning. To achieve the desired flexibility required for subsequent research on the construction and operation of such interchanges, many geometrical variables, as well as control variables for the resolution of conflicting vehicular movements, will be incorporated into the model.

In developing the overall logic for the merging model, an attempt was made to maximize the ratio of simulation time to computer time. To this end it was decided not to use the microscopic, time-slicing approach to simulation wherein the location of each vehicle is determined at the end of each of uniformly spaced small time increments. Instead, an essentially macroscopic, significant event simulation technique was chosen. In this approach, the model is broken down into several interconnected submodels of relatively uniform geometric configuration within which vehicles are aggregated. Individual vehicle characteristics, such as velocity, are determined only as vehicles cross boundaries between submodels.

The merging model was coded for the Philco 2000 computer in machine language. For high volume cases of around 7000 vehicles per hour, the simulation time to computer time is of the order of 24 while for the low volume cases of around 2000 vehicles per hour, the ratio is of the order of 57.

the numbers $f_i, r_i, \bar{x}_i, \bar{y}_i, i = 1, \dots, 18$ where f_i and r_i are the hourly freeway and on-ramp flow rates respectively corresponding to the i^{th} period; $\bar{x}_i$ and $\bar{y}_i$ are the averages of the freeway travel times and on-ramp travel times respectively for the i^{th} period.

Standardization was established by first setting the geometrical parameters of the model equal to those of the operational system. Second, the freeway and on-ramp cars were generated into the system by a random process having the

measured field rates. Other factors, such as curvature, vertical and horizontal alignments, grade, and lane width, which have not been used explicitly in the model, have been implicitly included in the estimates of vehicle-driver characteristics based on the judgment of the investigators. The only vehicle-driver performance characteristic that was determined experimentally and set in the model, as part of the standardization procedure, was the desired free flow velocity of freeway drivers. These were estimated using vehicle travel times

obtained with a stopwatch during very light traffic conditions when vehicles were able to weave with no apparent interference.

The initial condition for the computer model is that of no vehicles. The model is then run for four hours of simulated time cumulating separately the average travel times of freeway drivers and on-ramp drivers. An analysis of these cumulative averages shows they become very stable and hence may be assumed to be the exact mean values for travel times with negligible random error. Thus, the outcome of the simulation runs are the numbers ξ_i and $\eta_i, i = 1, 2, \dots, 18$

The present technique for collecting and analyzing field data is very crude, provides inaccurate results, and is restricted to interchanges for which the presence of a tall structure is available to provide the necessary vantage point for making observations. It is also very restricted with regard to collecting the required input information for better standardization. Hence, an extensive experimental study was undertaken to evaluate the feasibility of using ground based sensors and aerial photographic techniques for data collection. Also a study was made of data reduction methods.

SECOND PHASE FOR TRAFFIC STUDY

results it was felt that the model was behaving realistically. Based on these preliminary tailed significance probability of .21 was obtained. Similarly for on-ramp drivers an equal-cance probability of .61 was yielded. In this particular case, an equal-tailed significance $\xi_i - \mu_i = 0, i = 1, 2, \dots, 18$. Signed-Rank Test may be used on the numbers $\xi_i - \bar{x}_i, i = 1, 2, \dots, 18$ to test Also, these statistics can be assumed to be independent. Hence, the Wilcoxon Limit Theorem implies symmetry of the distributions of the statistics $\xi_i - \bar{x}_i$. Considering for a moment freeway drivers only, application of a Central Limit Theorem implies symmetry of the distributions of the statistics $\xi_i - \bar{x}_i$ and $\eta_i - \bar{y}_i, i = 1, 2, \dots, 18$. Tests of whether the relations $\xi_i - \mu_i = 0$ and $\eta_i - \tau_i = 0, i = 1, 2, \dots, 18$ hold for freeway drivers and on-ramp drivers respectively are based on the statistics $\xi_i - \bar{x}_i$ and $\eta_i - \bar{y}_i, i = 1, 2, \dots, 18$. Denote by μ_i and τ_i the true but unknown mean travel times of actual freeway drivers and on-ramp drivers respectively. where ξ_i and η_i are the exact mean travel times of simulated freeway drivers and on-ramp drivers respectively.

As a result of these studies, it is planned to use a 70 mm motion picture camera mounted on a helicopter which is hovering above the interchange. The data on the film will then be read using a semi-automatic film record reader and punched onto cards for further computer processing. Once this system is implemented it will be used for validating the merging model and all the subsequent models.

This system has many advantages, among which are these: (1) It will be possible to make validation studies at many interchanges; (2) The film record is permanent and contains all the information. Thus it may be used as often as it is necessary to get additional information; (3) The data are more accurate. Even at moderate frame rates it is possible to get nearly instantaneous velocities rather than averages over long time periods and reasonably good estimates of arrival times of the vehicles to the various submodel boundaries in the system.

Standardization of the computer model, when this system is implemented, will differ radically from that in the first phase version. First, this technique will make possible the determination of vehicle-driver characteristics that are required as inputs to model. Previously, many of these characteristics were based on the judgment of the investigators. Second, this scheme will make possible the measurement of arrival times and initial velocities at various points, number of lane changes, etc. These measurements will constitute the field data that will be compared to the computer model output.

With this setup it will also be possible to preload the model at time zero with the performance results of the vehicles that are initially in the system. The simulation model will then be run by inserting vehicles into the system at the observed entrance time with their observed initial velocities. The program logic, which contains many random elements, will then move the vehicles through the system and record the same items of information mentioned above.

The statistical test in this case will be essentially the same as in the first phase. The procedure is to first make a total of, say, n data runs, i.e., n helicopter flights. These flights will be conducted at a given interchange over a period of several days. If more than one flight is conducted on any one day, the time between flights will be long enough so that the runs are independent. The length of each run is estimated to be approximately six minutes. This is based on a frame rate of 2 per second using a 100' magazine with thin base film. This allows approximately 150 feet of film or about 695 exposures. Consider for the moment total travel time of freeway drivers for the merging model. For each of the n runs, the travel time for each vehicle which enters and leaves this system will be measured and averaged to give a number $\bar{x}_i$, $i = 1, 2, \dots, n$. Again, it may be assumed that $\bar{x}_i$ has a symmetrical distribution. The simulation model will then be run as described earlier in this section and the same average will be recorded. Call this average $\bar{y}_i^{(1)}$. The superscript indicates that this is the first of a sequence of n_i replications.

Finally, the average $\frac{1}{n_i} \sum_{j=1}^{n_i} \bar{y}_i(j)$ will be computed to yield a number $\bar{y}_i$ which,

because n_i is sufficiently large, will be the expected value, with negligible random error, for the random variable $\bar{y}_i$. The Wilcoxon Signed-Rank Test will then be applied to the differences $\bar{y}_i - \bar{x}_i$, $i = 1, 2, \dots, n$ to test if the means of the travel times are to the same. A similar procedure will be applied to on-ramp driver travel times and travel times through various submodels of the model.

An identical procedure will be applied to test whether the means of the velocities at various sub-model boundaries are the same for the operational

system and the simulation model. Thus, e.g., let $\bar{v}_i$ denote the average system exit velocity for all vehicles in the i th run. The Wilcoxon Signed-Rank Test may then be applied to the differences $\bar{v}_i - \bar{v}_i$, $i = 1, 2, \dots, n$, $\bar{v}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} \bar{v}_i(j)$, where $\bar{v}_i(j)$ is the average system exit velocity for all vehicles which enter and leave during the i th replication.

Some additional kinds of tests will occur for the second phase. These are based on asymptotic considerations. Explicitly, let the multidimensional point $\bar{z}_i$ be a vector of observations occurring during the i th data run. Functions $g(\bar{z}_i)$ occur that have technical meaning for traffic flow and also are believed to have approximately normal distributions when the number of observations in $\bar{z}_i$ is very large. An observation on $g(\bar{z}_i)$ is obtained for the operational system. A very large number of observed values of $g(\bar{z}_i)$ can be obtained for

the corresponding simulation model, so that its expected value ρ_i and its variance τ_i^2 are essentially known. The null hypothesis asserts that $Eg(Z_i)$ for the operational system equals ρ_i , and that the variance of $g(Z_i)$ for the operational system equals τ_i^2 . Then, with $g(Z_i)$ the value for the operational system,

$$\sum_{i=1}^n \left[g(Z_i) - \rho_i \right]^2 / \tau_i^2$$

has a null distribution that is approximated by the chi-square distribution with n degrees of freedom. Large values of this test statistic are

significant.

Use of this chi-square type statistic has the advantage that differences between the operational system and the simulation model for given i ($i = 1, \dots, n$) are not masked by cancellation against differences for other values of i . Also, approximate power function results can be developed without great effort. This should yield a suitable balance between the significance level and the probability of type II error for an "important" deviation from the null hypothesis.

It often happens that the variance of $g(Z_i)$ is larger than τ_i^2 , even when the simulation model provides a useful representation of operational systems. This is because there is ordinarily less allowance for statistical variation in a simulation model than in a corresponding operational system. Thus replacement of τ_i^2 in the chi-square statistic by an appropriate larger value can be

desirable. Incidentally, the power of chi-square test can be poor when the variance of $g(Z_1)$ is much smaller than the null value used, but this should seldom be the case in practice.

REFERENCE

1. Gafarian, A. V. and Walsh, John E., "Statistical Approach for Validating Simulation Models by Comparison with Operational Systems," The Proceedings of the Fourth International Conference on Operational Research, John Wiley and Sons, New York, 1969, pp. 702-705.

UNCLASSIFIED

Security Classification

DOCUMENT CONTROL DATA - R & D

Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)

1. ORIGINATING ACTIVITY (Corporate author)

2a. REPORT SECURITY CLASSIFICATION

UNCLASSIFIED

2b. GROUP

UNCLASSIFIED

SOUTHERN METHODIST UNIVERSITY

3. REPORT TITLE

Transportation example of statistical validation for simulation models.

4. DESCRIPTIVE NOTES (Type of report and inclusive dates)

Technical Report

5. AUTHOR(S) (First name, middle initial, last name)

John E. Walsh

A. V. Gafarian

6. REPORT DATE

April 30, 1970

8a. CONTRACT OR GRANT NO.

N00014-68-A-0515

b. PROJECT NO.

NR 042-260

d.

9b. OTHER REPORT NO(S) (Any other numbers that may be assigned
(this report)

63

9a. ORIGINATOR'S REPORT NUMBER(S)

13

7b. NO. OF REFS

1

10. DISTRIBUTION STATEMENT

This document has been approved for public release and sale; its distribution is unlimited. Reproduction in whole or in part is permitted for any purpose of the United States Government.

11. SUPPLEMENTARY NOTES

12. SPONSORING MILITARY ACTIVITY

Office of Naval Research

13. ABSTRACT

Statistical validation of simulation models has much interest in military applications. In particular, there is interest in transportation problems during emergency and standard situations. However, collection of suitable data is difficult and almost necessarily is limited to standard situations. Such a situation is examined in this paper. Vehicle traffic on a section of a freeway (expressway) near an on-ramp is simulated by a Monte Carlo model. Ability of the model to represent actual traffic flow is investigated by observing operational systems and making comparisons with results from corresponding simulations. Measures of effectiveness are used to compare statistics for operational systems with values of these properties for corresponding simulations (to see if the difference is significant). Separate comparisons are made for three kinds of statistics: 1) Average travel times for vehicles entering on the freeway; 2) Average travel times for vehicles entering on-ramp; 3) Some averages of velocities. Two phases occur, depending on the data for operational systems. Data for the first phase are crude and incomplete, so that correspondence of the model to an operational system is rough (vehicles enter observation on a random basis, etc.). The data for the second phase are to be complete and rather accurate. Then, the vehicles can enter observation at the same times, locations, and velocities as for the operational system. Also, more refined measure of effectiveness can be used. For the first phase, the basic comparison is between expected values of the statistics. Expected values and variances can be compared for the second phase.

(PAGE 1)

DD FORM 1473
1 NOV 65

S/N 0101-807-6811

Security Classification

UNCLASSIFIED

A-31408