NATIONAL BUREAU OF STANDARDS REPORT

10 624

SEARCH AND RESCUE SIMULATION MODEL
PHASE III REPORT
CALIBRATION/VALIDATION/INVESTIGATION

Technical Analysis Division
Institute for Applied Technology
National Bureau of Standards
for
United States Coast Guard

NBS

U.S. DEPARTMENT OF COMMERCE
NATIONAL BUREAU OF STANDARDS

NATIONAL BUREAU OF STANDARDS

The National Bureau of Standards [1] was established by an act of Congress March 3, 1901. Today, in addition to serving as the Nation's central measurement laboratory, the Bureau is a principal focal point in the Federal Government for assuring maximum application of the physical and engineering sciences to the advancement of technology in industry and commerce. To this end the Bureau conducts research and provides central national services in four broad program areas. These are: (1) basic measurements and standards, (2) materials measurements and standards, (3) technological measurements and standards, and (4) transfer of technology.

The Bureau comprises the Institute for Basic Standards, the Institute for Materials Research, the Institute for Applied Technology, the Center for Radiation Research, the Center for Computer Sciences and Technology, and the Office for Information Programs.

THE INSTITUTE FOR BASIC STANDARDS provides the central basis within the United States of a complete and consistent system of physical measurement; coordinates that system with measurement systems of other nations; and furnishes essential services leading to accurate and uniform physical measurements throughout the Nation's scientific community, industry, and commerce. The Institute consists of an Office of Measurement Services and the following technical divisions:

> Applied Mathematics—Electricity—Metrology—Mechanics—Heat—Atomic and Molecular Physics—Radio Physics [2]—Radio Engineering [2]—Time and Frequency [2]—Astrophysics [2]—Cryogenics.[2]

THE INSTITUTE FOR MATERIALS RESEARCH conducts materials research leading to improved methods of measurement standards, and data on the properties of well-characterized materials needed by industry, commerce, educational institutions, and Government; develops, produces, and distributes standard reference materials; relates the physical and chemical properties of materials to their behavior and their interaction with their environments; and provides advisory and research services to other Government agencies. The Institute consists of an Office of Standard Reference Materials and the following divisions:

> Analytical Chemistry—Polymers—Metallurgy—Inorganic Materials—Physical Chemistry.

THE INSTITUTE FOR APPLIED TECHNOLOGY provides technical services to promote the use of available technology and to facilitate technological innovation in industry and Government; cooperates with public and private organizations in the development of technological standards, and test methodologies; and provides advisory and research services for Federal, state, and local government agencies. The Institute consists of the following technical divisions and offices:

> Engineering Standards—Weights and Measures — Invention and Innovation — Vehicle Systems Research—Product Evaluation—Building Research—Instrument Shops—Measurement Engineering—Electronic Technology—Technical Analysis.

THE CENTER FOR RADIATION RESEARCH engages in research, measurement, and application of radiation to the solution of Bureau mission problems and the problems of other agencies and institutions. The Center consists of the following divisions:

> Reactor Radiation—Linac Radiation—Nuclear Radiation—Applied Radiation.

THE CENTER FOR COMPUTER SCIENCES AND TECHNOLOGY conducts research and provides technical services designed to aid Government agencies in the selection, acquisition, and effective use of automatic data processing equipment; and serves as the principal focus for the development of Federal standards for automatic data processing equipment, techniques, and computer languages. The Center consists of the following offices and divisions:

> Information Processing Standards—Computer Information — Computer Services — Systems Development—Information Processing Technology.

THE OFFICE FOR INFORMATION PROGRAMS promotes optimum dissemination and accessibility of scientific information generated within NBS and other agencies of the Federal government; promotes the development of the National Standard Reference Data System and a system of information analysis centers dealing with the broader aspects of the National Measurement System, and provides appropriate services to ensure that the NBS staff has optimum accessibility to the scientific information of the world. The Office consists of the following organizational units:

> Office of Standard Reference Data—Clearinghouse for Federal Scientific and Technical Information [3]—Office of Technical Information and Publications—Library—Office of Public Information—Office of International Relations.

---

[1] Headquarters and Laboratories at Gaithersburg, Maryland, unless otherwise noted; mailing address Washington, D.C. 20234.
[2] Located at Boulder, Colorado 80302.
[3] Located at 5285 Port Royal Road, Springfield, Virginia 22151.

NATIONAL BUREAU OF STANDARDS REPORT

NBS PROJECT 4314561 | **NBS REPORT** 10 624

SEARCH AND RESCUE SIMULATION MODEL
PHASE III REPORT
CALIBRATION/VALIDATION/INVESTIGATION

Stephen S. Karp

et al.

Technical Analysis Division
Institute for Applied Technology
National Bureau of Standards
for
United States Coast Guard

NBS

U.S. DEPARTMENT OF COMMERCE

NATIONAL BUREAU OF STANDARDS

PREFACE

This volume reports on the validation procedures applied to a Search and Rescue Simulation Model prepared for the United States Coast Guard. The model itself was developed by an inter-disciplinary team at the National Bureau of Standards with representation from the U. S. Coast Guard under MIPR Z-70099-0-01935. Complete documentation was provided in NBS Reports 10430 through 10436.

This validation report was prepared under the general supervision of Richard T. Penn, Jr. and Walter G. Leight. Technical project leadership was supplied throughout the course of the project, including development of the simulation and all validation efforts to date, by Stephen S. Karp, who is the principal author* of this report. Other participants in the validation effort included the following members of the Technical Analysis Division:

Linda Cummings, Jane Duberg, Joel Levy, Elizabeth Leyendecker, Marcia Maltese, Wayne Steele, and Michael Vogt.

U. S. Coast Guard participants included:

Paul D'Zmura, Thomas Matteson, and Gerald Underwood

Support services were furnished by the following members of the Technical Analysis Division:

Mary Abbott, Frances Jones, and Terrie Conrad.

---

*Special appreciation is due to Walter Leight for his guidance in the preparation of this report.

SARSIM VALIDATION REPORT

Abstract

The Search and Rescue Simulation (SARSIM) model developed for the U. S. Coast Guard has been subjected to a series of calibration runs, validation tests and investigative exercises. This report presents the results of the associated computer runs and their analyses. It also describes some sample exercises made to demonstrate the model's utility as a long-range planning tool. The report concludes with a summary of some significant findings made from these exercises; an assessment of the current SARSIM model; and recommendations for future model refinements.

TABLE OF CONTENTS

I. INTRODUCTION

1.1 Background

Due to the rapid and continuing growth of marine activity, especially in recreational boating, combined with the constraints imposed by national budget allocations, the Coast Guard has found it necessary to examine its readiness postures and operational policies and to project its future requirements for Search and Rescue (SAR) force levels on an integrated resource basis. This entails simultaneous consideration of aircraft, cutters, and shore stations (with their associated boats) to plan properly for the entire SAR mission. An earlier investigation of alternative methodologies for studying such a complex system led to the conclusion that only a large-scale simulation model could provide the desired results.

To provide meaningful answers to the problems faced by Coast Guard management, any simulation must faithfully represent current operations, yet also apply to a wide range of possible future modes. The heterogeneous characteristics of both clients (people or property in distress) and servers (Coast Guard resources, i.e., boats, aircraft, or cutters) present further complications. For example, distress cases occur essentially randomly over a non-uniform geographical area, and their needs vary considerably in both type and amount. Assistance can be rendered to clients by several types of resources, perhaps, each type constrained by its physical capabilities and the environmental conditions.

The Search and Rescue Simulation Model (SARSIM) has been designed as a management tool capable of answering a wide variety of questions, involving variations in such factors as:

(a) SAR workload --- The user may examine the ability of a planned or postulated SAR force to provide adequate service, and to explore the effects to be expected from changes in specified caseloads.

(b) Resource location --- The model permits investigation of the effects of adding, deleting, or relocating SAR facilities.

(c) Resource type --- Trade-offs may be examined with regard to cost-benefit relationships for alternative resource mixes.

(d) Resource performance criteria --- The influence of speed, endurance, or other parameters can be estimated.

(e) Manning levels --- Changes in the number of personnel assigned to given stations, or the shifts to which they are assigned, may be explored in relation to postulated SAR demands.

(f) Operating tactics --- These include the institution and location of patrols during busy hours in areas deemed likely to have high levels

of potential incidents, as well as the effects of employing alternative "server disciplines" and "interrupt rules.*"

1.1.1 The Simulation Model

The core of the simulation is a relatively sophisticated resource selection algorithm which assigns one or more "preferred" resources to each case entering the system, if at all possible. Assignments are based on the following factors:

(a) The client's requirements (e.g., search, personnel rescue, tow, etc.).

(b) The case's urgency, or severity.

(c) The capabilities of available resources to serve the client's needs.

(d) The environmental conditions at the time of the incident.

(e) The location of each capable resource in relation to the site of distress.

(f) The relative costs of responding by those resources which can supply satisfactory service.

In selecting the preferred resource(s), the algorithm considers both response time and cost for the set of

---

*These rules specify the degree to which neighboring stations interact with one another in responding to distress incidents.

capable resources: cost is dominant if a specified "maximum tolerable" response time can be met; otherwise, quickness in response governs the choice. If no capable resources are available at the time the notification of distress is received, consideration is given to interrupting ongoing service to a case of lower priority, or else the case is placed into a queue to await the availability of a capable resource.

The simulation model is divided into three functional modules: the Preprocessor, the Operational Simulator, and the Postprocessor. The Preprocessor is a set of FORTRAN programs which prepares a demand tape of distress cases from the historical SAR files. This tape contains a chronological ordering of cases which actually occurred in a given Coast Guard District, along with their historical attributes. The user may specify his choice of caseload, either overall or by selected criteria, such as specific case parameters; the Preprocessor then randomly generates a sequence of cases with realistic characteristics in conformance with any desired general scenario.

The Operational Simulator is the basic module of SARSIM. It is a discrete event digital simulation, written in SIMSCRIPT, which models resource assignment and service for each case. It calculates a set of summary statistics for standard output and can, when specified,

produce an output case tape for detailed analysis by the Postprocessor. The QUICK QUERY* information retrieval system is used in the Postprocessor to provide these analytical functions.

1.1.2 Project History

There have been three major phases in this project. The first, Conceptualization and Modeling, began with a study of the SAR system and investigated alternative methodologies for application to the Coast Guard management's long-range problems with regard to planning for the SAR mission. In Phase I a structure was developed for the simulation, and an analytic description was prepared for all the modules and basic algorithms in the model.

Phase II, Program Development, was concerned with program design, coding, and debugging of the model, based on the analysts' descriptions from Phase I. During this phase individual modules were debugged and simple test runs were made to compare simulation results with those derived from manual computations. In addition, interface debugging was accomplished with more complex test cases.

---

*Developed for the Economic Development Administration by Consolidated Analysis Centers Incorporated, Santa Monica, California, U.S.A.

This report is addressed to the work accomplished in Phase III, Validation and Exercise. The purpose in this phase was to calibrate several parameters within the model, to conduct validation tests, and to demonstrate SARSIM's utility by means of sample, investigative exercises of the model.

The reader desiring additional background information or greater detail concerning the SARSIM model is invited to consult appropriate portions of the seven-volume documentation, which is comprised of the following:

Volume I	Executive Level Documentation, NBS Report No. 10430.
Volume II	Analyst Level Documentation, NBS Report No. 10431.
Volume III	Programmer Level Documentation for "PREPROCESSOR", NBS Report No. 10432.
Volume IV	Programmer Level Documentation for "OPSIM", NBS Report No. 10433.
Volume V	Programmer Level Documentation for "POSTPROCESSOR", NBS Report No. 10434.
Appendix A	Flow Charts for Programmer Level Documentation, NBS Report No. 10435.
Appendix B	Program Listings for Programmer Level Documentation, NBS Report No. 10436.

Volume I, intended for executive level review, presents an overview of the model as well as an explanation of the simulation concept, including a sample simulation of a simple SAR situation.

Volume II, written at the analyst level, discusses the assumptions, limitations and design considerations of the model, and describes all the algorithms of SARSIM at the model (or analytic) level.

The remaining volumes contain the detailed descriptions of the various programs which comprise SARSIM. User's guides are also included with details of how to set up and operate each of these programs.

1.2 Purpose of Analysis Runs

When modeling almost any complex system or process, it is generally necessary to abstract and approximate the actual components and relationships within the system to insure manageability. Consequently, it is desirable for the modeler to verify that the model does indeed behave like the process it simulates under the full range of conceivable, realistic circumstances. However, in a system as complex as the Search and Rescue operations of the Coast Guard, the conceivable, interesting, realistic, alternative modes of operation include possible combinations of a wide range of values of model parameters, caseloads, resource levels, geographical configurations, etc., becoming

too numerous to itemize, let alone to vary completely in successive exercises of the simulation model. Moreover, even if all possible variations could be explored, conclusions as to the model's validity would be limited in that most of the input modes of operation would not have been historically experienced by Search and Rescue Forces, nor might they ever be.*

Since systems of this complexity can never be completely authenticated, the modeler's goal should be to find evidence which substantiates and corroborates the hypothesis that the model is valid. This implies that the results of each run of the model should increase the degree of confidence applied to the use of the model, realizing that total certitude of the model's accuracy is not attainable. (It is assumed here that neither contradictory nor inconsistent results will be observed and left uncorrected in the course of these runs.)

This process may be understood more clearly when it is separated into the several smaller steps which were carried out for SARSIM. The first step was <u>logic validation</u>

---

*In fact, one use of simulation is to determine, and then avoid, those allocations of resources which lead to undesired outcomes.

(or face validation), conducted during the conceptualization stage of the project before any computer programming had begun. It consisted of examining the logical processes inherent in the model to judge whether they seemed to be reasonable representations of the activities they were designed to simulate. The project staff was fortunate in having a Coast Guard officer assigned on a full-time basis to aid in the conceptualization process.* In addition, frequent meetings were held with potential Coast Guard managerial users so that their knowledge and experience would contribute to a more faithful and more useful model. This activity was concentrated, for the most part, in Phase I of the project.

The second step will be referred to as program validation (commonly referred to as "debugging") which assumed the following form:

- Debugging of Individual Modules -- This was a necessary first step since SARSIM contains a large number of program modules, written by several different programmers. Individual programmers checked out their own modules as they thought most suitable. When each module was

---

*Karp, S. S. and T. T. Matteson, "The Integrated Team Concept of Simulation Model Development," Proceedings, 1971 SCSC, July 1971.

checked out to the satisfaction of the responsible programmer, this phase was considered complete.

- Simple Test Runs vs. Manual Simulation -- Program results were compared with hand calculations in sets of a few to several simple test cases.
- Interface Debugging with Complex Test Cases -- As the sets of input test cases became larger and more complex, subtle "bugs", usually at the interfaces between program modules, were detected and corrected. During the final stages of this process, the test data became extremely complex, almost bizarre, as attempts were made to test out "dark corners" of the programs which would seldom be used during normal exercise of the model.

This program validation stage was concentrated in Phase II of the project.

The final stage of the formal validation process (for informal validation will continue each time the model is used) was model validation, which took part mainly during Phase III. This entailed running the computer-coded programs through a set of analysis runs designed to demonstrate the validity and usefulness of SARSIM.

The kinds of remarks which can be made about the validation results fall into three general categories.

The first pertains to adjustments of input parameters in a selected direction, with the direction of change of the output predicted by common sense. The second refers to the reasonableness of output levels when new parameter values are used, that is, changes being of the proper magnitude as well as being in the correct direction. Behavioral tests were made along these lines; the results are presented in Section 2.5.

The third type of validating results are considerably more useful than the first two, but much more difficult to obtain. These are tests of the model's ability to produce simulation results which replicate what occurs when the real SAR system operates under similar conditions. This constrains comparisons to scenarios which have actually occurred (or which will occur), and, further, to those in which an ample amount of data has been collected and is available. When model results can be compared to those obtained by the real process, this provides substantiating evidence of the model's validity. This evidence does not simply draw upon intuitive notions and common sense, but furnishes quantitative results which can be subjected to further statistical testing.

The following section outlines the computer runs of the model which were planned to use existing data in refining and demonstrating the model system's capabilities.

Chapter II then provides details of runs made to test the logical structure of the model and to calibrate internal parameters. Chapter III demonstrates the capabilities of the model by showing the results obtained from varying specific parameters to produce or describe situations of interest to Coast Guard decision makers. An overall summary of the validation experiments is presented in Chapter IV, including an evaluation of the Search and Rescue Simulation model, its capabilities, and accuracy, with recommendations concerning possible refinements for the future.

1.3 Planning of Runs

Exercises of the model were generally designed to fulfill a single, basic purpose. However, careful planning in the interest of minimizing costs frequently permitted the satisfaction of two or more requirements.

1.3.1 Types of Runs

The first set of runs was designed to test the consistency between the model and the real system. Although this purpose is basic to all runs, this initial set was intended solely to replicate results (e.g., resource utilizations, caseloads, number of cases waiting times longer than tolerance) experienced by the Search and Rescue System.

The second group of runs was intended to calibrate those parameters of the model which could not be accurately specified or measured. A good example of this is "tolerance" time that is, the maximum allowable time for a case to wait before it is serviced. Although clearly a matter of concern to the Coast Guard, explicit tolerance standards are not used operationally, but are model constructs to reflect qualitative practices. During the initial model development and debugging stages, values of these parameters were postulated as likely approximations to their real world analogs. In the analysis runs, these values were varied about the initial estimates and the resulting outputs were observed. The process was stopped when model outputs agreed fairly closely with historical statistics.

The third group of runs was designed to demonstrate the amount of error which could be expected if parameter estimates were incorrect or were to change in the course of time. The process is referred to as sensitivity analysis. Here output sensitivity to changes in tolerance times, costs, speeds, and resource levels, to name a few, were performed. (See Section 3.1.)

Finally, a fourth set of runs was designed to demonstrate the capabilities of the model. Given caseloads and modes of operation which might occur in the future, the

model provides an indication of the SAR system's ability to provide satisfactory service, as measured by resource utilization, the number of cases waiting excessively long for service, etc. (See Section 3.2.)

1.3.2 Measurement Criteria

One of the basic problems in designing runs is finding a yardstick against which to measure the model. Aside from the additional questions of criteria selection, which will be treated later, the approach selected was validation against historical performance. This gives rise to three problems. The first is the very practical problem of accessibility and "goodness" of the historical data. In the case of SARSIM, the data was generally accessible, but many man-months of effort were required to "clean" the data before it could be used. Even so, shortcomings still exist,to be discussed later.

The second problem with using historical data for validation is a philosophical one. Since history itself can be considered as but a single sample out of a universe of possible outcomes, it should not be sanctified as the sole criterion of validation. To overcome this limitation, additional caseloads were randomly generated, where bizarre predicaments were given a small probability of occurring.

The third problem arises from the fact that the model allows the system to be operated in ways which never

occurred historically, hence it becomes difficult to validate those portions of the model which assess outcomes based on hypothesized (that is, non-historical) inputs. Sufficient flexibility has been designed into SARSIM to permit representation not only of the existing SAR system and its operation, but also a wide range of hypothetical systems and operations for planning and investigative purposes. Consequently, extrapolation becomes acceptable: if simulation of the present system agrees well with actual outcomes, and if trends in the results appear reasonable and consistent when simulating various hypothetical systems, then confidence may be placed in the model's capabilities when applied to situations as yet not encountered. The term "investigative" runs will be used to describe such exercises of the model, intended to estimate the likely outcome of suggested changes in inputs.

Validation of the model should continue throughout its useful life. Whenever a change in the real SAR system has been instituted and transient effects have become negligible, the simulation model should be re-run with the newly-obtained "historical" data to verify whether the model remains valid or whether additional modifications are necessary.

Validation criteria must also be chosen. In theory, any simulation output which can also be derived independently from the historical data base might serve in this capacity, but some criteria are much more useful than others and some, in fact, may be improper to use in this particular context. For example, the time that a client has to wait from Coast Guard notification until the first resource arrives on scene, TWAIT, cannot be used as a validation measure in that its calculation depends on several items from the historical data bank. In addition to possible compounding of small errors in these elements, there may be a significant difference between reality and simulation in the calculation of waiting time for search cases which are classed as "overdue." These cases often arise in the evening when a boater is reported as late in returning from sea. The usual procedure, though not universal, is for the Coast Guard to conduct communications checks and make inquiries around the harbor area that evening; if no word is received, a search is instituted at dawn of the following day. As presently designed, however, the model always launches a single resource immediately, and commences full-scale search the following morning at sunrise. Other examples of output statistics which are invalid measures of simulation performance (if they are to be compared with the historical data)

are the number of cases occurring and the on-scene times, for these have been taken from past data as inputs to the simulation.

It is felt that the resource utilization indices serve as the best available validation statistics. Expressed in percentages, these indices combine two important simulation factors. In the first place, this index, especially when applied as a function of type of resource, indicates whether the resource selection algorithms are working properly. In addition, these indices reflect the times required to service a set of cases. These times, in turn, depend on resource speeds under different operational and environmental conditions; locations of cases, resources, patrols, and stations; assumptions concerning delays and time distributions made in the model; the probability that a given resource is in working condition when required (i.e., reliability); crew availability when needed; the sequencing of cases served and of the services performed; etc. The greatest degree of confidence can probably be placed in overall utilization indices, followed in decreasing order by indices aggregated by resource types, groups of stations, individually stations, and individual resources. (As will be described later, the historical data does not readily permit extraction of

utilization measures by time of day, or by shift, hence the indices computed for historical data are not directly comparable with simulation statistics for separate shifts.)

It should be noted that the utilization indices are calculated on the basis of the total number of resources of a given category, rather than the number of "ready" resources. Thus, for example, a station which has a complement of five boats of different types and two crews (i.e., at most two "available" boats) which expended 72 hours of resource time on SAR cases during a month will have a utilization index of 2%, based on all five boats, rather than 5%, based on the availability of only two crews. This method of calculation was employed since it would not otherwise be possible to obtain indices for resource types when crews are able to serve on two or more resource types.

It is also important to note that the utilization index is not used here as a measure of SAR system performance, but only as a check on the validity of SARSIM. There should be no misunderstanding,provided that comparable utilization indices for the simulation and for the historical data are calculated in the same way.

There are other criteria which might be used for validation, such as, for example, the number of cases with C-failure (i.e., cases which had to wait for service

longer than the specified tolerance time, based on the severity of distress). However, historical C-failure data cannot be depended on for accuracy due to questionable TWAIT data. Although statistics on queueing and service interruption are readily available, these are not very useful because of the relative rarity of occurrence of such incidents.

A discussion of mathematical treatments which might be applied for computation of a single "figure of merit" criterion, to be used in judging the performance in a SARSIM run, is included in Appendix A. This cannot be directly used for calibration or validation of the model, however, since a comparable figure is not used operationally. During the course of the study there was insufficient time to develop and apply the figure of merit concept for sensitivity analyses or demonstration runs, but it may with profit be used experimentally in future exercises of the model, and perhaps be modified to reflect operational assessments to a greater degree.

1.3.3 Specification of Runs

Four descriptive factors were specified as guides to the design of the experimental runs of the model:

(a) Selection of District: SARSIM was designed to be capable of simulating SAR operations throughout a Coast Guard District. For the initial validation runs, the First Coast Guard District, covering the area from

Rhode Island through Maine, was chosen as a representative example. This selection was influenced by several items of interest: its caseload mix, operating procedures and resources are fairly typical of most coastal districts; its caseload was fairly stable during the 1967-1969 period for which "cleaned" data is available; its data base served as background for the choice of many of the simplifying assumptions in the model; and a close working relationship had already been established between personnel at Coast Guard headquarters in Washington and those in the First District during the development and implementation of the Force Readiness Analytical Model, a precursor of SARSIM .

The use of First District data during the development and calibration procedures did, however, require that the model be exercised with other data bases to insure that SARSIM had not been tailored to fit only one district. The Seventh District (i.e., Florida, Georgia, South Carolina, and the Caribbean area) and the Thirteenth District (Washington, Oregon, and Northern California) were selected as differing significantly from the First District and also from one another. In the Seventh District, for example, the workload is essentially non-seasonal, with a high proportion of cases involving recreational boating. It is also a particularly large district, including both the Atlantic and Gulf Coasts of

the Southeastern U. S. and island areas in the Caribbean, thus contributing to substantial increases to average times required to service cases. The Thirteenth District is on the West Coast of the U. S. and is characterized by unusual shifting sand bar conditions, which could, conceivably, affect the applicability of SARSIM. Moreover, some major changes in aircraft stations have been contemplated for the Thirteenth District, hence the Coast Guard was interested in investigating the likely effects of such changes.

(b) Selection of Time Frame: Since more recent experience is better for validation purposes than is older data, FY-69, the last year for which "cleaned" data was available, was chosen for the analysis runs. Aggregated data for an entire year would not likely lead to instructive results due to wide fluctuations in demand during the course of a year, hence it was decided to investigate the summer (peak) season and the winter (non-peak) season separately. These two seasons are generally typified by different resource allocations, consonant with changes in demand.

(c) Sample Size: The number of cases of length or time simulated need not be large, primarily because the input caseload is usually fixed, containing cases in the same order and at the same times that they occurred

historically. The lack of randomness in inputs requires fewer runs to get representative results. Furthermore, since the output of the model can be checked against statistics from the historical data, confirmation of proper working of the simulation model with a given sample size can be readily ascertained. In contrast, the lack of reference data for comparison with investigative runs generally necessitates larger sample sizes.

For the First District the sample consisted of a full month of SAR activity (namely, July 1968), with a total of approximately 900 cases representing peak season activity, and about 260 cases for January, February, and March 1969 to represent a 3-month, non-peak season. These sample sizes proved to be adequate, and could be simulated in less than five minutes of UNIVAC 1108 computer time on the system at the National Bureau of Standards.

For investigative runs, where caseloads were randomly generated, the procedure consisted of creating a set of 10 samples, for one-month periods in peak season and 3-month periods in non-peak seasons. Each was generated with a fresh random number seed (see below), and a subset, to be exercised in OPSIM, was selected on the basis of summary statistics provided with the samples from the Preprocessor. (See Appendix A for a related theoretical discussion of determination of sample size.)

(d) Random Number Seed Variation: Nearly all of SARSIM's inherent randomness occurs in the Preprocessor, hence it was not expected that validation results would be sensitive to the choice or random number seeds when given a fixed input caseload. In OPSIM, the only random factors are the reliability of the selected resource (which exceeds 75% for most resource types) and changes in priority when on scene, which becomes important only when there is heavy queueing and interrupts take place. As expected, the differences among runs with different random number seeds were minor.

1.3.4 Listing of Significant Runs

Based on the details of selection given above, significant runs*were chosen; these are listed in Table 1.3.1, each identified by a six-character alphanumeric code. A further description of these runs and their associated analyses are presented in the following two chapters.

---

*Runs with errors ("bugs") or duplication of other results were eliminated.

Table 1.3.1: Analysis Runs of Spring 1971

Run #	Code*	Additional Description
13	01P1HA	Preliminary base run for District 1
14	01P1HB	All tolerances equal to 99 hours
19	01P1HC	Attend primary station assignments
20	01P1HD	Increased aricraft costs
24	01P1HE	All tolerances equal to 0 hours
26	01P1HF	All aircraft costs doubled
27	01P1HG	One-half the number of resources
28	01P1HH	Twice the number of resources and crews
29	01P1HI	Resource allocation policy number 1
30	01P1HG	Resource allocation policy number 3
31	01P1HK	Resource allocation policy number 4
32	01P1HL	Resource allocation policy number 5
34	01P1HM	Thirty-four resources
35	01P1HN	Same as 01P1HM but with new random number seed
36	01P1HO	Same as 01P1HH but with new random number seed
42	01P1HP	All tolerances doubled
43	01P1HQ	All tolerances octupled
44	13P1HA	Base run for District 13
45	01P1HR	Base run for District 1
47	13P1HB	All speeds increased by 10%
50	13P1HC	Changes in delay times and costs
51	13P1HD	Same as 13P1HC but with tow speeds equal 10 and 1
52	01P1HS	Deletion of two stations
53	01P1HT	Same as 01P1HS but with new random number seed
56	01N3HA	Winter run with cost, tolerance and speed changes
60	13P1HE	Changes in costs
61	01N3HB	Tow speeds equal to 10 and 12 knots
62	01P1DA	Randomly generated caseload scenario
63	01P1DB	Another randomly generated caseload
64	01P1HU	Caseload of August 1968
77	01P1HV	Tolerances increased by 50 percent
78	13P1HF	Speeds of advance decreased by 10%
79	13P1HG	Tow speeds equal to 4 and 6 knots
80	07P1HA	Base run for District 7
81	13P1HH	Reassignment of aircraft and their stations
85	01P1DC	1975 Forecase caseload
83	07P1HB	Caseload for August 1968
72	13P1HI	Caseload for August 1968

*A code aabcde implies the run was for district aa; b is P for peak-period or N for Non-peak c is the number of months the run simulated; d is H for Historical caseload input and D for Demgen caseload randomly generated; and e is a letter used to distinguish runs with the same first five character codes.

II. REFINEMENTS AND CALIBRATION RESULTS

2.1 Logic Changes

2.1.1 Preprocessor

Much of the validation of SARSIM was based on the use of historical case data as input, using actual times of arrival for cases presented in the original chronological order. As a result, it was felt that a special sub-validation should be undertaken to explore the random generation of cases in the DEMGEN portion of the Preprocessor. Furthermore, the mechanism originally designed for generating the arrivals was questioned relative to its ability to produce desired demand patterns.

Early test results indeed showed that a refinement to this mechanism was required. A modified algorithm was developed, and it, too, was validated separately from the remainder of SARSIM.

The sub-validation involved testing two different algorithms for generating random arrivals: the difference between them lies in the cut-off rule for long inter-arrival times (referred to as IAT's). The original design was characterized by a variable cut-off, limiting the chosen IAT to a maximum value of $3/\lambda$, where λ is the average number of arrivals per hour. (In other words, whenever the IAT value exceeds the cut-off value, it is reduced to $3/\lambda$.) A fixed cut-off value is established in the

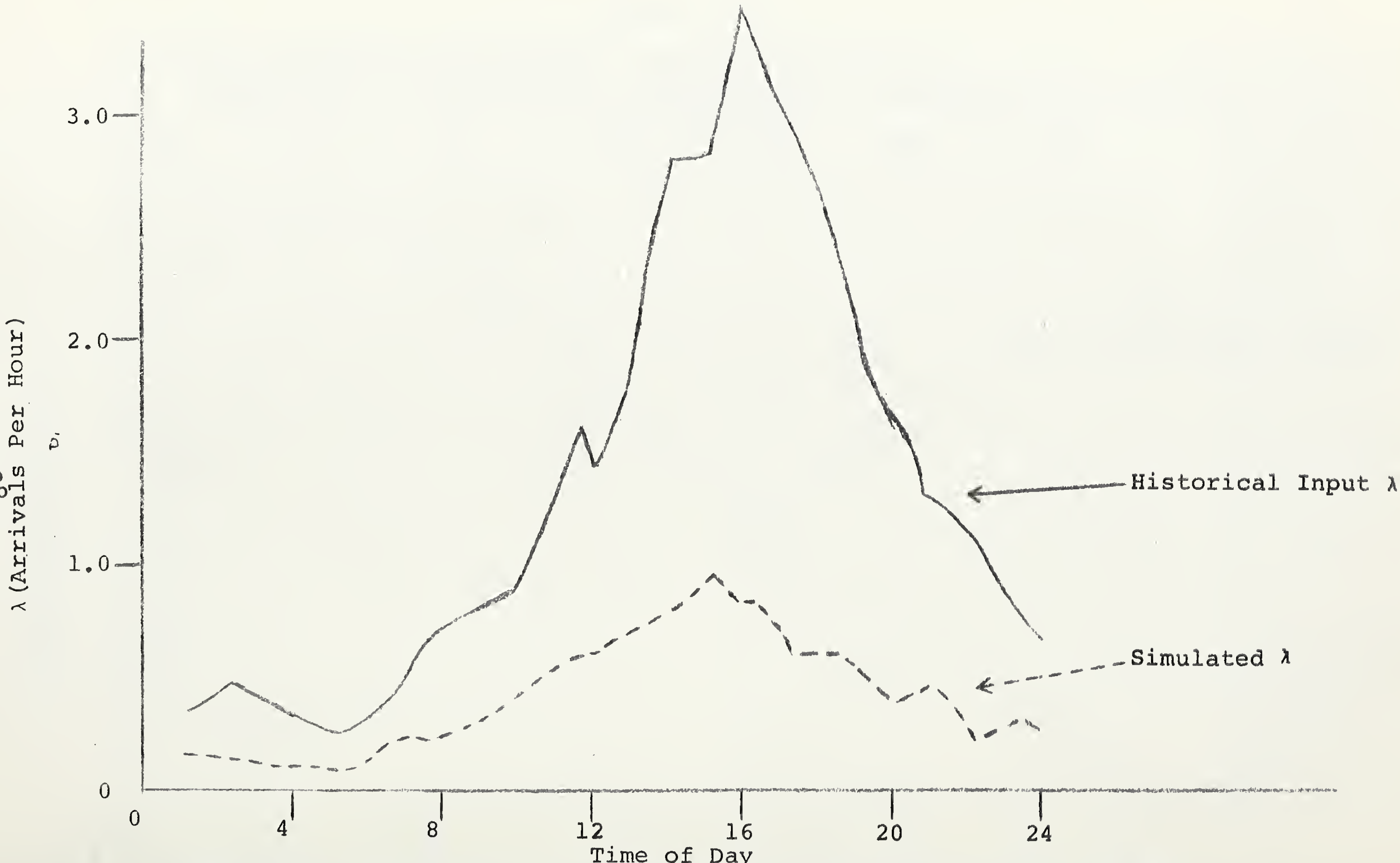
modified algorithm: whenever the value of the IAT exceeds 60 minutes, simulated time is advanced one hour, no new arrival is created, and a new random sample is drawn, using the newly-appropriate arrival rate, λ.

It was planned, for the sub-validation, to generate approximately 2500 cases, using historical hourly arrival rates, λ, for the FY-67-69 period for the First District and different random number seeds for the several runs made for each arrival mechanism. The simulated arrival rate was then calculated for each hourly interval during the day, with weekdays and weekend days considered separately; the results of calculations were compared with historical hourly rates for the same types of interval.

The first series of runs with the variable cut-off algorithm quickly demonstrated the unsuitability of this technique, as illustrated in Figures 2.1.1 and 2.1.2. The deficiency of the algorithm derives from the fact that even within the given (e.g. peak) season, the true SAR arrival pattern is highly variable. For example, the highest hourly arrival rate (namely λ = 3.186 arrivals per hour) is 43 times as great as the lowest (viz., λ = 0.075 arrivals per hour) during the peak season in the First Coast Guard District. The variable cut-off method does not adequately adapt to

Comparison of Historical and Simulated Hourly Weekday Arrival Rates for Variable Cut-off Method (Random Number Seed 1).

Figure 2.1.1



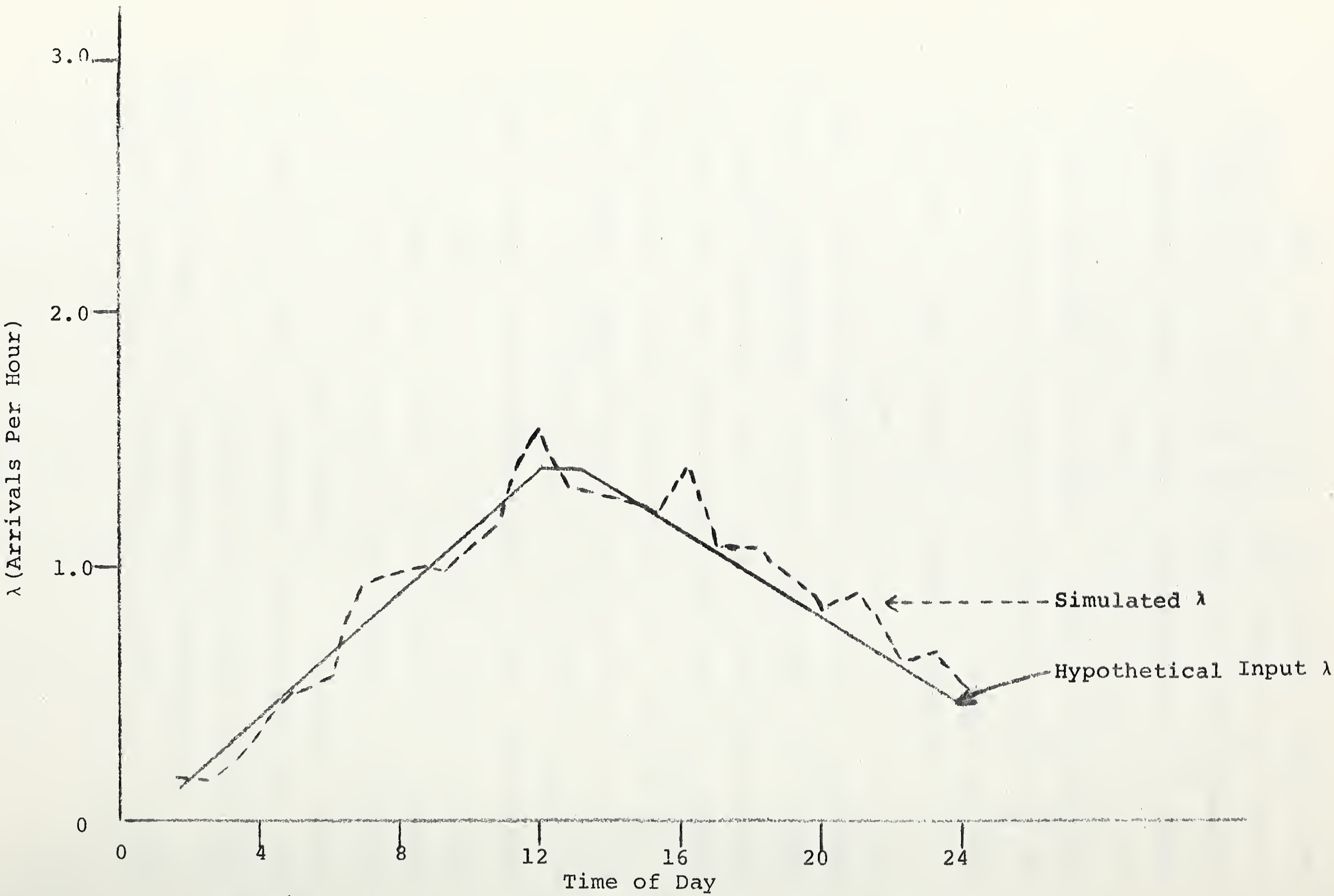
Comparison of Historical and Simulated Hourly Weekend Arrival Rates for Variable Cut-off Method (Random Number Seed 1).

Figure 2.1.2

such extreme variability. (To illustrate, when λ 0.075, the average IAT is 13.3 hours and the maximum allowable IAT is approximately 40 hours. If a small value of λ is encountered, several hourly intervals with much higher values of λ may be bypassed, seriously distorting the results. Indeed, the simulated values for the high-λ intervals were lower than their corresponding historical values by a factor of two or more.)

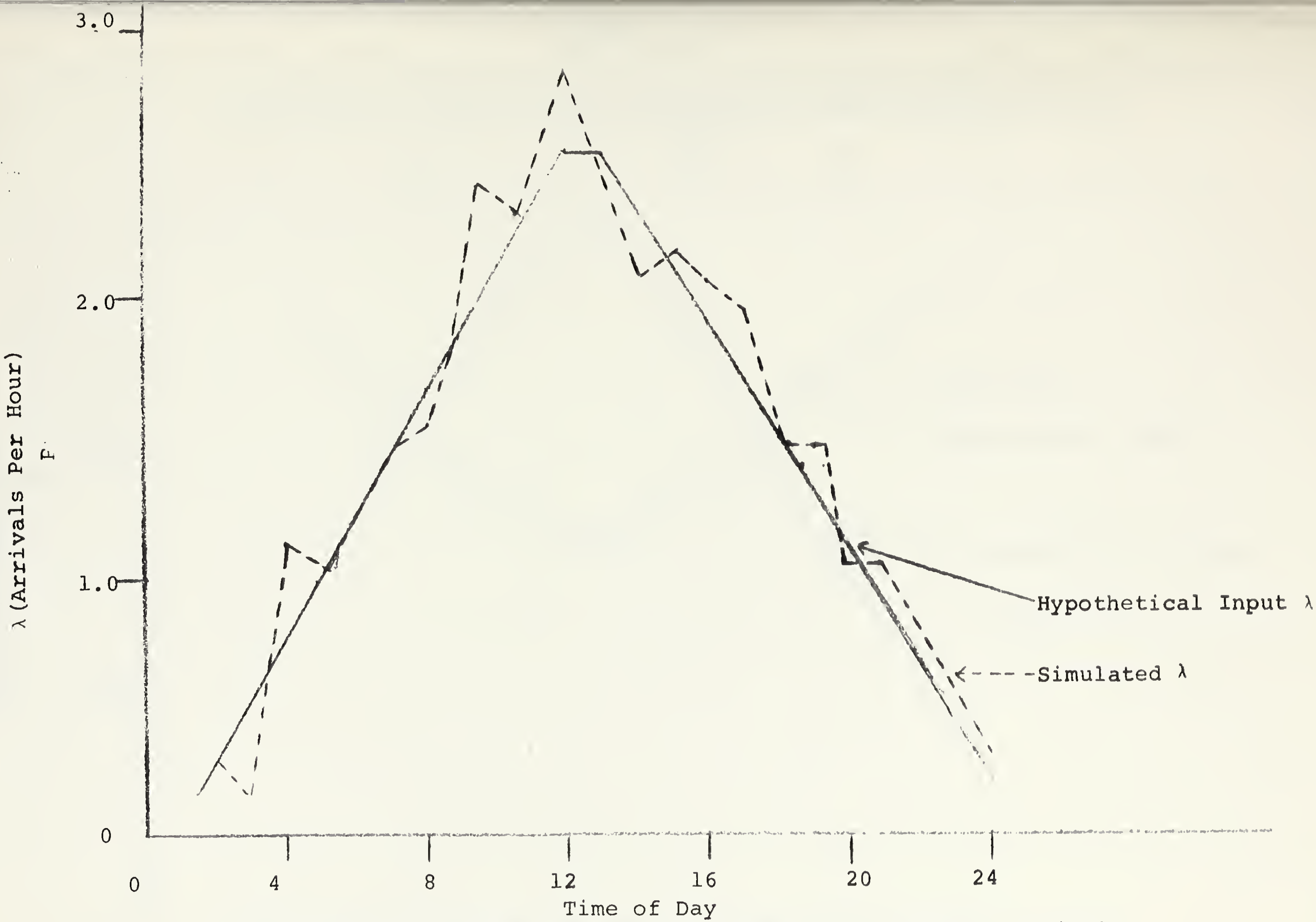
An additional series of runs was made to test the validity of the reasoning given above, that is, to show that the variable cut-off method would work if the arrival data were better behaved. A hypothetical arrival pattern was tested in lieu of the historical arrival pattern. A trapezoidal function was postulated with λ varied from 1 to 12 arrivals per hour on weekdays and from 2 to 24 arrivals per hour on weekends. Using two different random number seeds, the fits between the hypothetical (input) distributions and the corresponding simulated (output) distributions were very close. (See Fiugres 2.1.3 through 2.1.6.) These runs suggest that the variable cutoff method might be satisfactory for a wide variation in λ when each hourly λ is at least 1.0. This method might therefore be used for SARSIM with longer (than hourly) intervals.

The fixed cutoff results were much more satisfactory. Figures 2.1.7 through 2.1.10 show very good fits between



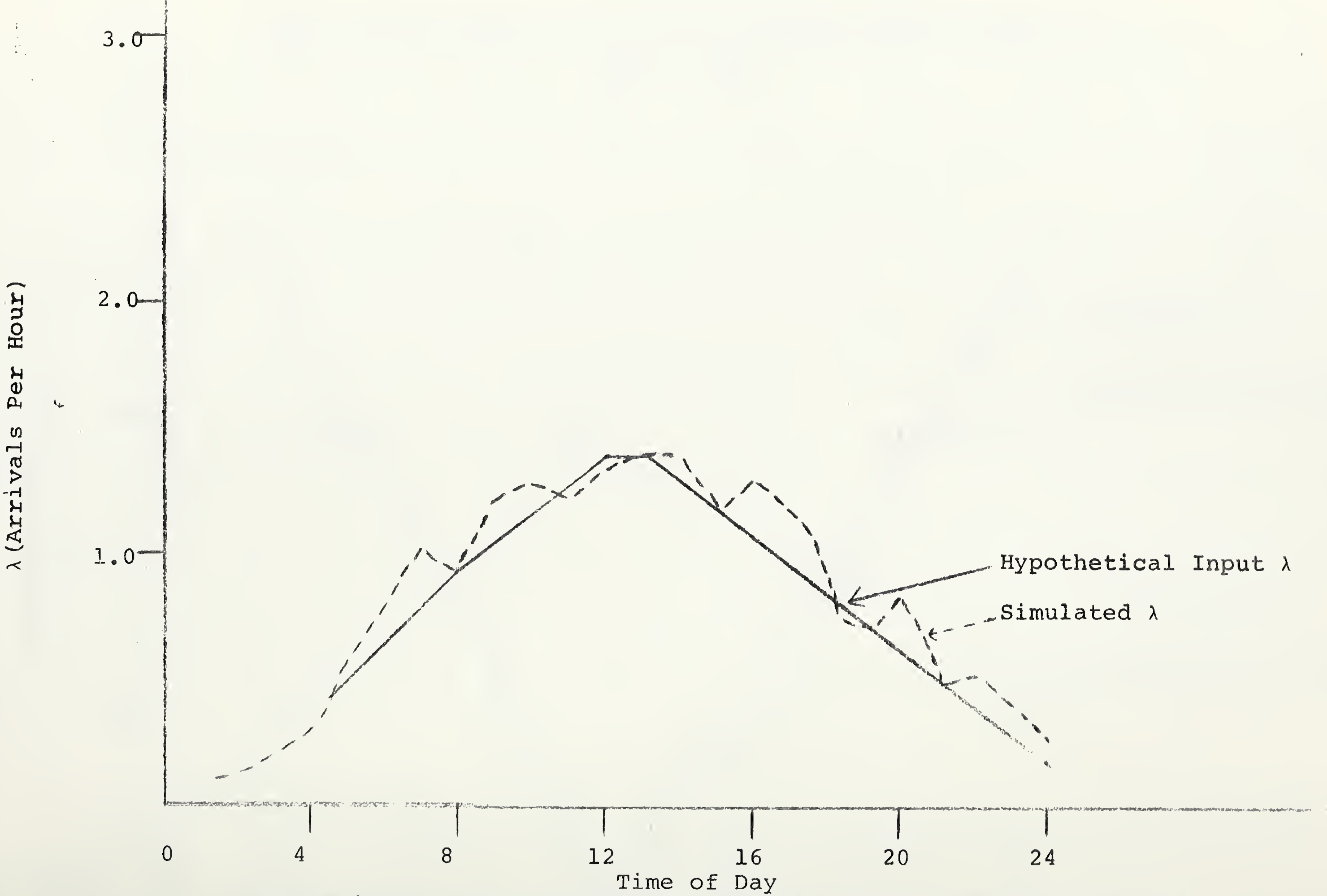
Comparison of Hypothetical and Simulated Hourly Weekday Arrival Rates for Variable Cut-off Method (Random Number Seed 1).

Figure 2.1.3



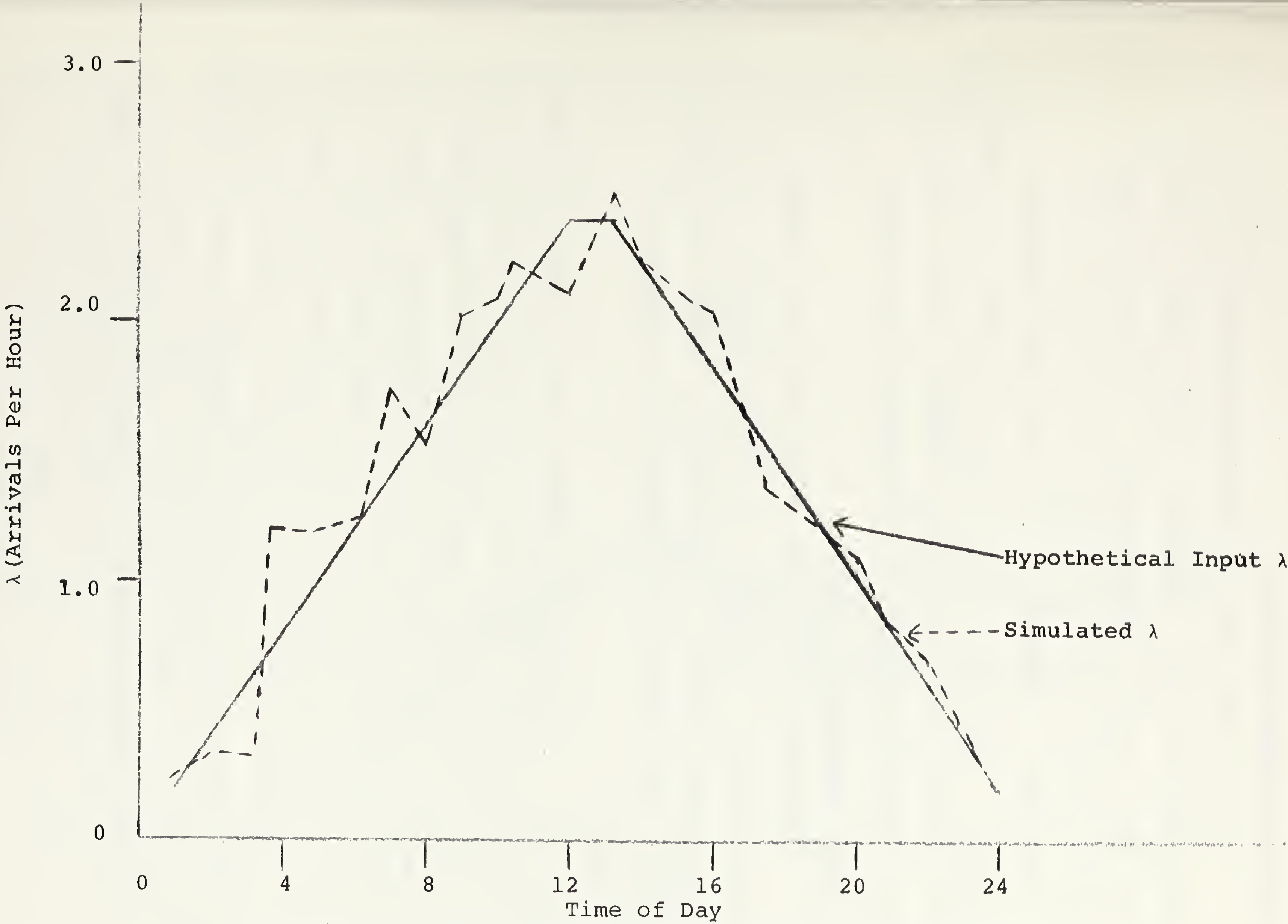
Comparison of Hypothetical and Simulated Hourly Weekend Arrival Rates for Variable Cut-off Method (Random Number Seed 1.)

Figure 2.1.4



Comparison of Hypothetical and Simulated Hourly Weekday Arrival Rates for Variable Cut-off Method (Random Number Seed 2.)

Figure 2.1.5



Comparison of Hypothetical and Simulated Hourly Weekend Arrival Rates for Variable Cut-off Method (Random Number Seed 2).

Figure 2.1.6

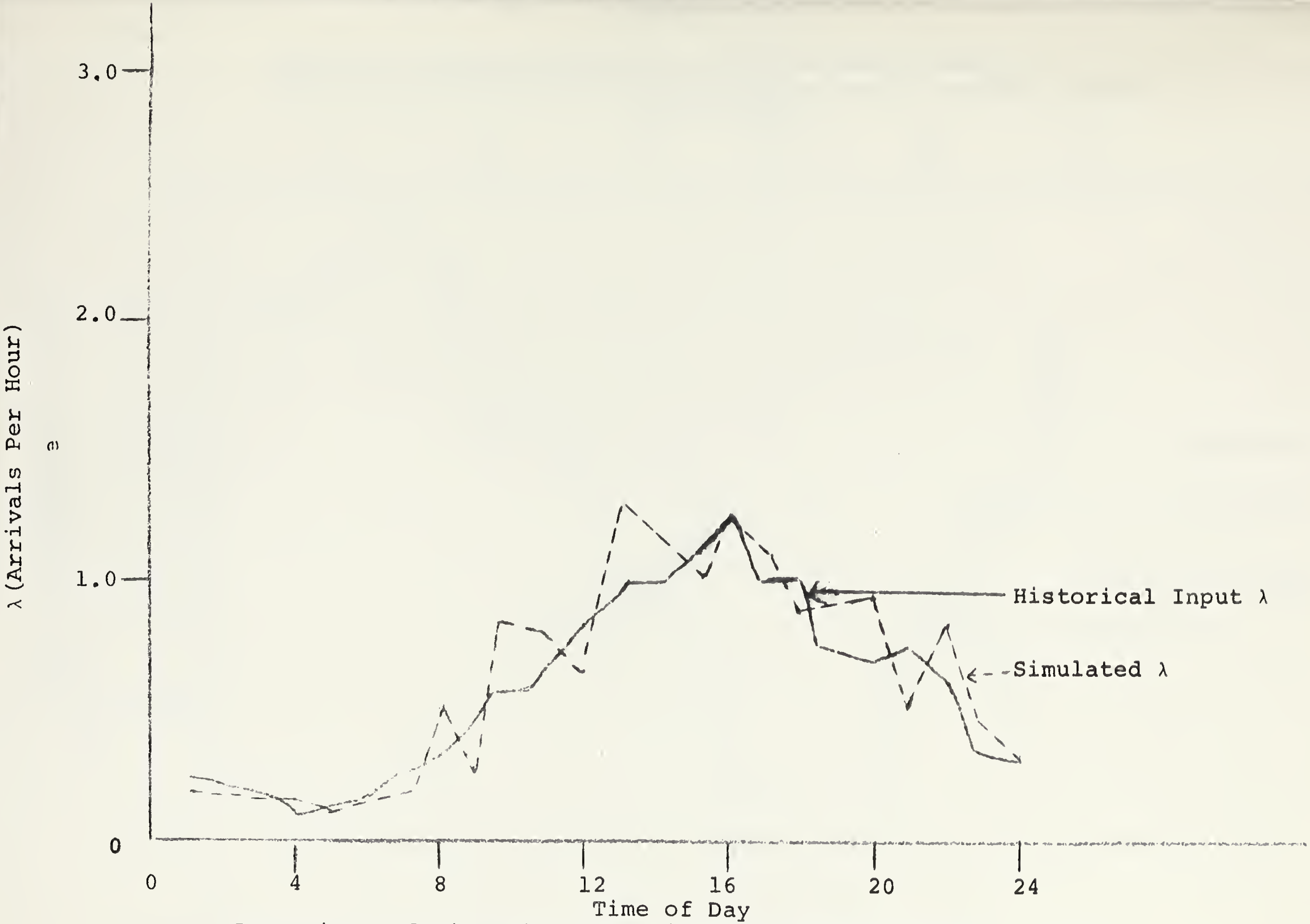
the historical arrival distributions and the corresponding simulated distributions.

Although the fixed cutoff method requires slightly more computing time (because a random sample is drawn at least once per simulated hour), it gives very good results without having to group arrivals into longer (than hourly) intervals. The fixed cutoff method is therefore considered to be a valid algorithm for representing the random arrival mechanism of DEMGEN, and was used for all subsequent DEMGEN runs discussed in this report.

2.1.2 Operational Simulator

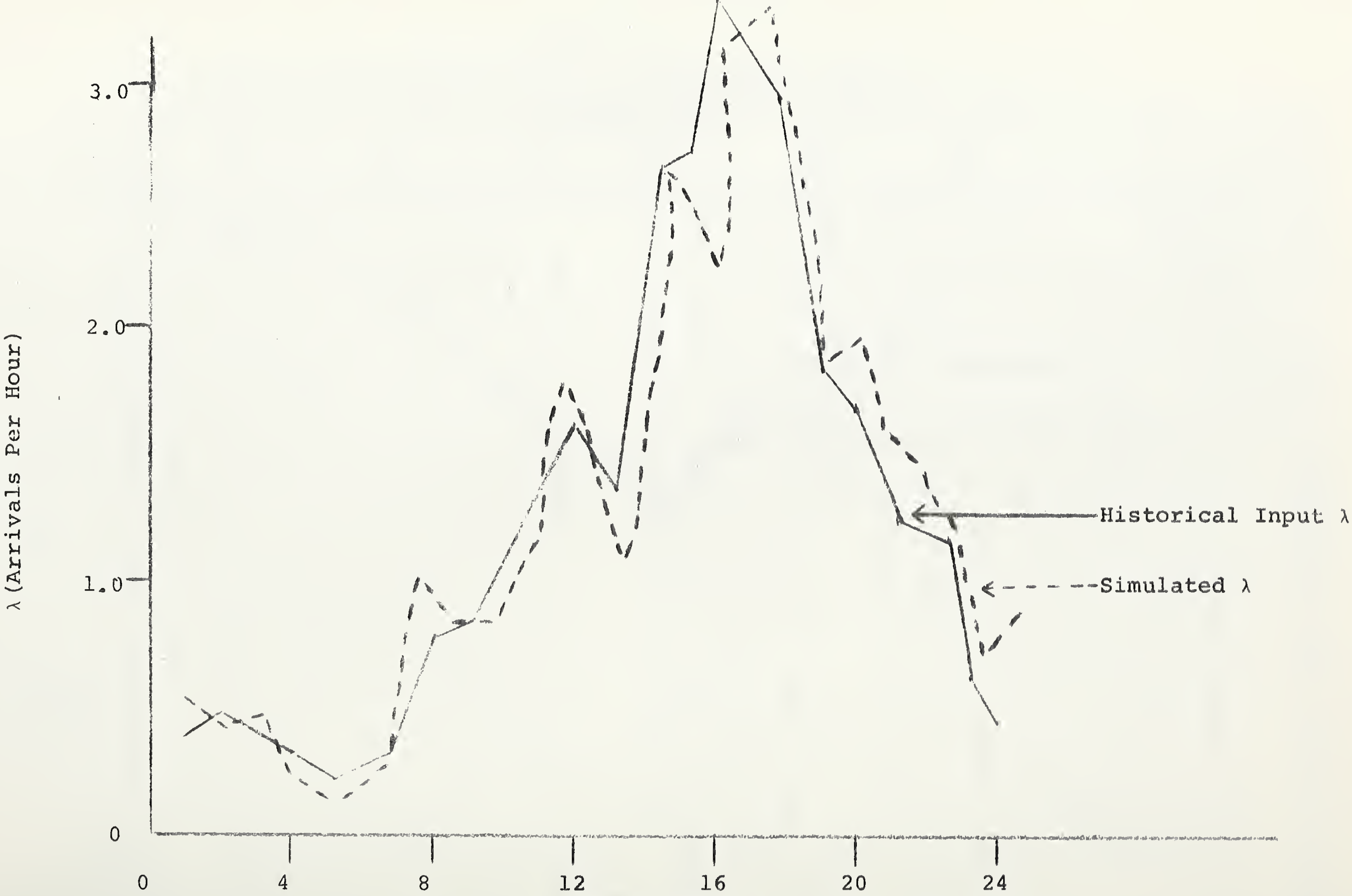
Experience gained with validation runs in Phase III led to several changes in the logical structure of the Operational Simulator (OPSIM) module. These are listed below, with explanations concerning the reasons for change:

(a) Incorporation of delay times for ready resources to get underway. Examination of the statistics on waiting times revealed that the modeling of resource departures immediately after notification was unrealistic. Delay values are now set by the user as attributes of the resource types, and apply only when resources are at their home stations (but not when on patrol or already servicing other cases). Since the resource selection process considers the delay when determining whether a given resource can meet the tolerance time for a case, its inclusion in OPSIM



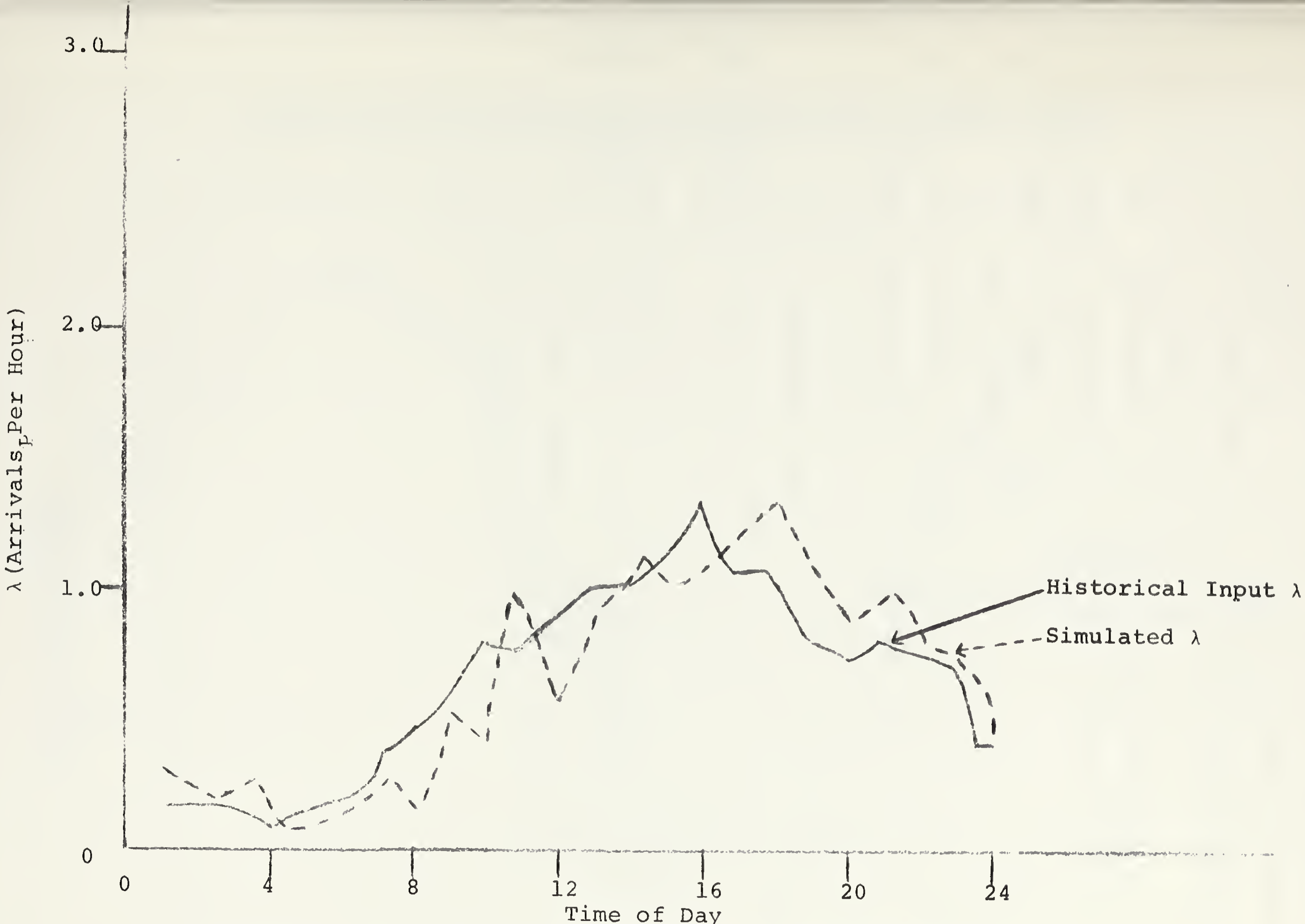
Comparison of Historical and Simulated Hourly Weekday Arrival Rates for Fixed Cutoff Method (Random Number Seed 1).

Figure 2.1.7



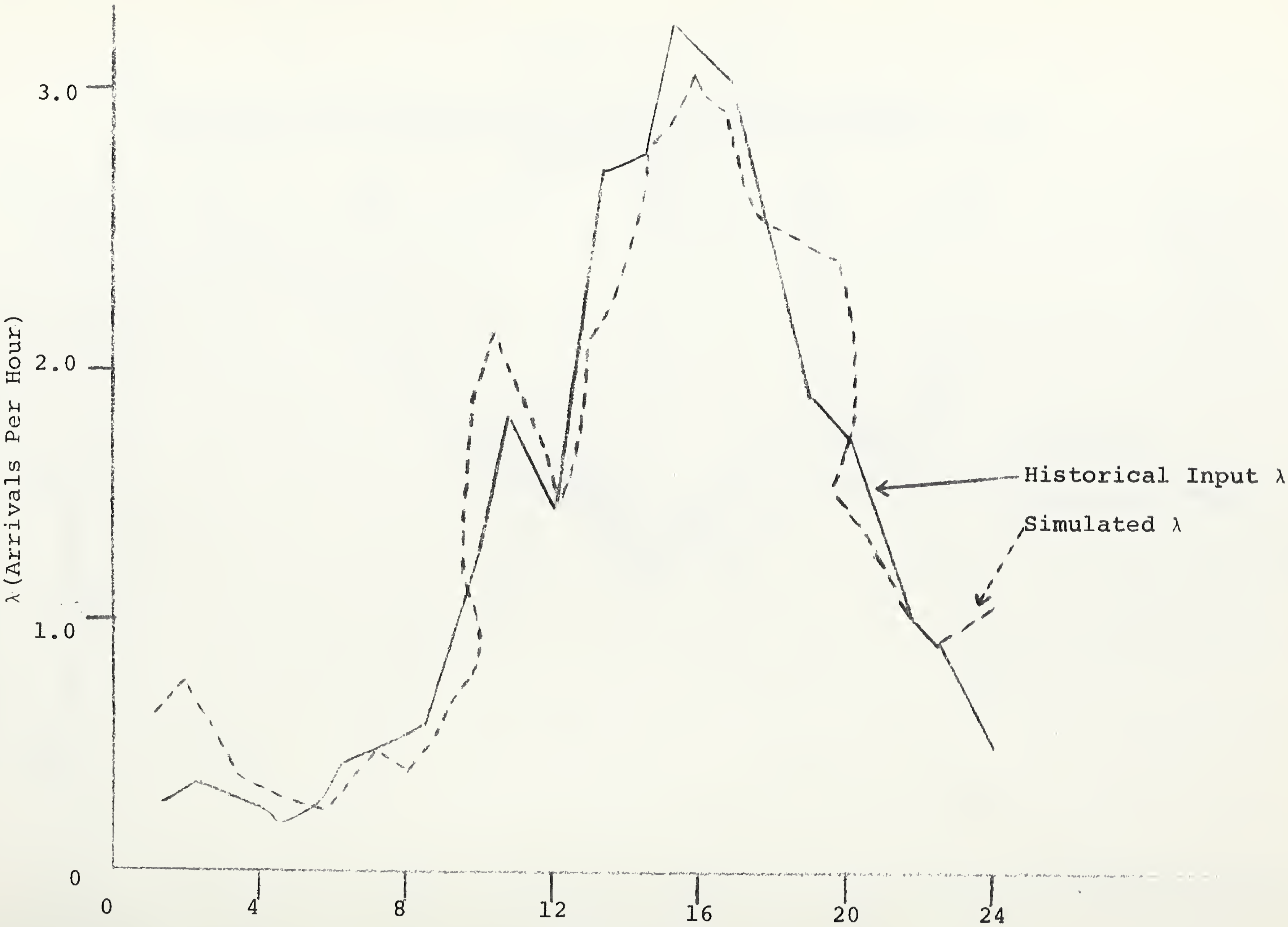
Comparison of Historical and Simulated Hourly Weekend Arrival Rates for Fixed Cutoff Method (Random Number Seed 1).

Figure 2.1.8



Comparison of Historical and Simulated Hourly Weekday Arrival Rates for Fixed Cutoff Method (Random Number Seed 2).

Figure 2.1.9



Comparison of Historical and Simulated Hourly Weekend Arrival Rates for Fixed Cutoff Method (Random Number Seed 2).

Figure 2.1.10

may affect results significantly. (However, consistent with the procedure for calculating historical utilization, delay is not incorporated into the resource utilization statistics produced by OPSIM.)

(b) Holidays are now treated in the same fashion as weekends, with weekday holiday crew levels raised to the (usually higher) weekend levels.

(c) Under some conditions of caseload, it is possible to assign a resource to search at long distance from base, thereby requiring most of the endurance time for transit to and from the scene and leaving only a small percentage of time for performing search on scene. This resource might then take an unreasonably long time to complete the search. Now the logic goes through the resource assignment section of the program each time a resource returns from search for refueling. As a result, the best currently available, capable resource is assigned to continue the remaining search need.

(d) Initial validation runs also evidenced incidents where aircraft and cutters travelled unrealistically long distances to serve cases. This stemmed from the original logic for examining the case queue: an idle resource was assigned to the first queued case which it was capable of serving. The range capabilities of aircraft and cutters generally qualify them to serve a large fraction of the cases which arise throughout the district, leading to a marked tendency to assign them to most queued cases which arise. To reduce such unrestrained simulated use of aircraft and cutters for queued cases, the user may now specify a threshold priority for aircraft or cutter assignment to a queued case. If this priority threshold is not met, a further check is made to determine whether, indeed, there are smaller vessels capable of serving the case. (In addition, if the queued need is for search, any capable smaller vessels must also satisfy an endurance check.) If there are no capable smaller vessels, the available aircraft or cutter is assigned despite the priority. However, if there are capable small vessels, the aircraft or cutter is not assigned, and the case remains in queue until a small vessel becomes idle (i.e., available).

(e) A number of problems concerning queueing were resolved by changes in logic. There were some infrequently-

occurring situations which allowed some cases to remain in queue unrealistically long; in fact, a case once remained in the queue for the entire simulation.

(i) One problem was spotlighted by a multi-resource case for which the first need had been serviced by a resource which was then assigned the responsibility of "covering" until another resource could arrive. When notification of the second need took place, the <u>only</u> resource capable of servicing this need was the resource on scene, which was "busy" covering! This caused the second need to remain in the queue for the entire simulation. The logic has been changed to prevent the recurrence of such situations: if, on notification, only one resource is revealed as being capable of serving the need, a check is made to determine whether that resource is covering on the multi-resource case of concern. If so, the covering resource is assigned to service the need. It may be observed that this change does not solve an extension of the same problem: if the notification for the second need occurs while the only capable resource is in the process of servicing a prior need, the second need is queued and the resource is assigned to cover when the first need has been served. As a further preventive, any resource assigned coverage responsibilities first determines

whether it can cover any queued need of the same case; if capable, it is assigned.

(ii) A related change causes each covering resource to examine the queue every TCHEK hours (a user input) to determine whether it is capable of serving any need in the queue. This relieves problems which can arise when a resource covering on one case is the only one capable of serving another case in the queue.

(iii) On one simulated occasion the non-operational status of the only resource which was capable of serving a given case forced that case to remain queued for a very long time. The modified logic now provides that an attempt be made to serve all queued cases whenever crews are changed.

(iv) Other cases may be relegated to the queue for long periods of time as a result of critical combinations of input data. For example, the option to avoid calling up standby crews may be selected. A case which could be served only by the resources from a single station could enter the system on, say, a Monday. If that station were manned only during one weekend shift, the case would remain in the queue throughout the week while waiting for a crew.

2.2 Statistical Modifications

2.2.1 Utility Program HSTAT

The basic approach in validating SARSIM was to compare simulated output (from OPSIM) with historical statistics. A FORTRAN utility program, called HSTAT, was written to compute the historical values from the historical CASE FILE district tape, which has been processed through programs MUTAPE and MUC130 for proper sorting and merging. (The same input tape is fed to OPSIM, which also requires program PCP to interrelate data, and either DEMGEN or HIST to select the cases for the time period in question.)

Since HSTAT was designed to be a simple utility program for validation purposes only, sophisticated switching options to designate the district, time period, resources, etc., were not incorporated. Approximately 14 data card changes and 10 to 20 coding changes are necessary for this purpose. In other words, HSTAT is not user-oriented, but should be run only by an experienced programmer. However, since other documentation does not exist for this utility program, a list of variables is given to assist the reader interested in operating HSTAT.

A listing of the required input variables follows:

NS = NUMBER OF STATIONS IN DISTRICT

NG = NUMBER OF GROUPS IN DISTRICT

PCT = UPPER LIMITS FOR CASE COUNT DISTRIBUTIONS

JPP = RESOURCE CONVERSION TABLE

IRG = BOAT(1) OR CUTTER(2) DESIGNATOR FOR FIRST 12 RESOURCE TYPES

IOP = OPFACT NUMBERS

NREST = NUMBER OF RESOURCES AT EACH OPFAC

NRES = NUMBER OF RESOURCES OF EACH TYPE

ISAR = SARSIN STATION NUMBER FOE EACH OPFAC

IRPG = NUMBER OF RESOURCES FOR EACH GROUP

IGP = GROUP NUMBER FOR EACH OPFAC

ISSN = WINTER(2) OR SUMMER(0) DESIGNATOR

INTNRS = TOTAL NUMBER OF RESOURCES IN DISTRICT

NOB = TOTAL NUMBER OF BOATS IN DISTRICT

NOC = TOTAL NUMBER OF CUTTERS IN DISTRICT

NWD = NUMBER OF WEEKDAYS IN PERIOD

NWE = NUMBER OF WEEKEND DAYS AND HOLIDAYS IN PERIOD

NDAYS = TOTAL NUMBER OF DAYS IN PERIOD

A listing of the output variables (in order of appearance) follows:

ITNC = TOTAL NUMBER OF CASES

ITFL = TOTAL NUMBER OF CASES WITH TYPE C FAILURE

NDAYS = TOTAL SIMULATED TIME

OVUTL = OVERALL AVERAGE UTILIZATION PER RESOURCE

UTLSH = AVERAGE UTILIZATION BY SHIFT

UTLR = AVERAGE UTILIZATION BY RESOURCE TYPES

URG(1) = AVERAGE UTILIZATION OF BOATS

URG(2) = AVERAGE UTILIZATION OF CUTTERS

ISAR = SARSIM STATION NUMBER

IOP = OPFAC NUMBER

NUMCAS = NUMBER OF CASES FOR EACH OPFAC

NOR = NUMBER OF NEEDS FOR EACH OPFAC

IFAILS = NUMBER OF TYPE C FAILURES FOR EACH OPFAC

AVST = AVERAGE WIAT(HRS.) PER CASE FOR EACH OPFAC

AVTV = AVERAGE TVEC (TIME TO VECTOR)(HRS.) PER CASE FOR EACH OPFAC

UTLST = AVERAGE UTILIZATION PER RESOURCE AT EACH OPFAC

VTM = AVERAGE WAIT - TOLERANCE (HRS.) PER CASE FOR EACH OPFAC

ITNN = TOTAL NUMBER OF NEEDS

TAVWT = OVERALL AVERAGE WAIT (HRS.)

TTVC = OVERALL AVERAGE TVEC (HRS.)

TAUTMT = OVERALL AVERAGE WAIT - TOL (HRS.)

UGP = AVERAGE UTILIZATION BY GROUP

TUD = WEEKDAY MEAN UTILIZATION

SDD = WEEKDAY STANDARD DEVIATION FOR UTILIZATION

CNTD = WEEKDAY CASE COUNT FOR UTILIZATION DISTRIBUTION

TWTD = TWAIT MEAN

SDWD = TWAIT STD. DEV. — WEEKDAYS

CNT4 = TWAIT CASE COUNT

TWTE = TWAIT MEAN

SDWE = TWAIT STD. DEV. — WEEKENDS

CNT1 = TWAIT CASE COUNT

TTVCD	=	TVEC MEAN	
SDVD	=	TVEC STD. DEV	WEEKDAYS
CNT5	=	TVEC CASE COUNT	
TTVCE	=	TVEC MEAN	
SDVE	=	TVEC STD. DEV.	WEEKENDS
CNT2	=	TVEC CASE COUNT	
TTMTD	=	TWAIT - TOL MEAN	
SDMD	=	TWAIT - TOL STD. DEV.	WEEKDAYS
CNT0	=	TWAIT - TOL CASE COUNT	
TTMTE	=	TWAIT - TOL MEAN	
SDME	=	TWAIT - TOL STD. DEV.	WEEKENDS
CNT3	=	TWAIT - TOL CASE COUNT	
TSRVE	=	SERVICE TIME MEAN	
SDSD	=	SERVICE TIME STD. DEV.	WEEKDAYS
CNT9	=	SERVICE TIME CASE COUNT	
TSRVE	=	SERVICE TIME MEAN	
SDSE	=	SERVICE TIME STD. DEV.	WEEKENDS
CNT7	=	SERVICE TIME CASE COUNT	
TRTD	=	RESOURCE USE MEAN	
SDRD	=	RESOURCE USE STD. DEV.	WEEKDAYS
CNT10	=	RESOURCE USE CASE COUNT	
TRTE	=	RESOURCE USE MEAN	
SDRE	=	RESOURCE USE STD. DEV.	WEEKENDS
CNT8	=	RESOURCE USE CASE COUNT	

Except for the few differences mentioned below, the statistics generated and formatted in HSTAT are similar to those of the OPSIM standard output:

(a) Failure A, Failure B, standbys called, standbys not used, number of queued cases, total interrupted needs, and normalized figure of merit are not available historically; hence cannot be calculated in HSTAT;

(b) Under "Utilization by Resource Types" in HSTAT, types 17 and 18 represent miscellaneous other boats and other aircraft respectively; that is, those not included in types 1-16.

(c) "Utilization by Shifts" is computed differently. In OPSIM, the time spent on a case is partitioned among all shifts during which resources are underway; in HSTAT, it is all allocated to the time period (shift) when the case began since more detailed data cannot be ascertained. <u>As a result, the historical utilization indices for shifts are not directly comparable with the corresponding indices from OPSIM</u>.

2.2.2 OPSIM Statistical Changes

In order to provide a more efficient means of obtaining validation data, statistical output was added to OPSIM during Phase III. (Details of calculation of output statistics are provided in Volumes II and IV of the SARSIM documentation.) To obtain the same results from Quick Query in the Postprocessor would take longer and be more costly.

(a) Under District statistics, Section II, utilization indices are presented for all cutters combined, all C-130's, and for all other aircraft combined. This allows the user to group resource types as desired and to examine aggregated utilization indices.

(b) Additional columns in Station Response, Section III, indicate the number of instances of queueing, the average time to vector (TVEC), average number of Type C failures, average time a case has to wait for service in excess of specified tolerance (TWAIT - TOL when non-negative), and a normalized figure of merit* for each station.

(c) A section was added for Group Response, showing combined statistics for stations within a user-defined group.

---

*See Appendix A for discussion.

(d) In Resource Utilization, Section IV, the number of needs to which each resource was assigned is listed alongside the utlization index for that resource.

(e) Distributions, Section VI, was added to provide histograms as aids to validation and inputs for various "goodness-of-fit" statistical tests. The daily utilization indices overall, for both weekdays and weekends, are output, as are the distributions of various critical case attributes, such as TWAIT, TVEC, TOL, TSVC, and Figure of Merit (all for both weekdays and weekends).

2.3 Calibration Results

2.3.1 Calibration Process

To provide insight into the procedures for calibrating and validating a complex model such as SARSIM, it may be useful to consider first a simpler example. Consider, for instance, a formula for estimating an individual's weight, w(a dependent variable), as a linear function of his height, h (an independent variable), that is,

$$w = ah + b. \qquad (1)$$

Before one can comment on the validity of this "model," the parameters a and b must first be specified. If a vector of heights

$$\underline{h} = (h_1, h_2, \dots h_n)$$

and the corresponding weights

$$\underline{w} = (w_1, w_2, \dots w_n)$$

are known for a sample set of n individuals, then the parameters a_o and b_o can be determined by using linear regression, provided that $n \geq 2$ (2 being the number of parameters to be estimated). The estimation equations are:

$$a_o = \frac{n \sum_{i=1}^{n} w_i h_i - (\sum_{i=1}^{n} w_i)(\sum_{i=1}^{n} h_i)}{n \sum_{i=1}^{n} h_i^2 - (\sum_{i=1} h_i)^2} \qquad (2)$$

and

$$b_o = \overline{w} - a_o \overline{h} \qquad (3)$$

where $\overline{h}$ and $\overline{w}$ indicate the mean values of the vectors $\underline{h}$ and $\underline{w}$, respectively.

Furthermore, this technique of linear regression readily permits estimation of how well the calibrated model describes (or "fits") the data on which it was calibrated. These estimates are initial indications of the model's validity.

Consider, next, an individual of height and weight w_{n+1}, that is to say, an individual not included in the original sample. Using Equation (1), one might compute this individual's weight from his height:

$$\hat{w}_{n+1} = a_o h_{n+1} + b_o$$

and calculate the error, ε, in this estimate as

$$\begin{aligned} \varepsilon &= \hat{w}_{n+1} - w_{n+1} \\ &= a_o h_{n+1} + b_o - w_{n+1} \end{aligned}$$

If ε is small relative to w_{n+1}, one would lean toward accepting the model as valid. On the other hand, large errors observed in a verification process would suggest that either (a) the model is invalid (that is, that the relationship between weight and height is not really linear); or (b) that the individual(s) used for the validation process was selected from a significantly different population than those used for calibration. For example, calibration parameters determined for male football players would hardly be expected to produce good estimates of the weights of female typists.

In the latter case (i.e., with improperly calibrated parameters),it is usually a straightforward matter to obtain additional data points, representative of the population for which the model is to be used, then to proceed with a re-calibration and revalidation. However, if the model itself proves to be invalid, one is faced with reformulating the relationship between height and weight and, perhaps, even adding other independent variables, such as age, sex, nationality, etc. A series of changes may be required, with the procedures of calibration and validation repeated until the reformulation becomes acceptable.

The process of calibrating and validating a model like SARSIM is analogous, but considerably more complex. First of all, there is no single dependent variable (such as weight) to describe the model. Rather, there is a set of statistics of interest, such as utilization of resources,average service time, figure of merit, waiting times, etc., and it is by no means obvious whether one should be willing to accept less close approximations to one variable in an effort to improve estimates of another.

Second, the number of independent variables and associated parameters is an order of magnitude larger than our example. Hence to obtain accurate parameter estimates, a much larger amount of sample data must be available. Compounding this problem is the fact that each data point requires the use of one to three months' worth of Coast Guard district performance data and a considerable amount of time and effort for preparation. As a result, the limited resources of time, money and data restricted the amount of calibration which could be performed.

Thirdly, SARSIM is a stochastic and highly nonlinear simulation model, rather than a closed form equation like (1). This implies that relationships cannot be inverted to provide closed form solutions of parameters such as (2) and (3). Instead, a technique of making educated guesses for the parameters, testing sample data

in the model observing results and (based on these results) revising parameter estimates had to be used.

The procedure proceeded essentially as follows:

(1) A particular data sample was selected using the criteria specified in Section 1.3. This might be, for instance, District 1 during July 1968. Parameters were extracted from Coast Guard records which described operations, such as number, location and costs of resource allocation policies and delay times required for resources to get ready to depart.

(2) The computer program HSTAT, described in Section 2.2.1, was exercised using caseloads of the period chosen to compute dependent variables such as utilization indices and service times.

(3) The simulation model was exercised using the same caseloads as HSTAT and the parameters described in (1).

(4) The dependent variables from the simulation run were compared to the dependent variables from HSTAT to measure the error in SARSIM predictions. If agreement was acceptable, the calibration procedure appeared to be completed and one could proceed to paragraph (7).

(5) If the disagreement was not small and appeared to be caused by an error in the model's logic, the model was altered to perform more like the process being simulated. Examples of these changes are contained in Section 2.1. After alterations, one would return to paragraph (3).

(6) If, alternatively, disagreement seemed to be caused by poor parameter values, new estimates were made and the model was exercised again as in paragraph (3). This interactive procedure continued until acceptable simulation outputs were obtained.

(7) At this point, the calibrated model would be checked for validity by rerunning both the calibrated model and DEMGEN on a similar data set (e.g., District 1 in August 1968, District 1 in July 1967, or a randomly generated caseload for District 1 which was condidered representative of a peak period month). If output proved unsatisfactory, model logic and/or parameter estimates were again reevaluated.

(8) A new data sample was selected, such as a different district or a non-peak period. Hopefully, parameters already calibrated would provide acceptable results for this new scenario. If in fact they did not, the procedure would be initiated again as described in paragraph (1). Such recalibrating was in fact necessary because of geographical differences in SARSIM performance.

2.3.2 Presentation of Graphical Results

It is not a simple matter to choose appropriate tests for demonstrating the validity of the SARSIM model. Most standard statistical tests which are used for hypothesis testing (e.g., goodness-of-fit tests) require either very large sample sizes or numerous re-runs of the model under the same input conditions to provide enough data for only a single test. For example, it may be recalled from Section 1.3 that the overall utilization index is the best single validation criterion, and that a month's worth of cases is about the minimum sample size adequate for the purpose. One simulation run yields only one overall utilization value, hence it might cost about $750 (and considerable time) to get a sample size of, say, 15, using different random number seeds for each run, thus permitting a single statistical test for one scenario with one set of operating conditions.

Furthermore, no one type of statistical test is sufficient for validation, for each has its own advantages and limitations. A given test applies to a single facet of the validation of a complex system. In addition, any statistical test is subject to some probability (which however, may be kept small) that acceptance or rejection of the hypothesis will be made in error.

It was therefore decided, in consultation with Dr. Joan R. Rosenblatt, Chief of the Statistical Engineering Laboratory at NBS, that several methods should be combined for presenting evidence of the validity of SARSIM,

with emphasis on graphical presentation, rather than sole reliance on statistical testing techniques. Graphical evidence is presented in the remainder of this section, with additional evidence put forth in Sections 2.3.4, 2.4, 3.1 and 3.2.

There are eight sets of graphs to represent significant results for three Coast Guard Districts (the First, Seventh and Thirteenth); three kinds of exercises (to wit, calibration, validation, and comparison between HIST and DEMGEN); and three time periods (July 1968, August 1968, and Winter 1969). The eight sets, which do not exhaust the possible combinations of elements represented, are identified in Table 2.3.2.1 below.

Table 2.3.2.1

SET	DISTRICT	EXERCISES	TIME PERIOD	RUN NUMBER	CODE
1	01	calibration	7/68	45	01PIHR
2	01	validation	8/68	64	01PIHU
3	01	calibration	W/69	61	01N3HB
4	01	HIST vs DEMGEN	7/68	*	**
5	07	calibration	7/68	80	07PIHA
6	07	validation	8/68	83	07PIHB
7	13	calibration	7/68	60	13PIHE
8	13	validation	8/68	72	13PIHH

* Run 42 vs. Runs (62 and 63) Averaged.

**Codes 01PHIP, 01PIDA and 01PIDB.

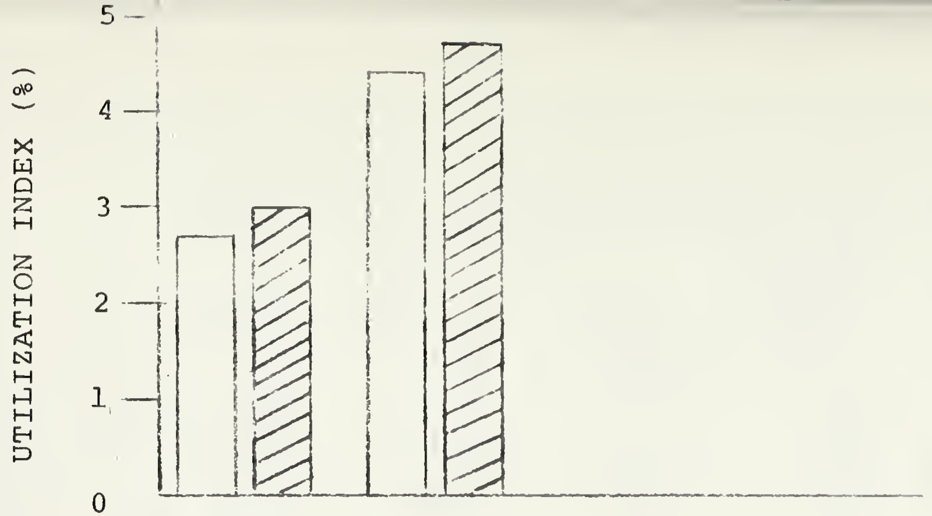
As shown, there was a calibration run to adjust the parameters as required for each district separately, using the appropriate July 1968 caseload scenario. A validation run then used the August 1968 scenario with parameters as established in the calibration run for the district. A non-peak (Winter 1969) run is presented for the First District, as well as a "Hist vs DEMGEN" exercise. This latter exercise was intended to verify that the selection of the July 1968 caseload was not a critical factor in these experiments. The adequacy of fit seen in the graphs of Set 4 indicates that any peak period month with approximately the same number of cases would likely provide similar results. This exercise also supplies evidence that SARSIM is not limited to the specific cases which actually occurred and which are listed on the CASE TAPE. The success achieved with a randomly generated caseload indicates that SARSIM adequately adapts to departures from fixed situations.

Within each of the eight sets there are eight graphs, labelled A through H, which compare simulated and historical outputs as listed in Table 2.3.2.2 below.

Table 2.3.2.2

Types of Graphs

GRAPH	REPRESENTATION
A	Average Utilization Indices for Weekdays (WD) and Weekends (WE).
B	Average WD and WE values for TWAIT, TVEC and TSVC.
C	Combined Resource Type and Overall Utilization Indices.
D	Frequency distributions for vector time (TVEC).
E	Frequency distributions for waiting time (TWAIT).
F	Frequency distribution for service time (TSVC).
G	Average Utilization Indices by Shift
H	Combined Station Group Utilization Indices



Comparison of historical and simulated utilization indices for First District in July 1968 (Run 45).

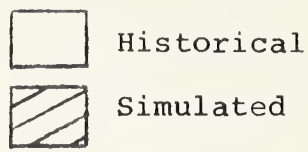
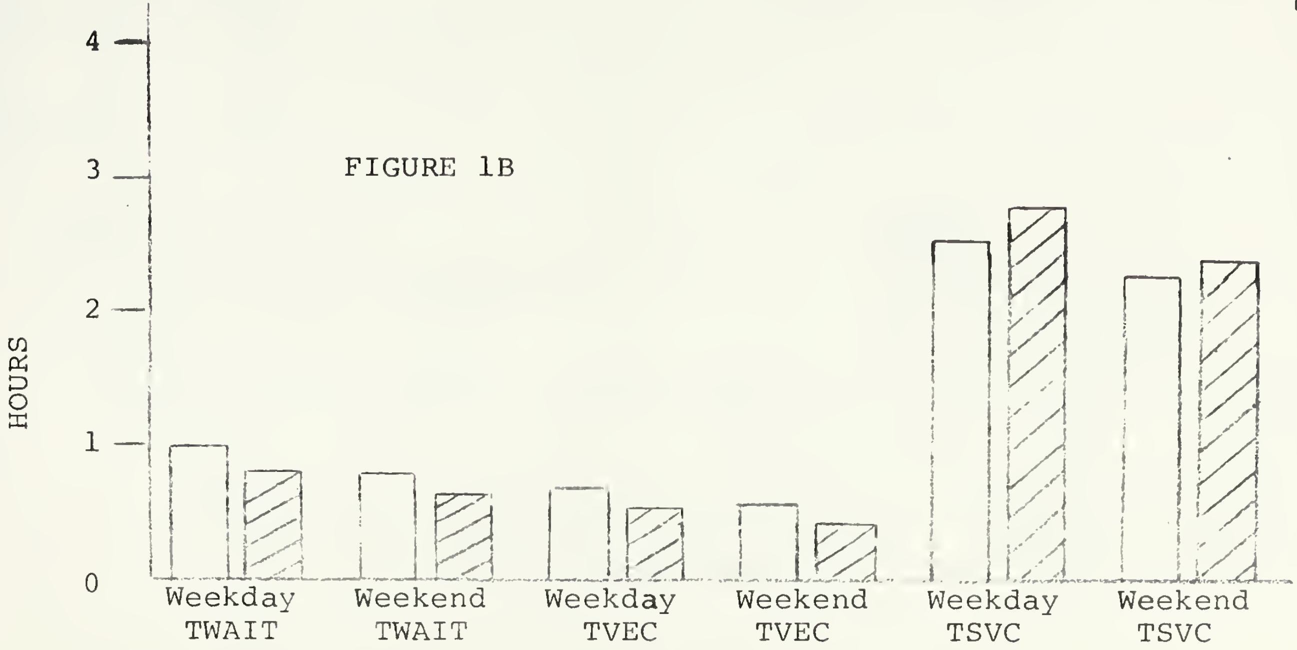
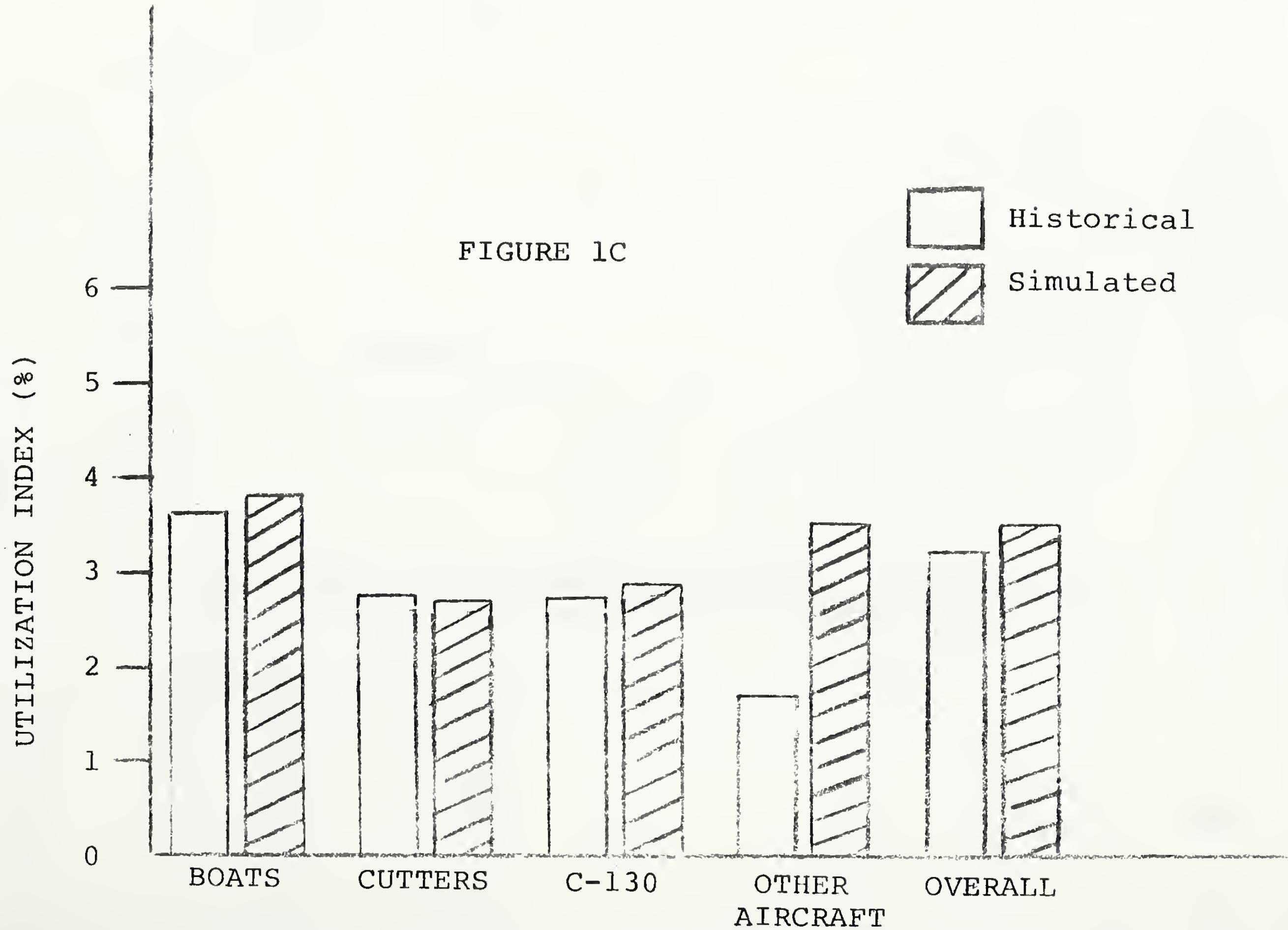


FIGURE 1B



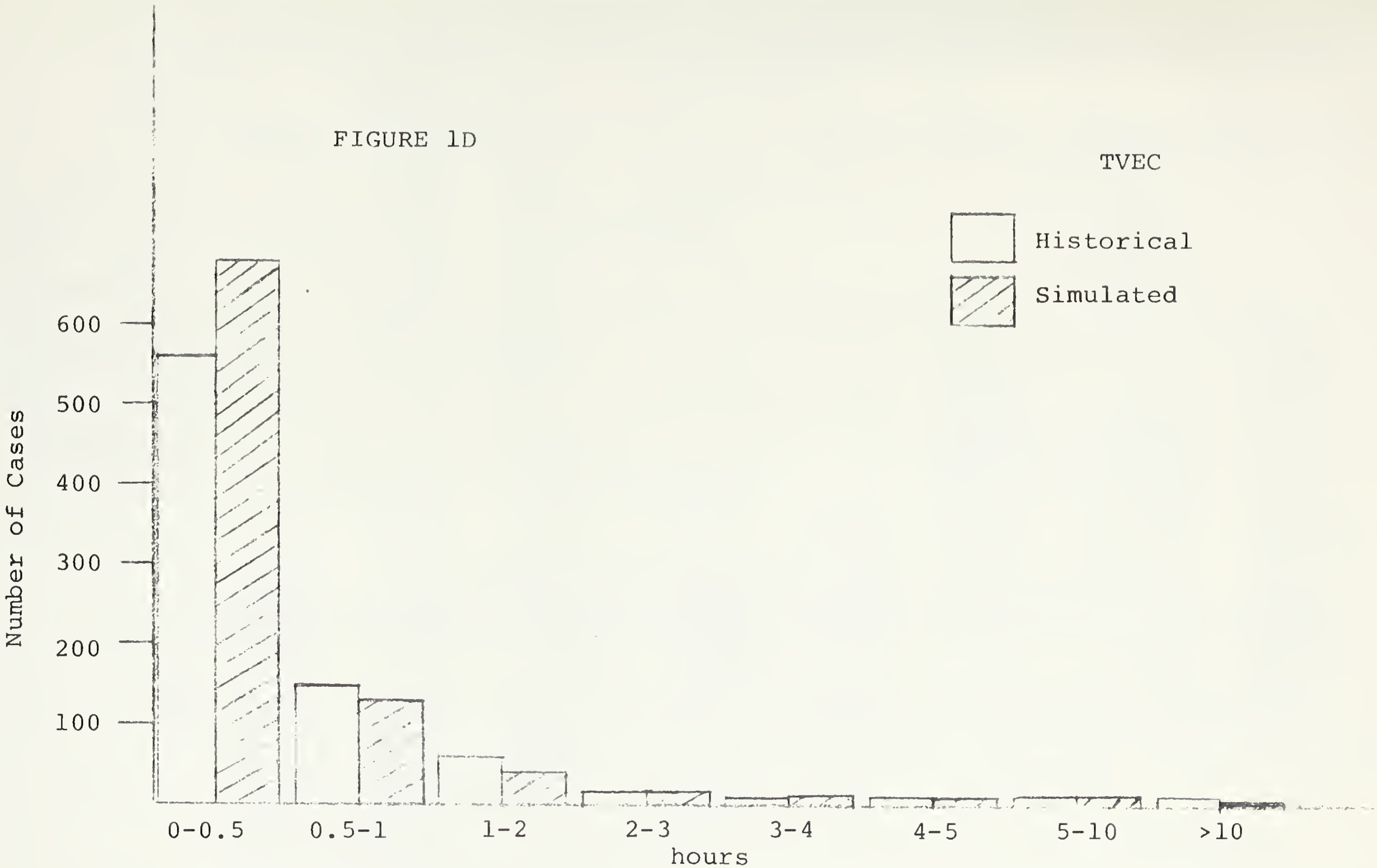
Comparison of historical and simulated output statistics for First District in July 1968 (Run 45).

FIGURE 1C



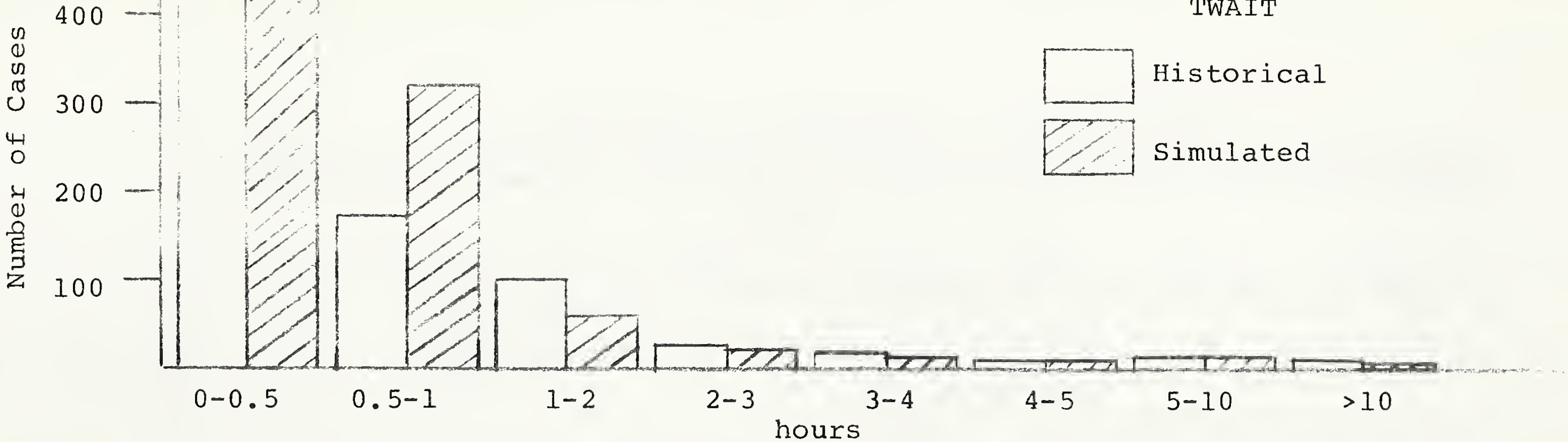
Comparison of combined Resource-Type Utilization Indices for First District in July 1968 (Run 45).

FIGURE 1D



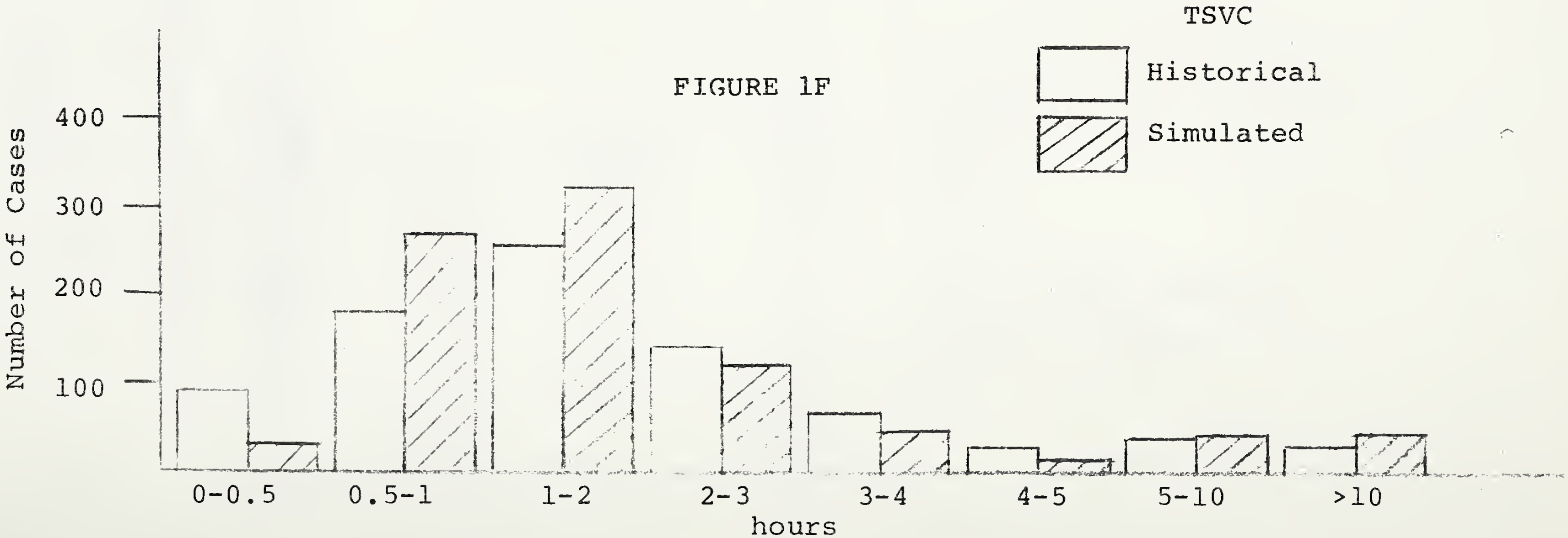
Comparison of numbers of historical and simulated cases for which vector times were within the indicated intervals for First District, July 1968. (Inequality of intervals should be noted.)

FIGURE 1E



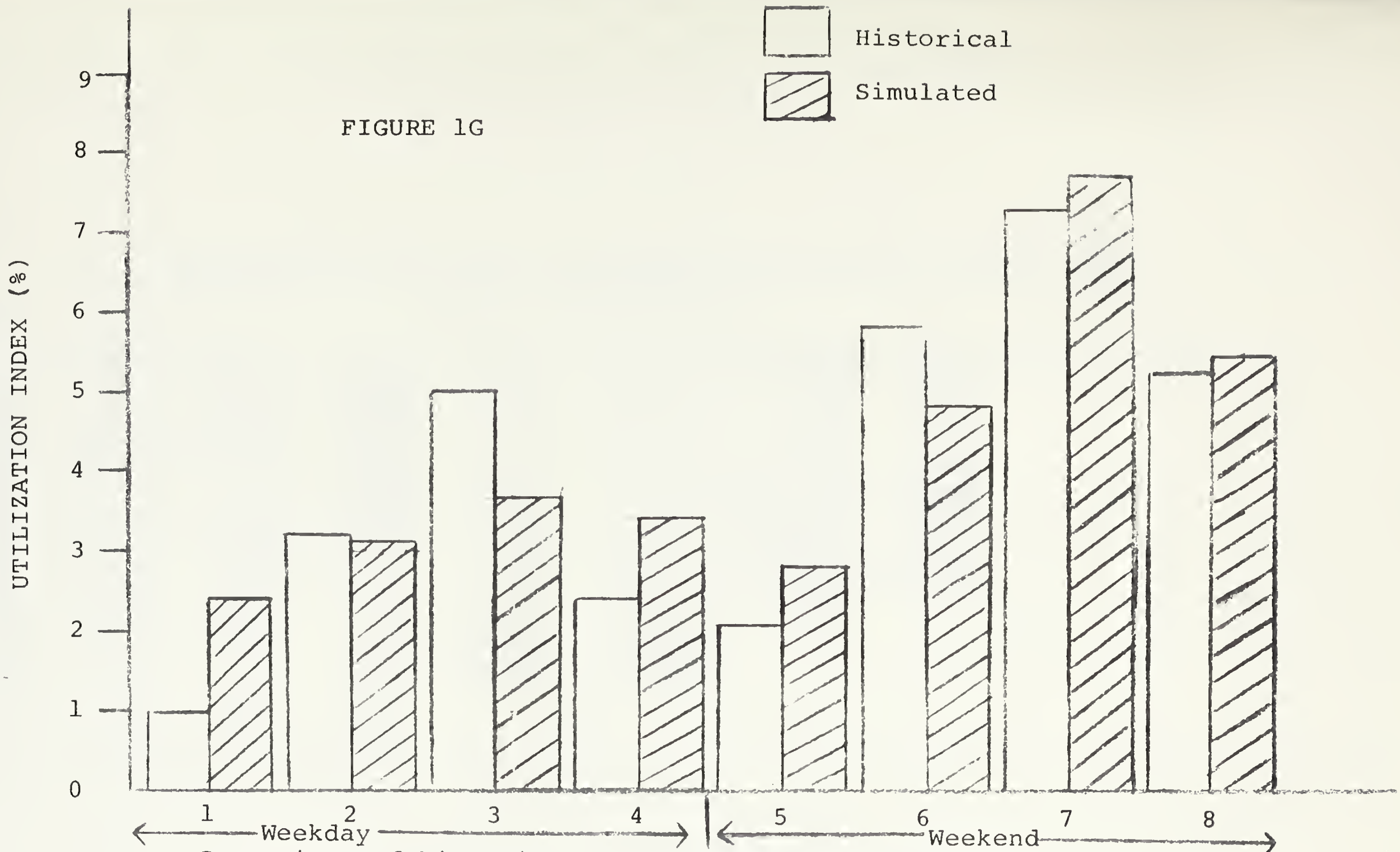
Comparison of numbers of historical and simulated cases for which waiting times were within the indicated intervals for First District, July 1968. (Inequality of intervals should be noted.)

FIGURE 1F



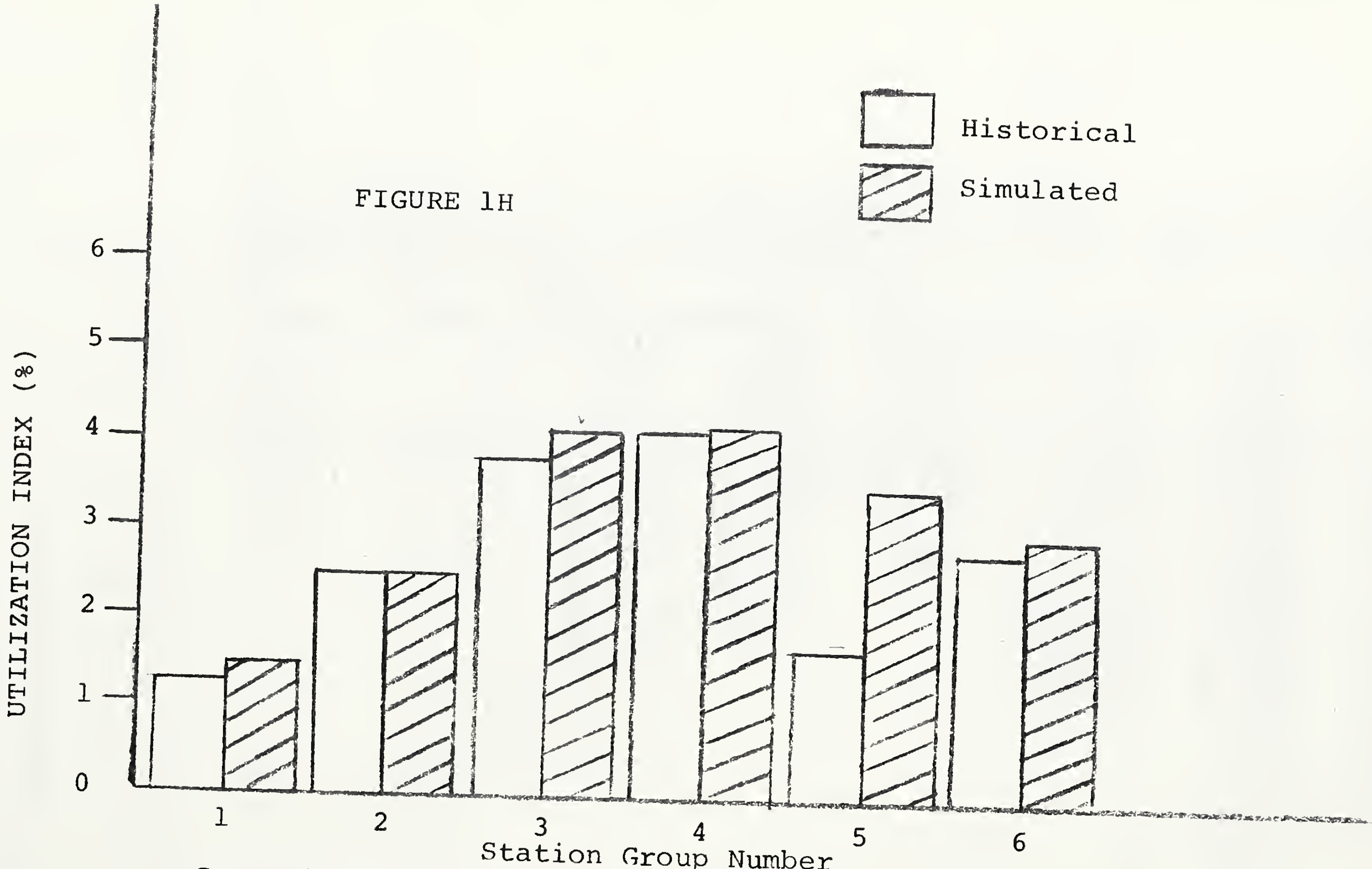
Comparison of numbers of historical and simulated cases for which service times were within the indicated intervals for First District, July 1968. (Inequality of intervals should be noted.)

FIGURE 1G



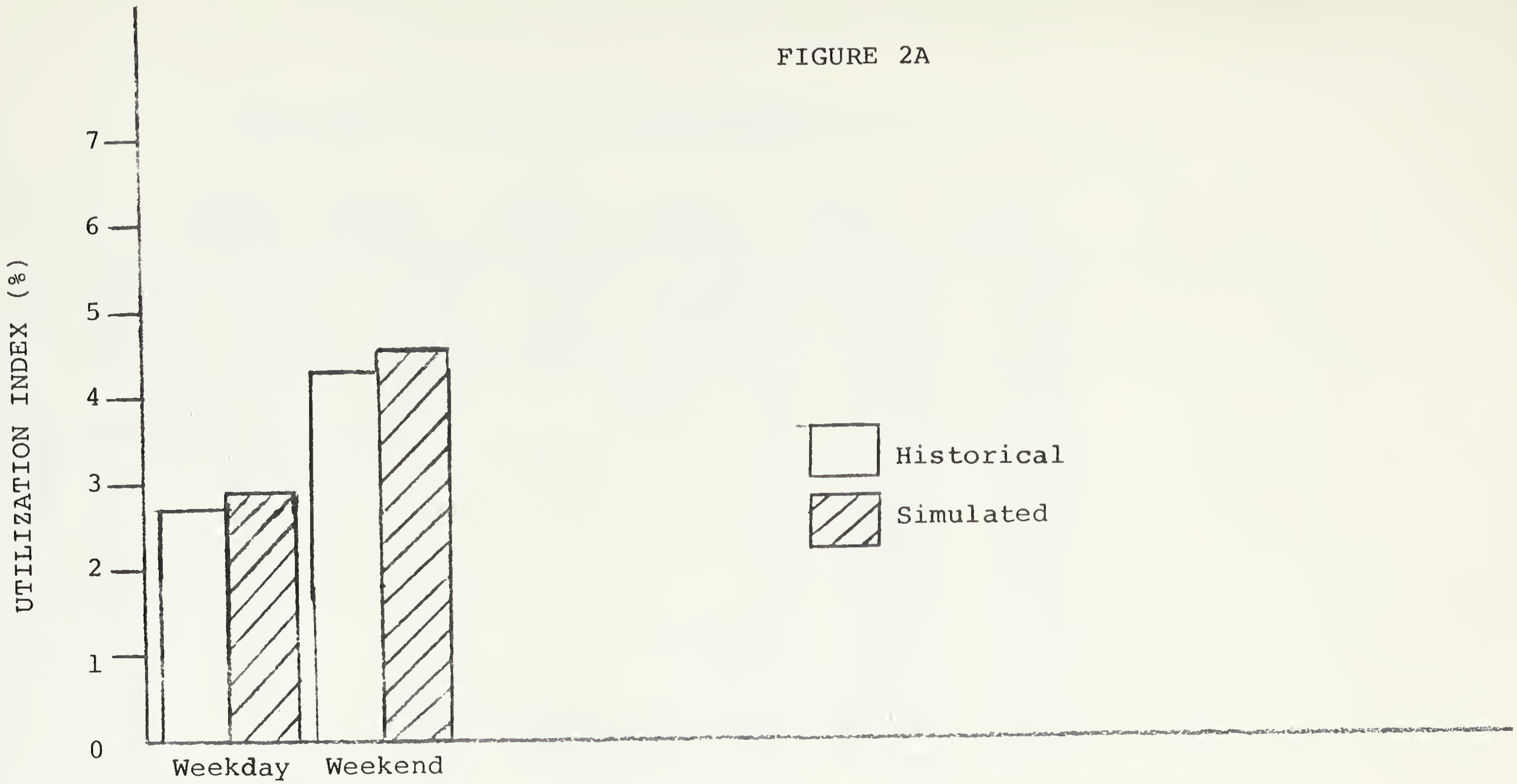
Comparison of historical and simulated utilization indices by shift for First District in July 1968 (Run 45). (Note: see comment in text regarding inapplicability of historical data in this context.

FIGURE 1H



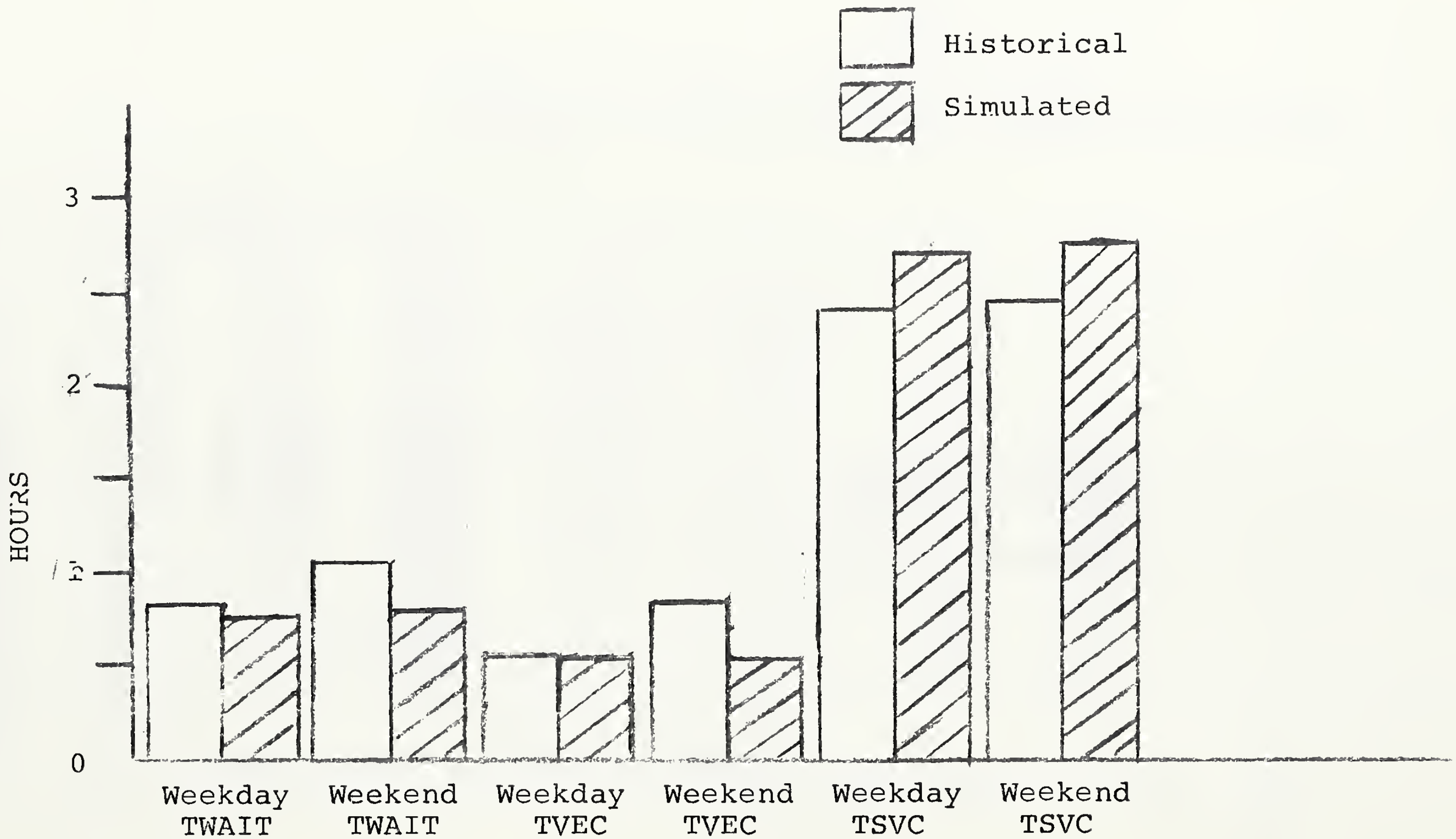
Comparison of historical and simulated combined station group utilization indices for First District in July 1968 (Run 45).

FIGURE 2A



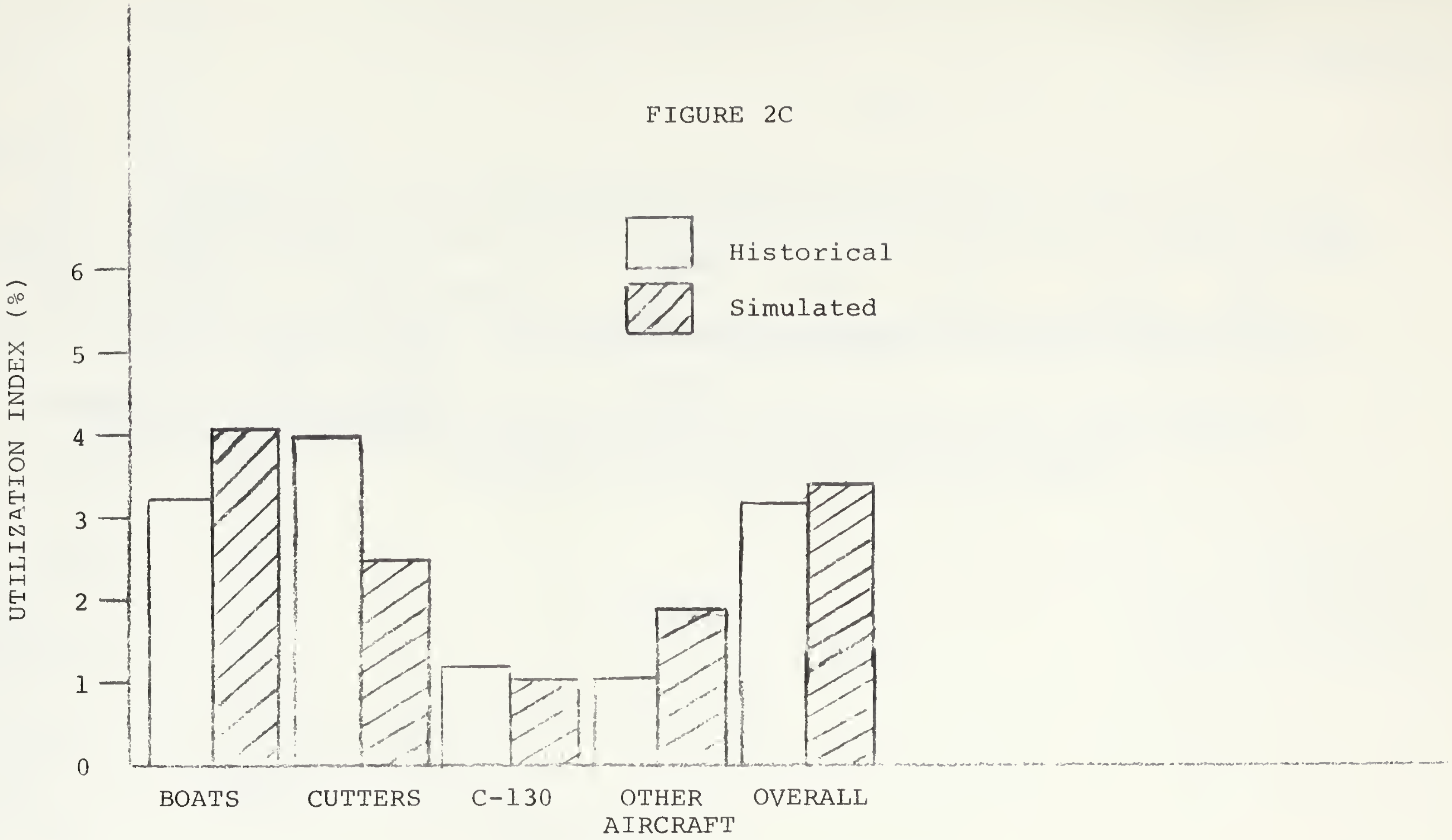
Comparison of historical and simulated utilization indices for First District in August 1968 (Run 64).

FIGURE 2B



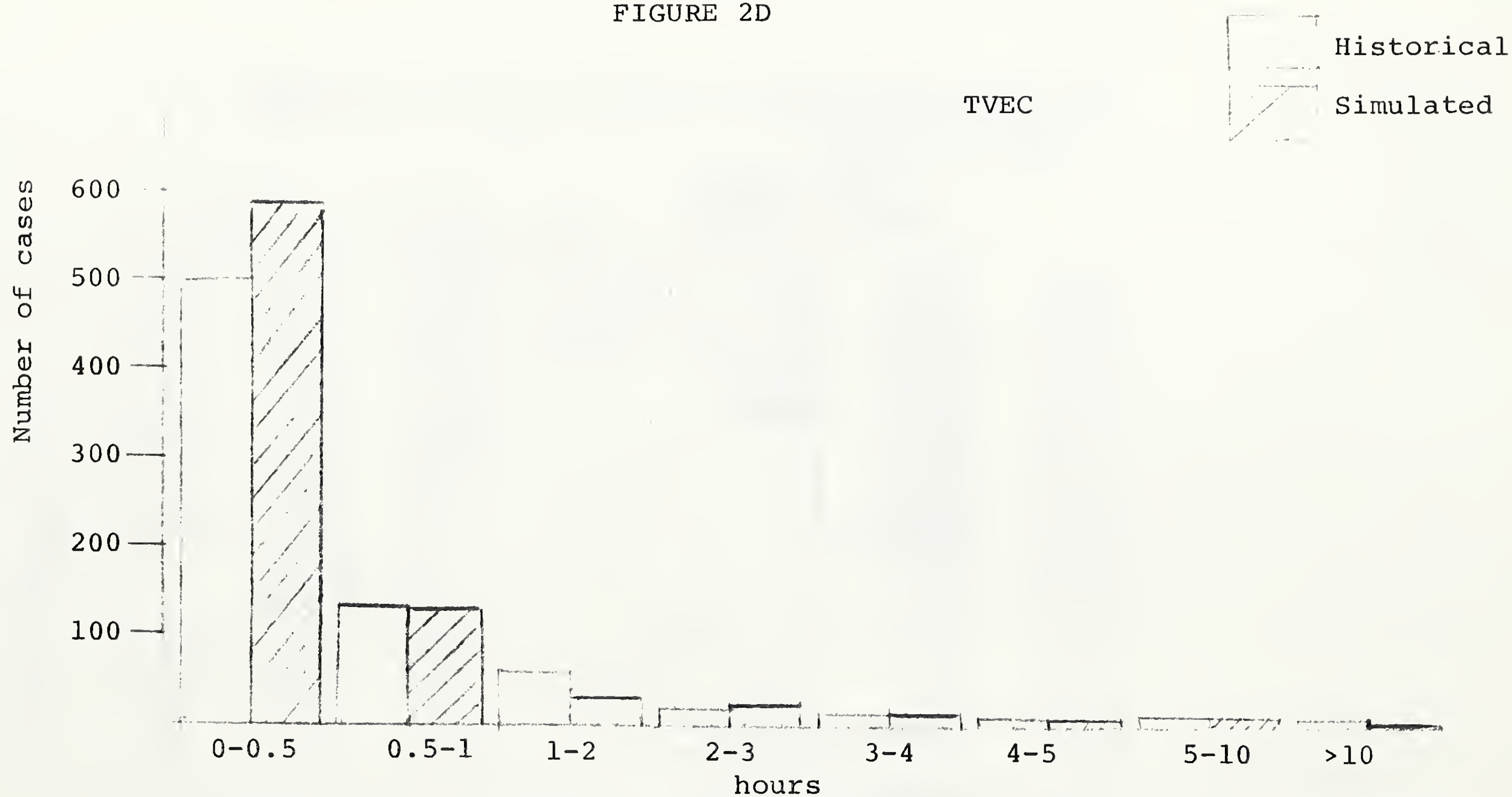
Comparison of historical and simulated output statistics for First District in August 1968, (Run 64).

FIGURE 2C



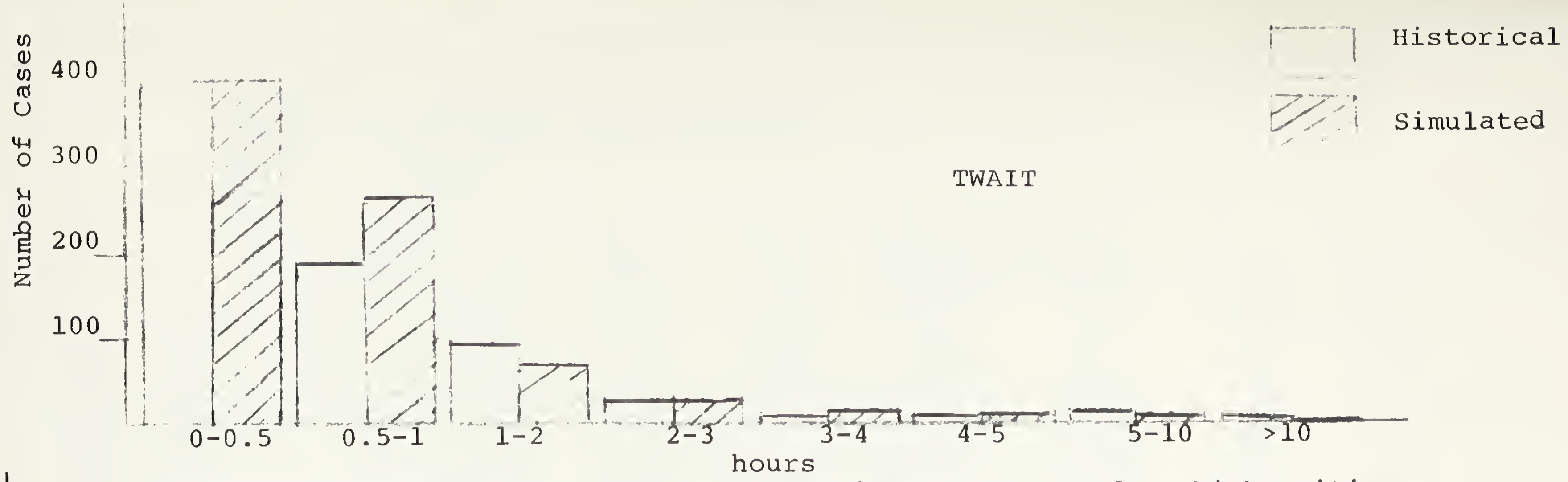
Comparison of combined Resource-Type Utilization Indices for First District in August 1968 (Run 64).

FIGURE 2D



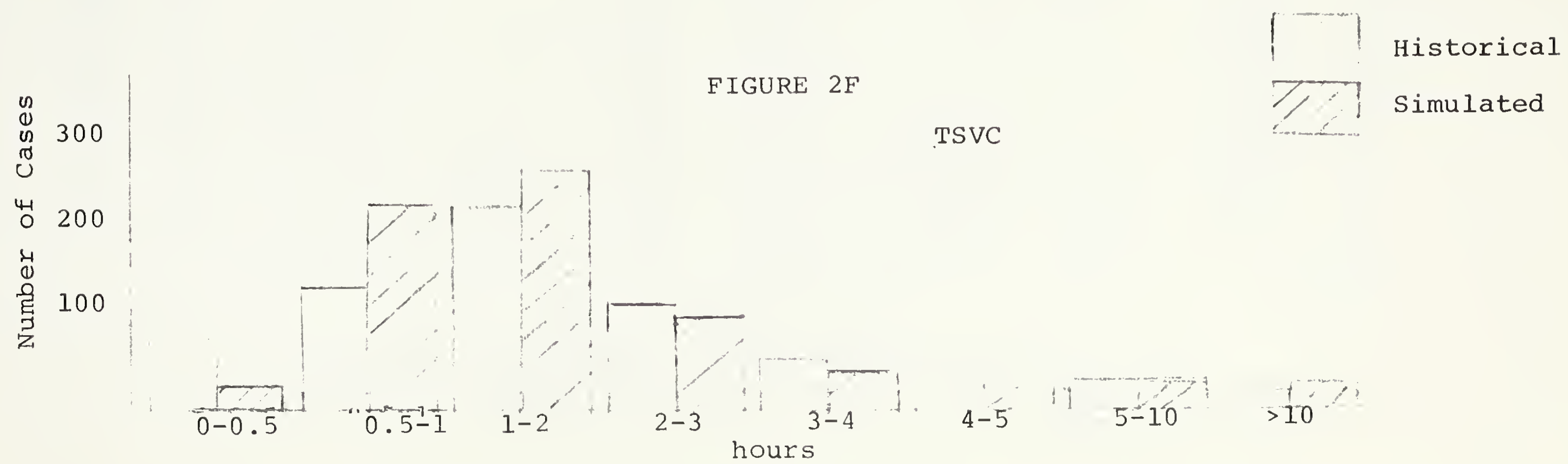
Comparison of numbers of historical and simulated cases for which vector times were within the indicated intervals for District 01, August 1968. (Inequality of intervals should be noted.)

FIGURE 2E



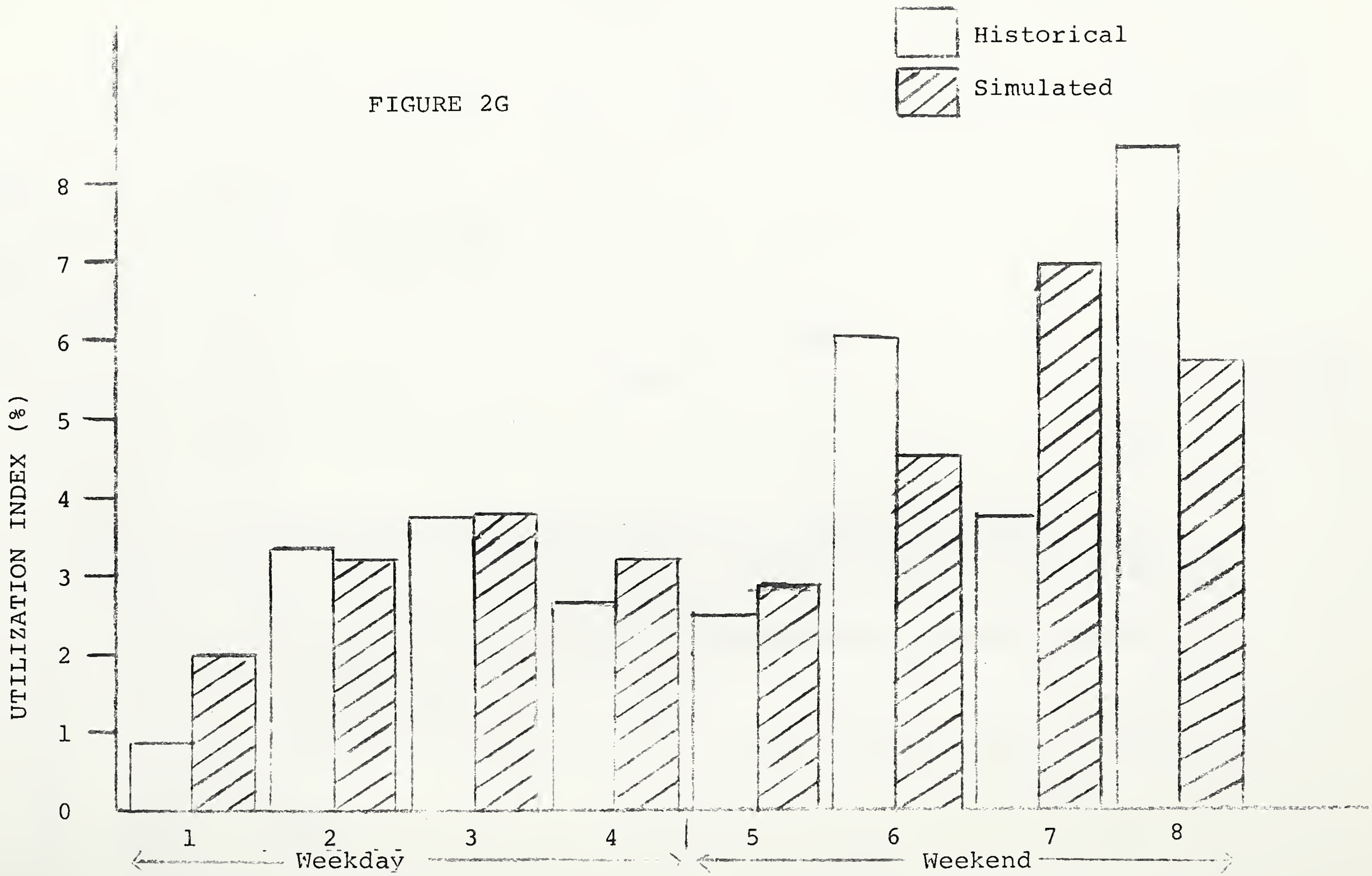
Comparison of numbers of historical and simulated cases for which waiting times were within the indicated intervals for District 01, August 1968. (Inequality of intervals should be noted.)

FIGURE 2F



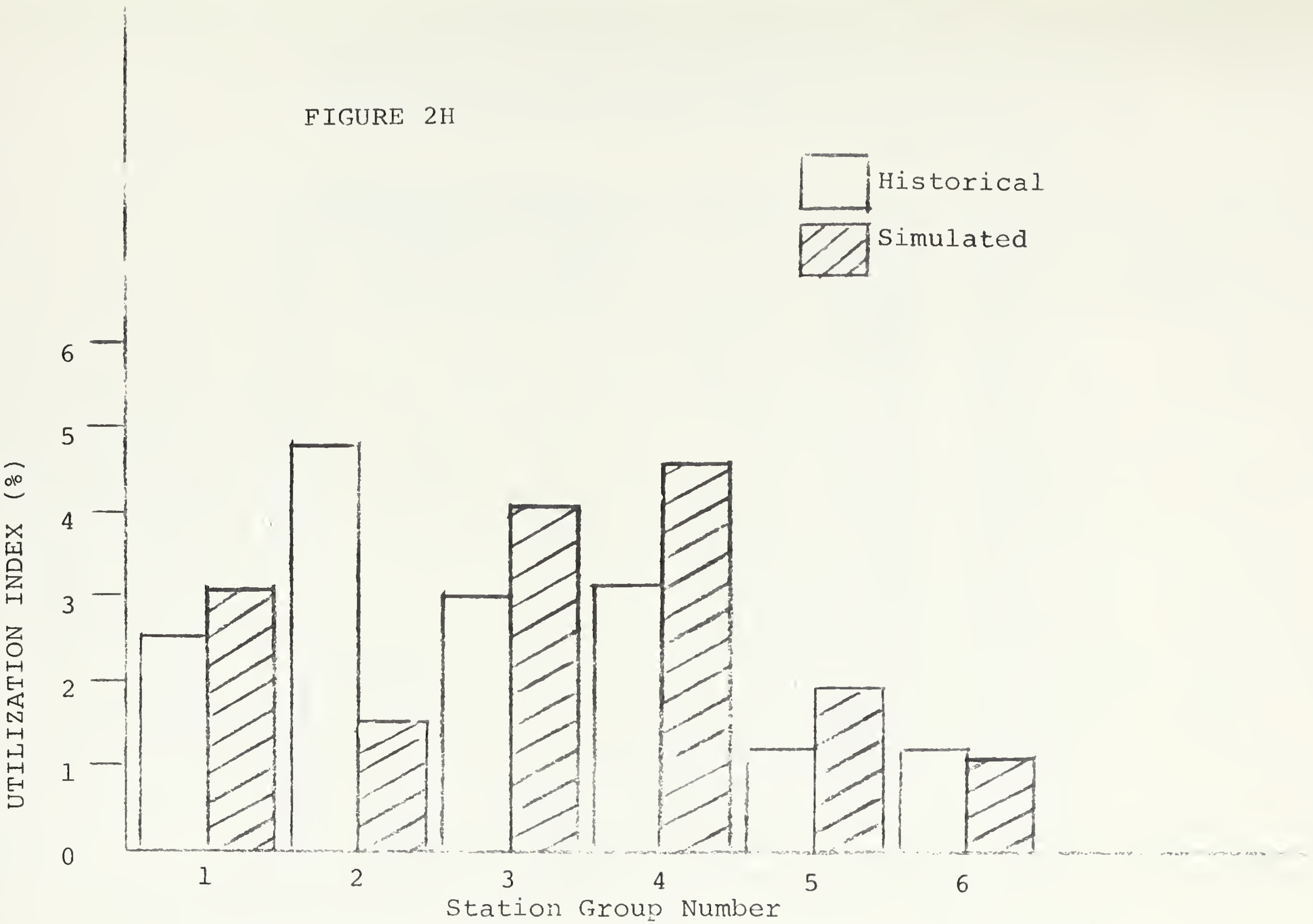
Comparison of numbers of historical and simulated cases for which service times were within the indicated intervals for District 01, August 1968. (Inequality of intervals should be noted.)

FIGURE 2G



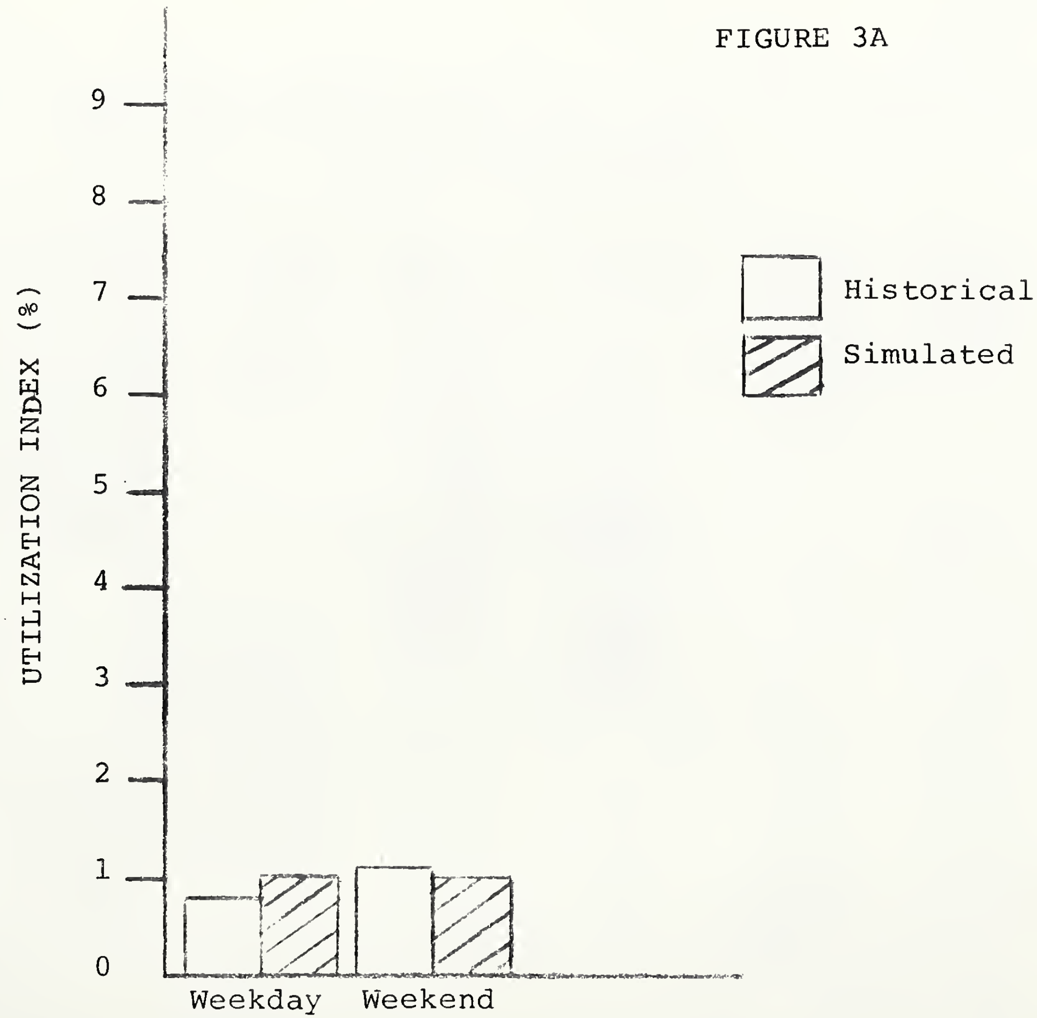
Comparison of historical and simulated utilization indices by shift for First District in August 1968. (Note: see comment in text regarding inapplicability of historical data in this context.

FIGURE 2H



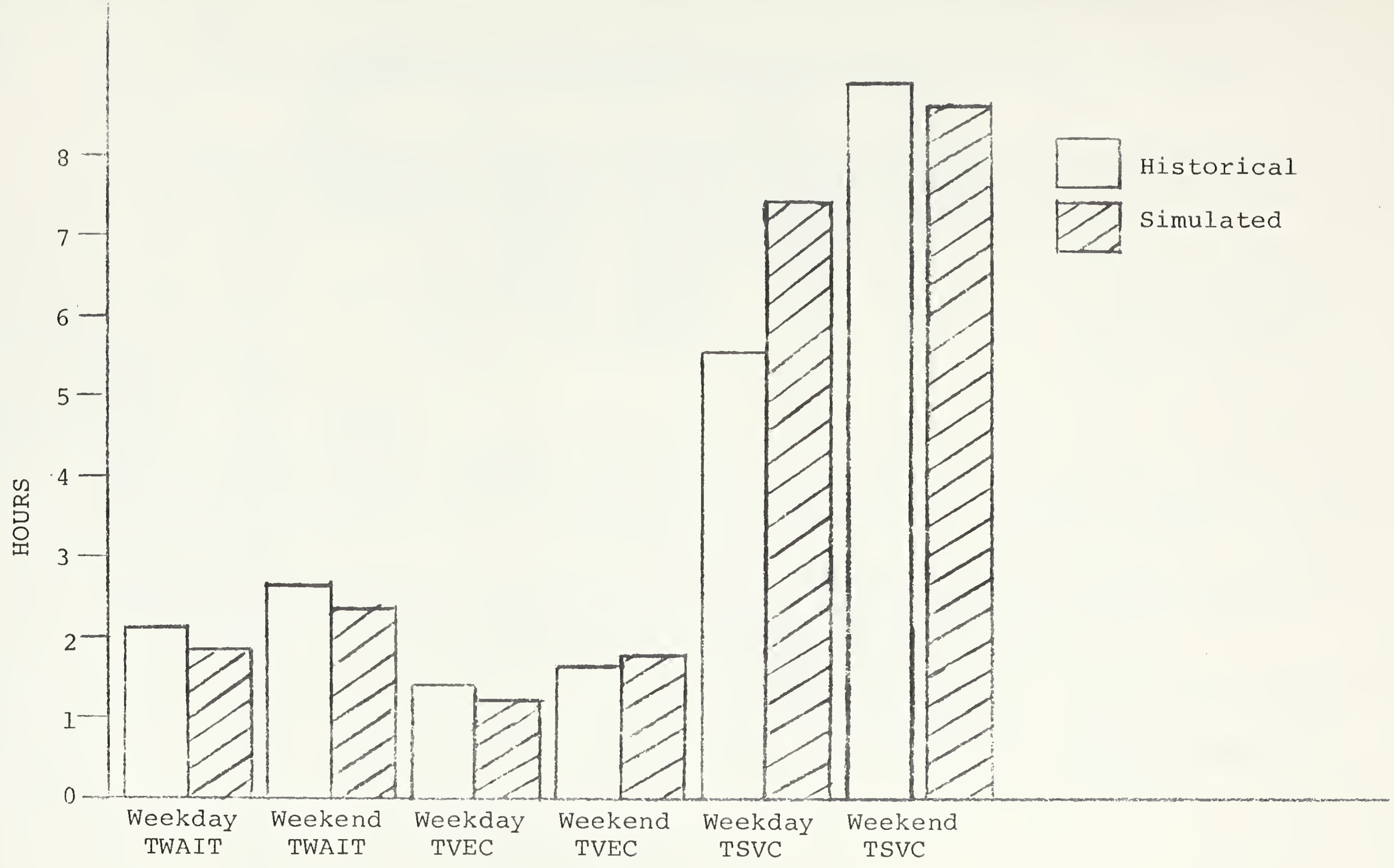
Comparison of historical and simulated combined station group utilization indices for First District in August 1968.

FIGURE 3A



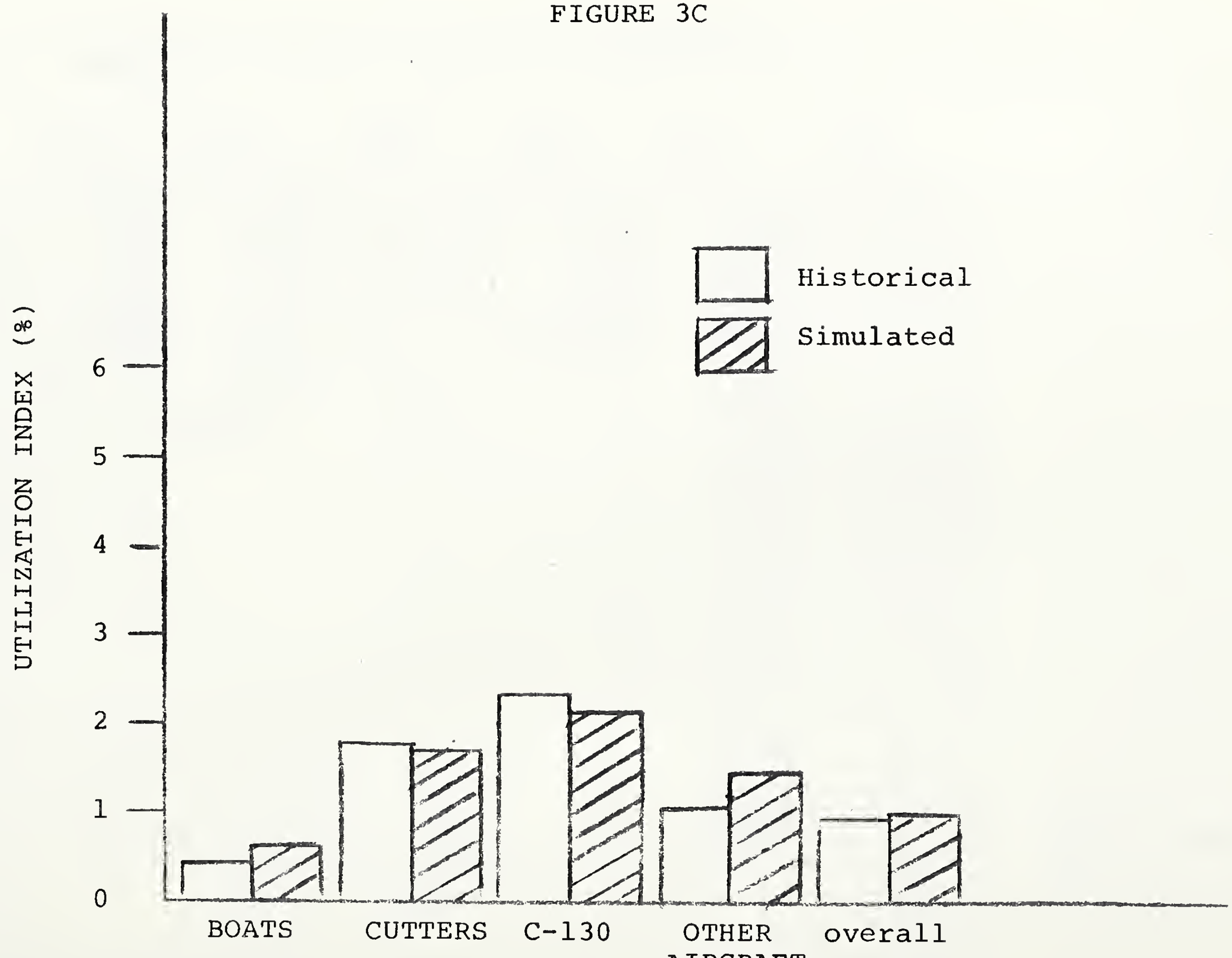
Comparison of historical and simulated utilization indices for First District in Winter 69, Run 61.

FIGURE 3B



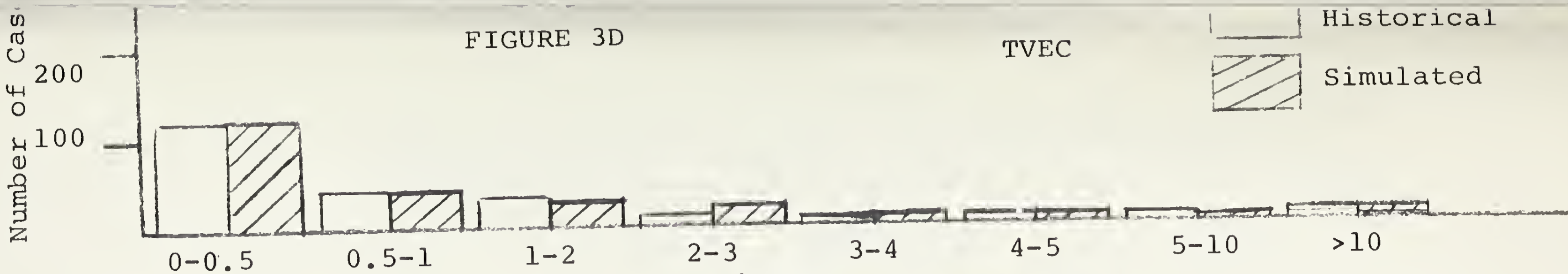
Comparison of historical and simulated output statistics for First District in Winter 69 (Run 61).

FIGURE 3C



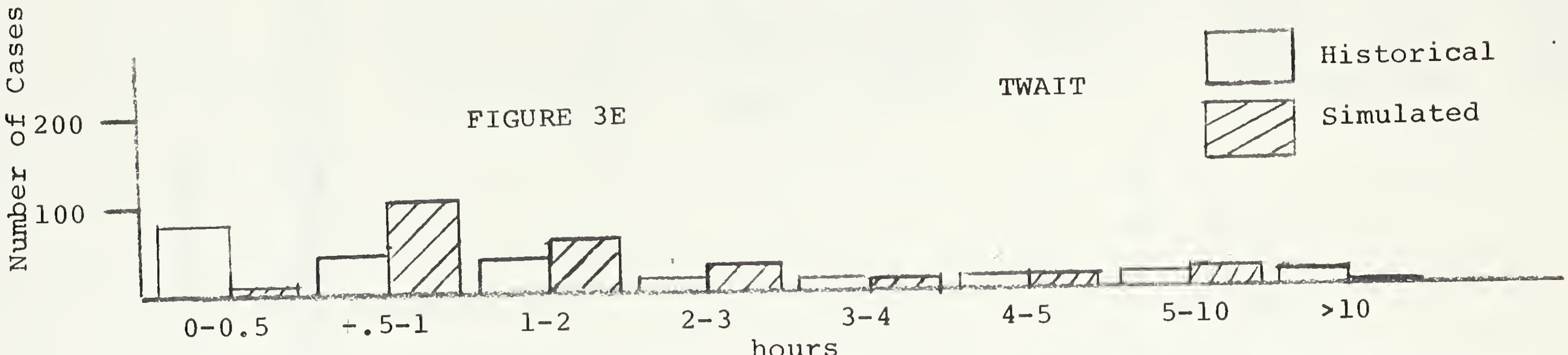
Comparison of combined Resource-Type Utilization Indices for First District in Winter 69 (Run 61).

FIGURE 3D



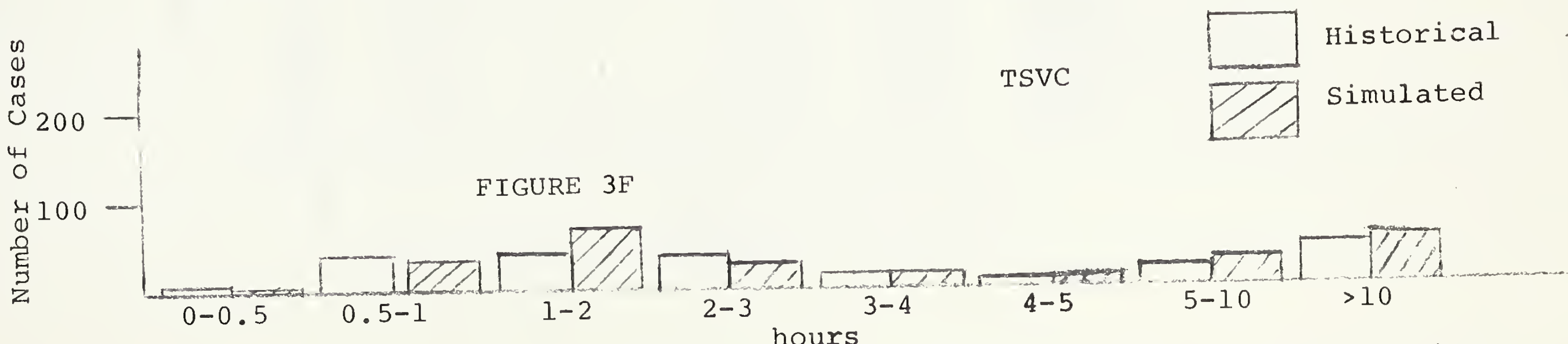
Comparison of numbers of historical and simulated cases for which vector times were within the indicated intervals for District 01, Winter 1969. (Inequality of intervals should be noted.)

FIGURE 3E



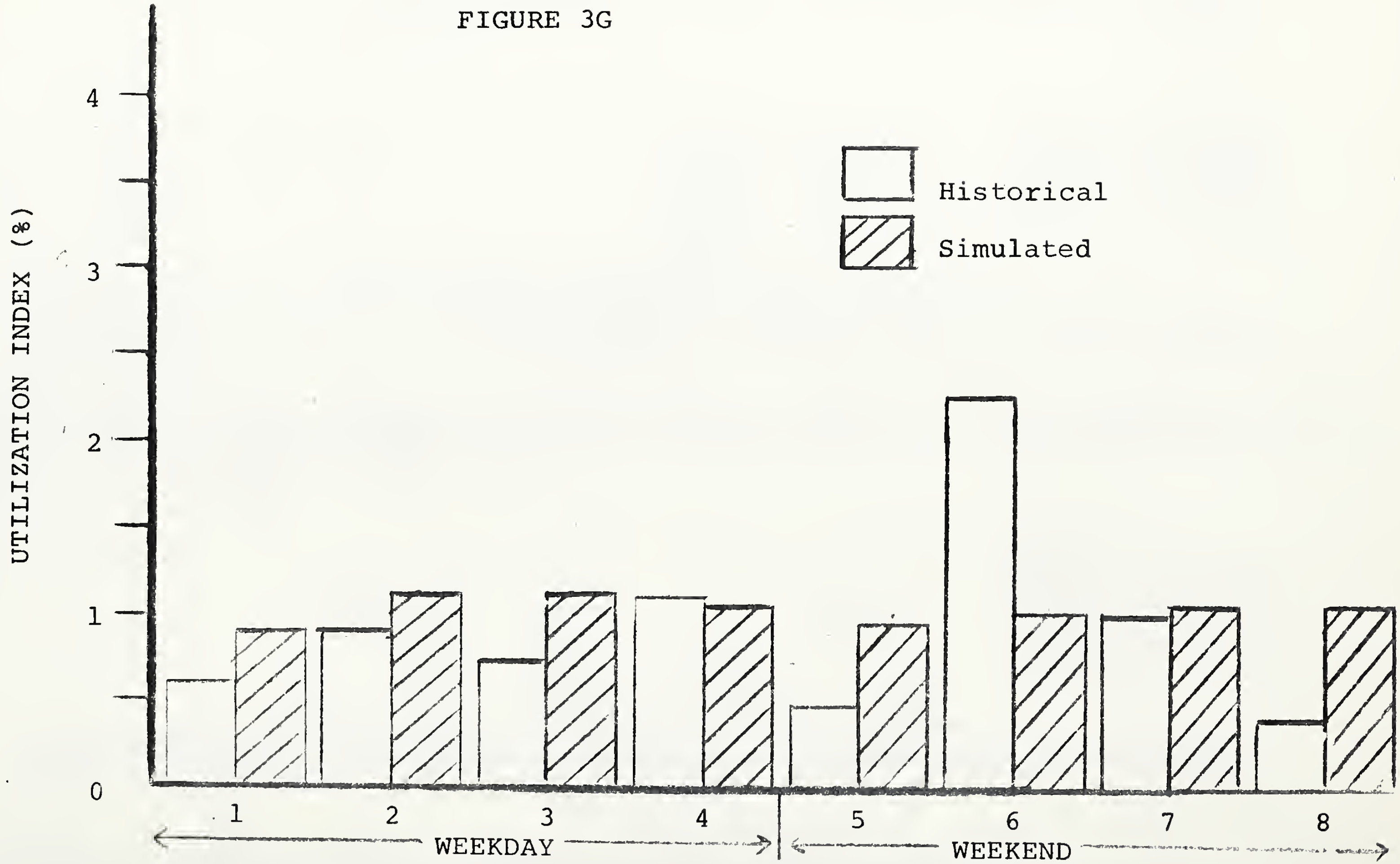
Comparison of numbers of historical and simulated cases for which waiting times were within the indicated intervals for District 01, Winter 1969. (Inequality of intervals should be noted.)

FIGURE 3F



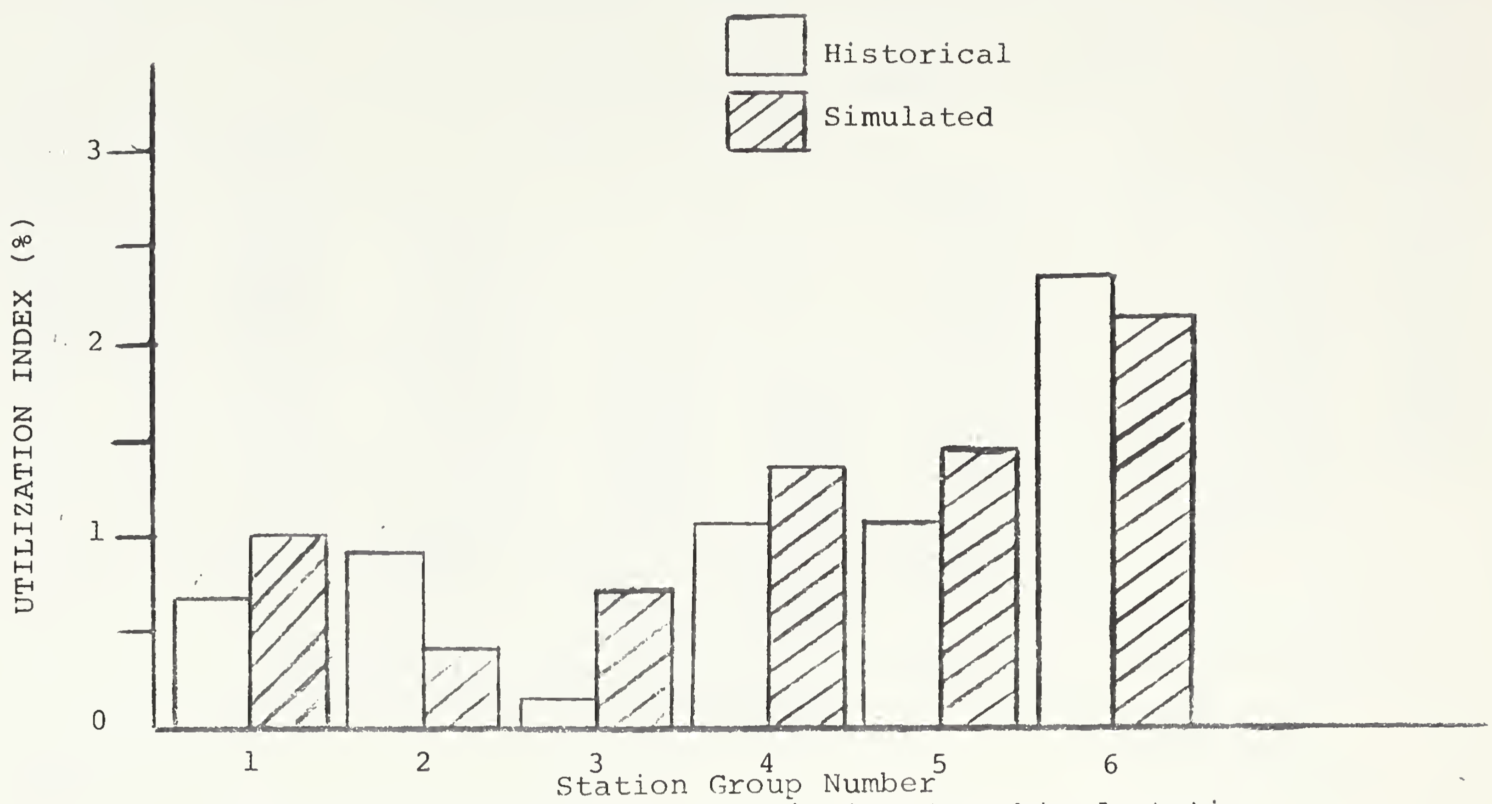
Comparison of numbers of historical and simulated cases for which service times were within the indicated intervals for District 01, Winter 1969. (Inequality of intervals should be noted.)

FIGURE 3G



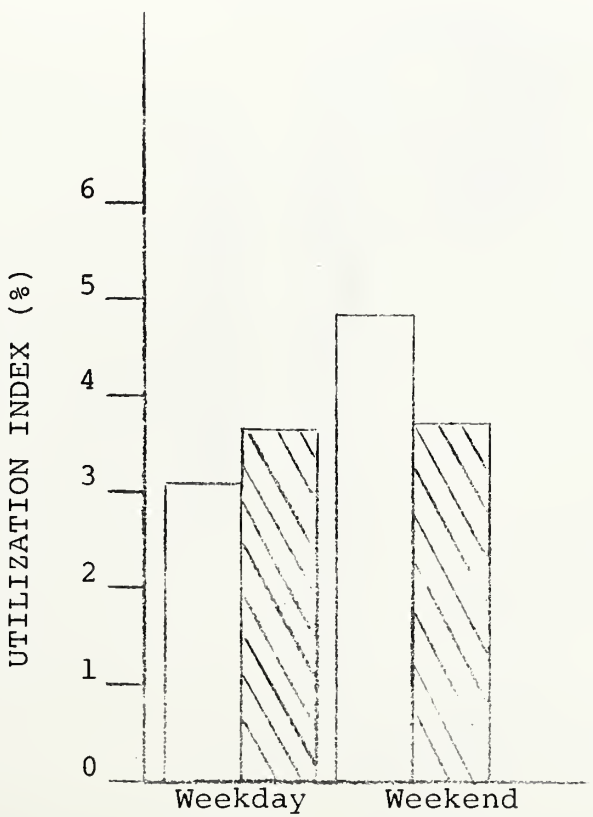
Comparison of historical and simulated utilization indices by shift for First District in Winter '69 (Run 61). (Note: see comment in text regarding inapplicability of historical data in this context.

FIGURE 3H



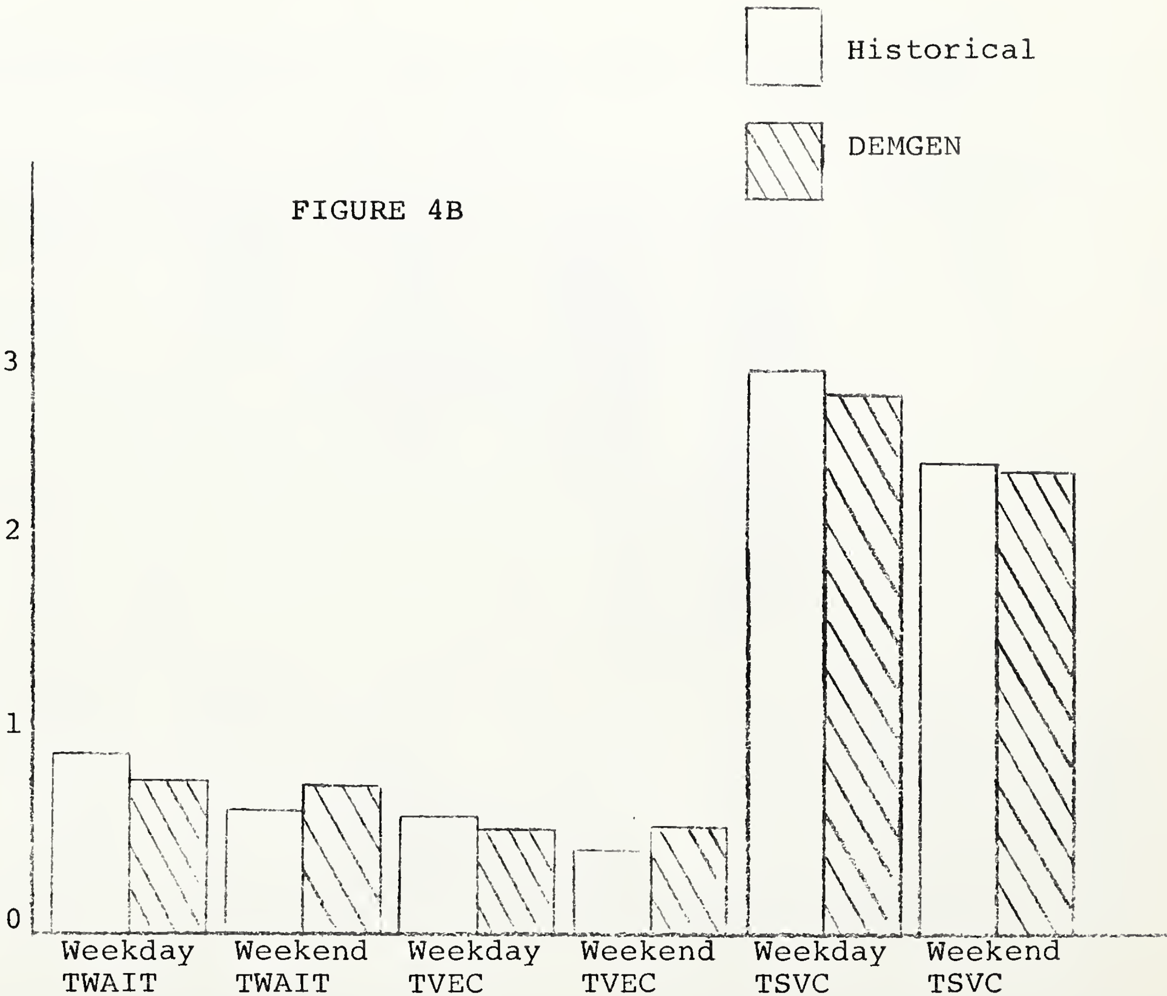
Comparison of historical and simulated combined station group utilization indices for First District in Winter, 1969.

FIGURE 4A



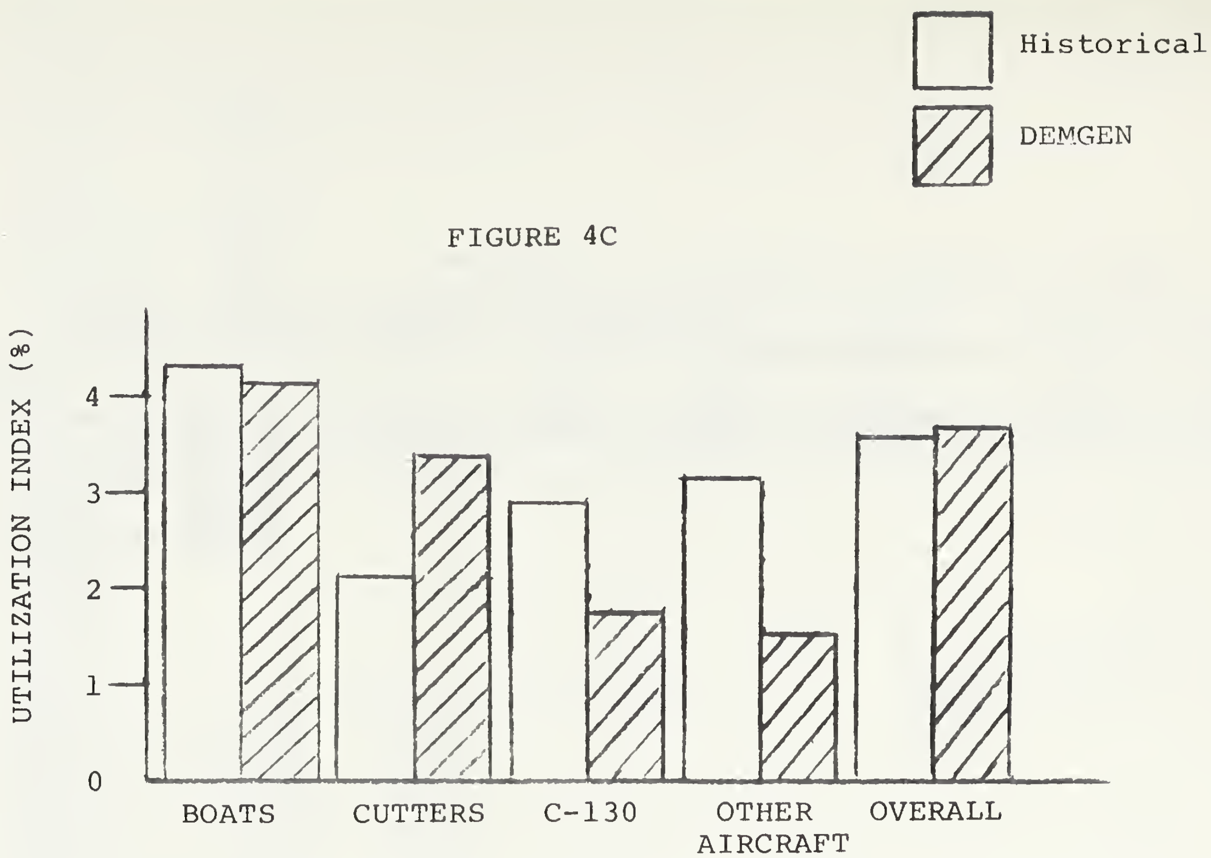
Comparison of historical and DEMGEN utilization indices for first District, July 1968.

FIGURE 4B



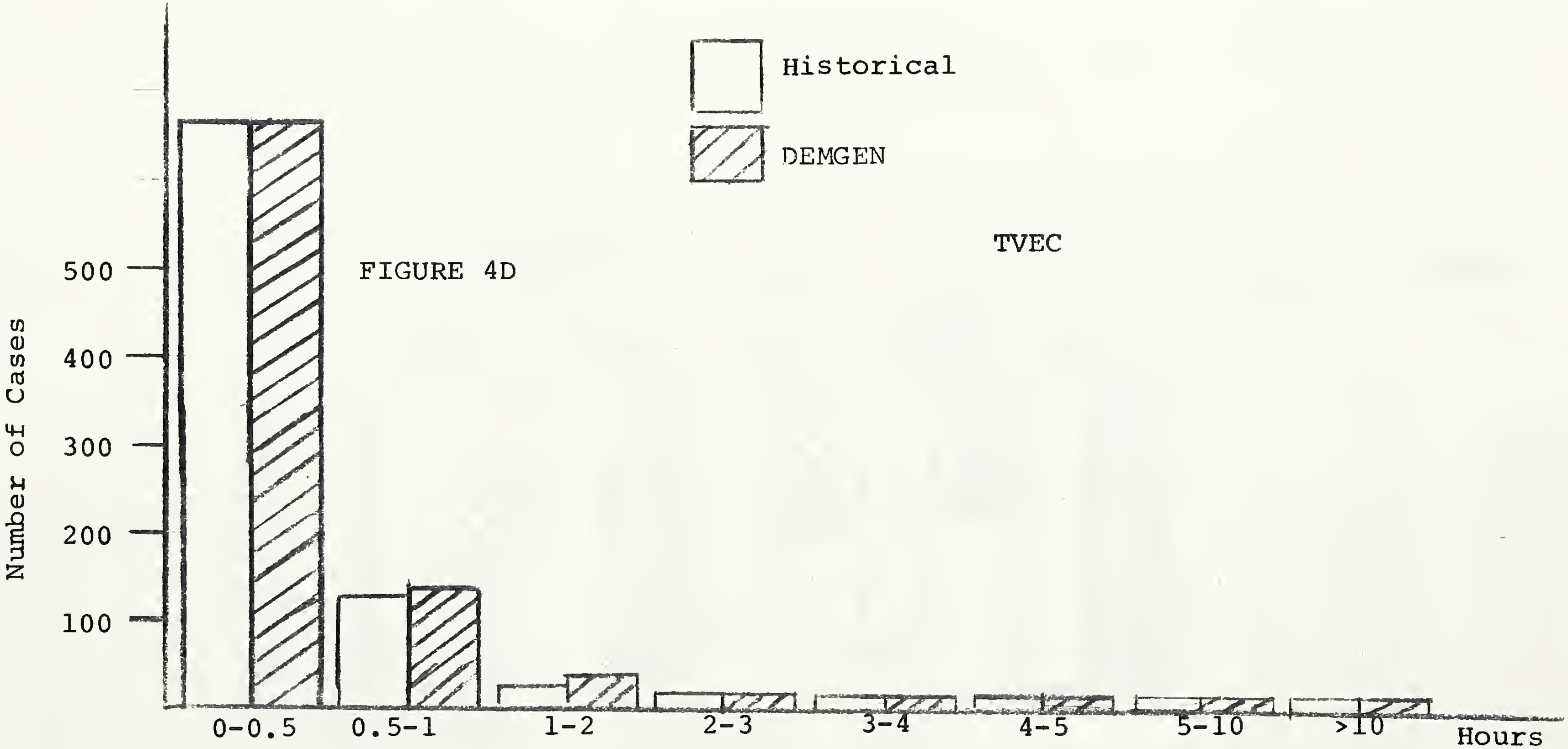
Comparison of historical and DEMGEN output statistics for First District, July 1968.

FIGURE 4C



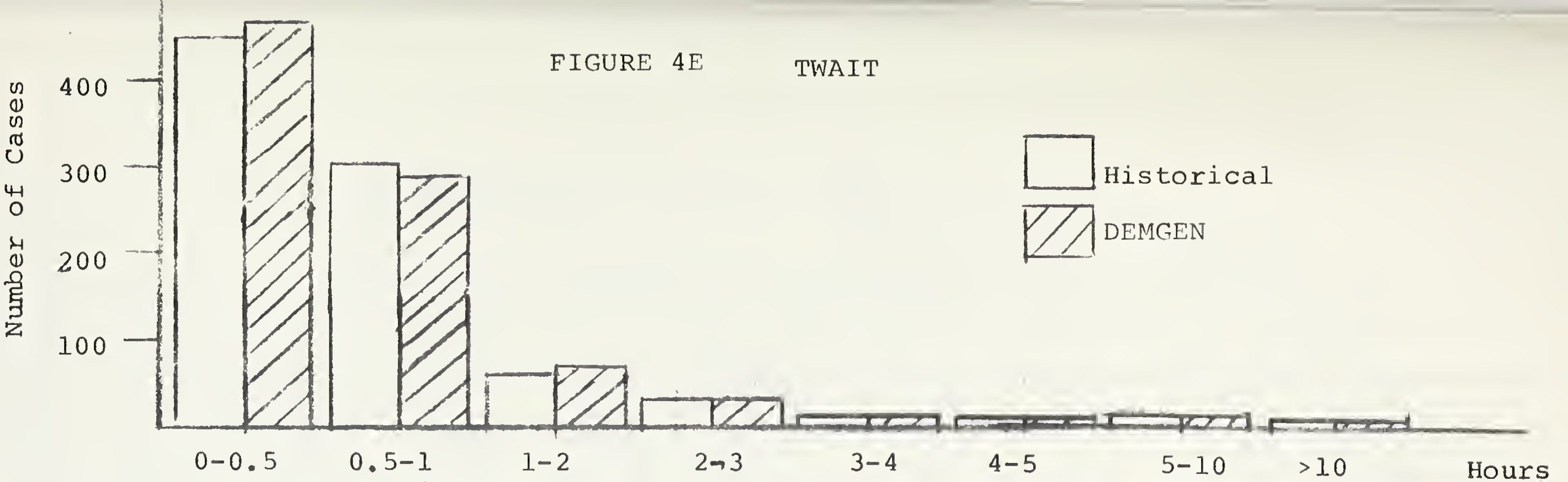
Comparison of combined Resource-Type Utilization Indices for First District in July, 1968. HIST vs. DEMGEN.

FIGURE 4D



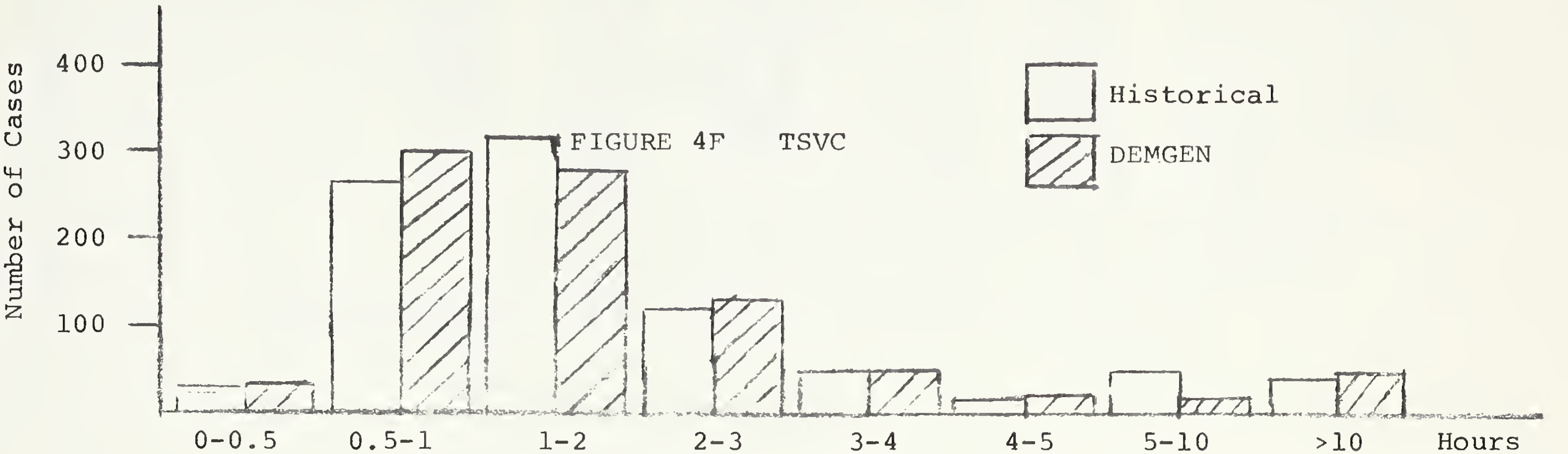
Comparison of numbers of historical and DEMGEN cases for which vector times were within the indicated intervals for First District, July 1968 (Inequality of intervals should be noted.)

FIGURE 4E TWAIT



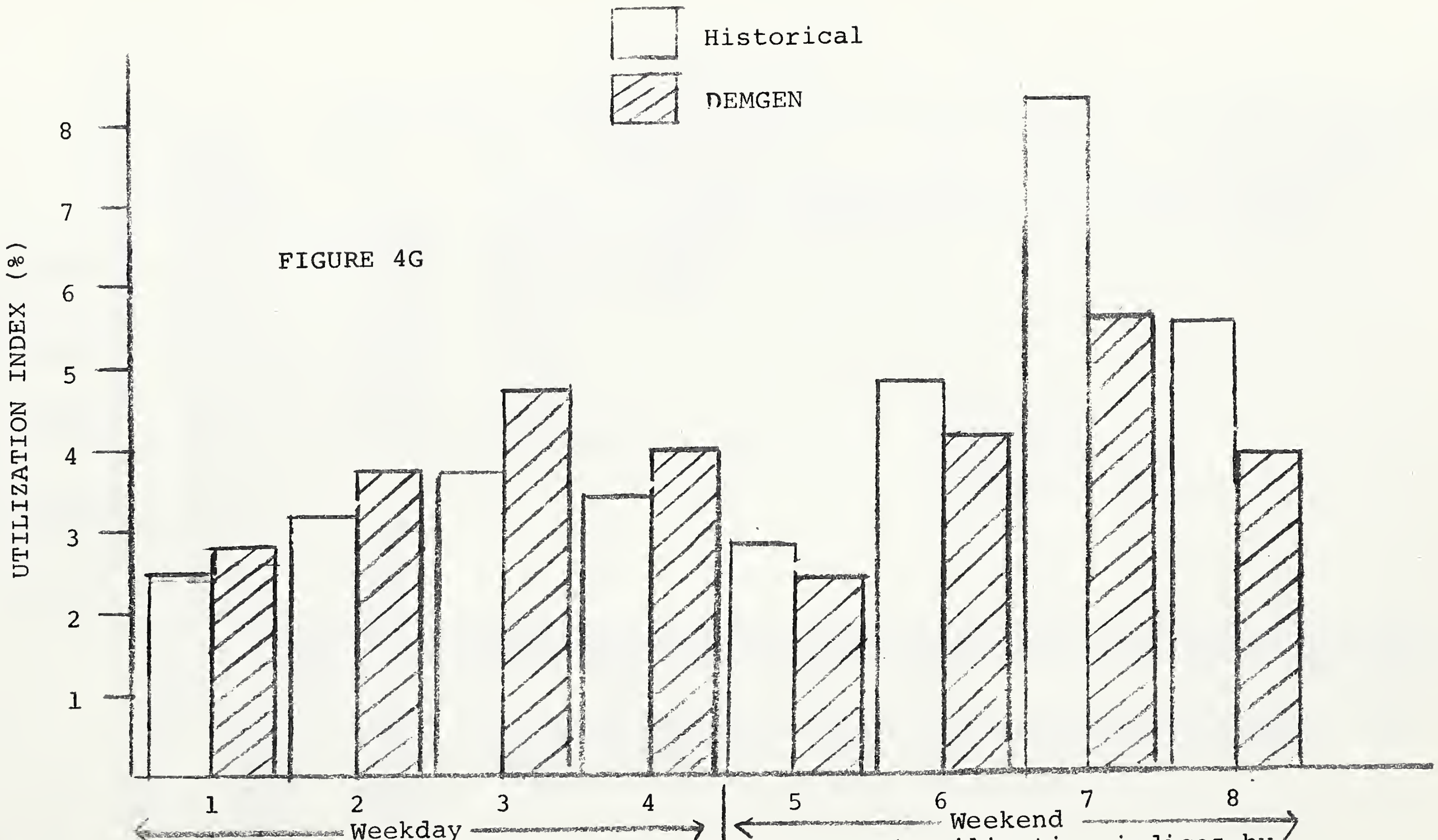
Comparison of numbers of historical and simulated cases for which waiting times were within the indicated intervals for First District, July 1968.

FIGURE 4F TSVC



Comparison of numbers of historical and DEMGEN cases for which service times were within the indicated intervals for First District, July 1968.

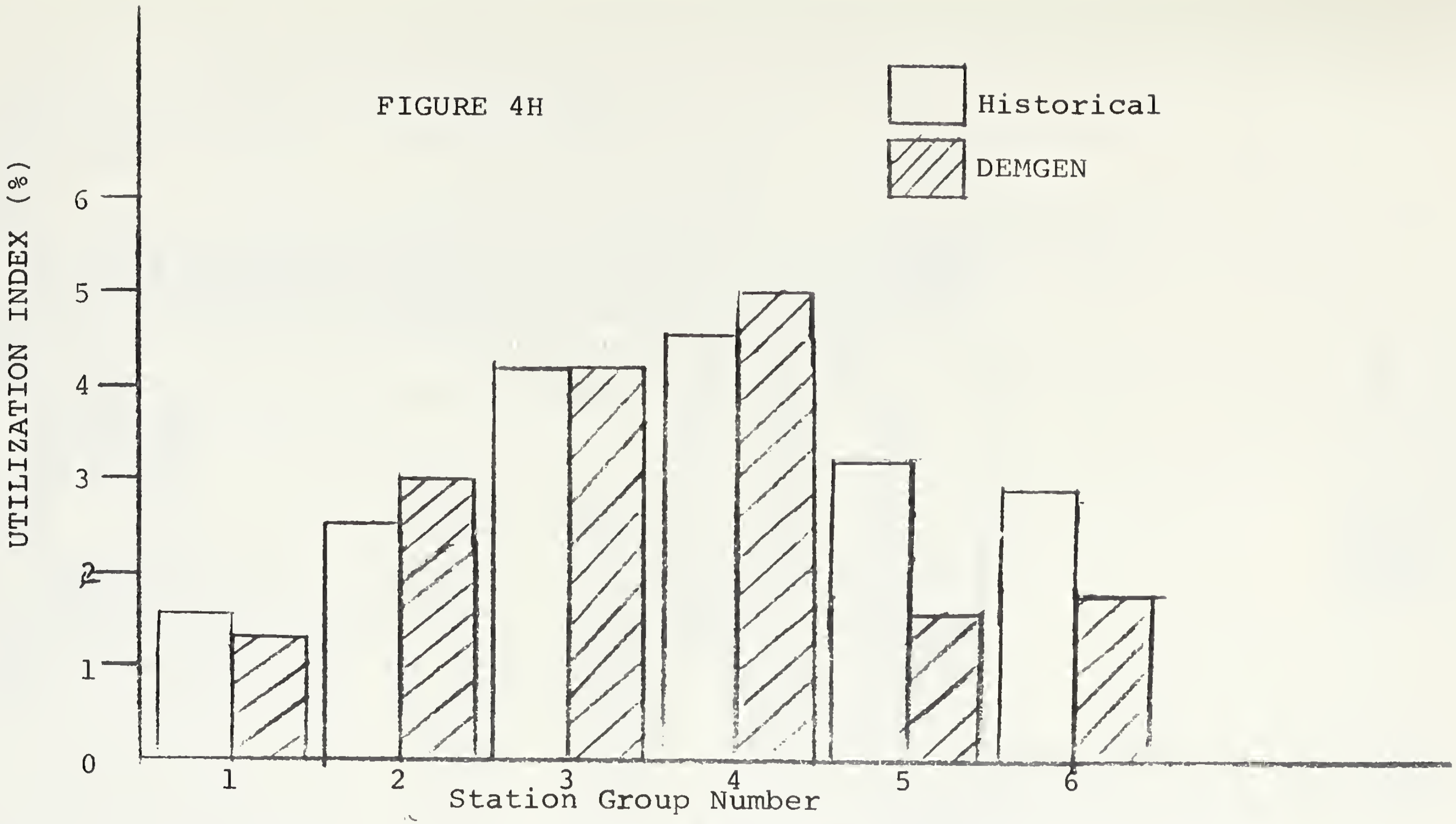
FIGURE 4G



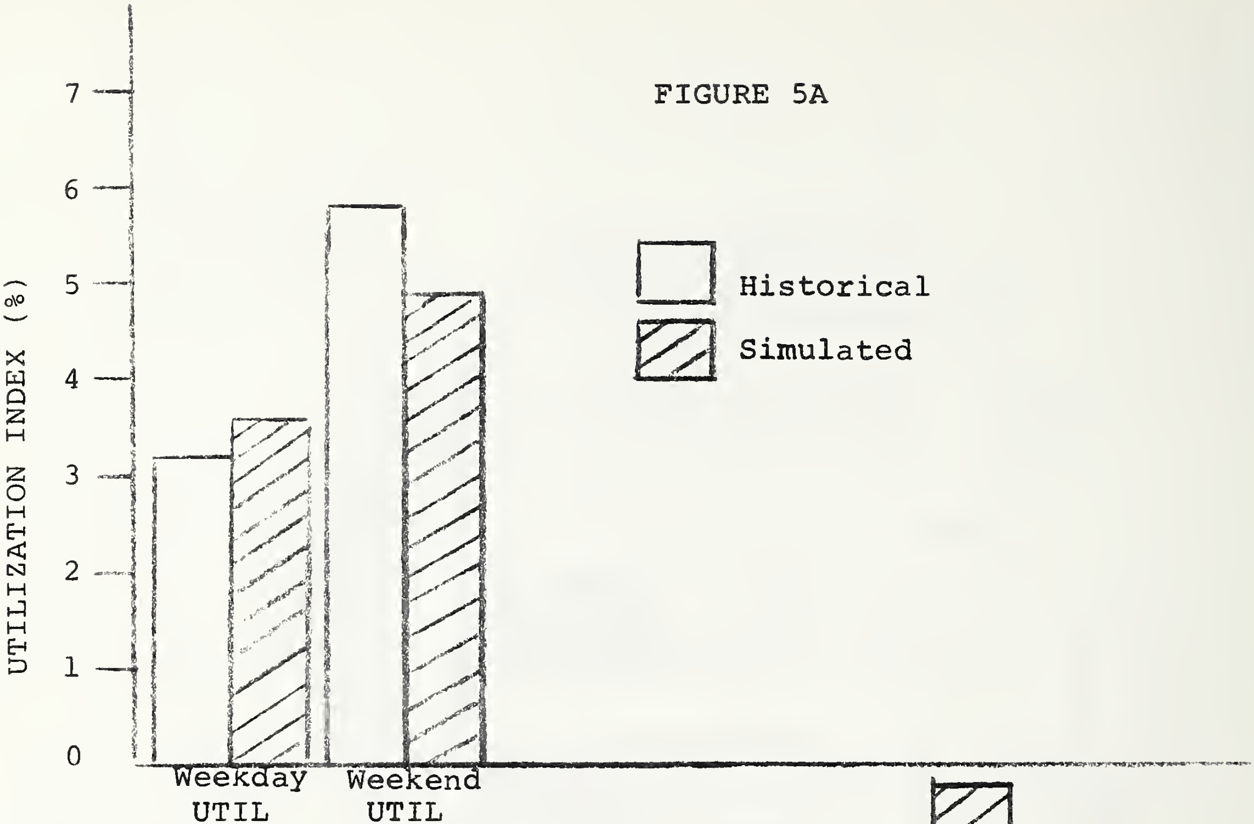
Comparison of historical and simulated utilization indices by shift for First District in July 1968. HIST vs. DEMGEN.

(Note: see comment in text regarding inapplicability of historical data in this context.)

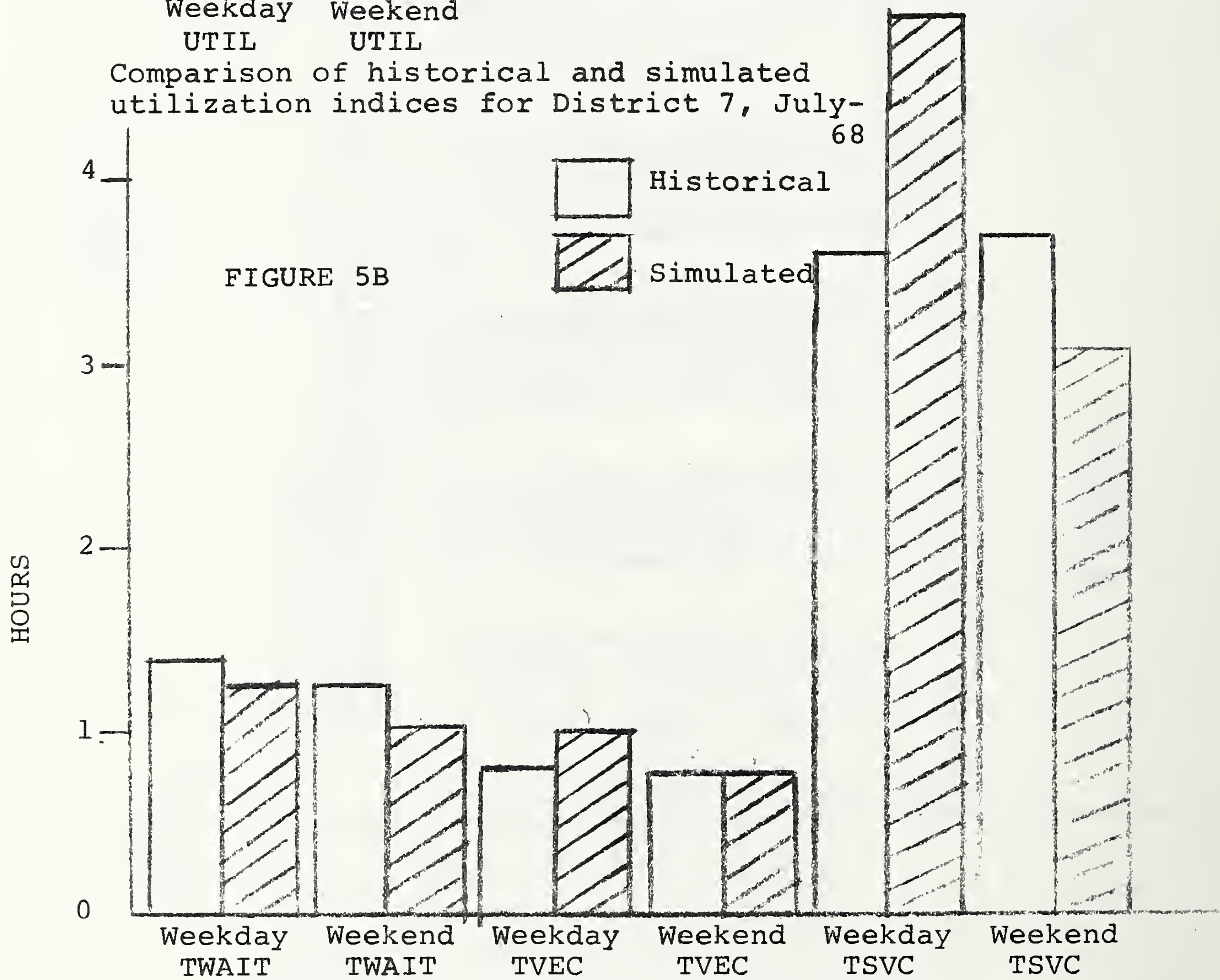
FIGURE 4H



Comparison of historical and simulated combined station group utilization indices for First District, July 1968 HIST vs. DEMGEN.

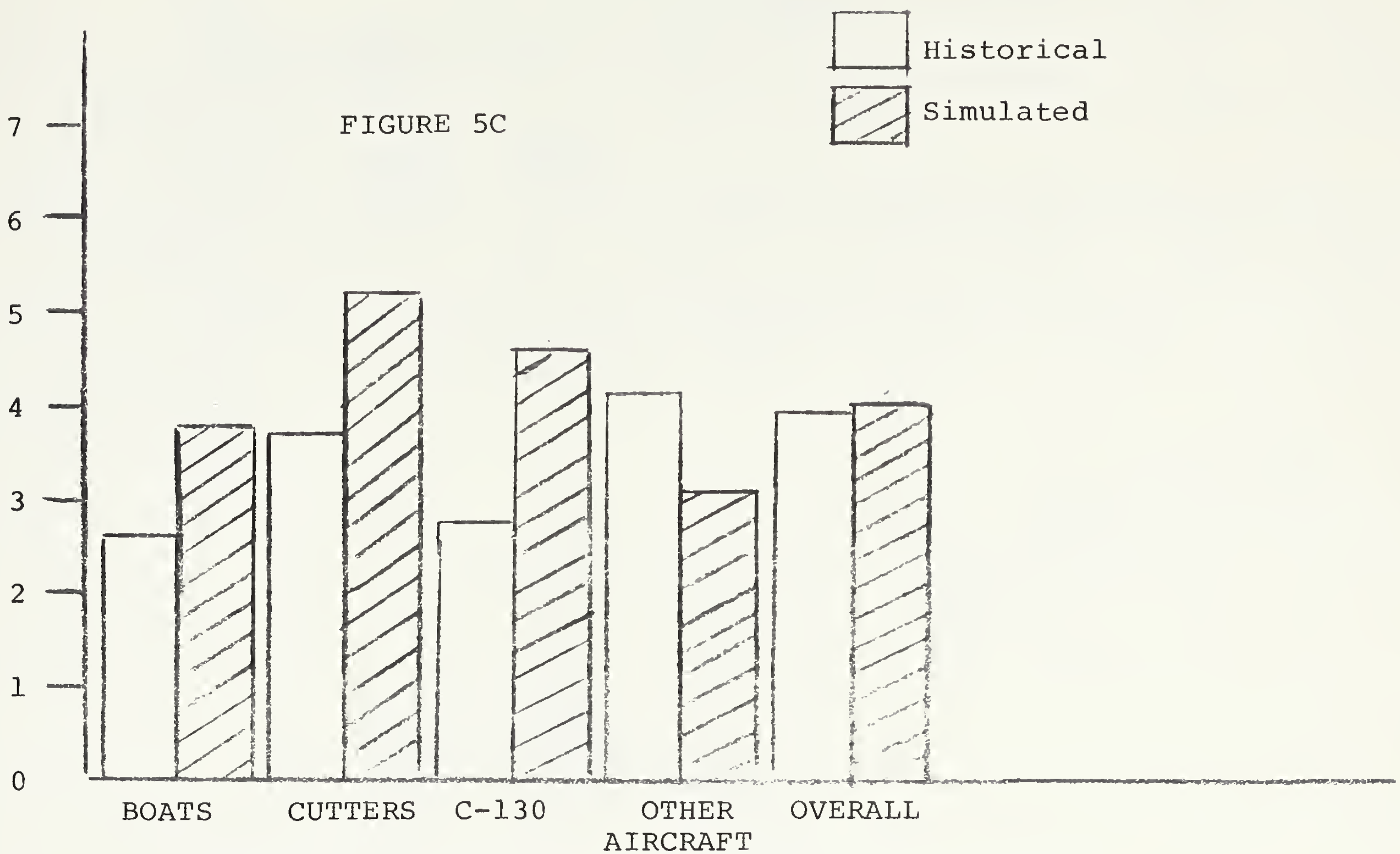


Comparison of historical and simulated utilization indices for District 7, July-68



Comparison of historical and simulated output statistics for District 7, July 1968.

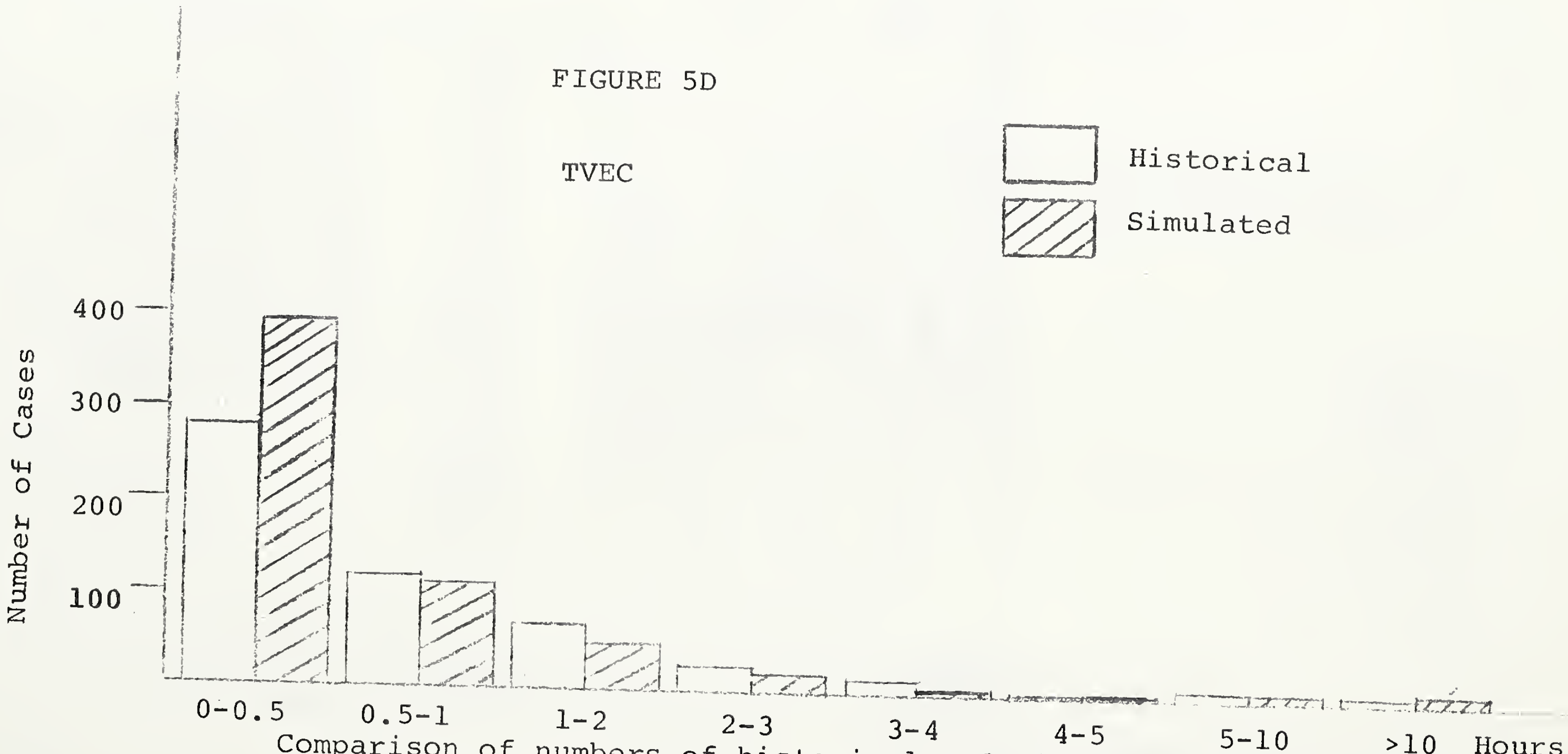
FIGURE 5C



Comparison of combined Resource-Type Utilization Indices for District 7, July 1968.

FIGURE 5D

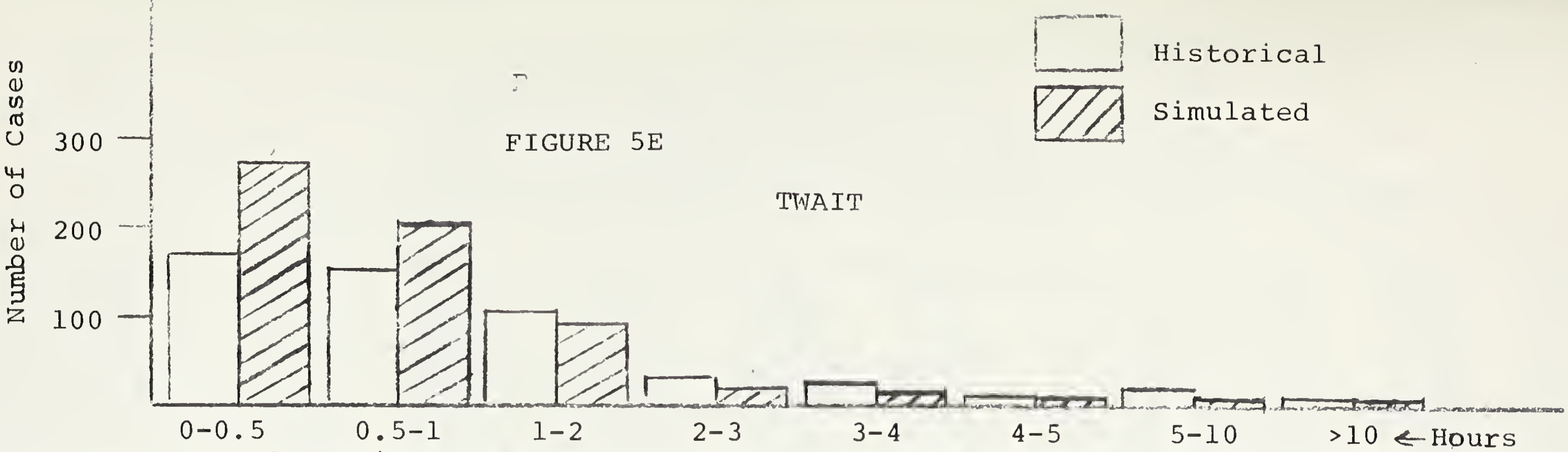
TVEC



Comparison of numbers of historical and simulated cases for which vector times were within the indicated intervals for District 07, July 1968. (Inequality of intervals should be noted.)

FIGURE 5E

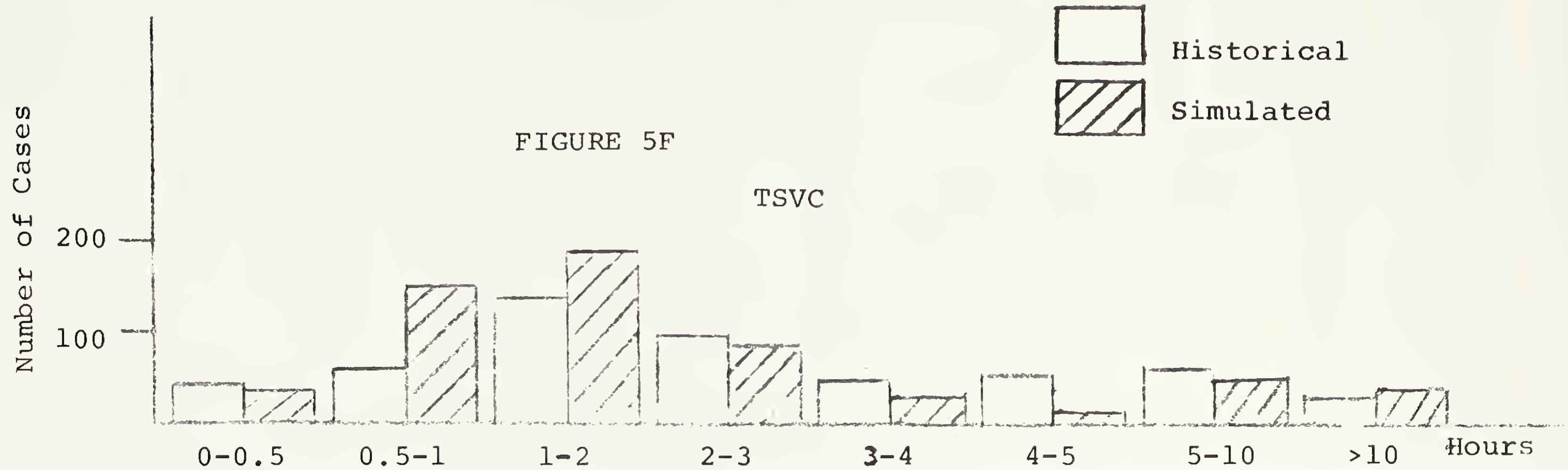
TWAIT



Comparison of numbers of historical and simulated cases for which waiting times were within the indicated intervals for District 07, July 1968. (Inequality of intervals should be noted.)

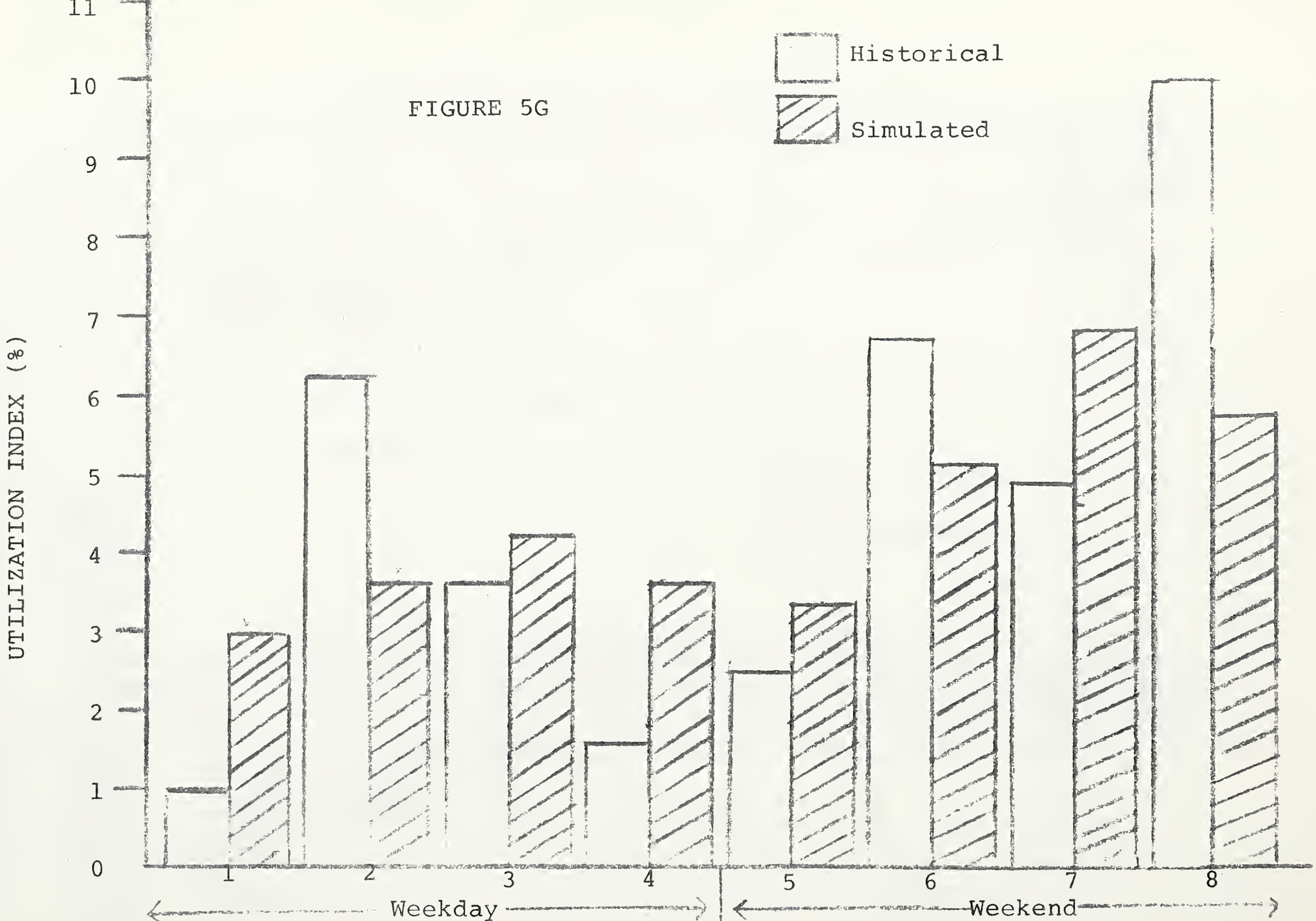
FIGURE 5F

TSVC



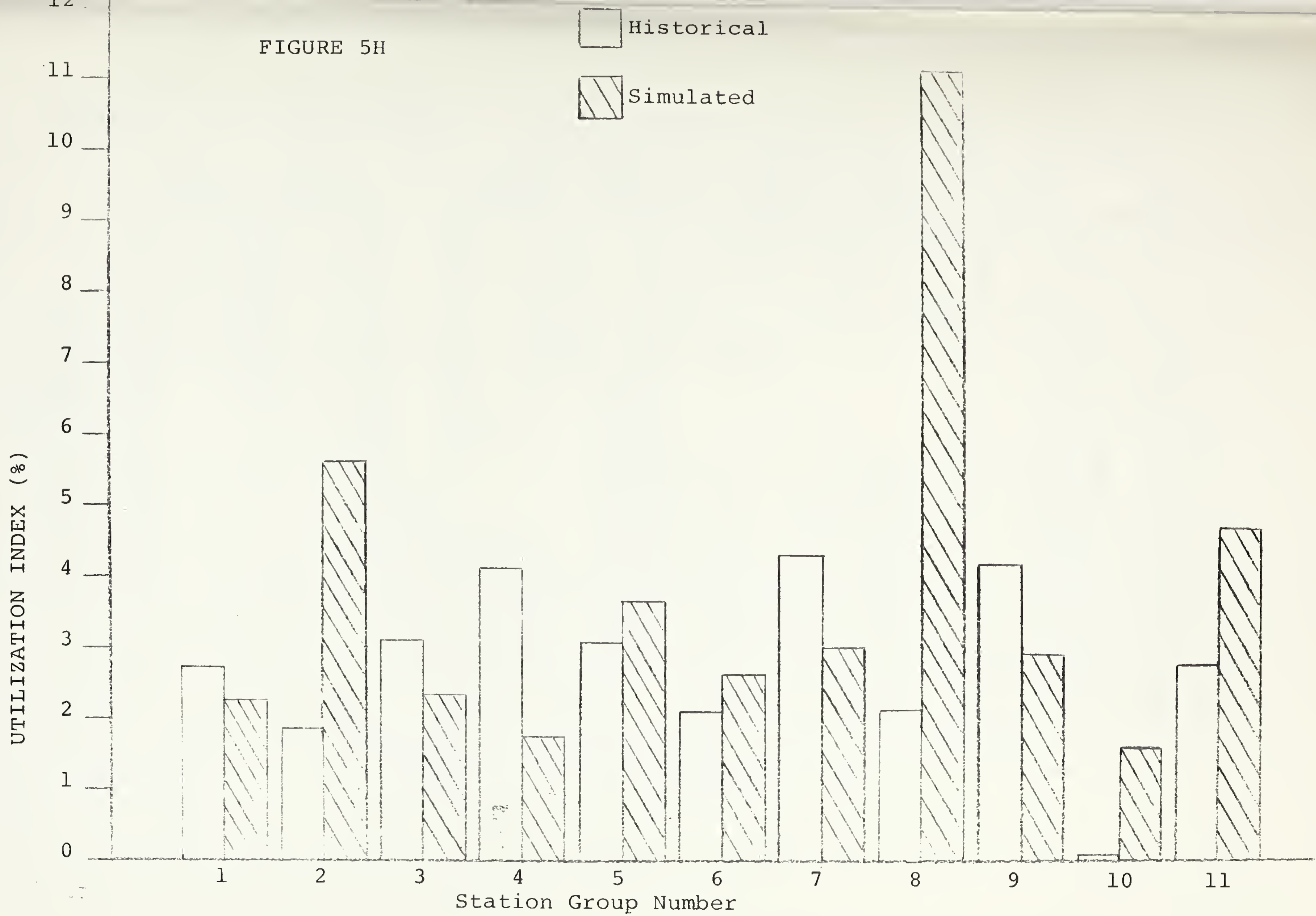
Comparison of numbers of historical and simulated cases for which service times were within the indicated intervals for District 07, July 1968. (Inequality of intervals should be noted.)

FIGURE 5G



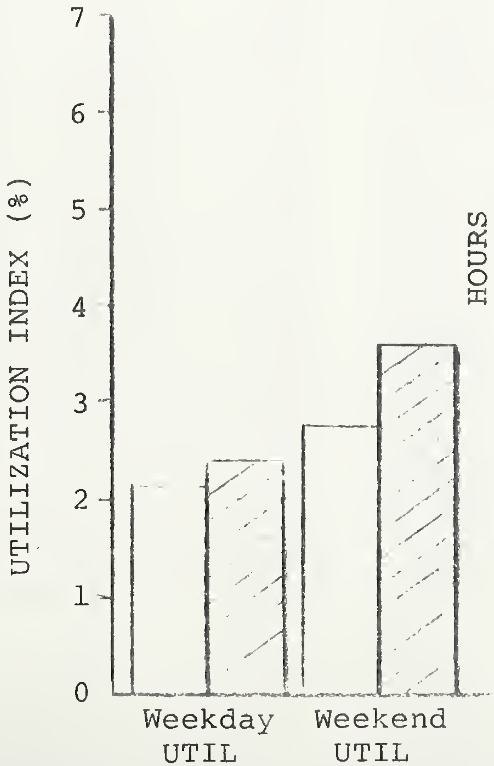
Comparison of historical and simulated utilization indices by shift for District 07, July, 1968.
(Note: see comment in text regarding inapplicability of historical data in this context.)

FIGURE 5H



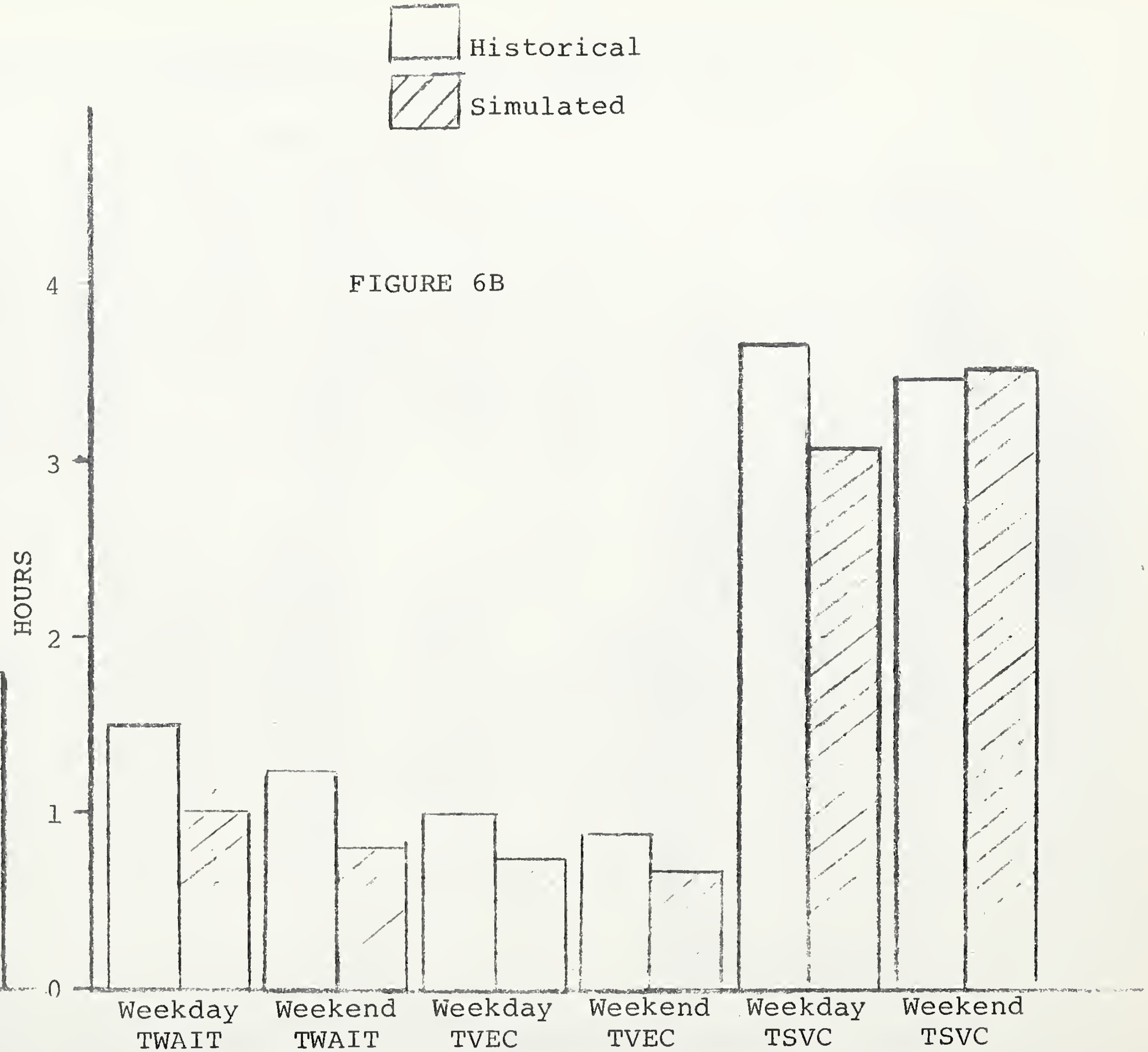
Comparison of historical and simulated combined station group utilization indices for Seventh District, July 1968.

FIGURE 6A



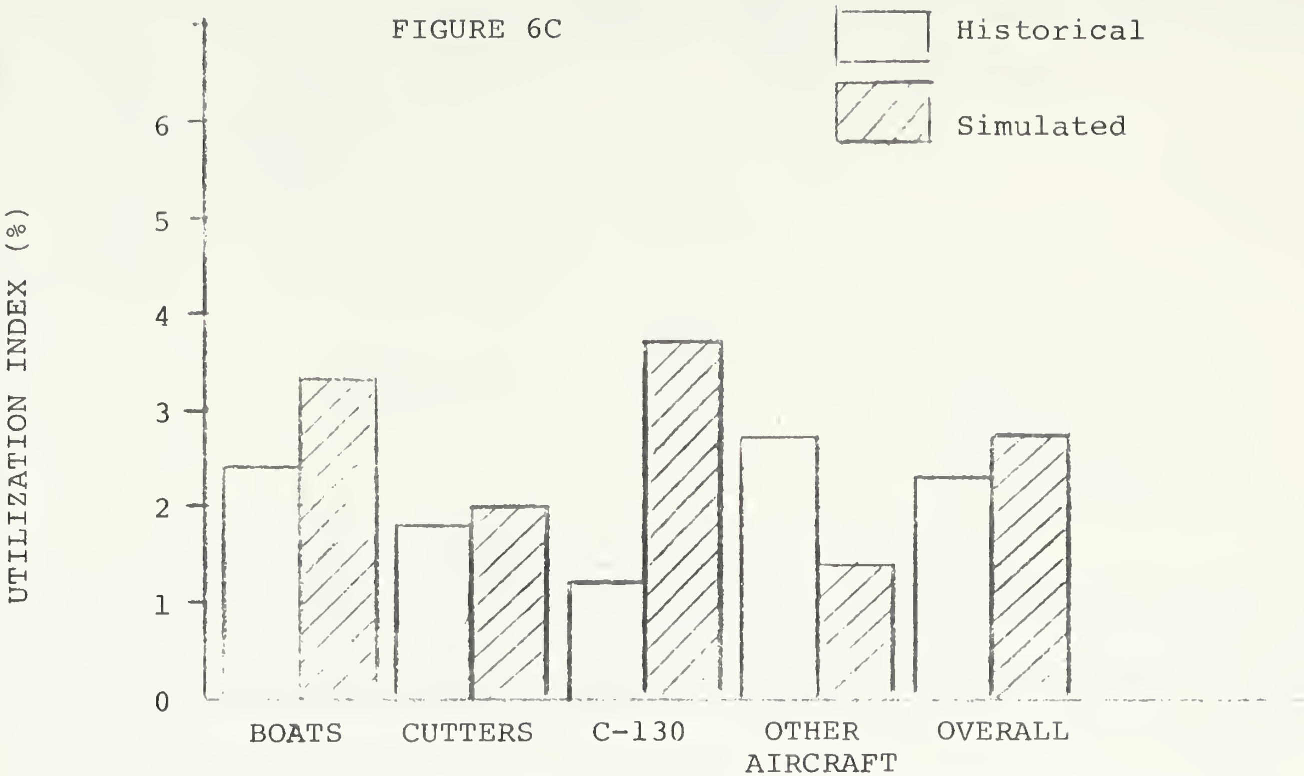
Comparison of historical and simulated utilization indices for District 7, August 1968.

FIGURE 6B



Comparison of historical and simulated output statistics for District 7, August, 1968.

FIGURE 6C



Comparison of combined Resource-Type Utilization Indices for District 07, August, 1968.

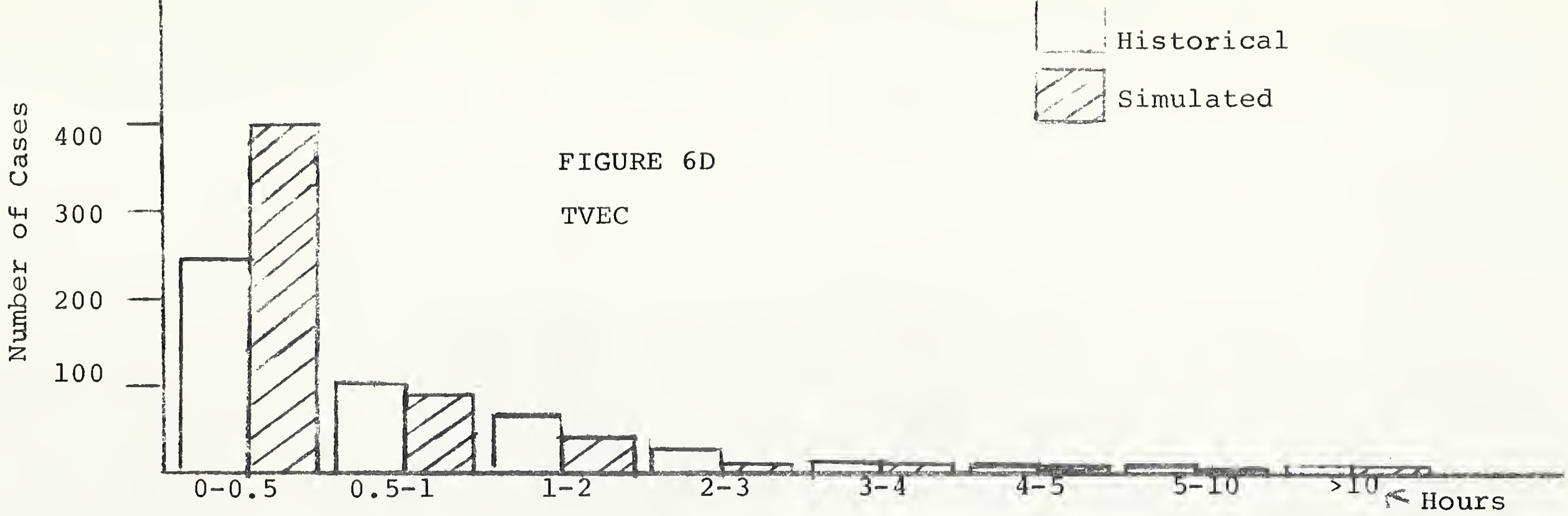


FIGURE 6D

TVEC

Comparison of numbers of historical and simulated cases for which vector times were within the indicated intervals for District 7, August 1968. (Inequality of intervals should be noted.)

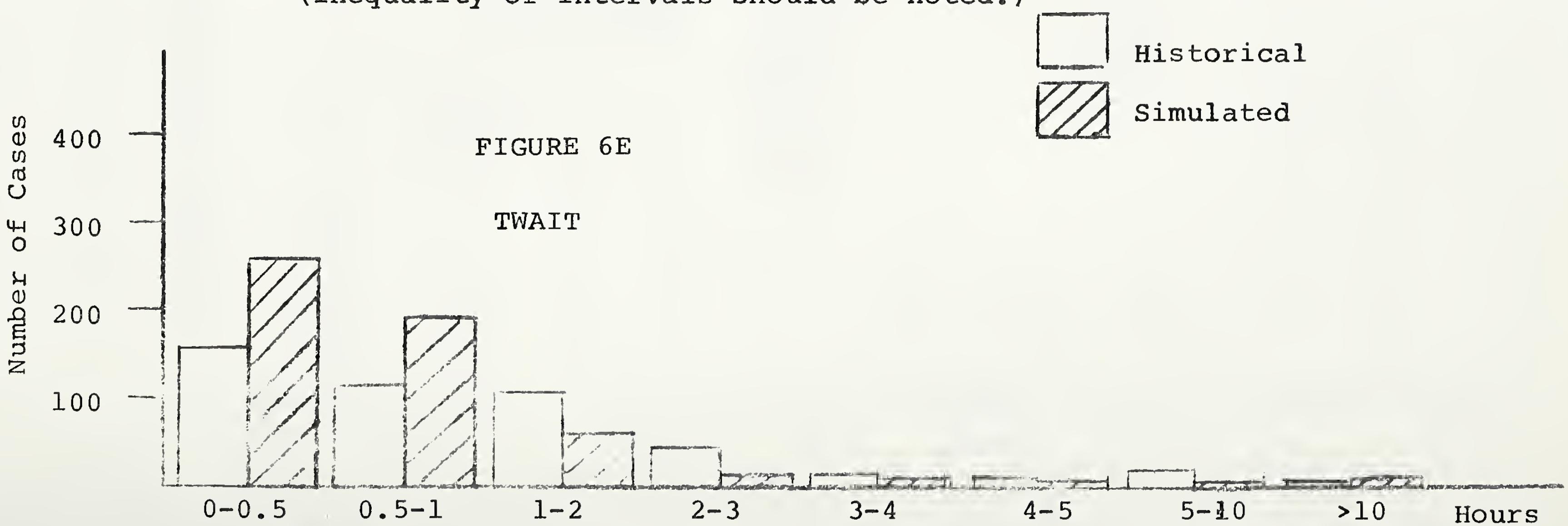


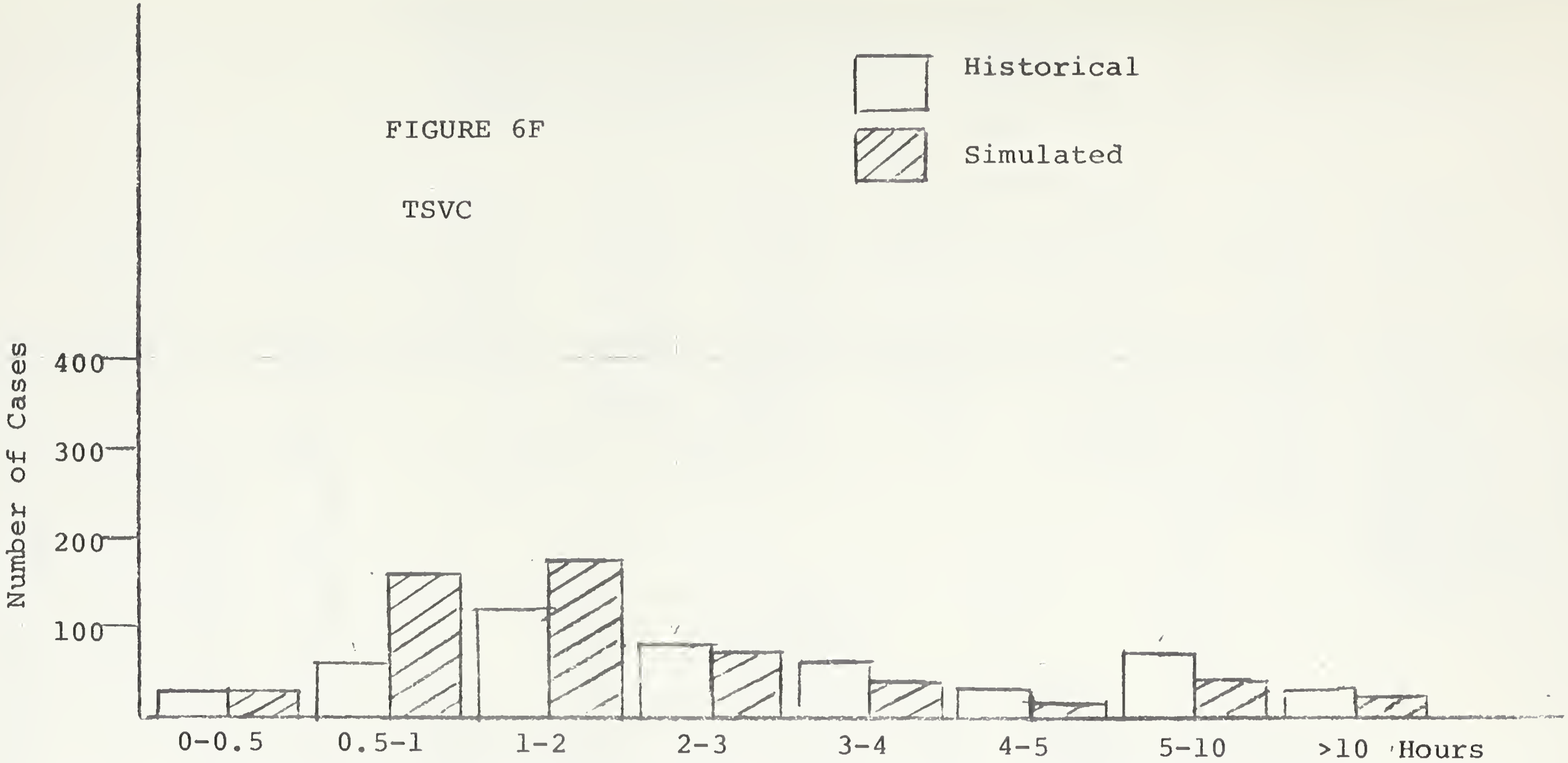
FIGURE 6E

TWAIT

Comparison of numbers of historical and simulated cases for which waiting times were within the indicated intervals for District 7, August 1968. (Inequality of intervals should be noted.)

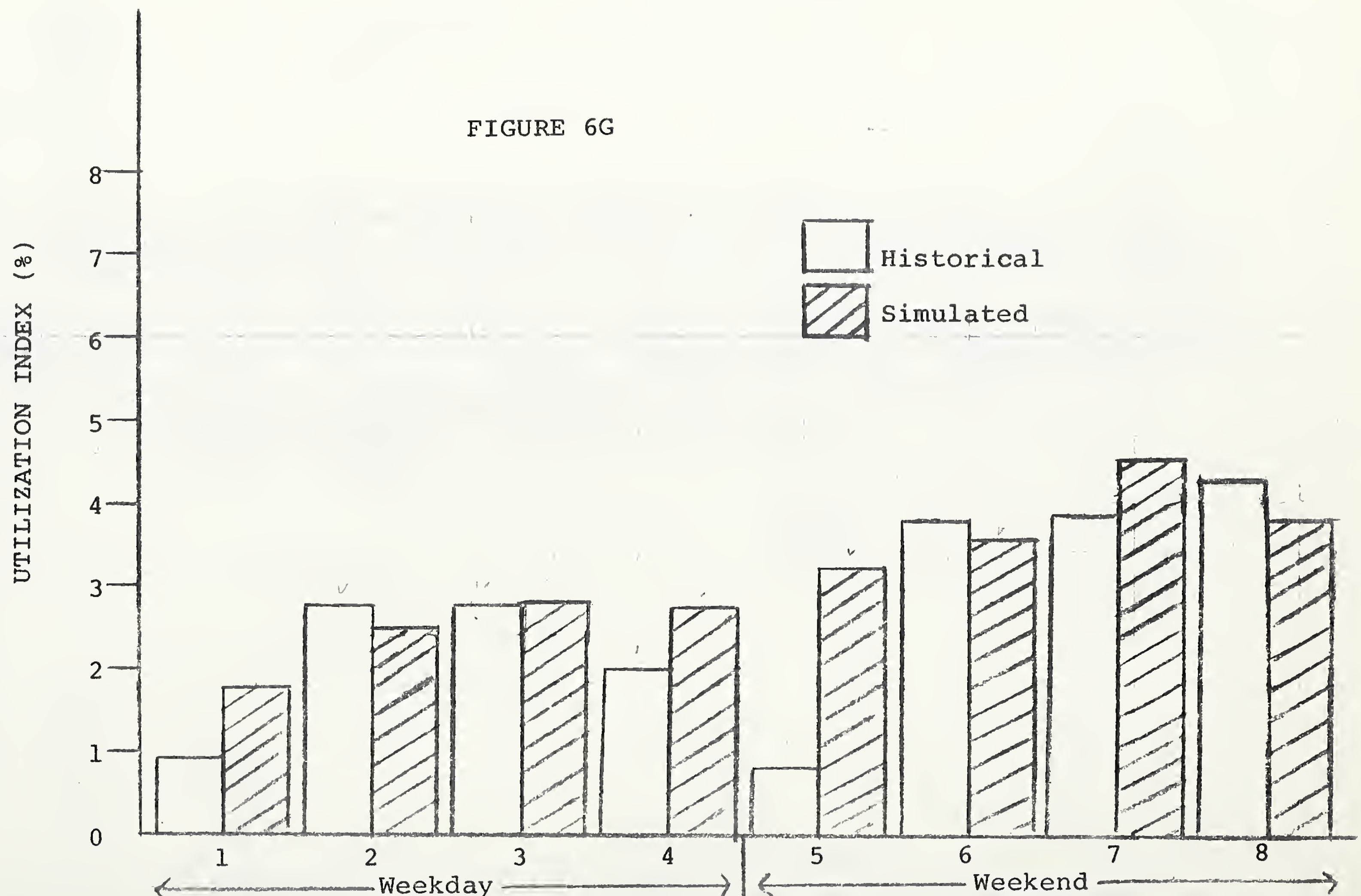
FIGURE 6F

TSVC



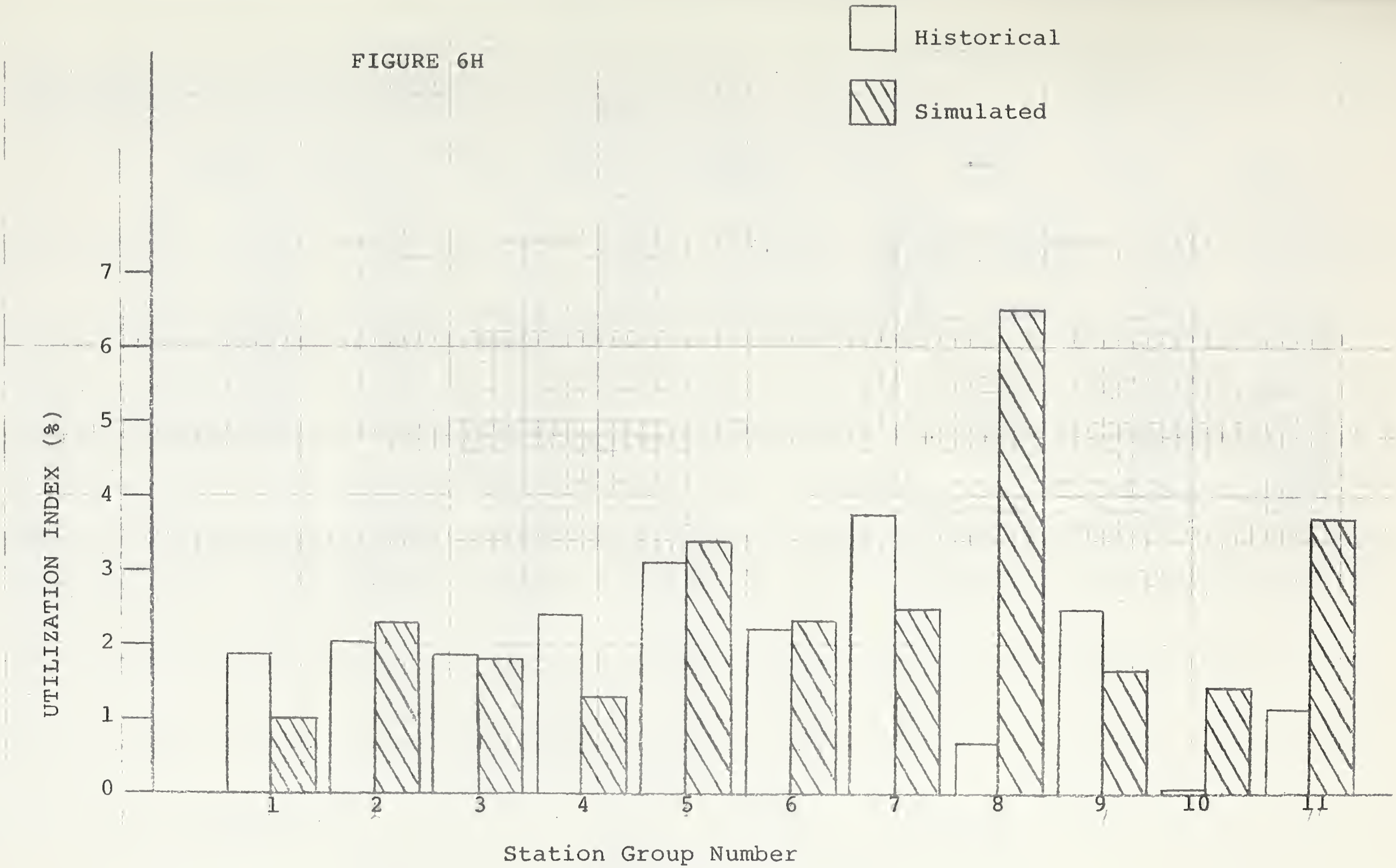
Comparison of numbers of historical and simulated cases for which service times were within the indicated intervals for District 7, August 1968. (Inequality of intervals should be noted.)

FIGURE 6G



Comparison of historical and simulated utilization indices by shift for District 7, August 1968.
(Note: see comment in text regarding inapplicability of historical data in this context.)

FIGURE 6H



Comparison of historical and simulated combined station group utilization indices for Seventh District, August, 1968.

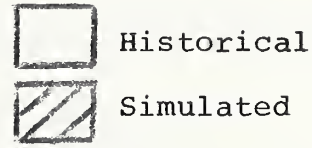
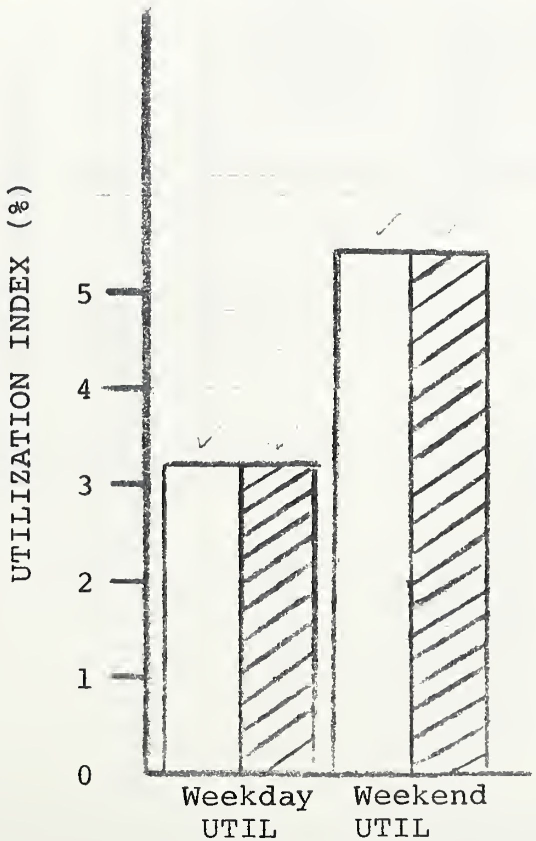
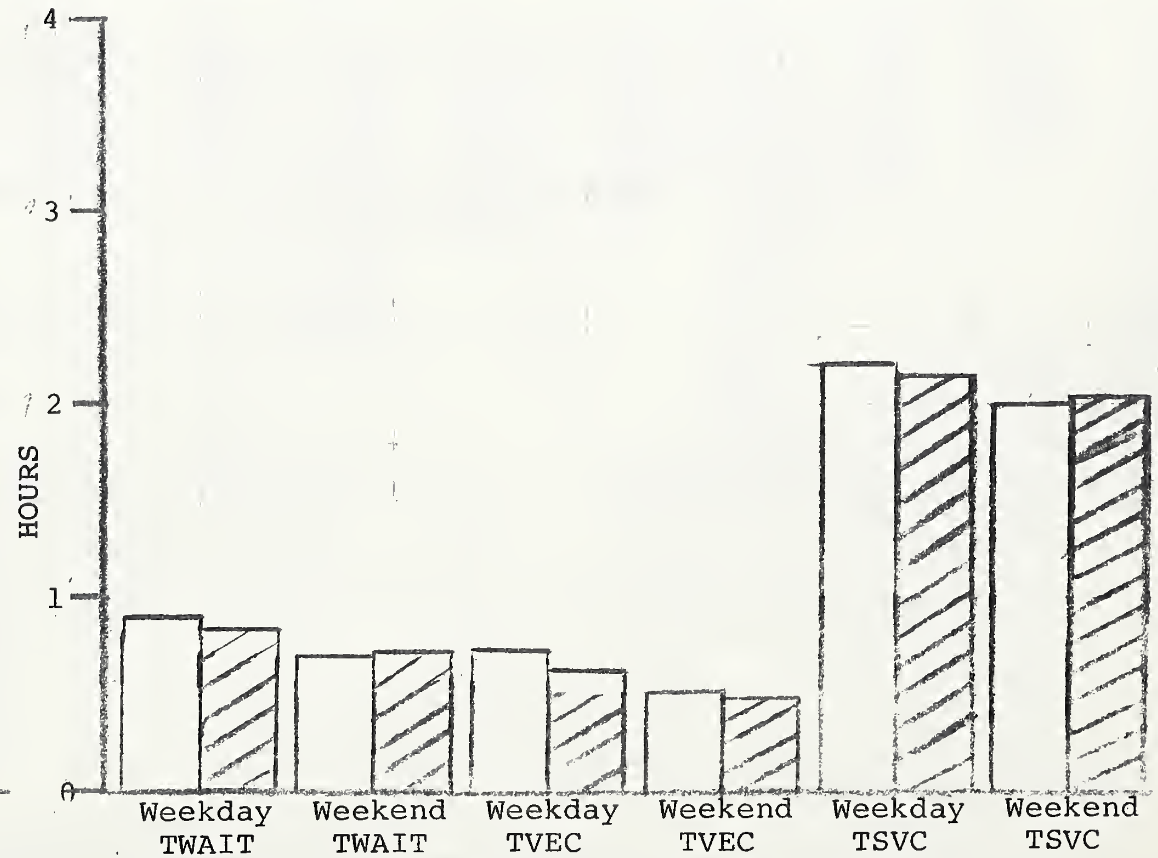


FIGURE 7A



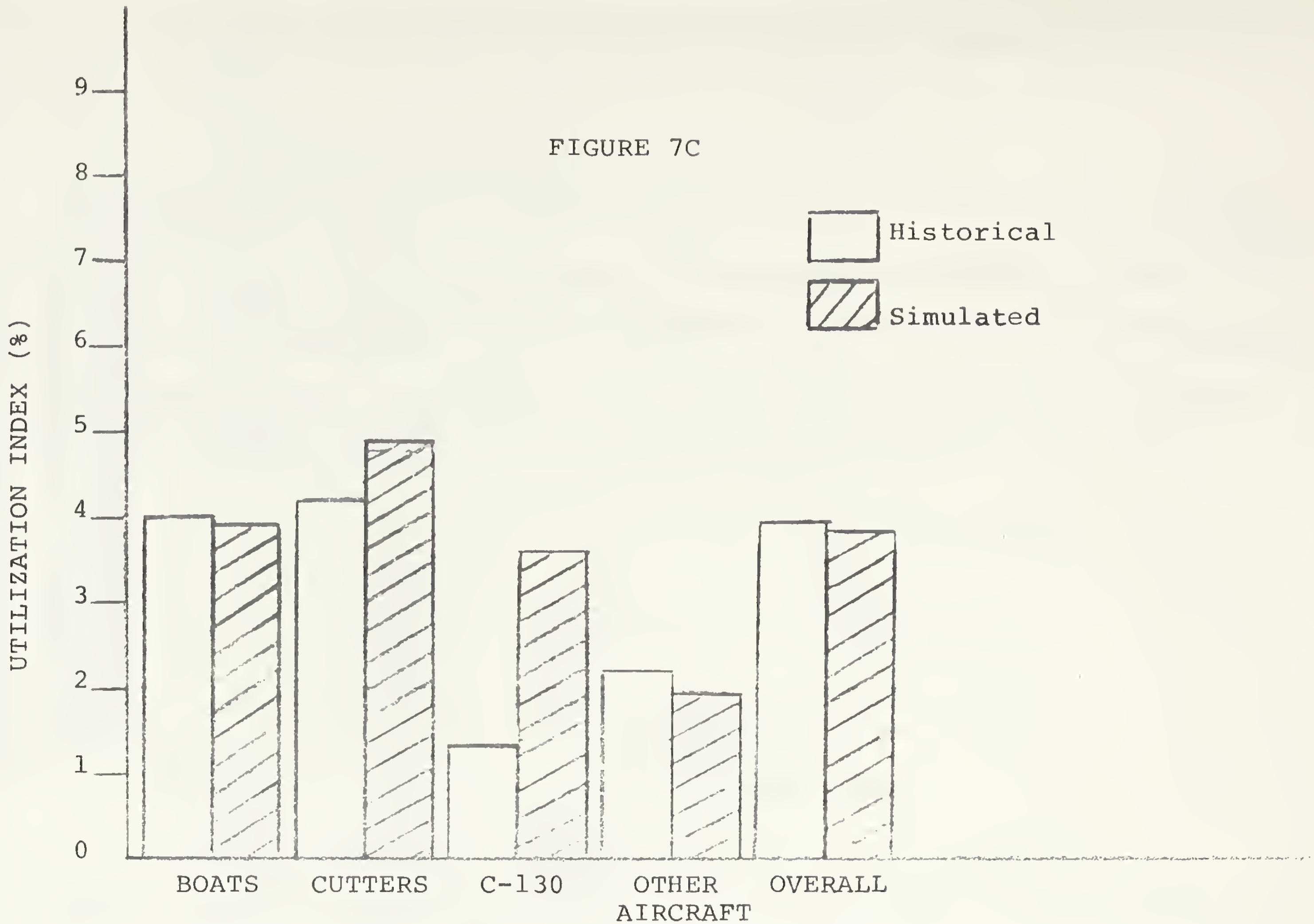
Comparison of historical and simulated utilization indices for District 13, July, 1968.

FIGURE 7B



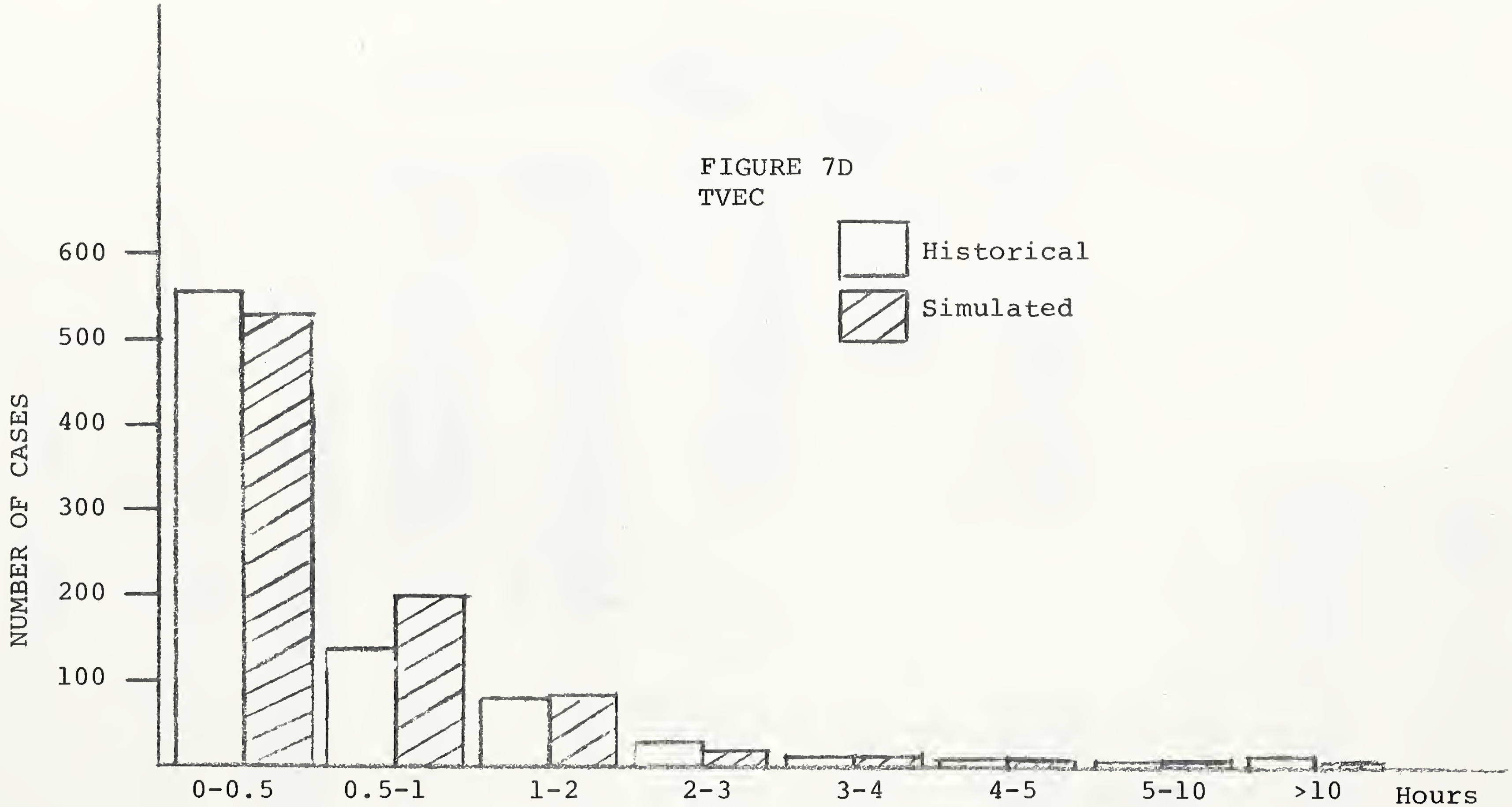
Comparison of historical and simulated output statistics for District 13 in July 1968.

FIGURE 7C



Comparison of combined Resource-Type Utilization Indices for District 13, July, 1968 (Run 60).

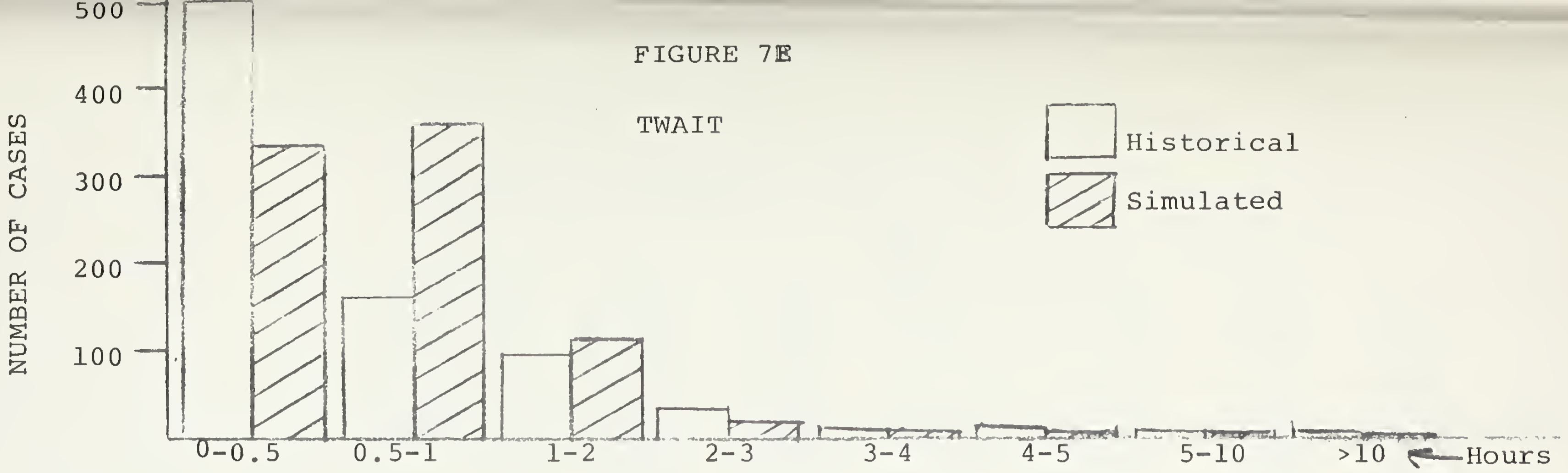
FIGURE 7D
TVEC



Comparison of numbers of historical and simulated cases for which vector times were within the indicated intervals for District 13, July, 1968. (Inequality of intervals should be noted.)

FIGURE 7E

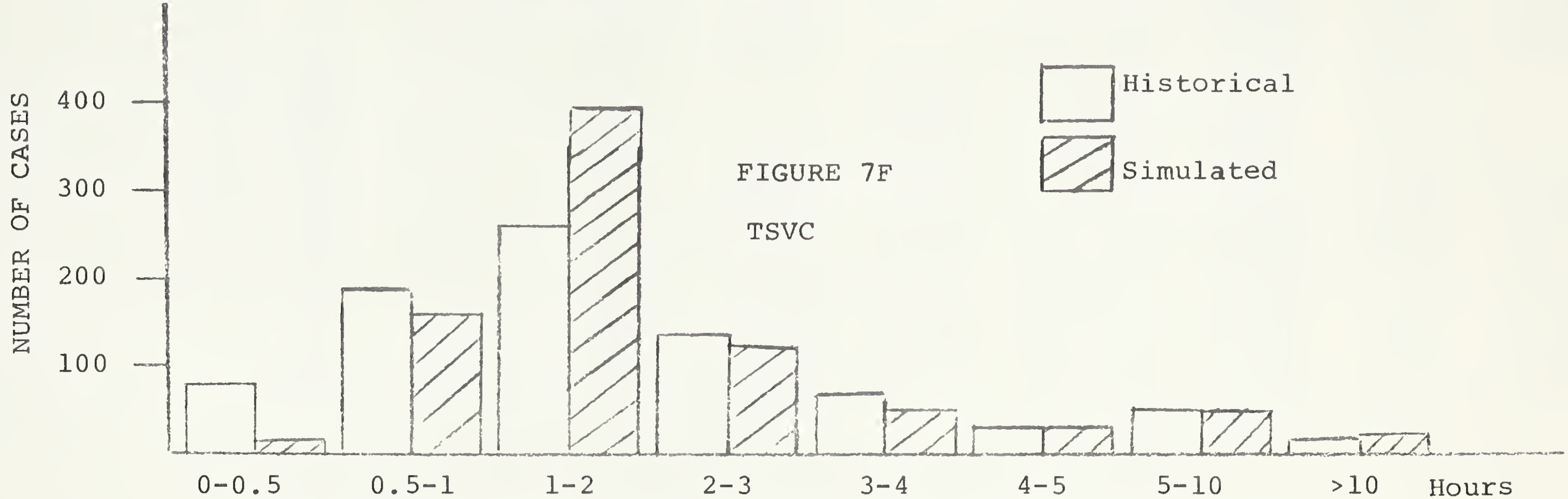
TWAIT



Comparison of numbers of historical and simulated cases for which waiting times were within the indicated intervals for District 13, July 1968. (Inequality of intervals should be noted.)

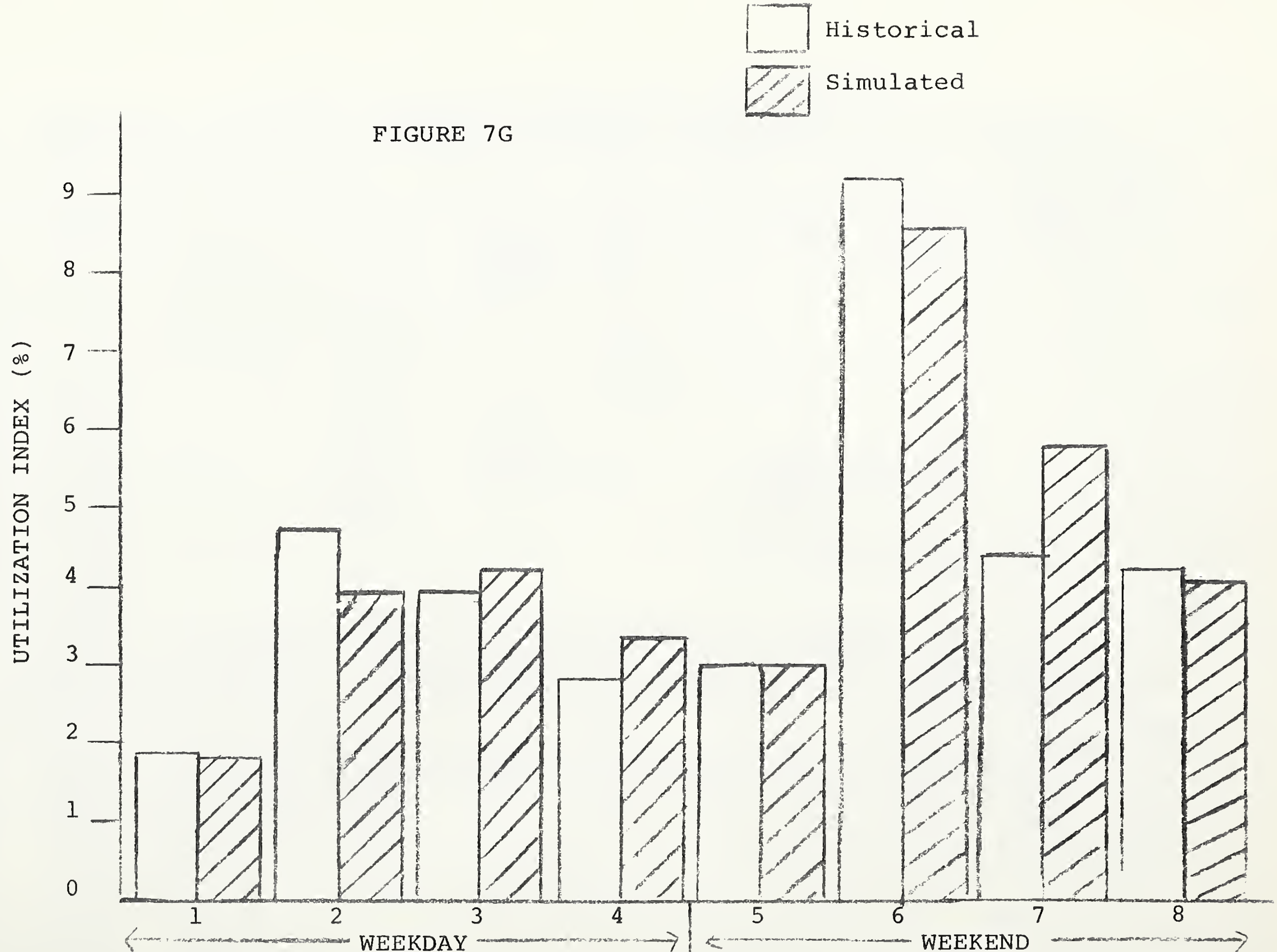
FIGURE 7F

TSVC



Comparison of numbers of historical and simulated cases for which service times were within the indicated intervals for District 13, July 1968. (Inequality of intervals should be noted.)

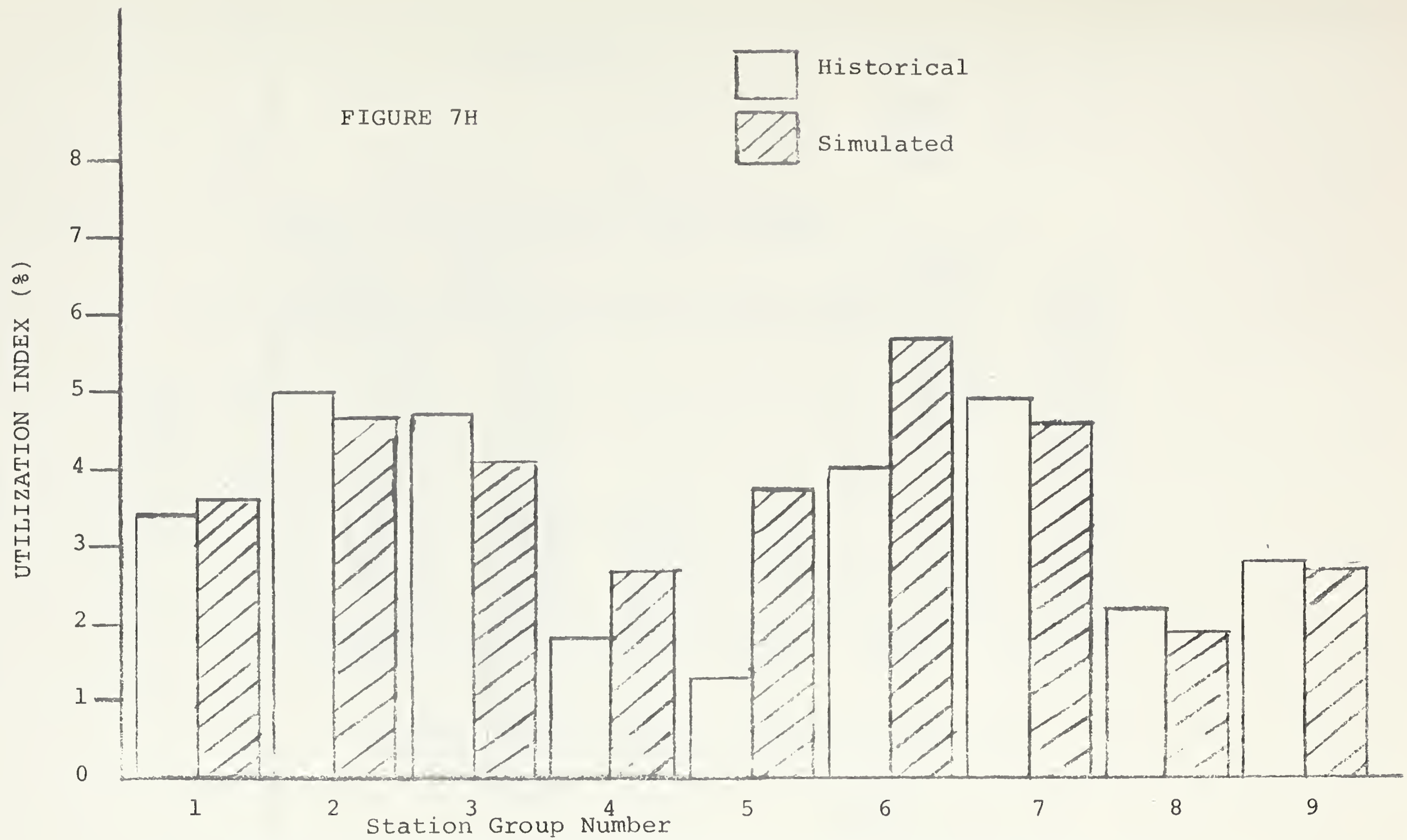
FIGURE 7G



Comparison of historical and simulated utilization indices by shift for District 13 in July 1968.

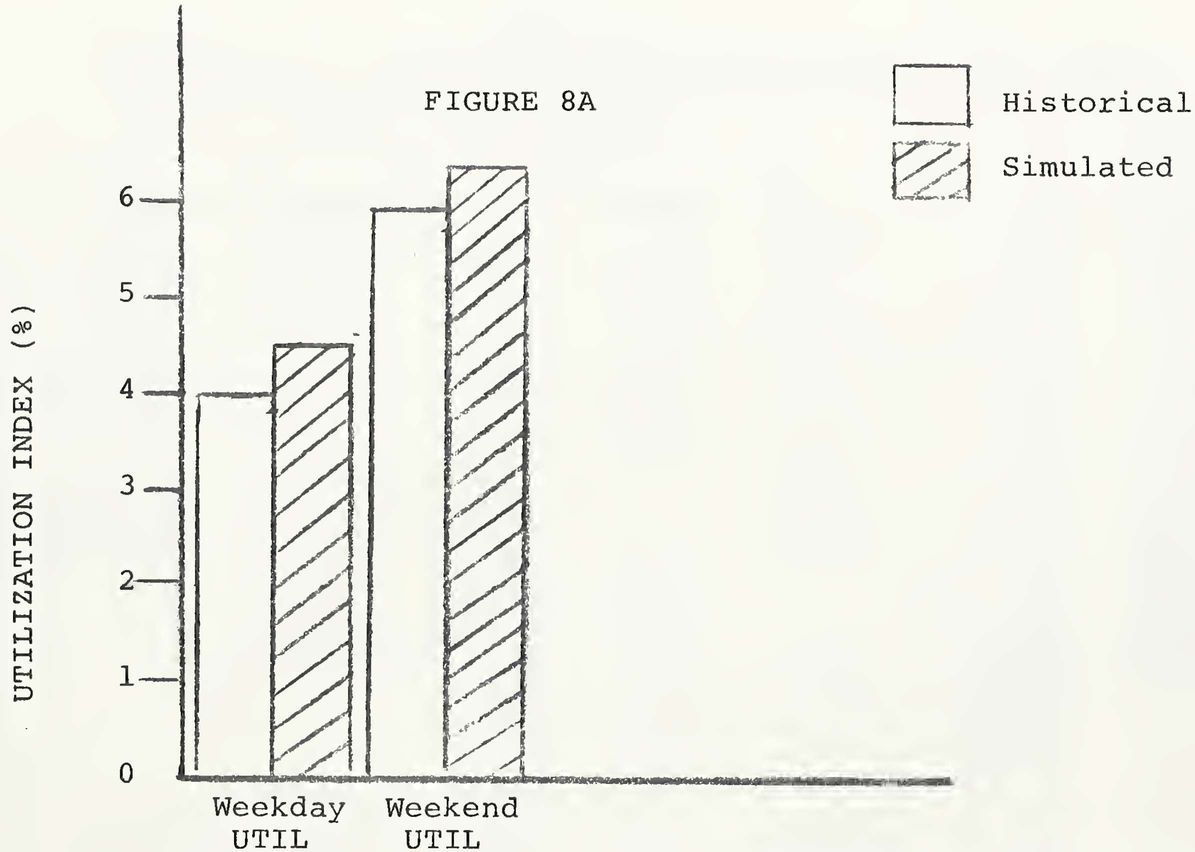
(Note: See comment in text regarding inapplicability of historical data in this context.)

FIGURE 7H



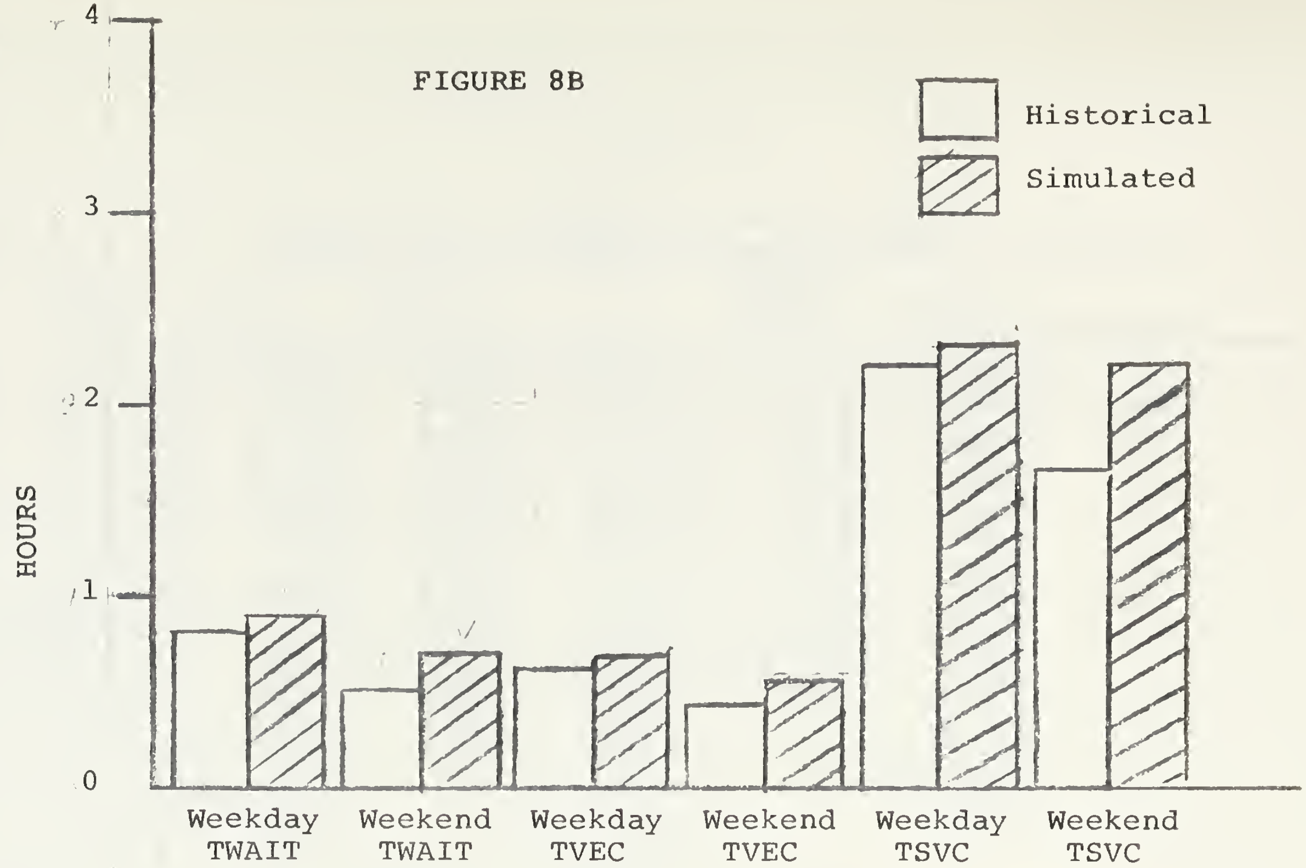
Comparison of historical and simulated combined station group utilization indices for Thirteenth District, July 1968.

FIGURE 8A



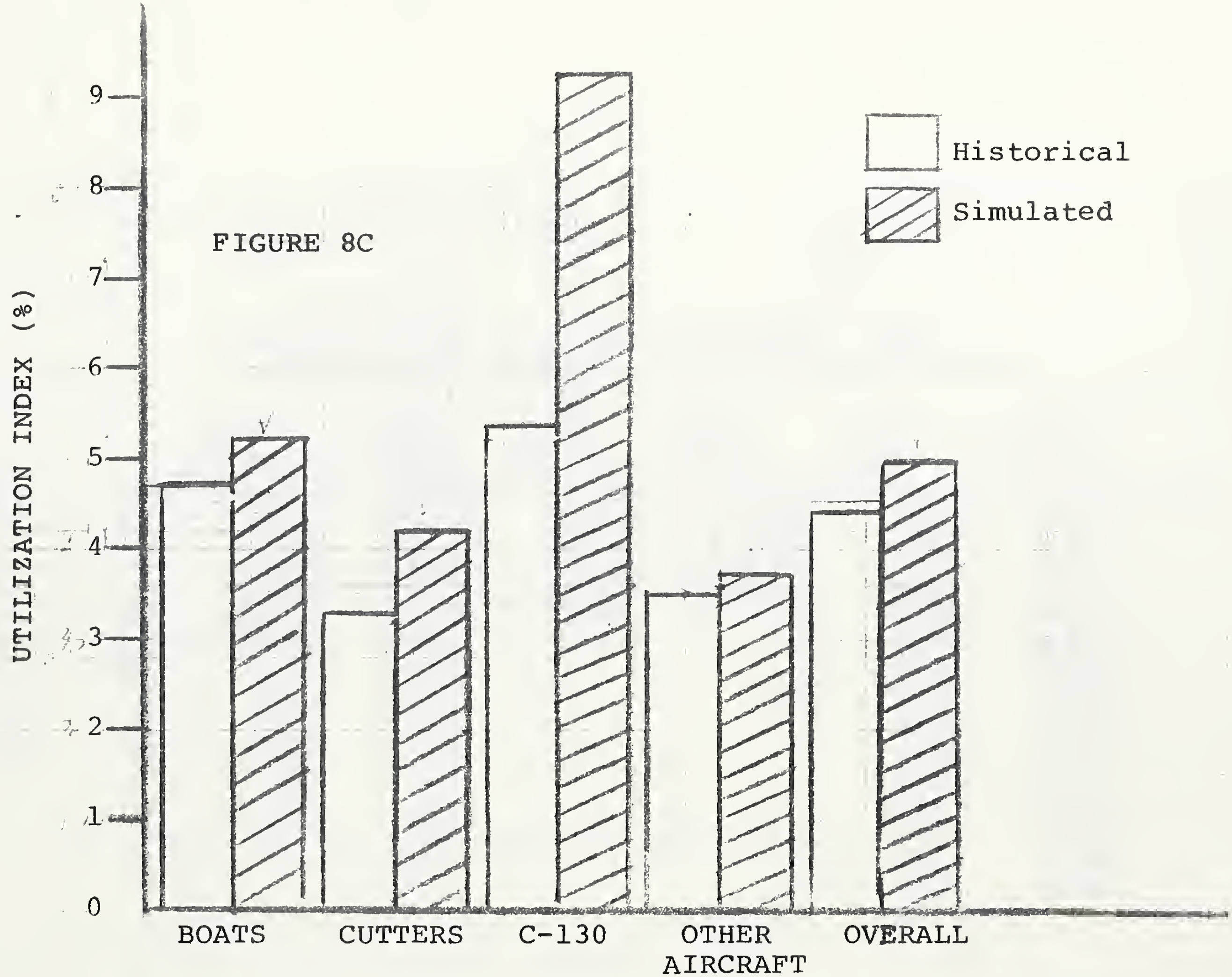
Comparison of historical and simulated utilization indices for District 13, August 1968.

FIGURE 8B



Comparison of historical and simulated output statistics for District 13, August 1968.

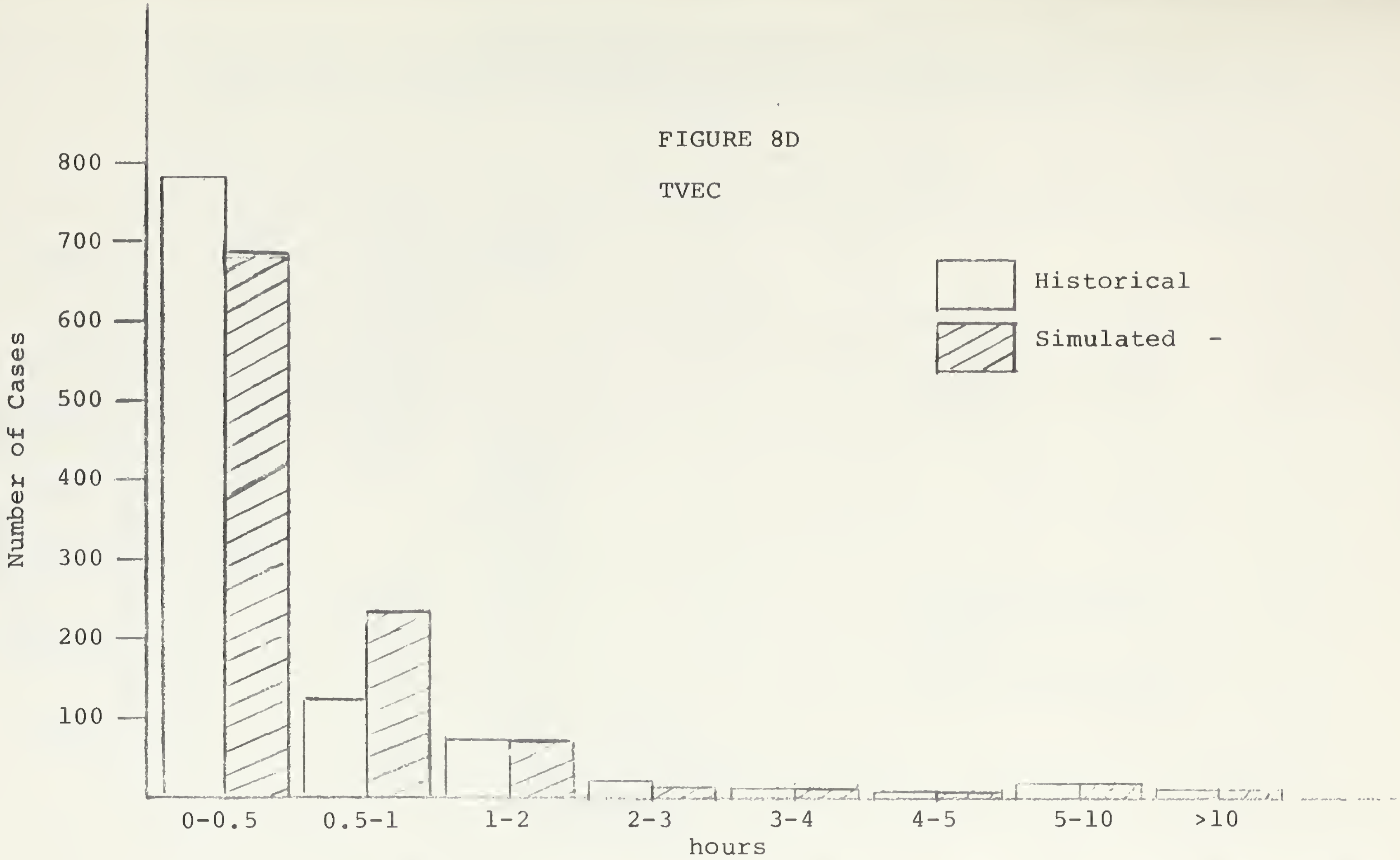
FIGURE 8C



Comparison of combined Resource-Type Utilization Indices for District 13, August 1968.

FIGURE 8D

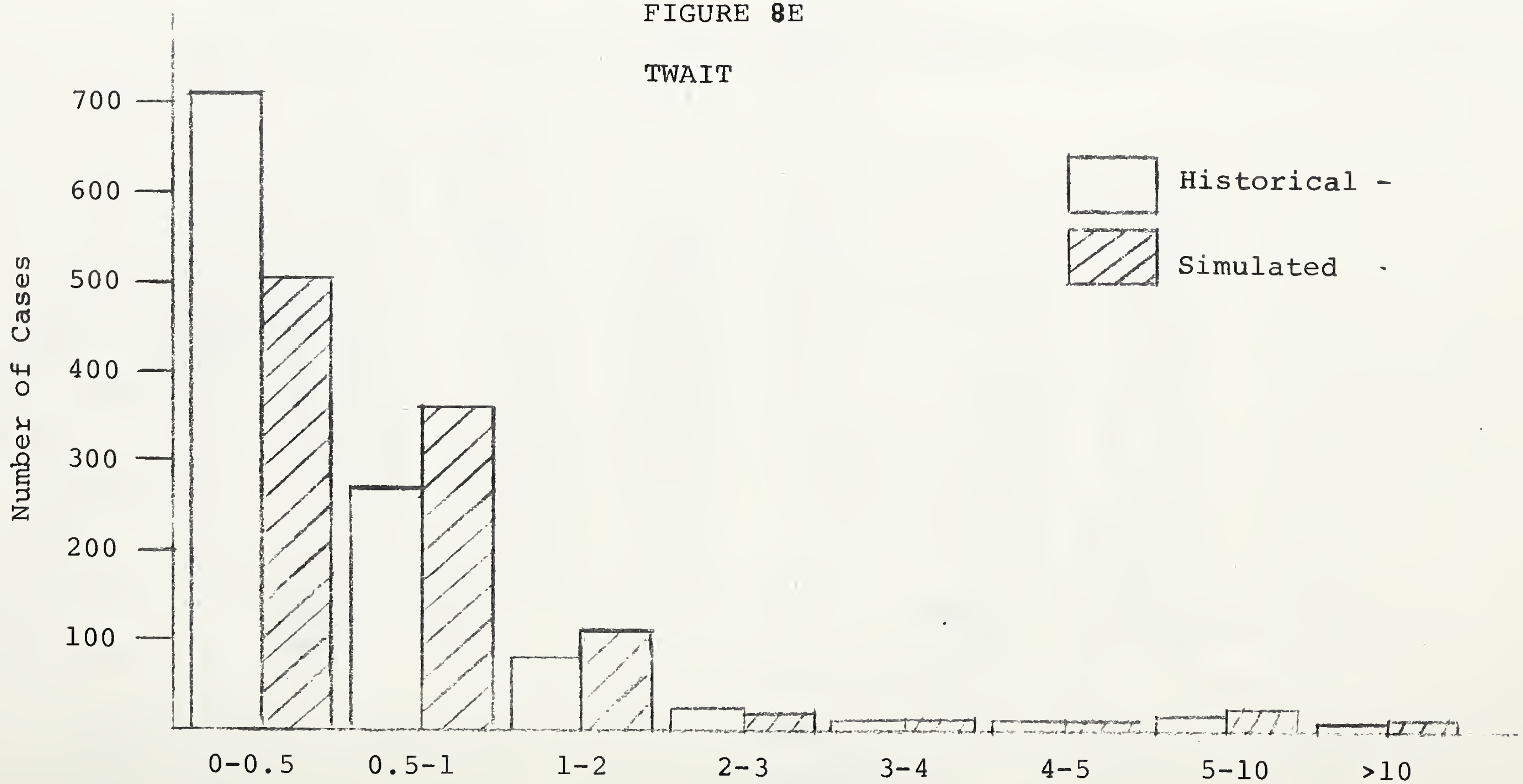
TVEC



Comparison of numbers of historical and simulated cases for which vector times were within the indicated intervals for District 13, August 1968. (Inequality of intervals should be noted.)

FIGURE 8E

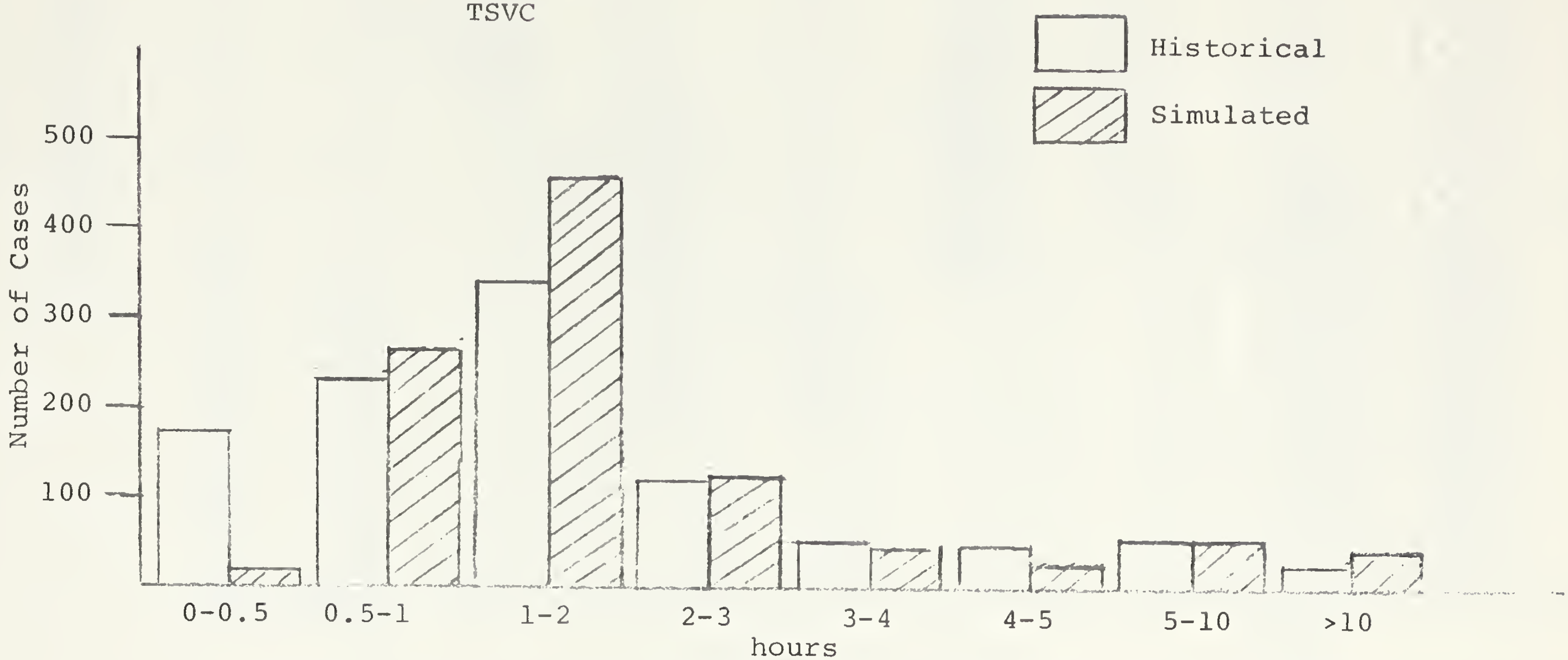
TWAIT



Comparison of numbers of historical and simulated cases for which waiting times were within the indicated intervals for District 13, August 1968. (Inequality of intervals should be noted.)

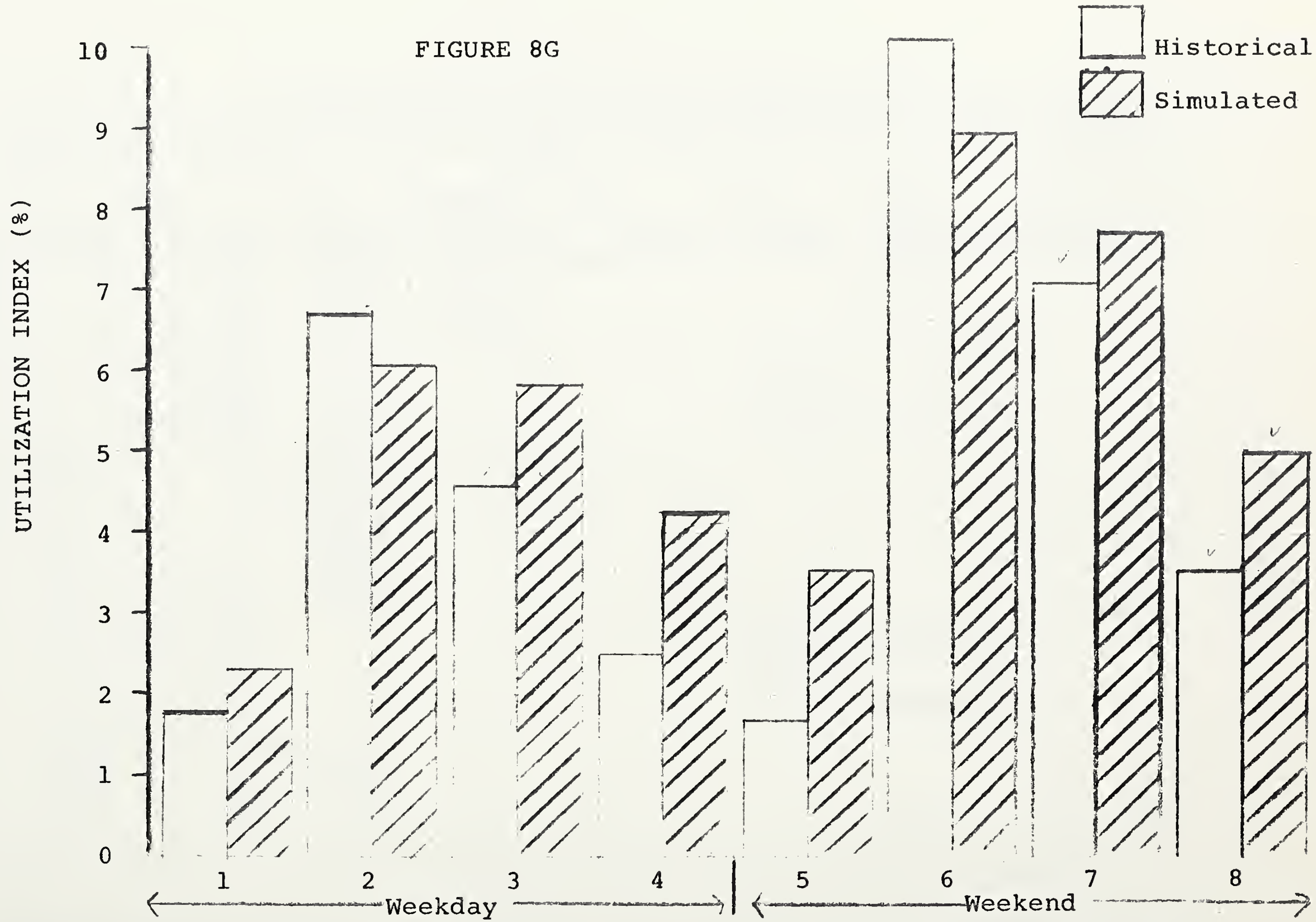
FIGURE 8F

TSVC



Comparison of numbers of historical and simulated cases for which service times were within the indicated intervals for District 13, August 1968. (Inequality of intervals should be noted.)

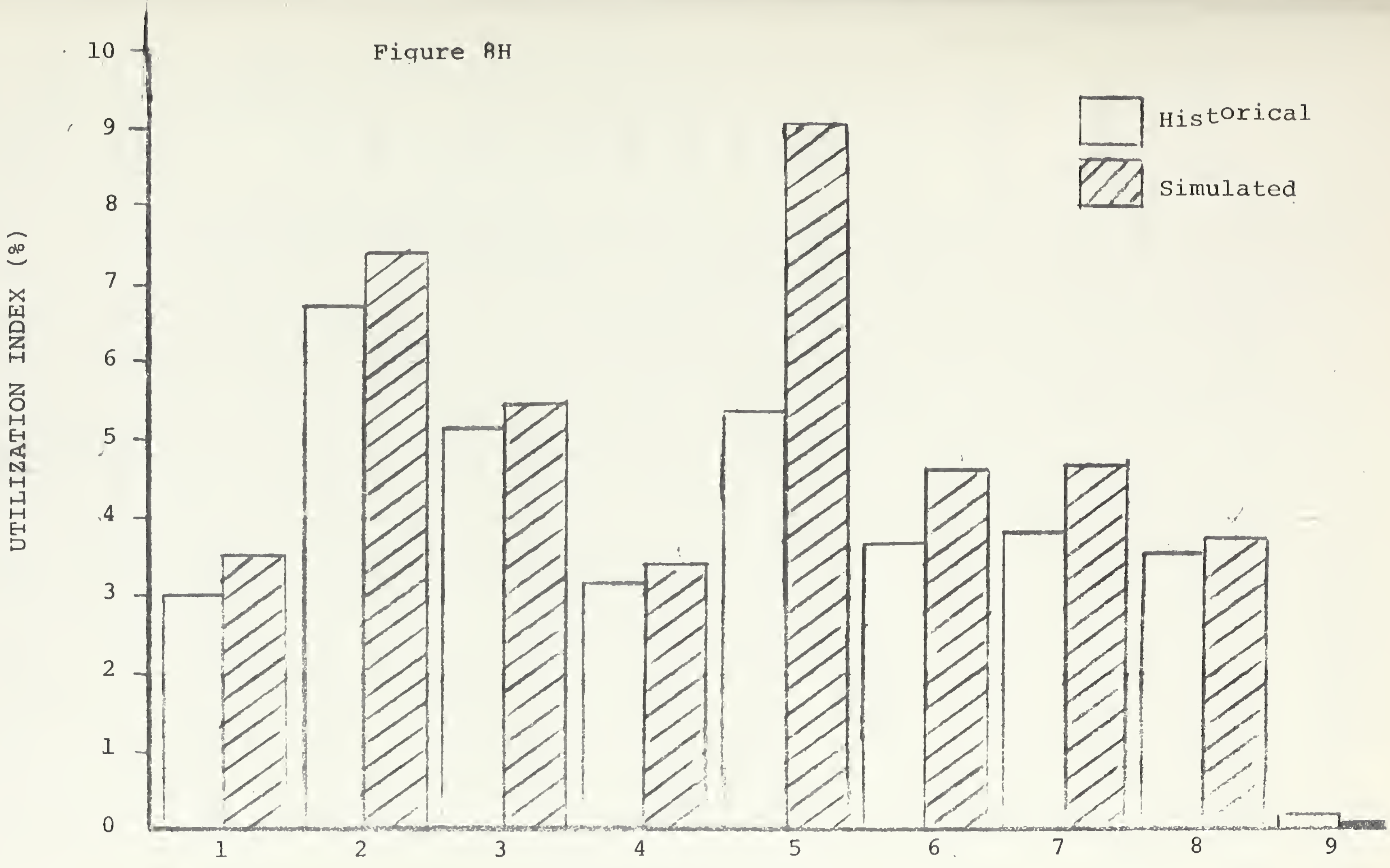
FIGURE 8G



Comparison of historical and simulated utilization indices by shift for District 13, August 1968.

(Note: see comment in text regarding inapplicability of historical data in this context.)

Figure 8H



Comparison of historical and simulated combined group utilization indices for Thirteenth District, August 1968.

The comparisons between simulated and historical outputs shown by the 64 graphs are, on the whole, suggestive of a good "fit", strengthening confidence in the simulation as a valid representation of the real SAR process. Noteworthy exceptions to the satisfactoriness of fit are itemized and discussed below, identified by the appropriate number and letter from Tables 2.3.2.1 and 2.3.2.2.

1C - The utilization index for "other aircraft" (i.e., helicopters and fixed-wing amphibians) from the simulation is about double the historical value. A possible explanation is that these resource types are not assigned as often as they might be in actual practice. The historical utilization indices for these aircraft could be replicated better in the simulation by increasing their "perceived" operating costs in the model. On the other hand, the values shown in the figure may be considered as normative values to illustrate that this resource type should be used more frequently for SAR. Since the July 1968 caseload was characterized by more search requirements than the typical peak period, and since SARSIM favors aircraft for long search cases, some of the discrepancy between historical and simulated utilization index in this figure

may be attributable to the selection of this particular scenario.

5,6- The results for the Seventh District are generally much poorer than those for the First and Thirteenth Districts, due mainly to the small amount of time available for calibration prior to the preparation of this report. The district is geographically quite large and, perhaps, better considered as two distinct sub-districts; additional efforts have been directed toward recalibration. (It may be noted that the variance for TVEC for the Seventh District is three to four times the variance in the same quantity in the other two districts.)

E - (All sets) -- The simulated values of TWAIT compare less favorably with their historical counterparts than do any of the other time-related outputs, as discussed in Section 1.3.

G - (All sets) -- The utilization indices for different shifts do not fit as well as other measured statistics. As discussed in Section 2.2.1, the difference in manner of calculation should preclude direct comparisons of shift utilization indices from HSTAT and OPSIM.

H - (All sets) -- Group utilization indices from the simulations do not always agree well with historical values, mainly because resources are temporarily transferred into and out of SAR groups. Many of such transfers are for operational reasons (e.g., assigning a SAR vessel to non-SAR activity, or vice versa), for scheduled overhauls, for unscheduled maintanance, etc. A large amount of effort would be required to account for all the shifts in assignments for each district and each time period, hence it was decided to accept the inaccuracies.

The exceptions listed above should not be over-emphasized: the comparisons shown in the 64 graphs are considered to be quite good. With additional calibration effort (as for the Seventh District), even closer results might well be obtained. It should be recalled that exact duplication of historical statistics is not to be expected since both the SAR system and the simulation model are quite complex. Moreover, little account can be taken of the human element, especially when decisions are based on personal preference and tradition, rather than consistent logic.

2.3.3 Results from Statistical Testing

2.3.3.1 Chi-Square test for Goodness-of-Fit

A test is available to determine the goodness of fit between simulation and actual data.* It involves the calculation of the chi-square (χ^2) statistic from the expression:

$$\chi^2 = \sum (f-f_o)^2/f$$

where f = theoretical (simulated) frequency,

and f_o = observed (historical) frequency.

The chi-square test may be used to test any of a variety of hypotheses: results expected on the basis of a given hypothesis (frequency of occurrence in this context) are compared with the results of observations. The computed value of χ^2 is interpreted by referring to printed tables which list maximum acceptable values for stated confidence levels. (Perfect agreement between "predicted" and actual values yields $\chi^2 = 0$; the larger the value of χ^2, the poorer the agreement.) If the computed value of χ^2 exceeds the tabulated, the disparity is considered to be too large to be ascribable to chance at the stated confidence level and the (null) hypothesis that the simulation represents the process is rejected. (Thus, for example, if the χ^2 calculated exceeds the published value at the 0.05 level, the probability that the anomaly is due to chance is less than 5%.)

---

*Herbert Orkin and Raymond Cotton "Statistical Methods," Barnes and Noble, New York City, 1967, page 109.

The chi-square test was conducted for several simulation runs for the First and Thirteenth Districts, using the histograms of various time attributes. The results of these tests were all satisfactory, as summarized below:

District	Period	Run	Parameter Tested	χ^2 value	Value for rejecting hypothesis at .05 level
1	August	64	TVEC	7.404	9.488
1	Winter	61	TVEC	0.868	9.488
1	Winter	61	TSVC	4.932	12.592
1	July	45	TVEC	8.875	11.071
1	Hist/ Demgen	*	TVEC	1.854	11.071
1	"	*	TWAIT	1.423	11.071
1	"	*	TSVC	12.039	14.067
13	July	60	TVEC	10.454	11.071

*Run 42 vx. Runs (62 and 63) Averaged.

Other areas of SARSIM output did not generally lend themselves to chi-square testing. First, application of the chi-square and other goodness-of-fit tests are generally limited to data for which histograms are available. For most of the system output from SARSIM, many runs under identical input conditions would be required to obtain such histograms. Of case outputs which can be tested this way, the best fits are obtained for TVEC.

As discussed in Section 1.3, TWAIT is not a good validation criterion because of the variability in "overdue" search cases. TSVC outputs generally involve large discrepancies between historical and simulated results for the 0-0.5 and 0.5-1.0 hour intervals. While these differences do not appear critical from the graphical output (see Section 2.3.2), the differences in these intervals, for such a large number of cases, would generally cause large contribution to χ^2. Before satisfactory chi-square tests could be applied to TSVC, it would be necessary to modify the statistical programs to discriminate more finely in the 0-1.0 hour interval.

The results of chi-square tests which were performed give additional evidence of the validity of the SARSIM model.

2.3.3.2 Sign Tests

The sign test is a popular non-parametric technique which serves as a short-cut for estimating goodness of fit. Although the simplicity of this test is attractive, its lack of "power" limits its usefulness. This test provides a basis on which to reject the validity of a model if the test fails.

When validating the Preprocessor, several runs of DEMGEN were made to test whether results under each demand category (e.g., the number of cases, time on scene, total miles searched, etc.) would cluster about their averages without excessive bias in either direction. This test was made several times and DEMGEN's data was acceptable each time.

Several runs of OPSIM were also checked against HSTAT output to determine whether there was any bias for such outputs as individual station utilization indices, individual resource type utilization indices, shift utilization indices, individual station TVEC's, TWAIT's, etc. All these tests were successful.

2.3.3.3 Difference of Proportions Test

A standard statistical Z - test can be applied to determine whether the porportion of all cases in the simulation served by any given resource type differs significantly from the proportion served historically by the same

resource type. The normalized z-statistic is calculated and compared to tabulated values for a pre-selected confidence level.*

Letting N be the number of cases and R the number served by the given resource type, the proportions of interest are:

$$p_H = \frac{R_H}{N_H} \qquad \text{and} \qquad p_S = \frac{R_S}{N_S}$$

where the subscripts H and S refer to Historical and Simulated, respectively. The normalized difference of proportions is

$$z = \frac{p_S - p_H}{\sigma}$$

where $$\sigma = \sqrt{\frac{N_H + N_S}{N_H N_S} \hat{p}\,\hat{q}}$$

$$\hat{p} = \frac{R_S + R_H}{N_S + N_H}$$

and $$\hat{q} = 1 - \hat{p}$$

This test has been performed with respect to selection of the C-130 for assignment and also for the "other aircraft" category. It was felt that the simulated assignment of aviation resources was not as closely matched to historical performance as was assignment of boats and of cutters. Therefore, if the test succeeded

---

*See Social Statistics, by Hubert M. Blalock, Jr. McGraw-Hill, New York, New York.

for air resources, it would undoubtedly be successful for the water-borne resources too.

For the two-sided hypothesis test with 0.05 probability of Type I error pre-selected as the confidence level, the critical value of Z (for rejection) is $\pm$ 1.960. As shown in the table below, the test was applied eight times; only one case ("other aircraft" in the First District, July 1968) yielded a critically high Z-value. All other results were satisfactory, supporting the validity of the resource selection algorithms.

District	Run	Period	Resource Type	Z-value	
01	45	July	C-130	0.2130	
01	45	July	Other Aircraft	3.5887	(Rejected
13	60	July	C-130	1.7548	
13	60	July	Other Aircraft	1.2289	
01	61	Winter	C-130	0.3089	
01	61	Winter	Other Aircraft	0.9769	
07	80	July	C-130	0.5711	
07	80	July	Other Aircraft	1.4209	

2.4 Calibration of the Model for Individual Districts

2.4.1 Need for Individual Calibration

As described in Section 2.3, runs of OPSIM have been made with input caseloads as they occurred historically (using the HIST tape from PREPRO) and varying other input initial conditions until the output resembled historical

results, as calculated by HSTAT. The original purpose of the calibration runs was to set internal parameters of the model while testing the logic and programming of OPSIM to insure its consistency with real world processes.

A base run was made for each of three test districts (the First, Seventh and Thirteenth). For each of these base runs, a set of "standard" values was selected for model inputs by knowledgeable individuals familiar with the geography, resources, and policies of the districts. The inputs included values for types of resource, speed and cost; the percentage of search area to be covered during the first day of search; towing speeds; etc. Over a series of experimental runs for each district, the standard values were modified until a good fit was achieved between OPSIM output and HSTAT.

As might be expected, the three districts which were calibrated at this time differed from one another in many respects, some of which turned out to be critical with regard to setting calibration parameters. The differences included geographical configurations, operating policies, prevailing weather conditions, traditions, and so forth. In the First District, for example, there was low utilization of aircraft historically. At some stations of the Thirteenth District there were bar patrols, and account of these had to be taken for the model to reflect actual practice validly.

The greatest difficulty was encountered in calibrating the Seventh District. As can be seen in the first two columns of Table 2.4.1, the base run doesn't match the historical results at all well. In an attempt to achieve a fit by calibration, a series of nine experimental runs was made until the simulation output resembled the historical. However, this necessitated varying some of the input values to unrealistic extremes. In other words, the model was being forced into compliance in a manner which reduced confidence in its validity. Accordingly, it was concluded that further analysis would be required to ascertain the peculiarities of this particular district and to permit more appropriate adjustment of the model's internal parameters.

This experience also led to the realization that each district would have to be examined so that its singularities might be identified and accounted for during its calibration.

2.4.2 Calibration of the Seventh District

In order to obtain a more realistic calibration of the Seventh District, a detailed analysis was made of the various outputs of the experimental calibration runs in an attempt to pinpoint district peculiarities which might be the cause of discrepancies. On careful review it appeared that there were three major factors which were somewhat interrelated: the size of the district, its separability into two disperate elements, and the treatment of cases occurring outside the nominal boundaries.

2.4.2.1 The Broad Extent of the Seventh District

The Seventh District consists of the coastal regions of Florida and Georgia and the area around the Greater Antilles, thus extending from the Gulf Coast of Florida to the Atlantic Coast of the Southeastern U. S. and the waters around Puerto Rico, the Virgin Islands, and vicinity. Since, conceivably, the simulation model might be stressed when applied to such a large area, an experiment was conducted with the nominal district limited to the coasts of Florida and Georgia. This was accomplished by rerunning the historical data through PREPRO with the more constrained district boundaries, then exercising OPSIM over this smaller area. The results are shown in the third column of Table 2.4.1 as Base Run 2. (These are comparable to the Historical Outputs shown by the right-hand figures in the first column of the table.)

2.4.2.2 Separability of Sub-Districts

The analysis of the smaller area described above suggested that, in effect, the Seventh District consists of two distinct components, namely, the mainland section and the Greater Antilles. This hypothesis was tested while exploring another technique for segregating the historical data on a geographical basis. Whereas Base Run 2 had been derived from artificially setting boundary lines for the district, the OPSIM input tape was now run through a utility program to divide the cases according to whether the stations which originally served them were in the mainland section or in the Greater Antilles. Base Run 3, shown in column 5 of Table 2.4.1, represents only the mainland portion of the input caseload, again comparable to the Historical Output figures to the right of the solidus in column 1.

2.4.2.3 Out-of-Area Cases

The significance of the separability of a district into two components is generally not apparent. Although the simulation model specifically allows for interaction among the stations of a district, most such interactions, both in the real world and in the simulation, are fairly well localized. Therefore, when sections of a district are well removed from one another, there is generally little opportunity for significant interaction.

In all districts, however, there are two circumstances which lead to noteworthy interaction, one real and the other apparent in the simulation. The first of these occurs when a Coast Guard resource, perhaps in transit far from its home station (e.g., on the way to the yard for repairs), is called on to service a case in distress near its then current position. For record purposes (and therefore reflected in the simulation input data), this out-of-area service is credited to a distant station, often outside the district's boundaries.

The second anomaly stems from the manner of treating, for simulation purposes, cases occurring anywhere outside the prescribed boundaries, albeit only by a short distance. These cases are arbitrarily reassigned to a location of (2,2) (i.e., two miles north and two miles east of the district's origin), thus creating a fictitiously larger caseload in this area; however, the simulation model generally operates to assign as primary station the one which historically serviced the case. (In the Seventh District, for example, 42 of the 605 cases in the historical calibration sample occurred outside the district boundaries, and 24 of these were historically served by stations in the Greater Antilles. Since the (2,2) location is close to Fort Lauderdale, Florida, the simulation effectively had resources from the Antilles travelling some 300 miles to provide search and rescue assistance!)

UTILIZATION INDICES (%)

		HISTORICAL OUTPUT*	BASE RUN 1 ENTIRE DISTRICT	BASE RUN 2 MAINLAND SECTION	BASE RUN 3	BASE RUN 4 ENTIRE DISTRICT
OVERALL		3.90/4.40	6.89	3.79	3.79	3.95
RESOURCE TYPE	1	4.88	8.87	7.38	7.24	7.43
	2	2.18	6.54	3.65	3.89	3.84
	3	.74	.32	.13	.12	.14
	4	7.18	3.27	5.05	5.05	5.05
	5	0	0	0	0	0
	6	0	0	0	0	0
	7	.03	0	0	0	0
	8	4.88	18.55	7.24	7.14	7.06
	9	0	7.18	13.32	11.37	13.43
	10	0	0	0	0	0
	11	2.51	9.45	.73	1.32	3.42
	12	0	0	0	0	0
	13	2.75	8.36	3.62	3.33	3.56
	14	5.61	5.88	2.86	3.07	2.99
	15	2.54	5.86	2.58	2.49	2.73
	16	0	0	0	0	0
COMBINED UTIL. INDEX:						
BOATS		2.70/2.91	4.95	3.66	3.67	3.74
CUTTERS		3.74/4.44	4.59	5.47	5.46	6.07
C-130		2.75/2.75	8.36	3.62	3.33	3.56
A/C		4.07/5.27	5.88	2.72	2.78	2.87
FAILURE C		66/66	98	66	65	68
OVERALL AVE. TVEC		.82/.82	1.39	.82	.81	.88
OVERALL AVE. TWAIT		1.33/1.33	1.79	1.09	1.09	1.15
DAILY UTIL. INDEX	(WD)	3.11/3.50	6.55	3.34	3.37	3.48
	(WE)	5.81/6.55	7.74	4.85	4.80	5.10
DAILY TWAIT	(WD)	1.39/1.50	1.92	1.04	1.06	1.07
	(WE)	1.25/1.21	1.52	1.09	1.08	1.20
DAILY TVEC	(WD)	.85/.86	1.47	.78	.79	.81
	(WE)	.77/.76	1.19	.81	.79	.92
DAILY TSVC	(WD)	3.61/3.63	8.11	4.14	4.24	4.30
	(WE)	3.70/3.68	4.29	3.33	3.28	3.74

*Note: Where two figures are given for Historical Output, the number to the left of the Solidus (/) applies to the entire district, whereas the right-hand figure is for the mainland section only. (Single figures are for entire district.)

As a more suitable approach to handling out-of-area cases, OPSIM can be exercised so that the closest station to a reported case of distress is assigned as the primary station, regardless of the station serving the case historically. This process increases computer running time, but only by a small amount. By means of this technique in both Base Runs 2 and 3, a more reasonable treatment of casual assignment of passing resources was effectuated. In addition, the (2,2) cases were culled from those two Base Runs to eliminate overloading stations in the Fort Lauderdale area while, at the same time, also deleting the simulated assignment of resources from very long range, which had been accountable for unusually long vector times and utilization indices.

2.4.2.4 Discussion of Results

Table 2.4.1 shows that Base Run 1 does not compare well with the historical output, but that considerable improvement was obtained in Base Runs 2 and 3 for the mainland section. One additional Base Run was therefore made for the entire district, assigning the closest station to the distressed case as the primary station and culling out any out-of-area cases; the results are shown in the last column of Table 2.5.1. These correspond sufficiently well with the historical output for the entire district that further calibration effort did not appear to be warranted.

2.4.3 Conclusions about Calibration

The problems encountered in calibrating the Seventh District lead to the following general conclusions about the calibration process:

(a) Each district must be examined for local peculiarities and calibrated to account for these. It may be expected that characteristics accounted for in one district will be manageable in the same fashion if encountered elsewhere.

(b) Size of a district does not, in itself, appear to be a problem for the simulation. The larger the district, however, the more likely it is that it consists of separable sub-districts.

(c) Components of a district can be simulated separately, without interaction with other sub-units. Several techniques have been found to separate the input data. It is also feasible, and is perhaps more realistic, to exercise the simulation over the entire district, but assigning the nearest station to a distress case to serve as the primary station.

(d) For calibration purposes, at least, it may be useful to exclude as "exceptional" those cases which were historically served outside the district's boundaries, rather than moving them fictitiously to other locations. (Note: other devices should be explored as alternatives

solutions to this problem, such as displacing out-of-area cases to the closest point within the district.)

(e) There is added evidence of the validity of the simulation model in that the programs were flexibly modified in several ways, including changes in PREPRO and in OPSIM during the calibration efforts for the Seventh District. These suggest that the basic logic and programming is sound.

2.5 Behavioral Validity Tests

Rather than relying solely on formal statistical tests for validation, as discussed in Section 2.3, more behaviorally-oriented tests were also introduced. Three such tests, intended to establish confidence, refer to "face validity", "variable-parameter validity", and "Turing-like validity."

Face Validity - "Is the surface or initial impression of a simulations realism and is obtained by asking people who know the real system (e.g., managers) to judge whether the model is reasonable. From the scientific point of view, this is not (a test of) validity at all and... (should be considered) a test of the reasinableness or credibility of the model."[1] There are two ways to

---

[1]Emshoff, J. R. and Sisson, R. L., "Design and User of Computer Simulation Models", MacMillan, London, 1970, p. 204.

"judge whether the model is reasonable," either on the basis of the structure and logic of the model or whether the outputs seem to be reasonable. As mentioned in Section 1.2, "logic validation" was performed successfully during Phase I of the project; this section discusses reasonableness of output.

Variable-parameter validity - "Sensitivity testing is a form of variable-parameter validity. In a sensitivity test one or more factors are changed to determine (a) if they affect the output and (b) if they help make the model produce results that match historical data more closely."[2] Such variable-parameter validity tests have been conducted by obtaining impressions from Coast Guard managers who know the operational system, rather than direct comparisons of simulated vs. historical outputs. This permits "validation" for situations for which no empirical data exist.

Turing-like validity - "Ideally a (validation) test should handle nonstationarity, compensate for noisy data, simultaneously evaluate a number of output measures and work for small samples. Does such a test exist? The answer is yes if one is willing to define test very broadly. The test is simple. Find people who are directly involved with the

---

[2]Emshoff, J. R. and Sisson, R. L., "Design and User of Computer Simulation Models", MacMillan, London, 1971, p. 205.

the actual process. Ask them to compare actual with simulation output. To make the test a little more rigorous, one might offer several sets of simulated data and several sets of actual data and see if the 'experienced' people can tell which is which. ...This test is sometimes attributed to Turing although Turing[3] actually was trying to find an operational definition of human intelligence when he suggested a similar procedure.

The behavioral validity tests conducted on SARSIM are conceptually a combination of these three types of tests. The thesis is that for SARSIM to be accepted as a valid model, Coast Guard managers should be able to judge how outputs of the model will change in response to changes in specified inputs. To anticipate the magnitude of change is hardly possible, for it is extremely difficult to consider all the factors and interactions which may influence system performance in a model as complex as SARSIM, especially for an individual not well-versed in the model. Gross measures of change for some areas of output should be predictable, however.

---

[3]Turing, A. M., "Computing Machinery and Intelligence," Mind, V-l. 59- (October 1950), pp. 433-460, reprinted in E. A. Feigenbaum and J. Feldman (Eds.), Computers and Thought, McGraw-Hill, New York, 1963.

[4]Van Horn, R. L., "Validation of Simulation Results", Management Science, Vol. 17, No. 5, January 1971, pp. 247-258.

As an exercise, eight SAR officers[5] were asked to predict the logical results of changing specified input parameters of SARSIM; their predictions were then compared with SARSIM results. They were presented with statistics from several base runs, were advised of changes in SAR planning (e.g., the closing of given stations), then they estimated whether the base statistics would increase slightly or greatly, decrease slightly or greatly or be unaffected. The exercise was designed to investigate only gross indication of magnitude of change, direction of change being stressed as more important than amount.

To maximize benefits, the test was administered on an individual basis by a person familiar with SARSIM, thus providing feedback to participants. The test administrator provided an effective substitute for extensive written background explanation, and at the same time he could indoctrinate several key Coast Guard managers with regard to SARSIM and how it might be used as an aid to decision-making.

A copy of the "questionaire" used for these behavioral validity tests is attached for reference, with the correct answers circled. The results of the exercise are very

---

[5]Essentially none of these had any previous knowledge of the SARSIM model.

encouraging, showing excellent agreement between expectations and simulation results. Of 318 questions answered by SAR professionals, only 34 responses were in disagreement with simulation results, most of these stemming from misinterpretation or misunderstanding of model constructs (e.g., tolerance times), rather than failure of model logic. In addition, Coast Guard officers often selected column 3 ("no change") regardless of the parameter being varied because they felt that the real SAR system would always select the same (i.e., fastest) resource, regardless of costs, tolerances, etc.

To sum up, this testing procedure was considered to be worthwhile, and the results contribute significantly toward establishing the validity of SARSIM.

In this questionnaire you are asked to utilize your experience in and knowledge of SAR and indicate what effect certain changes in SAR plans should have on the SAR operating statistics. The purpose is to determine how reasonably the SARSIM model is operating; i.e., do the results of these changes conform to the manner in which you predict the results? It is not a test of your knowledge; it is a test of SARSIM's validity. This form may be completed and returned unsighed. Only aggregate results will be tabulated. Answers are coded as numerals 1 through 5, their meaning defined below:

1. decrease of greater than 20%
2. decrease of 5% to 20%
3. no change (less than ± 5%)
4. increase of 5% to 20%
5. increase of greater than 20%

Some definitions of terms used in the tests follow:

Boats - included the following resources - 40'
30'
17'
44'
52'
36'
MSB/MRB

Attachment to Section 2.5

Cutters - includes resources as follows - 82' WPB

95' WPB

WYM/WYIL

MEC

HEC

Other A/C - all aircraft except C-130.

Tolerance - maximum acceptable time a client can wait for service from the time of incident notification.

Utilization (j) =

$$\frac{\text{Total Hours Underway on SAR (for category j resources) *100\%}}{\text{Days in Period * Hours in Day * Number of Res. (of cat. j)}}$$

$$= \frac{\sum_{j} \text{Hrs. Underway} * 100\%}{\sum_{j} \text{Hrs. Available}}$$

I. If the month of July is compared with January in CGD1, what differences would be expected?

	July	January (circle choice)				
A. Number of Cases	883	1	2	3	4	5
B. Overall Utilization	3.60%	1	2	3	4	5
C. TSVC: Average total time case is undergoing service (from notification to termination)	2.43 hrs	1	2	3	4	5

II. If the speeds of all CGD13 resources could be increased by 10%, how would the following SAR operating statistics be affected?

	Before Change	After change (circle choi				
A. TWAIT: The average length of time a case waits from time of notification to arrival on scene of first CG resource.	1.10 hrs	1	2	3	4	5
B. TVEC: The average vector or transit time, from departure of CG resource to arrival on scene.	0.79 hrs	1	2	3	4	5

Note: Correct answers are circled.

	Before Change	After change (circle choice)				
C. TSVC: Average time a case is undergoing service (from notification to termination)	2.53 hrs.	1	2	3	4	5
D. CFAILS: Number of cases not meeting tolerance.	128	1	2	3	4	5

III. If the operating costs used for resource selection comparisons of HU16 and HH52 A/C are increased 50%, how should the following be affected?

	Before Change	After change (circle choice)				
A. Utilization of HU16	4.91%	1	2	3	4	5
B. Utilization of HH52	3.96%	1	2	3	4	5
C. CFAILS: Number of cases not meeting tolerance	54	1	2	3	4	5

IV. If towing speeds in CGD13 could be increased from 7 knots for towing small vessels, and 9 knots for towing large boats, to 10 and 12 knots respectively, indicate changes expected on the following?

	Before change	After change (circle choice)				
A. Utilization of 40'	7.20%	1	2	3	4	5
44'	5.02%	1	2	3	4	5
82'/95'WPB	7.40%	1	2	3	4	5
HH52 (has no tow capability)	2.54%	1	2	3	4	5
B. Overall utilization (entire district)	4.31%	1	2	3	4	5
C. TVEC: The average vector or transit time, from departure of CG resource to arrival on scene	0.54 hrs.	1	2	3	4	5
D. TSVC: Average total time a case is undergoing service	2.33 hrs.	1	2	3	4	5

V. If the readiness posture (number of resources) of all CGD1 SAR resources were reduced by 1/2, how would the following data be affected?

	Before Change	After change (circle choice)				
A. Utilization of Boats	4.08%	1	2	3	4	5
Cutters	2.63%	1	2	3	4	5
C130	3.13%	1	2	3	4	5
Other A/C	3.21%	1	2	3	4	5
B. CFAILS: Number of cases not meeting tolerance	24	1	2	3	4	5
C. TWAIT: The average length of time a case waits from time of notification to arrival on scene of first CG resource	0.66 hrs	1	2	3	4	5
D. The number of times a CG resource was interrupted during a case to service another of higher severity	8	1	2	3	4	5

VI. If tolerances were changed from 4 hours for severity 1 cases, 3 hours for severity 2, 2 hours for severity 3, 1 hour for severity 4 and 1/2 hour for severity 5 to a value of 99 (unrealistically hours for all cases, what should the effect be on the following data?

	Before Change	After Change (circle choice)				
A. CFAILS: Number of cases not meeting tolerance	47	1	2	3	4	5
B. Utilization of Boats	3.95%	1	2	3	4	5
Cutters	3.65%	1	2	3	4	5
C130	2.98%	1	2	3	4	5
Other A/C	4.55%	1	2	3	4	5
C. Overall utilization	3.87%	1	2	3	4	5

VII.

If CGD1 should close Kennebec River and Fletchers Neck stations, how would the following change?

	Before Change	After Change (circle choice)				
A. Number of SAR cases at:						
Boothbay Harbor	20	1	2	3	4	5
Kennebec River	14	1	2	3	4	5
S. Portland	23	1	2	3	4	5
Fletchers Neck	11	1	2	3	4	5
Portsmouth Hbr.	14	1	2	3	4	5
B. Utilization of resources at:						
Boothbay Harbor	3.13%	1	2	3	4	5
Kennebec River	4.52%	1	2	3	4	5
S. Portland	1.88%	1	2	3	4	5
C. TSVC: Average total time a case is undergoing service	2.43 hrs.	1	2	3	4	5
D. CFAILS: Number of cases not meeting tolerance	24	1	2	3	4	5
E. Overall Utilization (entire district)	3.60%	1	2	3	4	5

III. DEMONSTRATIONS OF CAPABILITIES

3.1 Sensitivity Analysis

A manager who considers employing an analytical or simulation model should know how critically model results depend on model inputs. Knowledge about the sensitivity of output to input variations is valuable to the user, for considerable care must be given to estimating values of highly sensitive parameters for simulation runs.

Even greater use can be made of the results of sensitivity analyses. Suppose, for instance, that utilizations are found by simulation to be highly sensitive to some specific parameter, while "quality of service" measures are not. If the given parameter can be controlled by changing operating procedures, then significant cost savings might be obtained without reducing the service provided to the public.

This section describes sensitivity analyses for three different sets of input parameters: tolerance times, speeds of advance and towing speeds.

It may be recalled that the tolerance time for a case is the maximum allowable time from Coast Guard notification until the first resource arrives on scene. This quantity is not explicitly defined for the Coast Guard operational procedures, but was conceptualized by the modeler to represent the Coast Guard's intent to respond to all cases within an "acceptable" time limit.

Nominal values (i.e., one for each of the five severity levels) were established by Coast Guard personnel and proved to yield acceptable results when utilized in OPSIM. Tolerances were varied over a range of possible interest in four additional runs. Table 3.1.1 lists the values used and the code number for each run. As can be seen, each value is a multiple of the corresponding nominal value.

Table 3.1.1

Variations Made to Tolerance Times

Run		01P1HE	01P1HR	01P1HV	01P1HP	01P1HQ
Multiple of Nominal		0.0	1.0	1.5	2.0	8.0
Tolerance Time (in hrs.) for Cases of Severity	1	0.00	4.00	6.00	8.00	32.0 0
	2	0.00	3.00	4.50	6.00	24.00
	3	0.00	2.00	3.00	4.00	16.00
	4	0.00	1.00	1.50	2.00	8.00
	5	0.00	0.50	0.75	1.00	4.00

The effects of tolerance time variation on a selected set of model outputs are given in Table 3.1.2 and depicted in Figures 3.1.1 and 3.1.2.

Table 3.1.2

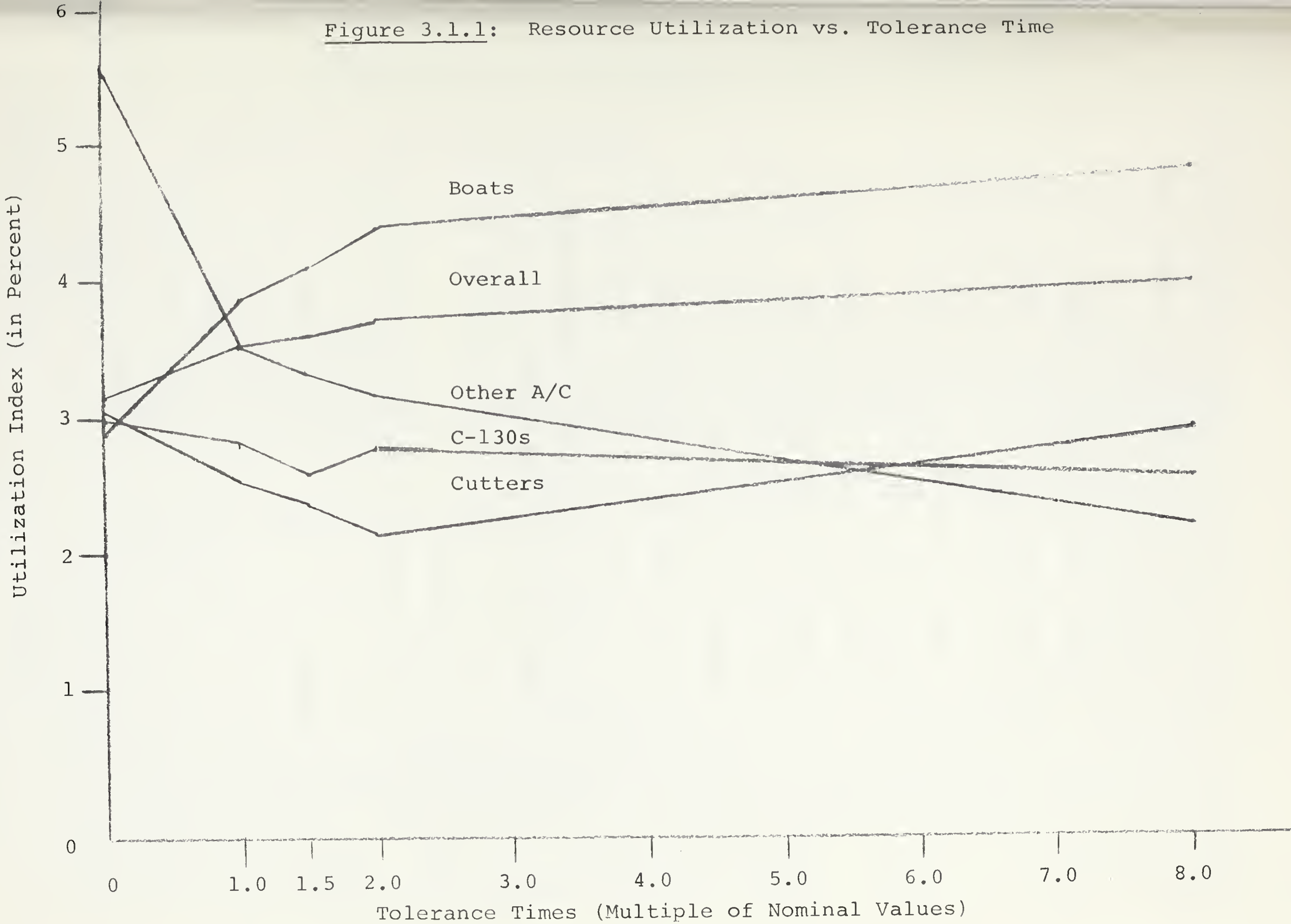
Effects of Tolerance Time Variations

Run	01P1HE	01P1HR	01P1HV	01P1HP	01P1HQ
Multiple of Nominal Tolerance	0.0	1.0	1.5	2.0	8.0
Overall Index *	3.16	3.49	3.53	3.60	3.81
Boat Util. Index*	2.94	3.85	4.06	4.27	4.43
Cutter Util. Index*	3.08	2.72	2.44	2.16	2.90
C-130 Util. Index*	3.00	2.88	2.70	2.89	2.55
Other A/C Util. Index*	5.77	3.49	3.32	3.18	2.15
Average TWAIT (hours)	0.73	0.77	0.81	0.85	0.80
Average TVEC (hours)	0.48	0.52	0.56	0.59	0.54
Average TSVC (hours)	2.38	2.68	2.60	2.64	2.80
Number of Type C Failures	880	59	39	24	3

*All utilization indices are expressed as percentages.

The variations in utilization indices shown in Figure 3.1.1 appear reasonable in that they can be justified by common sense arguments. For instance, if tolerances are <u>decreased</u> from their nominal values, one observes a decrease in overall resource utilization indices. This seems

Figure 3.1.1: Resource Utilization vs. Tolerance Time

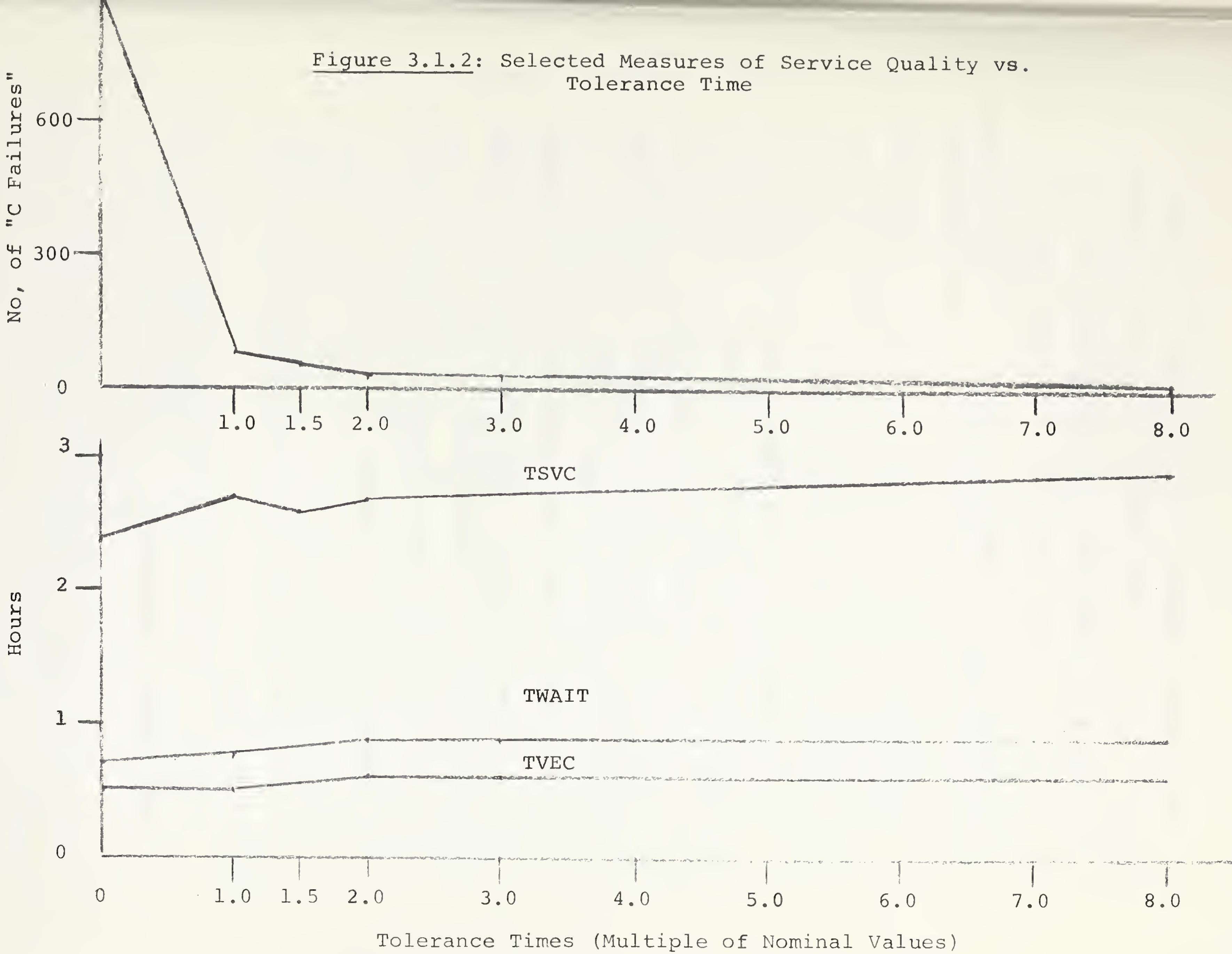


consistent in that low tolerance implies greater importance of quick service, hence faster resources are used more often; these also tend to require less time to finish their cases. Further evidence of this appears in the fact that cutter and aircraft utilization indices tend to increase while the utilization of slower (but less expensive) boats tends to decrease. The opposite effect appears when tolerance times are increased.

Figure 3.1.2 shows the effects of tolerance time variations on the quality of service provided. These effects are not critical, although the quality of service does decrease gradually as tolerance times increase. (The only measure which appears to be sensitive is the number of failures of type C, but this is deceptive since type C failure <u>is defined by the choice of tolerance time</u>. The relatively small change in TWAIT seems to confirm the observation that the failure C statistic is, at first glance, somewhat misleading.)

Basically two conclusions can be drawn from the results. The first, relating to model validity, is that changes in tolerance time seem to affect output statistics in the direction indicated by common sense arguments. The second, relating to subsequent use of the model, is that extremely precise stipulation of tolerance time is not necessary since the model does not appear to be highly sensitive to this parameter.

Figure 3.1.2: Selected Measures of Service Quality vs. Tolerance Time



Sensitivity analysis was also applied to speeds of advance. Table 3.1.3 shows values of Speeds Of Advance (SOA) used by the Coast Guard in seas with swells less than 5, 10 or 20 feet, depending on resource type, and speeds (SOA_2) for rough weather (i.e., sea swell greater than or equal to the stipulated limits). When exercised in OPSIM, these values produced satisfactory results except in the Thirteenth District, where slightly higher speeds seem to have been used historically.

To test sensitivity, all speeds were uniformly increased and decreased by 10%. A selected set of outputs are listed in Table 3.1.4 and depicted in Figures 3.1.3 and 3.1.4. Figure 3.1.3 illustrates two satisfying features. First of all, it appears that quality of service measures are not particularly sensitive to variations in speeds of advance. Secondly, the relationships exhibited appear to be linear within a region of the nominal values. This aids considerably in deciding on the proper speed,for if the slope of a given quality of service measure is s, then change in SOA of x will produce an sx change in that measure, provided that x is relatively small.

Figure 3.1.4 exhibits utilization indices which react to speed changes consistent with expectations. (Faster speeds allow resources to complete service earlier and therefore cut down on utilization and conversely for slower speeds.)

Table 3.1.3

Nominal Values of Speeds of Advance

Resource Type	SOA1 (in knots)	SOA2 (in knots)
UTB 40'	18	12
UTB 30'	16	10
UTB 17'	10	5
MLB 44'	14	8
MLB 42'	10	8
MLB 36'	8	5
MRB/MSB	10	6
WPB 82'/95'	18	12
WPB 82'/95P	18	12
YTM/YTL	12	8
WMEC	18	12
WHEC	17	12
C-130	165	165
HU16E	155	155
HH52A	75	75
HH3F	115	115

Table 3.1.4

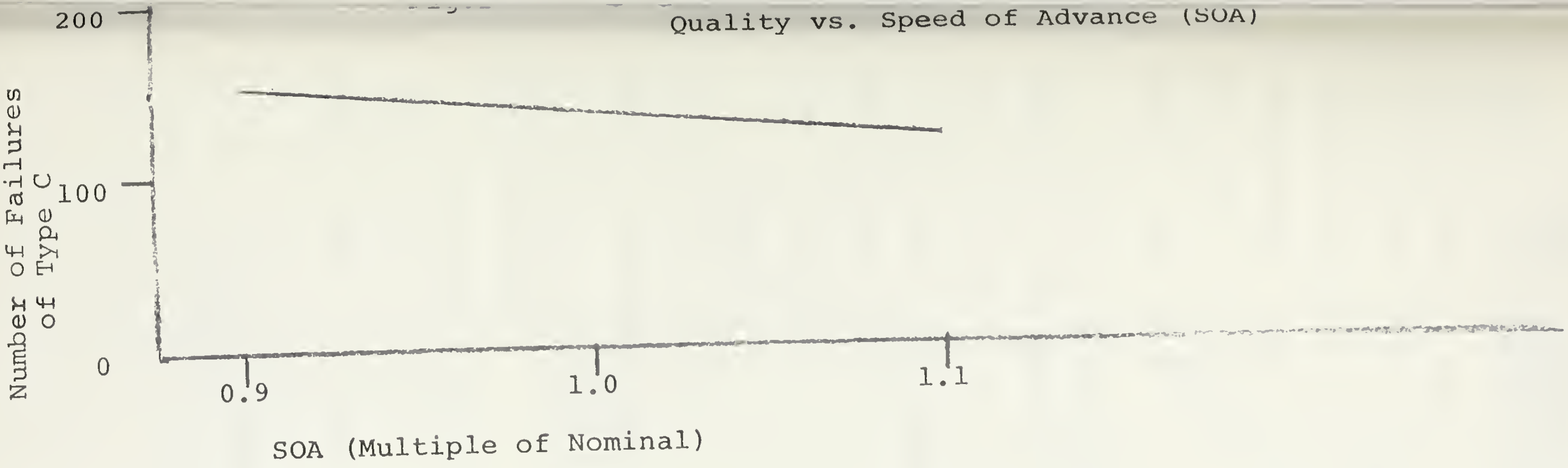
Effects of Changes in Speeds of Advance

RUN	13P1HF	13P1HA	13P1HB
SOAs*	.9N	1.0N	1.1N
Overall Util. Index**	4.92	4.70	4.32
Boat Util. Index**	5.05	4.75	4.56
Cutter Util. I Index**	6.77	6.84	5.09
C-130 Util. Index**	4.23	3.77	3.69
"Other Aircraft" Util. Index**	1.69	1.64	1.78
Number of "C" Failures	134	128	115
Average TWAIT (hrs.)	1.08	1.03	.99
Average TVEC (hrs.)	.79	.74	.62
Average TSVC (hrs.)	2.77	2.71	2.62

* Speeds shown are in terms of nominal values, N, as defined in Table 3.1.3 for various resources and sea conditions.

**All Utilization Indices are expressed as percentages.

Quality vs. Speed of Advance (SOA)



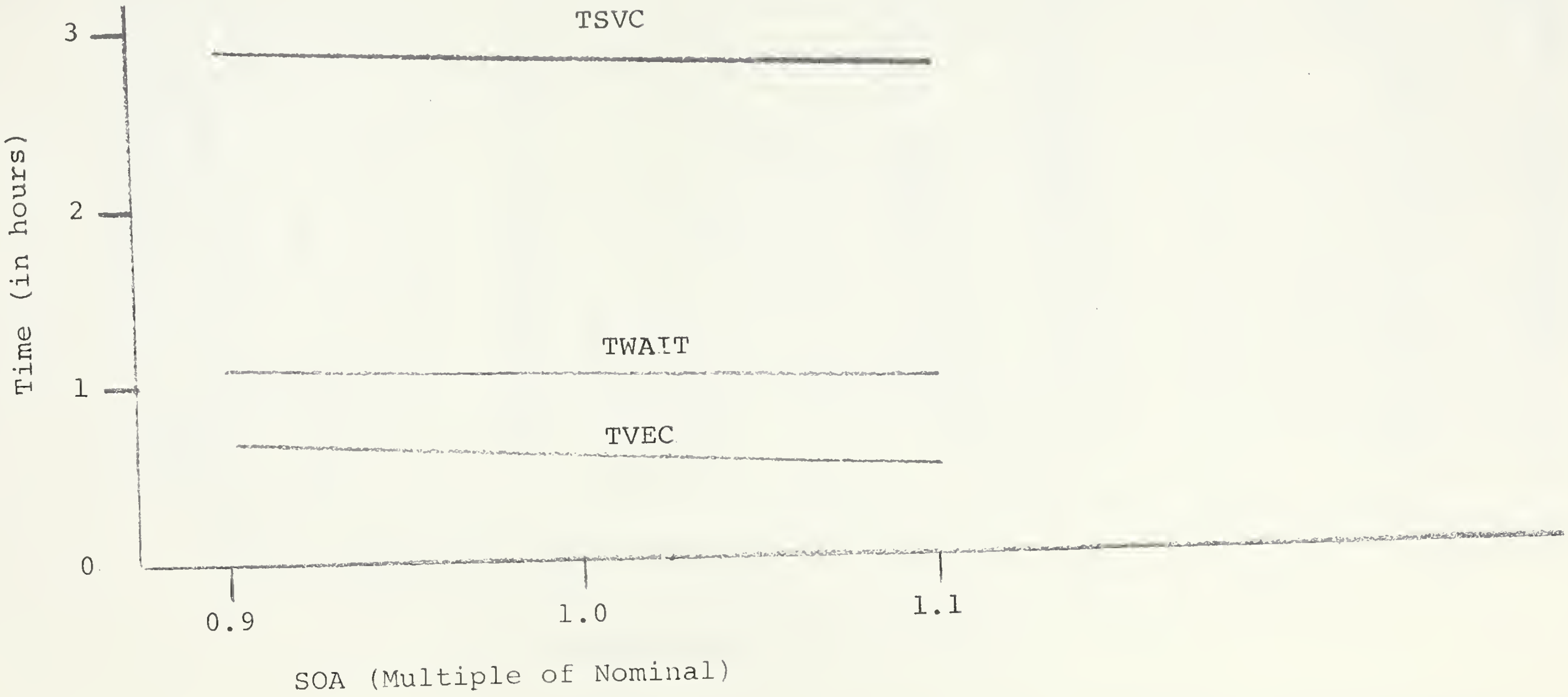
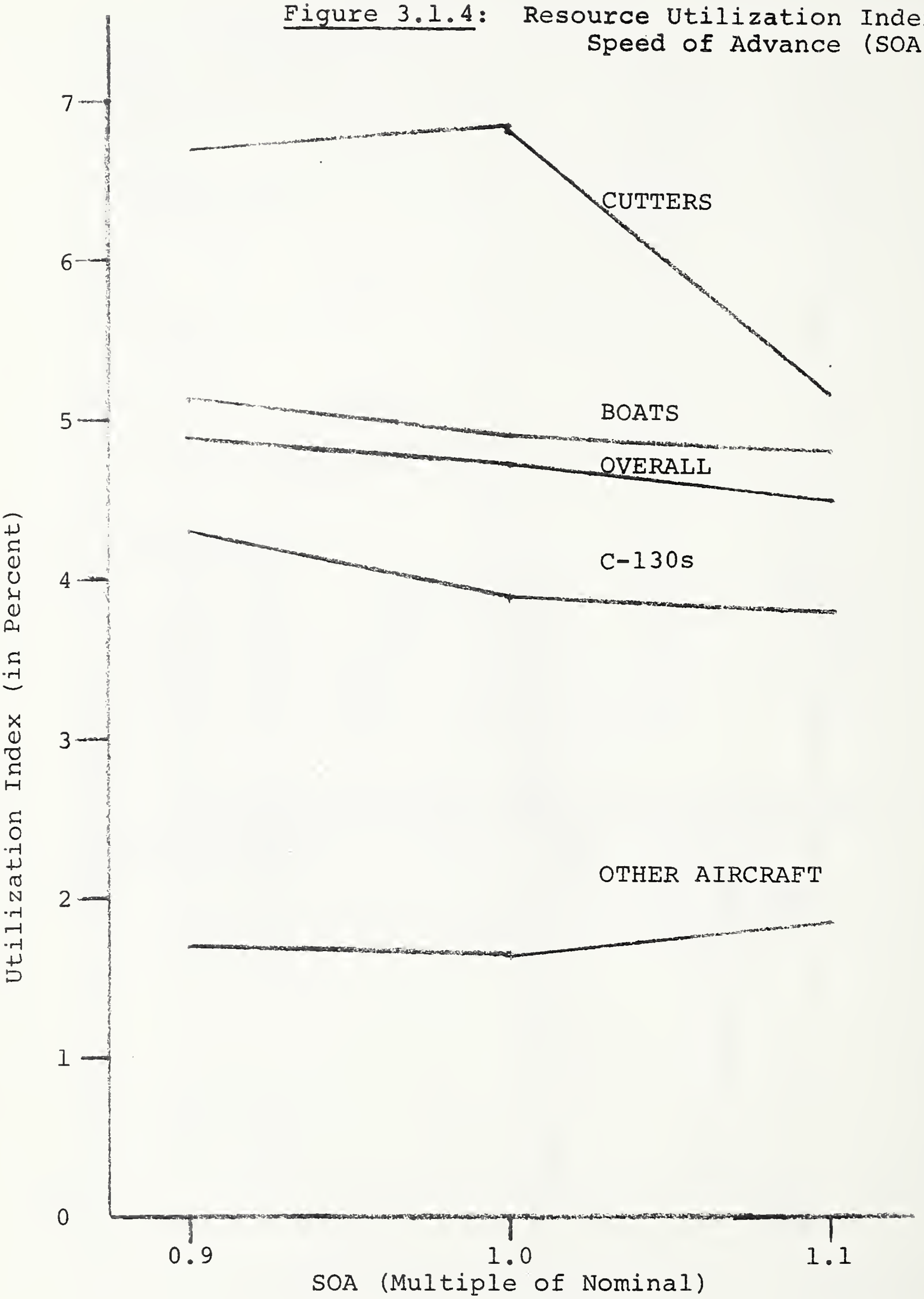


Figure 3.1.4: Resource Utilization Index vs. Speed of Advance (SOA)



A marked decrease in cutter utilization index occurs as speeds are increased above nominal values. This may be attributable to the fact that higher speeds allow boats to service many cases which previously could only be handled within tolerance time by cutters. Since the boats are considerably less expensive to operate, they are selected more often, accounting for a slower decrease in boat utilization index than the overall average.

Cutter utilization does not increase when speeds are decreased from nominal values, probably because tolerance cannot be met at the slower speeds, hence aircraft must handle some of the cases otherwise assigned to cutters. The increase in C-130 aircraft utilization indices, greater than the average overall indices, supports this.

Conceivably, then, operating costs might be reduced or the level of service increased by the introduction of faster resources. This is a matter of marine engineering technology, and not a question to be addressed by this SAR simulation.

The final sensitivity analysis shows results of varying towing speeds. It may be recalled that towing speeds (TSP) are limited to two discrete values, depending on the size of the disabled vessel to be towed. For OPSIM, nominal values suggested by the Coast Guard were 7 knots for towed boats shorter than 26 feet and 9 knots for longer boats.

For the sensitivity analysis TSP's were increased and decreased by 3 knots; selected output values are shown in Table 3.1.5 and Figures 3.1.5 and 3.1.6. As figure 3.1.5 shows, utilization indices for surface resources definitely increase as towing speeds are decreased, while utilization of air resources remains essentially unchanged. This is an expected result since lower tow speeds require surface resources to be out on duty more of the time even if serving the same number of cases.

It is interesting that the C-130 utilization index also decreases slightly when towing speeds are increased (even though aircraft are not permitted to tow vessels in SARSIM). This probably occurs because boats and cutters finish servicing their cases sooner, allowing them to handle non-tow cases which would otherwise be assigned to C-130's. Since cutter utilization decreases at a slower rate than boat utilization, it is apparent that cutters are now handling most of these cases.

As might be expected, quality of service measures which describe the Coast Guard's ability to respond to a given caseload are not highly affected by changes in towing speeds. (See Figure 3.1.6.) However, total service time decreases significantly as towing speeds are increased, a consistent result.

Table 3.1.5

Results of Tow Speed Variations

Run	13P1HG	13P1HC	13P1HD
Towing Speed (TSP) knots	4-6	7-9	10-12
Overall Util. Index*	5.20	4.31	3.88
Boat Util. Index*	5.52	4.34	3.84
Cutter Util. Index*	6.39	6.12	5.64
C-130 Util. Index*	3.20	3.25	2.96
Other Aircraft Util. Index*	1.83	1.86	1.89
Number of Type C Failures	94	94	86
Average TWAIT (hrs.)	.88	.87	.85
Average TVEC (hrs.)	.61	.62	.61
Average TVSC (hrs.)	3.12	2.53	2.24

*All Utilization Indices are expressed as percentages.

Figure 3.1.5: Resource Utilization Index vs. Tow Speed

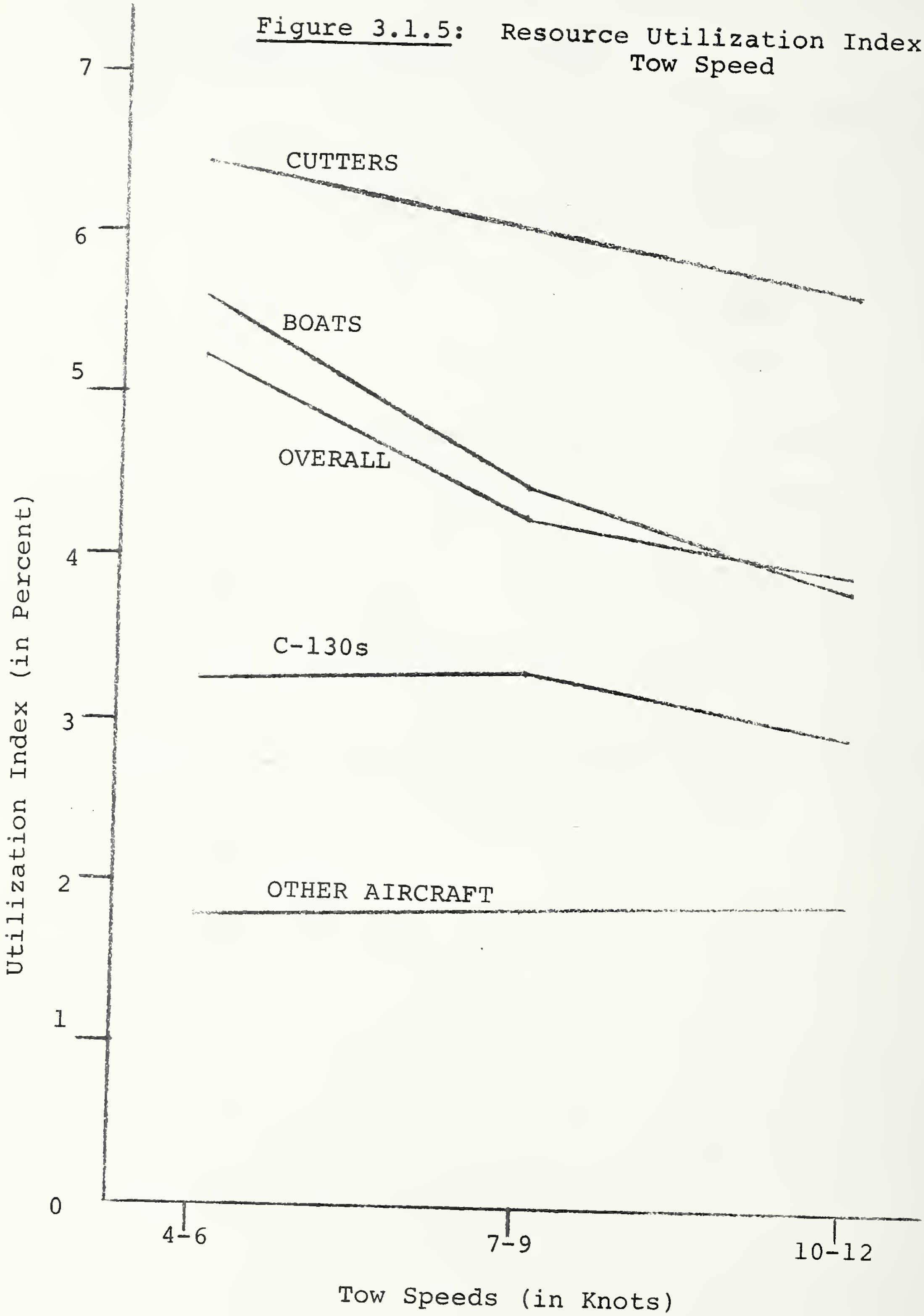
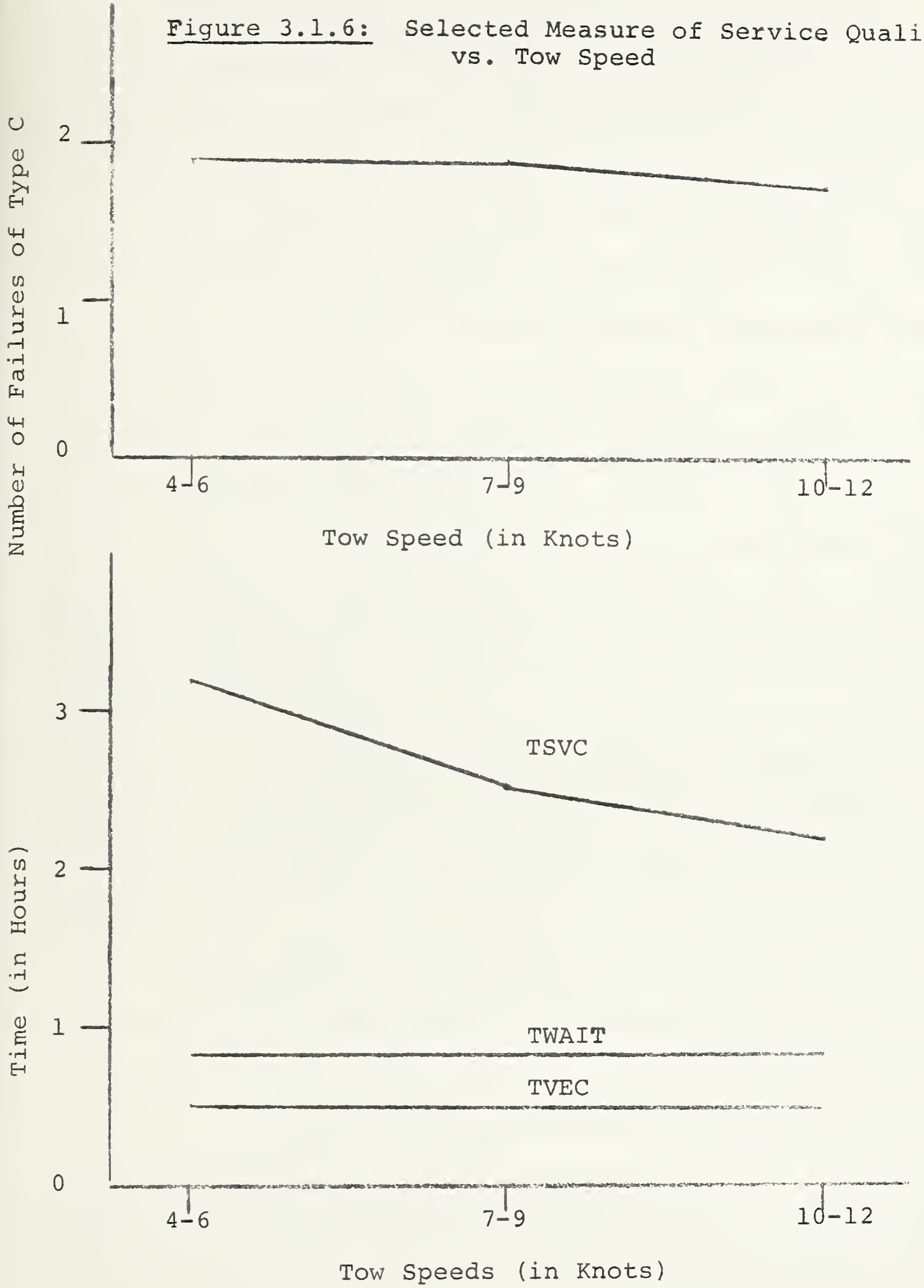


Figure 3.1.6: Selected Measure of Service Quality vs. Tow Speed



Other analyses similar to those presented above could have been performed, but the constraints on time and money during Phase III limited the number of sensitivity runs which could be made. The three analyses described serve to demonstrate what can be done. Further, each of these sensitivity analyses, in behaving as expected, contributes additional evidence of SARSIM validity.

3.2 Summary of Investigative Exercises

In an effort to demonstrate the value of the SARSIM model as a planning and predictive tool for Coast Guard decision-makers, and to give further evidence of the model's validity for situations for which no empirical data exist, the model was run using a variety of operating strategies and various scenarios designed to provide useful information related to current and future SAR system performance. Towards this objective, an attempt was made to answer the following questions:

1) How would overall system effectiveness be altered should the Coast Guard decide to use different operating policies when determining the resources to be assigned to each case?

2) What would be the effect of maintaining all the stations currently in a district, but with only one-half the number of resources? One-quarter? Twice as many?

3) What repercussions would result from eliminating stations which appear to be inefficient or underutilized?

4) Would system performance improve if helicopters replaced fixed-wing amphibious aircraft and if they were relocated throughout a district in an effort to provide

better coverage?

5) If present stations are not changed and resource levels and operating policies remain the same, will the demand predicted for the system in 1975 or subsequently be handled adequately?

The remainder of this section treats each of the questions in order and, it is hoped, provides information to serve the purposes of substantiating the reasonableness of SARSIM's behavior. It may also provide insights to the Coast Guard as to the overall effects of proposed policy alternatives.

3.2.1 Effects of Resource Assignment Policy (RAP) Variation

A set of runs was conducted using the July 1968 First District caseload as a base case for examining the sensitivity and differences due to variation in server disciplines (RAP's). Selected statistics for five RAP's are shown in Table 3.2.1; differences are not pronounced. However, the low utilization indices and infrequent incidence of queueing suggest that the system was not under sufficient stress.

This set of runs does reveal some points of interest. Under these low utilization postures, the policies which are more autonomous (i.e., RAP's 2, 4, and 5 which have less interaction among adjacent stations than do RAP's 1 and 3) show slightly faster responses to cases and have lower average

Run Number	Code	Server Discipline	Abbreviated Description of RAP Sequencing*	No. of Cases Waiting Longer than Tolerance	No. of Queues	No. of Interrupts	Average Time to Vector (Hrs.)	Average Waiting Time (Hrs.)	Average Service Time (Hrs.)	No. of Standbys	Average Utilization Index (%)
29	OIPIHI	RAP 1	P, A, PI, AI, Q	27	21	3	.36	.82	2.74	5	3.66
26	OIPIHF	RAP 2	P, PI, A, AI, Q	24	NA	6	.56	.82	2.81	3	3.61
30	OIPIHJ	RAP 3	(P+A), (P+A)I, Q	30	19	1	.57	.81	2.79	4	3.84
31	OIPIHK	RAP 4	P, PI, Q	24	18	6	.54	.79	2.80	2	3.63
32	OIPIHL	RAP 5	P, Q (No Interrupt)	25	22	0	.54	.80	2.81	3	3.65

*P Examine idle resources at Primary station
A Examine idle resources at Adjacent station
PI Examine Interruptible resources at Primary station
AI Examine Interruptible resources at Adjacent station
(P+A) Examine idle resources at Primary and Adjacent stations together
(P+A)I Examine Interruptible resources at Primary and Adjacent stations together
Q Queue the case

Table 3.2.1 Effects of Resource Assignment Policies

utilizations. This is probably due to the fact that resources at the primary stations are closer to their caseload than are the resources at the adjacent stations. On the other hand, these more autonomous RAP's exhibit longer average service times to complete their cases, probably due to longer waiting for additional resources to be assigned to multi-resource cases. Thus at low activity levels (or high resource levels), the more autonomous policies result in a slightly-faster initial response to the average case, but longer times are required to complete these cases because of limited interaction with adjacent stations.

3.2.2 Resource Level Variation

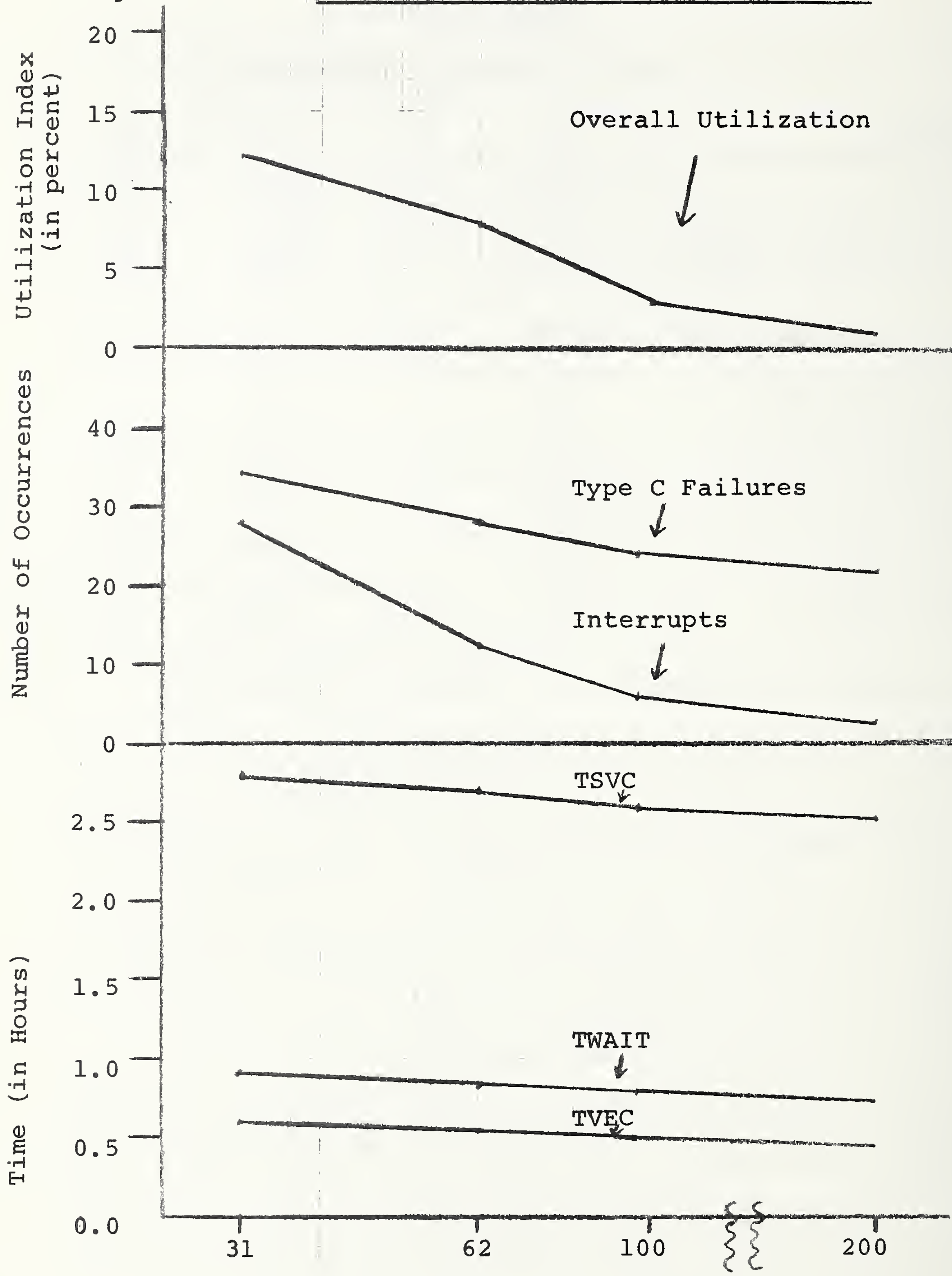
The greatest potential use of SARSIM will undoubtedly lie in evaluating system performance under various conceivable levels of resource allocations. A variation of approximately three binary orders of magnitude was investigated, using the First District's July 1968 demand scenario as the base for comparisons. Table 3.2.2 and Figure 3.2.1 show the results of these runs. The 100% point represents the base condition, i.e., the 109 resources which were operational at that time. The 200% point thus represents a doubling of both resources and crews, using the same mix (i.e., relative proportions) of resource types. The 62% point represents an attempt to halve the number of resources, but with each station having at least

Table 3.2.2

Effects of Resource Level Variation

Run Number	35	27	26	36
Code	O1P1HN	O1P1HG	O1P1HF	O1P1HO
Percent of Normal Resources	31%	62%	100%	200%
Overall Utilization Index	12.44%	5.98%	3.61%	1.75%
Number of Interrupts	29	11	6	2
Number of Failure C's	33	28	24	24
Av. TVEC (hrs.)	0.66	0.58	0.56	0.50
Av. TWAIT (hrs.)	0.94	0.85	0.82	0.75
Av. TSVC (hrs.)	2.85	2.78	2.68	2.63

Figure 3.2.1: Effects of Resource Level Variation



Resource Level (Percentage of First District Allocations, July 1968).

one resource and crew assigned at all times. The 31% point represents half the total number of resources of the previous level, but without any restriction as to providing at least one resource per station.

As can be seen from the figure, all the measures of system performance show similar reactions to change in resource level: a rather flat, essentially log-linear response from 200% to 62%, and a greater slope between 62% and 31%. This tentative relationship is encouraging, for it suggests that consideration might be given to reducing SAR resources (with concomitant monetary savings), provided that the levels of performance continue to be adjudged as satisfactory in the light of predicted demands.

3.2.3 The Closing of Two Shore Stations

SARSIM may profitably be used to investigate the likely outcome of variations in the allocation of SAR resources. The following demonstration example explores the prospective closing down of two shore stations in the First District, namely Kennebec River and Fletcher's Neck. Historically, both stations have been characterized by relatively light case loads: in July 1968, for example, Kennebec had 19 SAR cases and Fletcher's Neck only 10. Each station had a single 40' boat and a single crew at all times, and 80% of the Fletcher's Neck cases were single resource tows.

The fundamental question to be addressed is whether the SAR system's performance would have been significantly affected if these two stations had been closed, subject to other, minor, variations. Any changes in system output must be examined, of course, not only from the point of view of utilization of SAR resources, but especially with regard to service to clients.

For this demonstration example, the crews and equipment from the two closed stations were either redeployed or considered to be phased out. One of the 40' boats was assigned to South Portland Base; in addition, a single crew was assigned to Portsmouth Harbor and another to Boothbay Harbor at all times except weekends from 0800 to 2400, when two crews were assigned to each. (In the base case, there were two crews at Portsmouth Harbor and one at Boothbay Harbor at all times.) For both the base case and the experimental run, the input demand scenario was the historical case load for July 1968, with the same cases which occurred then in their actual order of occurrence.

It might be anticipated that, with the two stations closed, the loads from those stations would be handled by the neighboring stations in fairly equal proportions, taking into account that Base South Portland is between the two "closed" stations. As shown by the station measures listed in

Table 3.2.3.1, most of the cases were indeed picked up by the adjacent stations, but far from equally. Examination of the detailed output of the simulation reveals that most of the cases previously coming into Kennebec were served in the experimental run by Boothbay Harbor, while Base South Portland picked up many of those previously assigned to Fletcher's Neck; only two of the Fletcher's Neck cases went to Portsmouth. This appears to be due to the geographical distribution of the stations and to the locations of the incidents previously served by Fletcher's Neck.

There were three cases in the base run which were not handled by the three remaining stations in this area for the experimental run; these were served by the covering air stations and cutters.

As viewed from the perspective of Coast Guard operations and service, Table 3.2.3.1 shows that the closing of the stations leads to an increase in the number of cases queued (from 7 to 16), with most of the additional queueing taking place at Base South Portland; there were, in addition, two interrupts occurring at Boothbay Harbor, accounting for 2 cases there being queued. For the group of stations as a whole, however, there were no additional Type C failures: even though more cases were placed in queues, they could still be serviced within tolerance times.

Table 3.2.3.1 shows a number of other statistics for the group of stations of interest, but, as can be seen by examination of the table, the changes from the base run to the experimental run are generally small. (For example, note the small increase in utilization index despite the added load per resource assigned.)

Included among the more important statistics of the table are those which demonstrate adequacy of service from the client's viewpoint. Thus, it can be seen that, for the group of stations taken together, the average time from Coast Guard notification of distress until a resource arrives on scene to provide service rose by almost 10 minutes (see TWAIT), an increase of approximately 20% when averaged over all clients served in that area. However, there is a rather strong suggestion that this average increase was not spread over all clients, but, rather, may have been concentrated on those cases which were farthest from the servicing stations: as can be seen, the average time in transit, TVEC, increased by more than 6 minutes. In other words, the bulk of the extra time waited on the average was contributed by the time to reach the scene.

The pertinent statistics for the district as a whole are shown in Table 3.2.3.2. It may be observed that the total number of Type C failures went up by 2 from the base run to

STATION MEASURES

STATION		No. of cases	No. of needs	Failure C	No. of Q	No. of Int.	Avg Hrs TVEC	Avg Hrs TWAIT	Avg Hrs CFAIL	Standbys Calls/ Unpro	Standbys Used	Util. Index %
	Before	20	33	1	7	0	.35	.58	.16	2/2	0	3.13
Booth Bay	After	32	50	1	9	2	.56	.79	.16	6/5	1	5.72
	Chng.	+12	+17	0	+2	+2	+60%	+37%	0	+4/+3	+1	+80%
	Before	14	20	0	0	0	.41	.66	.0	0/0	0	4.52
Kennebec	After	0	0	0	0	0	0	0	0	0/0	0	0
	Chng.	-	-	-	-	-	-	-	-	-	-	-
	Before	23	37	1	0	0	.50	.75	2.09	1/1	0	1.88
Base South Portland	After	31	45	2	7	0	.93	1.33	1.43	11/8	3	3.09
	Chng.	+8	+8	+1	+7	0	+86%	+77%	-32%	+10/+7	+3	+59%
	Before	11	13	1	0	0	1.66	1.91	2.04	0/0	0	8.75
Fletcher's Neck	After	0	0	0	0	0	0	0	0	0/0	0	0
	Chng.	-	-	-	-	-	-	-	-	-	-	-
	Before	14	19	0	0	0	.28	.53	0	0/0	0	2.54
Portsmouth Harbor	After	16	18	0	0	0	.40	.65	0	0/0	0	1.92
	Chng.	+2	-1	0	0	0	+43%	+23%	0	0/0	0	-41%
	Before	82	122	3	7	0	.566 h	.811 h	1.43 h	3/3	0	3.67
TOTAL/ AVG's	After	79	113	3	16	2	.672 h	.972 h	1.01 h	17/13	4	3.92
	Chng.	-3	-9	0	+9	+2	+.106 h or +6.36 m	+.161 h or +9.66 m	-.42 h or -.25.2 m	+14/+10	+4	+6.83%
							+19%	+20%	-29			

h=hours
m=minutes

TABLE 3.2.3.1

DISTRICT MEASURES

	BASE RUN		EXPERIMENT		CHANGE	
Average Utilization Index* Overall	3.60		3.71		3.05	
Combined Util.Index and Needs Serviced	Index	Needs	Index	Needs	Index	Needs
" Boats	4.27%	1139	4.33	1132	1.4%	-7
" Cutters	2.16	54	2.43	58	12.5	+4
" C-130	2.89	25	2.98	25	3.13	0
" A/C	3.18	67	3.28	70	3.15	+3
Total Number of Type C Failures	24		26		+2	
Total Standby Crews used	4		8		+4	
Avg. TWAIT (hrs.)	.81		.83		-.02 (or 1.2 mins.)	
Avg. C-FAIL TWAIT-TOL (hrs.)	2.04		1.58		-.46 (or -27.6 mins)	
Number of Queued Cases	19		30		+11	
Number of Interrupts	15		14		-1	
Avg. TVEC (hours)	.56		.57		.01 (or .6 mins.)	

TABLE 3.2.3.2

the experimental run, but this increase occurred outside the group of stations of interest. The total number of cases in queue went up significatnly, but there was one less case interrupted, and the average waiting time throughout the district went up by only a minute. Interestingly enough, the average waiting time in excess of tolerance (i.e., TWAIT-TOL) was reduced by almost a half-hour; although there were two more Type C failures, the average "failed" client waited much less time for service.

Table 3.2.3.2 shows that the additional standby crews used in the smaller area of interest were the only such throughout the district. Utilization indices show small changes for the district as a whole, except for the cutters (which took on increased cases).

These results suggest that the economies potentially derivable from closing the two selected stations would have been warranted insofar as utilization of resources and operation of remaining stations were concerned. Service supplied to clients is altered somewhat: the degree to which this is considered acceptable is subject to user discretion.

3.2.4 Aircraft Modernization and Relocation

As another example of the investigative use of SARSIM, the effects of redistribution and modernizing air resources within a district has been examined. The basic run was made for the Thirteenth district, the Washington and Oregon coast, again using the historical resources and caseload demand for July, 1968. In the base configuration air resources were stationed at Port Angeles, Washington, where 3 HU16E's (fixed-wing amphibious medium-range aircraft) and 3 HH52A's (single engine helicopters) were available, and at Astoria, Oregon, where 2 HH52A's were available. A total of 865 cases occurred in the district for that month.

This base simulation run indicated a relatively high number of type C Failures (91 out of 865 cases) and a low aircraft utilization index (1.92%). The Coast Guard had been considering modernizing the air resources, especially the replacement of the old amphibious HU16E's, and shifting the geographical distribution southward to provide better coverage.

A test run was then made with these modifications in the air resources: Port Angeles was left with only the 3 HH52A's; the 3 HU16E's were taken out of service; Astoria's older helicopters were replaced by 3 HH3F's, newer dual engine turboprop helicopters; and Coos Bay, Oregon, which historically had only three small boats, was provided with 3 HH52A's in

addition to the boats. Although the number of aircraft in the district was only increased by one, the replacement set had, in general, more capability, longer endurance, and better geographical distribution.

Running the simulation with these new resources and the July 1968 demand scenario brought the total number of Type C Failures to 83 (down 9%) and increased the aircraft utilization index to 3.04%. Other utilization indices (all of which are expressed in percentages) changed as follows:

Run Number	60	81
Code	13P1HE	13P1HH
	BASE RUN	EXPERIMENTAL RUN
Boat Utilization Index	3.93	3.58
Cutter Utilization Index	4.90	5.29
C-130 Utilization Index	3.68	2.75
Other Aircraft Util. Idex	1.92	3.04
Overall Util. Index	3.86	3.74

It is also noteworthy that the modifications which were simulated reduced the requirement for expensive C-130 services in the Thirteenth District by 25% (from 3.68% to 2.75%) during

this experimental run. Thus some cost savings, in terms of fewer C-130's, might have been possible; these could have offset some of the additional costs of procuring and operating the three new HH3F's and the additional HH52A.

The average service time decreased slightly for both weekdays (from 2.29 hrs. to 2.28 hrs.) and weekends (from 2.07 hrs. to 2.03 hrs.). The average time to vector to a case remained unchanged at 0.61 hours, as did the average time a case waited for service, 0.82 hours.

At Port Angeles, where air resources were cut in half, the cases handled by air resources decreased by only 5, from 36 in the base run to 31 in the experimental run, indicating that the original 6 resources were probably not all necessary. At Astoria, 3 HH3F's were able to handle 20 cases in the experimental case, as opposed to 11 in the base case. The three new HH52A's at Coos Bay handled 16 cases in the experimental scenario.

The overall utilization index decreased over 4% (from 3.86% to 3.74%) with the addition of one extra resource which should only account for a 1.3% decrease since 76 resources were simulated in the base run. Thus the addition of this one aircraft influenced the overall utilization three times as much as would be expected from adding an average resource

to this district.

If the changes in performance revealed by the simulation are, indeed, considered to be acceptable by the Coast Guard, additional runs and analyses, with other demand scenarios, would be worthwhile.

3.2.5 The Effect of Increasing Demand

An early test was conducted to ascertain the reaction capabilities of the simulation model when demand alone is significantly changed. Using growth in demand as forecasted in studies prepared elsewhere for the Coast Guard, July 1975 demand conditions were simulated for the First Coast Guard District while holding resources at their July 1968 levels. It may be noted here that a more complete investigation of the effects of changes for decision-making would require a systematic and detailed series of changes, observing effects before selecting modifications for additional trials.

This section offers only a summary sketch of the exploration of demand-only changes, where the simulated caseload rises to 1317, in place of the historical 883 cases in July 1968. A more detailed discussion, based on an investigation of reallocations within the same district, is provided in the following section.

A priori, one might expect such effects as an increase

in demand to result in increased utilization of resources (i.e., higher indices), more Type C failures and more queued cases, congestion effects at specific locations (such as Point Allerton), and increased waiting times for clients.

The increased caseload resulted in the following shifts in utilization index (in percentages) by type of resource:

RESOURCE TYPE	JULY 1968	JULY 1975
Boats	3.85	5.53
Cutters	2.72	4.35
C-130	2.88	2.47
Other aircraft	3.49	4.39

All resources belonging to the district show higher utilization indices, as expected. The decrease in C-130 (a non-district resource) utilization results from a shift in service of but a single case (from C-130 to district resources), perhaps stemming from a random determination of maintenance status which is inherent in the model.

The number of Type C failures (i.e., failure of any Coast Guard resource to arrive on scene within specified tolerance time) increased from 59 to 92. This increase is essentially in proportion to the increase in number of cases and may simply reflect constant fraction of hard-to-reach cases while the

system is still unpressured, even at the level postulated for 1975.

The data on the runs also reveal that if a standby crew had been called whenever all crews were dispatched from a station, 124 recalls would have been required in 1968, but these crews would have serviced only 4 cases. For the 1975 caseload, 214 recalls resulted in service to only 2 cases. This further suggests that a critical service situation has not been reached.

The simulated number of times it was necessary to interrupt service of a case to respond to a need of higher priority rose from 9 to 23 with the increased load. This rate of interrupting service is sufficiently low to suggest that service only to the lowest priority cases was interrupted. Verification of this supposition is the kind of inquiry which takes advantage of the features of POSTPRO.

Surprisingly, the quality of simulated service at Point Allerton, the busiest station, did <u>not</u> decay. The number of cases served increased from 133 (with 210 needs) to 264 (with 380 needs), but the average time from notification to resource arrival on scene did not change, holding at 0.45 hours. This situation is explainable by an examination of queueing formulas which consider the number of available servers. This examination

reveals that a unit with higher numbers of servers has considerably less impact for a given percentage increase rate of demand. Thus, quality of service at Point Allerton, with 3 around-the-clock ready crews, did not decay with the increase in caseload. Examination of the waiting times in the northern group of the district, where single ready resources are the rule, indicates a modest increase in client waiting time, thus tending to confirm the previous analysis.

A detailed examination of change in waiting times on a station-by-station basis indicates that the growth in cases occasioned only modest change in service at any station, but there was an increased strain placed on the crews involved. For example, at Point Allerton day-time weekend crews were underway about one-third of the total time they were on duty, during daylight hours under the heavier load. This points up the necessity of contemplating not only quality of service to customers, but also usage of crews and hardware,in arriving at a decision as to the appropriateness of a given operational arrangement.

This brief analysis indicates that the model responds generally as expected. It points up the need for careful analysis to explain the comparative results of two runs. Of course, much more detailed information can be developed, and

should be, in considering actual or contemplated force deployments. The runs described herein (#45 and #85 of Phase III) provide positive evidence that SARSIM is appropriately responsive to growth in demand.

The following section illustrates a more detailed investigative run entailing exploratory reallocations within the same district.

3.2.6 Analysis of an Entire District

A final example of SARSIM capabilities is provided by a demonstration exercise which was conducted with senior Coast Guard personnel during a briefing on Phase III. A First District simulation was set up with data corresponding to the June 1971 configuration of SAR resources and an input caseload of approximately 1300 cases, the high level prediction of likely caseload for 1975. The purpose of the exercise was to demonstrate how unacceptable situations might be revealed by simulation, thereby providing insight into desirable courses of action, leading to another simulation run to analyze the effects of any changes made.

The results of the base run for the demonstration were presented to the audience,with the more obvious problem areas highlighted. Some general types of change, as well as some specific suggestions, were proposed. Some of these were

accepted, some modified, and some replaced by more attractive changes, yielding a composite set of input alterations which seemed to offer the greatest promise of improving outputs. Base run results, changes made, and follow-on results are discussed below.

The procedure used to analyze the runs was to examine the output first on a district-wide basis, then by groups, and finally by specific stations. Tables 3.2.6.1 through 3.2.6.3 summarize DISTRICT, GROUP, and STATION data for the two runs. As can be seen from the tables, there are four significant output features which point toward necessary changes, namely Type C failures, the utilization index, total cases queued, and the number of standbys called. Three additional items of interest are the average TVEC, average TWAIT, and total number of interrupts.

TABLE 3.2.6.1

DISTRICT OUTPUT STATISTICS

	RUN 1	RUN 2
Number of Type C Failures	90	75
Average Utilization Index (%)	4.86	4.60
Number of Queued Cases	18	8
Number of Standbys called	259	187
Average TVEC (hours)	.57	.52
Average TWAIT (hours)	.80	.76
Number of Interrupts	29	20

TABLE 3.2.6.2

OUTPUT STATISTICS, BY GROUPS

	No. of Type C Failures		UTIL INDEX (%)		AVE. TVEC (hrs)		AVE. TWAIT(hrs)	
	RUN 1	RUN 2	RUN 1	RUN 2	RUN 1	RUN 2	RUN 1	RUN 2
S. W. Harbor	6	6	1.32	1.41	.63	.67	.88	.91
Portland	9	8	1.63	1.68	.43	.44	.68	.69
Boston*	11	7	7.05	6.99	.33	.31	.56	.53
Woods Hole*	29	23	7.80	7.29	.67	.64	.91	.88
District*	14	14	2.52	2.25	.64	.39	.84	.64
C-130	7	3	3.10	2.06	1.15	1.31	1.54	1.60
Patrols*	14	14	28.08	27.84	3.03	3.09	3.11	3.09

* Potential problem sites warranting further attention.

TABLE 3.2.6.3

OUTPUT STATISTICS, BY STATIONS

GROUP BOSTON STATIONS

	No. of Type C Failures		UTIL. INDEX (%)	
	RUN 1	RUN 2	RUN 1	RUN 2
Merrimac River	3	2	5.14	4.25
Gloucester	1	1	7.23	6.62
Cape Cross	1	0	7.62	5.99
Boston	1	1	1.23	1.21
Pendant	0	0	0	0
Pt. Allerton*	5	3	12.14	15.60
Scituate*	0	0	13.55	12.16

GROUP WOODSHOLE STATIONS

	No. of Type C Failures		UTIL. INDEX (%)	
	RUN 1	RUN 2	RUN 1	RUN 2
Cape Cod Canal	2	2	5.17	4.97
Race Point	3	2	8.53	7.46
Chatham	2	1	8.12	8.80
Woods Hole	1	1	5.45	4.33
Pt. Jackson	2	3	12.46	10.55
Pt. Bonita	1	1	8.89	2.40
Brandt Point*	3	4	18.14	18.85
Castle Hill	3	4	9.00	11.18

*Potential problem sites warranting further attention.

Table 3.2.6.3 (continued)

DISTRICT-WIDE UNITS

	No. of Type C Failures RUN 1	RUN 2	UTIL. INDEX (%) RUN 1	RUN 2
Portland - 3 HEC	0	0	0	0
Portsmouth - 2 MEC	0	0	.17	2.66
Boston - 6 HEC	1	3	1.09	.88
Cape Cod as HH3*	10	9	6.45	3.05
Cape Cod as HH 52	3	0	8.04	3.18

PATROLS

	No. of Type C Failures RUN 1	RUN 2	UTIL. INDEX (%) RUN 1	RUN 2
AO	6	5	28.55	26.31
AV*	8	7	27.62	29.39

*Potential problem sites warranting further attention.

The four groups chosen for more detailed examination were BOSTON, WOODS HOLE, DISTRICT UNITS, and the PATROLS. Within each group, the individual stations(s) with the greatest potential problems (e.g., high utilization index or many Type C failures) were singled out as indicated in the tables and listed below:

GROUP BOSTON:	PT. ALLERTON
	SCITUATE
GROUP WOODS HOLE	BRANDT POINT
DISTRICT:	CAPE COD AIR STATION (AS)
PATROL:	AV Patrol

Tables 3.2.6.4 through 3.2.6.8 show the changes made in resource allocations and pertinent output statistics for the base and follow-on runs; the detailed changes to the inputs were as follows:

PT. ALLERTON: The 44' and 17' boats were replaced by an additional 40' boat (a reduction in total number of boats at the station), but an additional ready crew was assigned during daytime shifts on weekends.

SCITUATE and BRANDT PT: Changes were planned for both of these stations including relocation of the AV patrol near Brandt Point. However, when all changes for the district were examined together, it was

decided to leave these stations unchanged and to observe the effects there caused by other changes elsewhere.

CAPE COD AS: Two new-type 200-knot fixed wing aircraft were added with 1 ready crew assigned to it on all shifts.

AV PATROL: Relocate from Woods Hole to LAT 41° - 15' LONG 70° - 27' to reduce travel time (TVEC).

The intent of the proposed changes was to improve overall performance in the district as reflected by the selected statistical parameters. Results, summarized below, indicate that this was achieved.

1) The overall district figures for Type C failures, average utilization indices, numbers of queued cases, numbers of standbys called, average transit time, average waiting time, and interrupts each decreased to an extent appreciated by the Coast Guard participants in the demonstration/experiment.

2) The figures for Type C failure and utilization index for the problem groups both decreased.

3) Individual station Type C failures, utilization index, cases queued, and other parameters all generally decreased.

TABLE 3.2.6.4

POINT ALLERTON

UNIT ANALYSIS

NUMBER OF CASES: 271/281*

Number of Type C Failures	Number of Cases Queued	Resource Type	On Board Allowance	Resource Utilization Index (%)	Standbys Called	Standbys Used
5/3	3/0	44 UT 40 UT 17 UT	1/0 3/4 1/0	3.28/0 1.56/1.56 .47/0	60/19	1/0

CREW ANALYSIS

	Weekday				Weekend			
	00 - 08	08 - 16	16 - 20	20 - 24	00 - 08	08 - 16	16 - 20	20 - 24
Number of Ready Crews	3/3	3/3	3/3	3/3	3/3	3/4	3/4	3/3
Station Utilization Index (%) (12.14/15.60)	5.26/ 6.67	9.93/ 13.05	11.21/ 14.01	13.96/ 17.41	10.63/ 13.16	24.04/ 31.93	31.06/ 44.50	15.18/ 15.18

*NOTE: Figures to left of solidus (/) apply to base run; figures to right reflect changes made and the results of changes.

TABLE 3.2.6.5

SCITUATE

UNIT ANALYSIS

NUMBER OF CASES: 117/108*

Number of Type C Failures	Number of Cases Queued	Resource Type	On Board Allowance	Resource Utilization Index(%)	Standbys Called	Standbys Used
0/0	2/0	44' UT 40' UT 17' UT	1/1 1/1 1/1	7.16/5.26 26.12/24.03 7.24/7.09	14/15	0/0

CREW ANALYSIS

	Weekday				Weekend			
	00 - 08	08 - 16	16 - 20	20 - 24	00 - 08	08 - 16	16 - 20	20 - 24
Number of Ready Crews	2/2	2/2	2/2	2/2	2/2	2/2	2/2	2/2
Station Utilization Index (%) (13.55/12.16)	12.83/ 10.66	10.96/ 12.14	14.16/ 14.98	11.66/ 10.07	4.52/ 4.56	11.76/ 7.93	37.48 23.95	31.20/ 30.34

*NOTE: Figures to left of solidus (/) apply to base run; figures to right reflect changes made and the results of changes.

TABLE 3.2.6.6

BRANDT POINT

UNIT ANALYSIS

NUMBER OF CASES: 40/34*

Number of Type C Failures	Number of Cases Queued	Resource Type	On Board Allowance	Resource Utilization Index(%)	Standbys Called	Standbys Used
10/4	0/0	44' UT 40' UT	1/1 1/1	12.12/12.55 24.07/25.09	0/0	0/0

CREW ANALYSIS

	Weekday				Weekend			
	00 - 08	08 - 6	16 - 20	20 - 24	00 - 08	08 - 16	16 - 20	20 - 24
Number of Ready Crews	2/2	2/2	2/2	2/2	2/2	2/2	2/2	2/2
Station Utilization Index(%) (18.14/18.85)	17.62/ 20.41	12.60/ 15.40	19.34/ 20.59	25.68/ 26.32	8.05/ 7.00	17.15/ 15.02	37.31/ 32.37	30.21 24.62

*NOTE: Figures to left of solidus (/) apply to base run; figures to right reflect changes made and the results of changes.

TABLE 3.2.6.7

CAPE COD AIRSTA

UNIT ANALYSIS

NUMBER OF CASES: (HH3 - 32/17; HH52 - 32/14; Fixed Wing 0/59)*

	Number of Type G Failures	Number of Cases Queued	Utilization Index (%)	Standbys Called	Standbys Used
HH3(3)	10/9	0/0	6.45/3.05	2/3	0/0
HH52A(3)	3/0	0/0	8.04/3.18	3/3	0/0
FIXED WING	-/2	-/0	-/8.86	-/30	-/0

CREW ANALYSIS

		00 - 08	08 - 16	16 - 20	20 - 24	00 - 08	08 - 16	16 - 20	20 - 24
Number of Ready Crews	HH3	1/1	1/1	1/1	1/1	1/1	1/1	1/1	1/1
	HH52	1/1	1/1	1/1	1/1	1/1	1/1	1/1	1/1
	FIXED WING	1	1	1	1	1	1	1	1
Station Utilization Index(%)	HH52	4.87/ .46	9.59/ 3.83	9.47/ 5.62	5.82/ 2.79	6.53/ 1.72	6.30/ 4.21	16.97/ 8.23	15.86/ 1.68
	HH3	2.92/ 1.81	6.45/ 3.22	8.68/ 3.03	6.65/ 1.25	5.92/ 2.59	7.79/ 2.45	16.76/ 14.24	8.55/ 6.48
	FIXED WING	-/5.35	-/10.38	-/8.71	-/9.95	-/8.26	-/10.27	-/13.79	-/10.31

*NOTE: Figures to left of solidus (/) apply to base run; figures to right reflect changes made and the results of changes.

TABLE 3.2.6.8

TABLE 3.2.6.8

AV PATROL

UNIT ANALYSIS

NUMBER OF CASES: 10/12*

Number of Type C Failures	Number of Cases Queued	Utilization Index(%)	Standbys Called	Standbys Used
8/9	0/1	27.62/29.39	0/0	0/0

CREW ANALYSIS

	00 - 08	08 - 16	16 - 20	20 - 24	00 - 08	08 - 16	16 - 20	20 - 24
Number of Ready Crews	1/1	1/1	1/1	1/1	1/1	1/1	1/1	1/1
Station Utilization Index(%)	31.95/ 39.72	29.02/ 26.59	39.80/ 34.75	29.38/ 29.04	18.61/ 17.03	9.44/ 23.35	27.43/ 38.17	33.87/ 34.34

*NOTE: Figures to left of solidus (/) apply to base run; figures to right reflect changes made and the results of changes.

It may be noted that such discrepancies or unexpected changes as increases in BRANDT POINT Type C failures and utilization index warrant additional examination of the district and, possibly, additional changes. It is recognized, of course, that many short-cuts were taken for the sake of an effective demonstration of the model's capability. An actual analysis of a district would utilize considerably more information, and a much more detailed examination of individual cases. POSTPRO might well be called on for a more detailed investigation. An essential precaution is to watch for and avoid unintentionally degrading performance at one location while attempting to improve performance elsewhere.

IV. CONCLUSIONS

It was the objective of the SARSIM effort (which has now progressed through three phases) to develop a tool to allow Coast Guard decision-makers to consider the consequences of alternative courses of action proposed for the Search and Rescue program area. The preceding sections of this report and the documentation of the model (available in seven other NBS Reports) indicate that this objective has now been largely achieved.

Operation of the model through an extensive series of runs indicates that it responds to changed conditions in a logically correct manner, and that the magnitude of the observed changes appear reasonable. Not unexpectedly, the validation process has also shown that the model must be carefully calibrated separately for each Coast Guard District. This reflects the fact that real-world SAR operations vary from one geographic region to another, depending on local conditions and procedures.

Simulation runs have also shown that the model provides an ability to assess the effects of trade-offs between classes of resources -- cutters, aircraft and shore stations. The model thus can serve as an integrated planning tool.

Experience to date suggests that careful analysis is necessary to probe the reasons for changes in level of

service with changed input conditions. Given the complexity of the simulation model, it appears to be highly advantageous for an experienced analyst to interface between the managerial decision-maker and the model. This does not at all gainsay the desirability of management understanding of the nature and operation of the model but rather, provides the opportunity for enhanced comprehension and deeper probing. It should be noted in this context that reasonable explanations have been found for all runs made thus far, but the answers have not always been obvious. (This leads to a subsidiary conclusion, namely that experience in using the model is a necessary supplement to the available complete and extensive documentation to afford effective analysis of results. Overlap in tours of Coast Guard analysts would be highly desirable to minimize costly relearning.)

Future explorations of alternatives for resource allocations should include systematic examination of the effects of all inputs which yield acceptable levels of service to the public. It can be expected that repeated applications of the model will improve the user's ability to estimate outcomes and increase efficiency. (This reinforces the need for continuity.) Of course, the model will never provide criteria for choosing appropriate levels

of service to be provided but only shows the consequences of prospective decisions. The evaluation of these consequences in conjunction with selected levels of service and the attendant costs remains the function of the SAR mission manager. His choice may be based, at least in part, on model output, but cannot be derived mechanistically from the simulation.

At its present state of development, the Search and Rescue Simulation Model (SARSIM) incorporates a number of compromises necessary in the trade-off between complexity and realism. As a result, SARSIM is considered to offer broad strategic indicators with considerable realism, although some statistical outputs may not be completely realistic when examined in exceedingly small detail. It is accordingly recommended that only vital changes be made for improving fine details, and that several months of operational experience should precede any major changes. This would not restrict the use of the model since its flexibility is high, permitting the user to change force deployments, force capabilities, manning levels, station locations, and operational strategy without necessitating changes in basic model structure. Lesser changes, such as modifications in standard model output, would be worth considering as needs for improvement are identified.

The most immediate operational usage of the model should of course, be limited to the districts for which calibration has been accomplished. Simultaneously, efforts should be applied to calibrating parameters in districts not yet exercised, using calibration parameters from roughly comparable regions as guides for setting initial values. In particular, priority should be given to those localities where budget decisions relative to force levels are most urgently required.

The Coast Guard might also deem it worthwhile to invest in a smaller model designed to investigate and sound the feasible set of operationally acceptable force levels. This would permit reserving the exercise of SARSIM for in-depth exploration of the situations and allocations of greatest interest. (The possibilities of economies potentially obtainable by this procedure have been advanced in a proposal submitted in November 1970.)

Finally, it cannot be stressed too strongly that the entire developmental and validation effort was possible only through the complete cooperation of Coast Guard management. There was a continuing, integrated team effort by individuals from the Coast Guard and the National Bureau of Standards. The end-product of these efforts is a proven simulation model, ready for operational use.

APPENDIX A

Mathematical Insights for Application to SARSIM

1. The Nature of a Figure of Merit

A single index, or figure of merit (FOM), is often employed to describe the level of performance of a complex system. For search and rescue, the FOM may be used to indicate how a given set of resources provides service, with high values reflecting inadequacy, although not necessarily inefficiency of operation. It may be observed that, by common practice, figure of merit increases with less adequate performance (service in this context), and thereby reflects unsatisfactoriness in the attribute measured. One must keep in mind that low figures represent greater merit.

The figure of merit lends itself to summation over a series of cases, stations, and time, and also to calculation of average values.

2. Application to SAR

Consistent with ideas developed in Volume II of the SARSIM documentation, a figure of merit can be defined to reflect the degree of adequacy of service by the Coast Guard to units in distress. In particular, as will be clear from the definitions to be provided, timeliness in arrival of a SAR resource is the essential consideration. In particular, no attempt will be made here to assess the quality of service provided after the resource has arrived on scene.

The figure of merit to be assigned to each case requiring assistance depends on the urgency (or severity) of the case, and on the amount of time which elapses from receipt of notification of distress by the Coast Guard until a suitable resource arrives. To be more precise, a "tolerance time" is established to reflect for each stipulated level of severity a maximum acceptable waiting time: if a server arrives within this pre-set tolerance time, the figure of merit is zero, indicating that service was provided in timely fashion. If a resource does not arrive within the specified tolerance time, a penalty is assessed, the amount of penalty depending both on the severity of the case and the client's waiting time in excess of tolerance.

Consideration may also be given to other pertinent factors, such as the distance from shore of the distressed

unit. Thus, for example, it may be desirable to access lesser penalties when the server cannot possibly arrive within tolerance time (because of the case's distance and the resource's maximum attainable speed) than would be given for an equivalent wait when the client is located within acceptable time-range of service, but suffers delay due to other factors.

3. Mathematical Formulation for FOM

In mathematical terms, the figure of merit for SARSIM will be defined as $G(t,q)$, where $t \geq 0$ is the waiting time in excess of tolerance (from time of receipt of notification by the Coast Guard until the first server arrives on scene) and q is a variable with domain any of a set of qualitative characteristics of the case. By definition,

$$G(t,q) = 0 \quad \text{for} \quad t \leq 0$$

and $G(t,q)$ is monotonically increasing in t (i.e., $G(t,q)$ may level off, but does not decrease with increasing t).

One simple form which satisfies these conditions is :

$$\begin{aligned} G(t,q) &= a(q) + b(q) * t \text{ for } 0 \leq T(q) \\ &= a(q) + b(q) * T(q) \text{ for } f \geq T(q) \end{aligned}$$

where $a(q) \geq 0$ and $b(q) \geq 0$ are suitable selected functions of case severity and $T(q)$ is an upper bound on excess time, as defined later.

Clearly, there is infinite choice for combinations of parameters in this single type of formulation. The essential criteria for utility of the FOM are that it be easily derivable from case data and that it appropriately reflect different levels of performance.

The specification of values for $a(q)$, $b(q)$ and $T(q)$ may be contingent on categorization of cases, where fewer classes simplify computation, storage, and comprehensibility but where greater flexibility and utility may stem from

increasing the extent of partitioning. The effect of partitioning is compounded when the figure of merit is to be applied to a number of cases, not all of which are in the same category. Within a given class, the figure of merit is computed by:

$$D = \sum_i G\ (t_i,\ q_i)$$

where each case is indexed and the summation is over indices of cases in the specified class. For total system performance,

$$D(I) = \sum_{i \varepsilon I} G(t_i,\ q_i)$$

where I is an index set of all cases.

Interpretation is facilitated if D(I) is normalized, especially when it is desired to compare situations concerning time periods of different length. Normalization can be accomplished with respect to length of time or number of cases covered by I.

4. Initial Candidate FOMs

Four formulae were initially defined as figures of merit, two of which treat all cases alike, and two of which partition cases into two classes, based on s severity level. Class Q_1 is defined as the set of all cases of severity level 1 or 2, and Q_2 includes cases of severity levels 3, 4 and 5.

The first four candidates for figures of merit are established as follows:

Formula I

$a(q) = 1,\ b(q) = 0$ for all q

In this formulation, each case with missed tolerance is penalized one unit, independent of case severity level* and excess time over tolerance.

Formula II

$a(q) = 1,\ b(q) = 0$ for $q \in Q_1$

$a(q) = 2,\ b(q) = 0$ for $q \in Q_2$

Again, only the fact that tolerance is missed (and not by how much) is taken into account. However, cases of severity levels 3, 4 and 5 are assessed twice the penalty of levels 1 and 2.

---

*It should be noted, however, that the tolerance time which was missed is itself dependent on severity level.

Formula III

$a(q) = 0,\ b(q) = 1,\ T(q) = 0.5$ for all q.

Here the full emphasis is placed on waits longer than one-half hour, regardless of severity. In effect, this formula accumulates all times waited in excess of the stipulated tolerance, and a single very long wait has as much effect on the total as a very large number of small excesses.

Formula IV

$a(q) = 0,\ b(q) = 1,\ T(q) = 0.5$ for $q \in Q_1$

$a(q) = 0,\ b(q) = 2,\ T(q) = 0.5$ for $q \in Q_2$

As in the preceding formula, only time in excess of a half hour contributes to the figure of merit, but the higher severity levels are double the importance of the lower.

5. Application of Formulae I-IV

The four candidate formulae defined above were applied to data from Quick Query run 602, based on OPSIM run number 7 for the First Coast Guard District. There were 881 cases, divided almost equally into eleven subsets. The resultant values of figure of merit are presented in Table A-1, where comparisons in magnitude should be made within columns.

As a means of facilitating comparison, the subsets have been grouped, by formula, into sets of nearly equal size and in order of increasing FOM, as shown in Table A-2.

Subset	Cases	Formula I	Formula II	Formula III	Formula IV
1	1 - 80	3	6	.64	1.28
2	81 - 160	3	6	.04	.08
3	161 - 240	7	11	.51	.87
4	241 - 320	1	2	.01	.03
5	321 - 400	4	7	.21	.41
6	401 - 480	6	9	.49	.52
7	481 - 560	3	4	.51	.52
8	561 - 640	5	8	.43	.62
9	641 - 720	2	4	.01	.03
10	721 - 800	4	8	.68	.54
11	801 - 880	6	9	.68	.71

Table A-1: Values of Figure of Merit Calculated by Four Formulae

Formula I	1-2 4,9	3 1,2 7	4-5 5,8 10	6-7 3,6 11
Formula II	1-4 4,7 9	5-7 1,2 5	8 8,10	9-11 3,6 11
Formula III	0-.10 2,4 9	.11-.45 5,8 10	.46-.55 3,6 7	.45-1 1,11
Formula IV	.0-.10 2,4 9	.11-.52 5,6 7	.53-.75 8,10 11	.76-1 1,3

Table A-2
Quartile rank position by Figure of
Merit as Calculated by Four Formulae

As would be expected, the different natures of the four formulae are manifested in the differences in groupings, some of which are more striking than others.

6. Extreme Values

As is suggested under the definition of Formula III, it is possible for an individual case with a high excess over tolerance to dominate the figure of merit for a set of cases whenever $b(q) > 0$. In order to limit the contribution of penalty from a single case, a maximum waiting time in excess of tolerance, $T(q)$, can be set as a function of severity; the value of the FOM is then limited to that attained when $t = T(q)$. Thus:

$$G(t,q) = q(q) + b(q) * t \quad \text{for } 0 < t \leq T(q)$$
$$= a(q) + b(q) * T(q) \quad \text{for all } t \geq T(q)$$

In many instances an extremely long wait for the first server to arrive stems not from lack of capable resources or failure of the system to react properly, but solely as a function of case parameters (e.g., far from shore). Another technique for bounding the figure of merit and for limiting the impact of these external problems is to adjust the parameters in the linear formulation - or, perhaps, resort to slightly more complex formulae. (The latter approach has been reserved for subsequent investigation since it could not be implemented during the SARISIM validation phase.) The method chosen and applied thus far is the selection of values for figure of merit parameters so that, for given values of severity or tolerance time, cases farther from shore do not incur as much penalty as those closer in despite equal waits in excess of tolerance.

7. Formulae Incorporating Distance Effects

To include the effects of distance from shore in the FOM calculations, cases can be segregated (for example) into those within 20 miles of shore and those farther away. (It can be observed from Figure A-1 that less than half of the cases in the sample* used occurred beyond 20 miles from shore. This particular numerical break-point was chosen on the basis of a comparison of speeds of current Coast Guard resources and tolerance times for cases of differing severity levels. Distance from shore in excess of 20 miles generally precludes arrival within tolerance time.)

Class Q_3 is defined as the set of all cases within 20 miles of shore and Q_4 as the set of all cases beyond 20 miles Referring to the earlier definitions of Q_1 (i.e., severity levels 1 and 2) and Q_2 (i.e., severity levels 3, 4 and 5), additional classes may be defined as follows:

$Q_5 = Q_1 \cap Q_3$ (that is, all cases of severity level 1 or 2 within 20 miles of shore)

$Q_6 = Q_2 \cap Q_3$ (severity 3, 4 or 5 within 20 miles)

$Q_7 = Q_1 \cap Q_4$ (severity 1 or 2 beyond 20 miles)

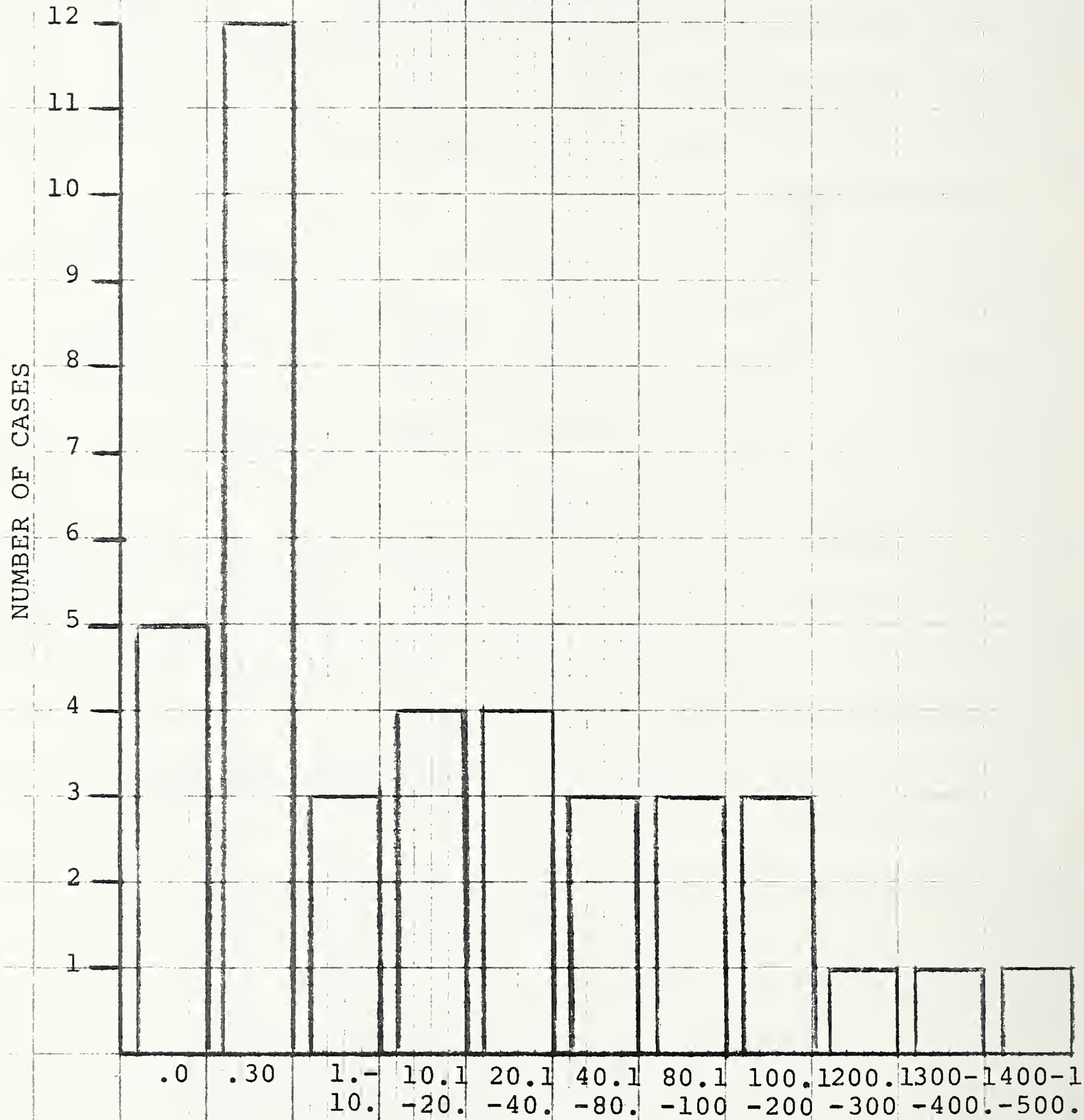
$Q_8 = Q_2 \cap Q_4$ (severity 3, 4 or 5 beyond 20 miles)

These definitions give rise to the following additional FOM formulae:

---

*The sample includes only those cases which were not served within tolerance time.

Figure A-1: Number of Cases which occurred at Various Distances From Shore in a Sample*of 40 Cases.



*See footnote on previous page.

Formula V

$a(q) = 0. \ b(q) = 2, \ T(q) = 10$ for $q \in Q_3$

$a(q) = 0, \ b(q) = 1, \ T(q) = 10$ for $q \in Q_4$.

Here (as in Formulae III and IV) time in excess of tolerance is accumulated over all cases; however, the penalty is doubled for the closer cases and, in any individual case, the maximum penalty is that assigned to a wait of 10 hours beyond tolerance time.

Formula VI

$a(q) = 0, \ b(q) = 2, \ T(q) = 10$ for $q \in Q_5 \cup Q_8$

$a(q) = 0, \ b(q) = 3, \ T(q) = 10$ for $q \in Q_6$

$a(q) = 0, \ b(q) = 1, \ T(q) = 10$ for $Q \in Q_7$.

This formula, too, accumulates penalty in terms of time in excess of tolerance, with a cut-off at 10 hours beyond tolerance. Maximum penalty is incurred for $q \in Q_6$, that is, high severity cases close to shore, and least penalty is assessed when $q \in Q_7$, or to low severity cases far out.

Using the same data as in Section 5, Formulae V and VI have been followed to obtain the results shown in Table A-3. As in Table A-2, the FOMs by subset have been grouped into four sets by increasing values, shown in Table A-4; the results presented earlier for Formula IV are included for comparison.

Subset	Cases	Formula V	Formula VI
1	1 - 80	.90	1.66
2	81 - 160	.04	.08
3	161 - 240	.58	.94
4	241 - 320	.01	.03
5	321 - 400	.30	.50
6	401 - 480	.78	.80
7	481 - 560	.52	.53
8	561 - 640	.87	1.05
9	641 - 720	.02	.04
10	721 - 800	.33	.61
11	801 - 880	1.09	1.13

Table A-3: Values of Figure of Merit Calculated by Formulae V and VI

	.0-.250	.251-.50	.51-.80	.81-+
Formula V	2,4 9	5,10	3,6 7	1,8 11
	.0-.30	.31-.70	.71-1.0	1.1-+
Formula VI	2,4 9	5,7 10	3,6	1,8 11
	.0-.10	.11-.52	.53-.75	.76-1.
Formula IV	2,4 9	5,6 7	8,10 11	1,3

Table A-4: Quartile Rank Positions by Figure of Merit Calculated by Formulae V, VI and IV (from Table A-2).

8. Figure of Merit Computed in SARSIM Runs

Several statistics based on the figure of merit have been incorporated in the OPSIM calculations. Under "Station Response" in the OPSIM output are: (1) Average Positive TWAIT - Tolerance (abbreviated as ATW here), and (2) Normalized Demerit (abbreviated as NDM here). These statistics are related to the FOMS computed by Formulae III and VI respectively.

Average Positive TWAIT - Tolerance is obtained by computing for each case $G_{III}(g_j, t_j)$ according to a modification of Formula III in which no upper bound $T(q)$ is specified. The values $G(q_j, t_j)$ for all cases served by a given station are summed and the total divided by the number of cases served by that station.

NDM is obtained for each case by computing $G_{VI}(q_j, t_j)$ according to Formula VI again with no upper bound specified. For the district being simulated, $\Sigma G_{III}(q_j, t_j)$ and $\Sigma G_{VI}(q_j, t_j)$ are obtained by summing over all cases served. The ratio, $S = \dfrac{\Sigma G_{III}(q_j, t_j)}{\Sigma G_{VI}(q_j, t_j)}$ is then computed. If I is the index set of cases served by a given station, let $D(I) = \Sigma_{i \varepsilon I}\, G_{VI}(q_j, t_j)$ and N be the number of cases served by that station. For the given station, $NDM = \dfrac{S*D(I)}{N}$, where S is a scaling factor.

In the computer print-out for an OPSIM run, Note 1 under Station Response provides the value of the scaling factor and Note 2 specifies the coefficients used in the Formula VI FOM for the three classes into which the case parameter set has been partitioned.

Scaling allows more direct comparison of ATW and NDM. For a given positive t, the coefficients in $G_{VI}(q,t)$ used for different classes of cases alter the relative significance of different cases, but these coefficients also tend to raise G_{VI} summed over a given class of cases above the value of G_{III} summed over the same cases. Scaling is designed to eliminate the latter effect while preserving the former.

For a given station, ATW > NDM indicates that the cases occurring at that station during the period represented in the simulation were, on the average, those with case parameters in the set with the smaller coefficient assigned by Formula VI. On the other hand, if ATW < NDM, the cases on the average had case parameters in the set with larger coefficients assigned by Formula VI. Thus through scaling, the factors used in defining Formula VI (i.e., severity and distance from shore) are reflected in the comparison of NDM with ATW.

9. Examples

To illustrate the foregoing some of the output of OPSIM run 56, a simulation of the first Coast Guard District during the winter, are reproduced below:

Station	ATW	NDM
8	0.12	0.22
9	1.94	2.48
17	0.25	0.32
18	23.81	15.25
33	1.91	3.68

The Scaling factor for this run was: S=0.64

There are forty seven stations in this district, but tolerance was not exceeded for any of the other stations. (Since this was a winter simulation, most stations were able to meet all calls within tolerance.)

For each station, the ratio $r = \frac{NDM}{ATW}$. is an index of the seriousness of the cases which resulted in failures, where seriousness is here interpreted in accordance with the weighting coefficient of Formula VI. A comparison of these indices is shown below:

Station	r
8	1.83
9	1.28
17	1.28
18	.64
33	1.93

Another example, from the simulation of July 1968 cases in the First District, is presented on the following page.

It should be emphasized that the figure of merit cannot be examined in isolation, but much other data is also available from any OPSIM run. The figures of merit do, however, offer a useful index of cases which have failed to get (relatively) prompt service.

Three tables follow to illustrate typical data which can be supplied by the POSTPRO (Quick Query) for cases with failure type C. These have been culled from Quick Query run 6021 from OPSIM run #7 for the First District.

Station	ATW	NDM	r
4	6.77	5.04	.74
7	0.11	0.17	1.55
9	0.13	0.20	1.54
10	0.05	0.08	1.60
11	0.13	0.21	1.62
12	0.01	0.02	2.00
14	0.10	0.16	1.60
15	0.02	0.03	1.50
16	0.01	0.01	1.00
18	0.12	0.19	1.58
20	0.02	0.03	1.50
21	0.02	0.02	1.00
22	0.06	0.10	1.66
23	0.01	0.10	1.00
24	0.15	0.23	1.53
25	0.02	0.03	1.50
26	0.07	0.11	1.57
31	1.31	1.75	1.34
32	0.04	0.06	1.50
33	0.82	1.07	1.30
34	0.57	0.60	1.05
35	0.01	0.01	1.00
37	0.02	0.02	1.00

The scaling factor for this run was: S = 0.519

Seq. No. (Tolerance Code QQP-3)	Severity (Case attributes)	Time to Tolerance	Wait Time for 1st arrival (QQP-4)	t	Miles Offshore (Case Attribute Definition)
3	5	.02	.02	.00+	20.00
33	5	.02	.64	.64	315.00
39	4	.04	.180	.14	.30
84	5	.02	.05	.03	21.00
88	5	.02	.03	.00+	.30
121	4	.04	.04	.00+	.30
200	5	.02	.02	.00+	.00
201	5	.02	.04	.02	90.00
203	5	.02	.35	.33	25.00
222	5	.02	.02	.00+	8.00
237	1	.17	.24	.07	74.00
238	1	.17	.23	.06	.30
239	1	.17	.14	.02	75.00
316	5	.02	.04	.01	85.00
352	4	.04	.15	.12	200.00
355	5	.02	.04	.02	.00
386	1	.17	.18	.01	.25

Seq. No. (Tolerance Code QQP-3)	Severity (Case attributes)	Time to Tolerance	Wait Time for 1st Arrival (QQP-4)	t	Miles Offshore (Case Attribute Definition)
373	3	.08	.14	.06	.30
401	1	.17	.36	.20	14.00
418	1	.17	.35	.18	147.00
437	1	.17	.26	.09	.25
439	5	.02	.04	.02	250.00
455	4	.04	.04	.00	.30
457	5	.02	.02	.00	.30
482	5	.02	.03	.01	8.00
523	1	.17	.49	.32	470.00
546	1	.17	.34	.18	128.00
572	1	17.	.38	.21	.25
579	1	.17	.21	.04	16.00
587	5	.02	.03	.00	.00
601	4	.04	.08	.04	.30
607	4	.04	.18	.14	.30
677	5	.12	.02	.00	40.00

Seq. No. (Tolerance Code QQP-3)	Severity (Case attributes)	Time to Tolerance	Wait Time for 1st Arrival (QQP-4)	t	Miles Offshore (Case Attribute Definition)
703	5	.02	.03	.01	.30
761	4	.04	.25	.21	83.00
785	3	.08	.10	.02	.30
787	3	.08	.12	.03	.00
793	5	.02	.03	.01	.00
811	5	.02	.03	.01	.30
812	5	.02	.02	.00	.20
819	1	.17	.48	.31	3.00
821	5	.02	.04	.02	.55
825	1	.17	.26	.09	.25
830	1	.17	.41	.25	22.00

10. Some Mathematical Aspects of Risk

In terms of the figures of merit discussed in this appendix, two similar, but not identical, concepts of risk may be formulated with respect to SAR. The problems of finding numerical values for the risk function are discussed in the following section. For a given time interval, a stochastic process $S_1(t)$ is specified for cases requiring assistance. Another stochastic process, $S_{2(m,n)}(t)$ may also be specified to describe service when a resource of type m is assigned to a case of type n, given a set of resources and a policy for assigning resources to cases as they occur.

For each case $(j_i$, denote by t_i the amount of time in excess of tolerance before the first server arrives on scene, if positive. (If the waiting time does not exceed the tolerance, $t_i = 0$.) As before, $G(q_i, t_i)$ is the figure of merit; $G(q_i, t_i) = 0$ for $t_i = 0$ and is monotonically increasing for $t_i > 0$. For a period of time from 0 to T, $D(T) = \sum G(q_i, t_i)$ where the summation is over all cases occurring in the period.

For the conditions specified above, further define: $F(x) = \Pr\{D < x\}$. For $\bar{x}$ a positive number, $1 - F(\bar{x})$ is the probability that D is not less than $\bar{x}$ and can be called the risk of attaining level $\bar{x}$.

Alternatively, we can define $G_o(x) = \Pr\{\max\ [G(q_i, t_i)] < x\}$. For $\bar{x}$ a given positive number, $1 - G_o(\bar{x})$ is the probability that $\max\ [G(q_i, t_i)]$ is at least $\bar{x}$. This, too, might be called the risk associated with level $\bar{x}$.

The functions F and G_o depend on four factors: (1) $S_1(t)$, (2) $S_2^{(m, n)}(t)$, (3) the inventory of resources, and (4) the resource assignment policy. Conditions affecting these functions, F and G_o, find expression in one or more of the factors listed.

11. Estimating Risk from a Simulation

Given the four factors listed above for a time interval [0,T], it is not always easy to obtain the functions $F(x)$ and $G_o(x)$. Even for a small number of resources and stochastic processes $S_1(t)$ and $S_2^{(m,n)}(t)$ of simple form, it may not be possible to get tractable expressions. These functions or their values may possibly be obtained for specified x by simulation.

It is assumed that $S_1(t)$ and $S_2^{(m, n)}(t)$ are not both deterministic processes, but that there exist mutually exclusive events such that the probability of each of these events is positive. If this assumption does not hold, then one run of the simulation is sufficient to assure whether the figure of merit does or does not attain a given value.

In setting up and using a simulation there are two classes of questions. The first is: do the data fed into and generated by the computer correspond to a description of the events of the processes being simulated? If so, a second question must be posed: How many times does the process have to be repeated (i.e., run on the computer) in order to verify any given assertion with respect to the process?

Let E be an event, one among those used to define risk in Section 10, and let p denote the probability that

this event will occur on a given day. The four factors listed in Section 10 are assumed to be the same on all days under consideration for this problem. It is also assumed that there is no cary-over from one day to the next in terms of caseload, and that the process $S_1(t)$ and $S_2^{(m,n)}(t)$ for any sequence of days are completely independent.

The following argument applied to general random sampling with replacement can be found in W. Feller - <u>An Introduction to Probability Theory and Its Applications</u>, Vol. I pp. 189 - 190. It applies to this problem under the assumptions given above. In n independent simulations of the process, let n_1 be the number of the simulations in which E occurred. The expected value of n_1/n is p. The interval $n_1/n \pm k\sqrt{p(1-p)/n}$ will be such that it contains p with a probability, depending on k, which can be read off the table of the normal curve with sufficient accuracy. k is the number of units of standard deviation; the table of the normal curve enables one to determine the area under that curve contained in the part between $u-k\sigma$ and $u+k\sigma$, where u is the mean and σ the standard deviation. [While the size of this interval depends on an unknown p, its maximum is $n_1/n \pm k/2\sqrt{n}$. Further <u>a priori</u> information about p may narrow the size of the interval.] Thus, wishing an interval estimate of p of

size d that will cover p with probability s, we put $2k \sqrt{p(1-p)/n} = d$. In place of the unknown $p(1-p)$ we put the maximum value of that expression, 1/4. This gives us $n = (k/d)^2$ as the number of times it will be necessary to run the simulation in order that, with probability as least s, the interval $n_1/n \pm d/2$ should contain p.

Of course Feller's warning, "The practical difficulty is usually to obtain a representative sample of any size," may also hold for simulations.