THE USE OF SIMULATION IN THE DEVELOPMENT AND EMPIRICAL VALIDATION OF ANALYTIC MODELS FOR EMERGENCY SERVICES

Edward J. Ignall[1], Peter Kolesar [2]
and Warren E. Walker [3]

[1] Columbia University; [2] City College, CUNY;
[3] The New York City Rand Institute

ABSTRACT

Simulation models are generally costly tools to use in systems analysis. Whenever applicable, one prefers to use a simple analytic model. However, in many cases, the conditions assumed by solvable analytic models do not hold in the real world. But a simulation can be used to suggest an approximate model and to determine how good an approximation an analytic model is. We show how simulations of New York City's fire and police operations have been used to develop and validate simple analytic models which are now being used to determine the deployment of resources in these two services.

I. INTRODUCTION

A simulation model of a large and complex system can be a very useful, but time-consuming and costly tool to use in systems analysis. Whenever applicable, one prefers to use a simple analytic model yielding closed form algebraic expressions relating system inputs and outputs. However, in many cases, the conditions assumed by solvable analytic models do not hold in the real world and more realistic models are too complex to solve--hence simulation. Typically, simulations of complex systems are used to provide specific numerical estimates of performance under specified conditions. This is use of the simulation as a "pilot plant." However, a simulation also can be used to suggest an analytic model or validate one. If the analytic model provides an adequate approximation, it can be used more economically instead of the simulation for future analyses. Use of simulations oriented toward the development and testing of other mathematical models is analogous to experiments carried out by physical scientists in their development of new theory. It is this sometimes overlooked use of simulation models that we focus on in this paper.

At The New York City-Rand Institute, we have developed large-scale simulation models of the fire operations of the New York City Fire Department and of the patrol activities of the New York City Police Department. Although the fire simulation presently models one borough of New York City and the police simulation presently models one police precinct, insights obtained from each of the simulations have been used to produce and verify several analytic models having City-wide applicability.

We discuss three such models in this paper. In one case, an analysis of police patrol car allocation, the simulation and analytic models were constructed in parallel. One of the chief reasons for building the simulation was to determine how well and under what conditions an analytic queueing model represented the real world. In the second case, the estimation of fire engine response times, an analytic model was suggested after the simulation was written, and special simulation runs were made to confirm its validity. In the third case to be discussed, prediction of the number of fire engines dispatched to an alarm, the analytic model was suggested (and verified) by an analysis of simulation runs which had already been made for other purposes.

II. A MODEL FOR ALLOCATING POLICE PATROL RESOURCES

In [6] a queueing model is proposed to represent the activities in a police department dispatching center. A patrol car is dispatched immediately to answer a call for service arriving at the center if one is available; otherwise, the call is queued. The queueing model is the simple M/M/N priority model of Cobham [3]. The assumptions underlying the model are not all satisfied in the operating environment of the New York City Police Department. The basic assumptions are that calls arrive according to a stationary Poisson process, that service times are independent and exponentially distributed, and that each call is served by a single patrol car.

While call arrivals for any short interval are approximately Poisson, call rates, even during a single 8-hour tour of duty, are not constant. Service times are not exponentially distributed, and include the time required for a car to travel to an incident, which depends on the number of cars available to dispatch. Moreover, a call may be served by more than one patrol car.

As a result, although we wanted to use the stationary queueing predictions of the simple M/M/N model to analyze deployment options for the NYPD, we first had to verify that, despite the above-mentioned variants from the model, it still produced predictions of sufficient accuracy. To make appropriate tests, we wrote a detailed police patrol simulation of a single police precinct [5]. The simulation included all the complexities mentioned above as well as others and used actual call histories in the precinct for arrivals and service times. We compared simulation results to those obtained from the queueing model with the same average call rate and the same average service time. The results, described below, were close enough to give us and the Police Department confidence that the queueing model could be used instead of the simulation model to analyze some important deployment problems.

Based on the call rate, average service time, and number of servers, the queueing model gives the probability distribution of the number of calls being serviced and the number waiting to be dispatched. From these probabilities a great deal of information about the performance of the system can be obtained. For example, suppose N patrol cars are on duty and let P_i (i = 0, 1, 2,...) be the probability that there are i calls for service in the system (being served and waiting to be served). Then, two of the quantities which can be calculated are:

(1) The probability that all N patrol cars are busy (q_N):

$$q_N = \sum_{j=N}^{\infty} P_j$$

(2) The average time a call will spend in queue before being dispatched ($\bar{D}$):

$$\bar{D} = 1/\mu \sum_{j=N}^{\infty} (j + 1 - N)\, P_j$$

where $1/\mu$ is the average service time.

To test the usefulness of this model we calculated these quantities as functions of N and compared them to results obtained from the simulation model. The 71st Precinct in Brooklyn was chosen for study because a particularly rich set of data on its operations was available. We analyzed records of all calls for service received within the precinct during the months of August and September 1972. The calls were aggregated by their time of occurrence into the three shifts or "tours" worked by the policemen: Tour 1 - midnight to 8 a.m.; Tour 2 - 8 a.m. to 4 p.m.; Tour 3 - 4 p.m. to midnight. An average call rate was determined for each tour. The average service time was found to be approximately the same for all tours. The queueing model was used to analyze conditions for tours 1 and 3 with different numbers of cars on duty (N ranged from 4 to 12). The simulation was run for the same values of N, using as input the actual stream of calls for service for a given tour (the input stream for the simulation of a given tour was compared by concatenating all of the calls received during that tour for the months of July and August 1972. For example, when simulating tour 3, the last call before midnight on one day would be followed by the first call after 4 p.m. on the following day).

A comparison of the results from the simulation and queueing models is given in Figs. 1 and 2. Figure 1 shows the percent of time that all patrol cars are busy (relationship 1). The results are remarkably similar, with the queueing predictions being consistently slightly lower than the simulation results. (We predicted this difference because the simulation makes multiple car dispatches while the queueing model assumes that one car is sent to each call.) Figure 2 plots the average queueing delay (relationship 2) as a function of the number of patrol cars on duty. The results, again, are quite close.

The Police Department accepted the fact that the queueing model does represent a reasonable approximation to the dispatching and service activities of the patrol force, and have begun using it as an aid in determining the number of patrol cars to assign to duty during each tour in each police precinct.

III. A MODEL FOR ESTIMATING FIRE ENGINE RESPONSE TIMES

In [4] Kolesar and Blum derive an inverse square root relationship between the average distance traveled by emergency units responding to calls for service and the number of locations in the region from which they respond. The relationship was derived rigorously under several idealized mathematical conditions¹. It was derived for an infinitely large region in which the units are located either uniformly on a grid or purely at random, and in which calls for service are distributed homogeneously in space, while emergency vehicles travel along simple response paths. However, to have practical usefulness, it was important to show that the relationship provided a reasonable approximation to actual average response distances under more general and realistic conditions.

An existing simulation of New York City fire fighting operations was used to test the validity of the model for fire engine responses. Details of the simulation design are given in [1] and some of its other uses for policy analysis are described in [2]. Two hypothesized relationships were to be tested by simulations:

(1) The expected response distance of the closest available unit (ED(N)) to a fire alarm occurring when there are N available engines in a region of area A, is given by

$$ED(N) = K\sqrt{A/N}.$$

(2) If there are n engines located in a region of area A and if, on the average, b are busy, then the average first engine response distance to all calls for service in the region is approximately given by

$$\bar{D}(n) = K\sqrt{A/(n - b)}.$$

(Note: $\bar{n}$-b is the long-run average of N and $\bar{D}(\bar{n})$ is the long-run average of ED (N).) Since n is a major policy variable under management's control, this relationship is of more general interest than (1). It should also be noted that, if the square root relationship in (1) holds, then (2) cannot also hold exactly since $E(\sqrt{x}) > \sqrt{E(x)}$. But, by means of simulation, we wished to determine if (2) was a good approximation.

Verifying these relationships by gathering empirical data from Fire Department operations would be an extremely difficult task. In fact, verifying (2) would require the Department to vary the number of units operating in an experimental region at different times - an unthinkable procedure if the changes meant using so few companies that lives and property were endangered. Instead, by means of simulation, these tests could be made safely and economically without any modifications to actual Fire Department operations.

The fire simulation models operations in the Bronx, one borough of New York City. Relationship (1) was verified by recording results of operations in a small, high-incidence region in the South Bronx. Several simulation runs were made at alarm rates of 9.4 and 15 alarms per hour in this region. The number of engine companies assigned to the region was held constant at 21 for all runs. The number of ladder companies was varied from 13 to 19. Relationship (2) was verified using simulation results of operations in the whole Bronx. Some of the simulation runs used had already been made for other purposes and some were specially tailored for this analysis. During each simulation run the program recorded, among other statistics:

- Response distance of the first ladder and first engine arriving at an alarm in the South Bronx, together with the number of ladder and engine companies available in the retion at the instant the alarm occurred (to examine relationship (1)).
- Average first engine and first ladder response distances to all alarms in the Bronx and the average number of Bronx engine companies and ladder companies available during the course of the entire simulation run (to examine relationship (2)).

Relationship (1)

To analyze the relationship

$$ED(N) = K\sqrt{A/N}$$

we concentrate on the results of two simulation runs. The first was run with 13 ladder companies in the region at an alarm rate of 9.4 alarms per hour. The second was run with 19 ladder companies and an alarm rate of 15 alarms per hour. In each case, the response distances for all responses which we made when N ladders were available were averaged for each value of N from 0 to 13 or 19. These averages are plotted against N in Figs. 3 and 4. To test the validity of the square root relationship the curve

$$ED(N) = \alpha/\sqrt{N}$$

was fit to the data using least squares regression. An inspection of the graphs shows that the resultting curves fit the data well. The value of R^2 from the regression is .914 for the 9.4 alarms per hour curve and .764 for the 15 alarms per hour curve.

Relationship (2)

Figure 5 is a graph of overall simulated average response distance for closest ladders over all alarms vs. the average number of ladder companies available in the Bronx over the course of the run. Each simulation run produces one point on the graph (while each simulation run produced a whole curve for Figs. 3 and 4). The number of ladders assigned to the Bronx varied from 12 to 31 in these runs.

Two curves were fit to the data using least squares regression

(a) $\bar{D} = \alpha/\sqrt{\bar{N}}$

(b) $\bar{D} = \alpha\bar{N}^{\beta}$

If relationship (2) is valid, (a) should give a good fit and the value of β in (b) should be close to -1/2. An inspection of the graph shows that (a) provides a good fit and that there is little difference between the curves for (a) and (b). The value of β is -.542 which is "close to" -1/2.

Uses of the Square Root Model

The distances which responding fire units must travel to reach fire alarms is an important measure of the service being provided by the Fire Department. By being able to predict the average response distance in a region as a function of alarm rate, the number of units assigned to the region and other measurable parameters of the region, allocation policies can be evaluated quickly without the use of simulation [4]. Among the many questions which the Fire Department of New York has already used the square root model to answer are:

- What will be the effect on response distance of removing a company from a region?
- How should the number of units on duty be varied over the day (as the alarm rate varies) to maintain a given average response distance in a region?
- How many fire units will be required in the future under projected alarm rates to maintain desired average response distances?

IV. A MODEL FOR PREDICTING THE NUMBER OF UNITS SENT TO A FIRE ALARM

In New York City the dispatching of fire companies to an incoming alarm is governed by information provided on the "alarm assignment card", associated with the fire alarm box closest to the alarm. The first line of an alarm assignment card contains the names of the three closest engine companies and the two closest ladder companies. The traditional policy for alarms turned in by box had been to send whichever first line companies were available, "special calling" companies further down on the card if necessary to assure a response of at least one engine and one ladder. As a result of this policy, (which we call a "New York" dispatching policy), the number of engines and the number of ladders actually sent to a box alarm is a random variable which depends on the availability of fire companies in the area surrounding the box at the time the box is pulled. (For example, as many as three engines or as few as one engine might in fact be dispatched).

We were concerned with predicting how the actual number of units dispatched depended on the alarm rate and the number of units stationed in the region; that is, how the number dispatched depended on the average unit availability. By analyzing some fire simulation runs which had been made for other purposes, we were able to derive a simple relationship between the number of units sent to incidents in a region for which a "New York" dispatching policy is used, and the average unit availability in the region.

We will discuss the use of the simulation in deriving and verifying this relationship for a "New York 2" policy--the traditional dispatching policy for ladders. In similar ways, we have derived and tested relationships for other dispatching policies.

Let a be the average unit availability in a given region; i.e., "a" is the average fraction of the time a unit is available. Define p (n, a) as the probability that n units are dispatched to an incident at which the New York 2 policy is used during a time period in which the average availability is a.

We found that we obtained a good fit to the simulation data by using:

$$(*) \quad P(2, a) = a^2$$

$$P(1, a) = 1 - a^2.$$

This relationship would be true if

(1) The average availability were the same for every unit in the region, and

(2) The event that any particular unit is available were independent of the status of all other units.

Neither of these is true, yet the relationship appears to hold to a good approximation. In fact, the validity of the relationship (*) was discovered in the course of attempting to see how poor a^2 was as an estimate of p (2, a).

The relationship was developed from a set of simulation runs which had originally been made to test the effects at different alarm rates of both adding new companies to an area and modifying the dispatching policy (see [2]). As mentioned before, the simulation models the Bronx and, for data collection purposes, it was partitioned into two regions: the South Bronx, a very high activity region (which we shall call region 2), and the rest of the Bronx (region 1). Nine simulation runs were analyzed in which four different alarm rates and five different allocations of fire companies were used.

For each simulation run we calculated, for each ladder company, the fraction of time it was available. We then obtained the average of these availabilities over all the ladder companies in the region and called this average a. For all incidents in the region for which the response policy was New York 2 we found p(2, a) the proportion that received two ladders as their initial dispatch (these data had been part of the normal output from all simulation runs). The results in Table 1 show that p(2, a) was very close to a^2. In particular, in the first simulation run, based on an observed availability in region 2 of .898 we would predict by relationship (*) that two ladders would be sent to 80.6 percent of all box alarms. In fact, two ladders were sent to 80.7 percent of the box alarms.

The relationship was then validated by analyzing the results on a different set of simulations which had also been run previously but had not been examined during the derivation of the relationship. The original objective of these runs had been to study the effects of matching the number of fire engines on duty more closely to the time-varying alarm rate. Results of four simulations were analyzed and compared to the results predicted by the relationship (see Table 2). On the basis of this comparison we were able to conclude that the relationship we had derived did provide a useful approximation to the actual field dispatching behavior (especially in the high alarm rate region--region 2). We have, therefore, begun to use this relationship (and others so derived for different dispatching policies) instead of the simulation to analyze the effects of various Fire Department deployment options on the number of units dispatched.

RESULTS OF POLICE PATROL SIMULATIONS TO VERIFY QUEUEING MODEL

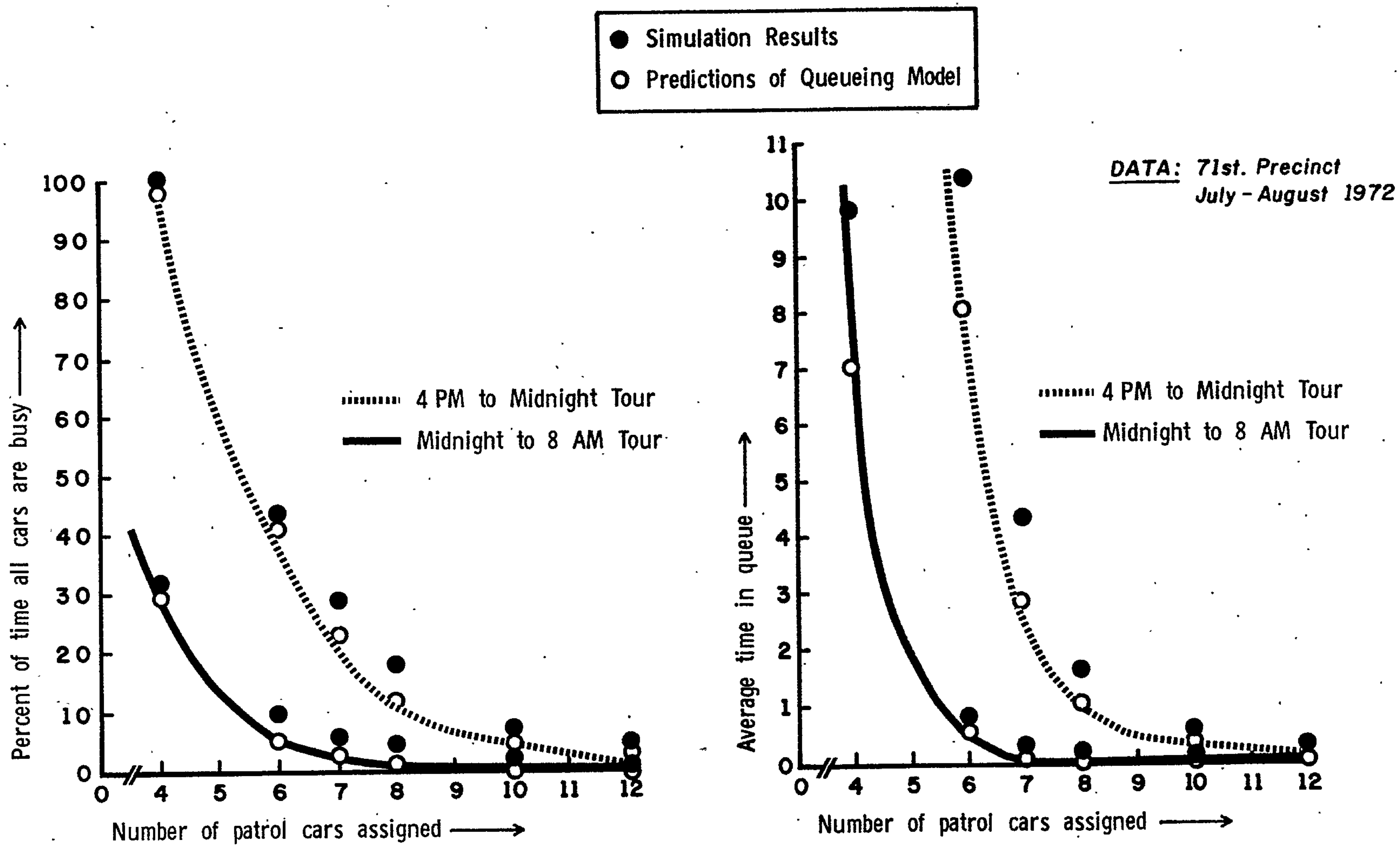

FIGURE 1: Patrol car available vs. number of cars assigned

FIGURE 2: Delay in queue vs. number of patrol cars assigned

RESULTS OF FIRE SIMULATIONS TO VALIDATE THE SQUARE ROOT RELATIONSHIPS

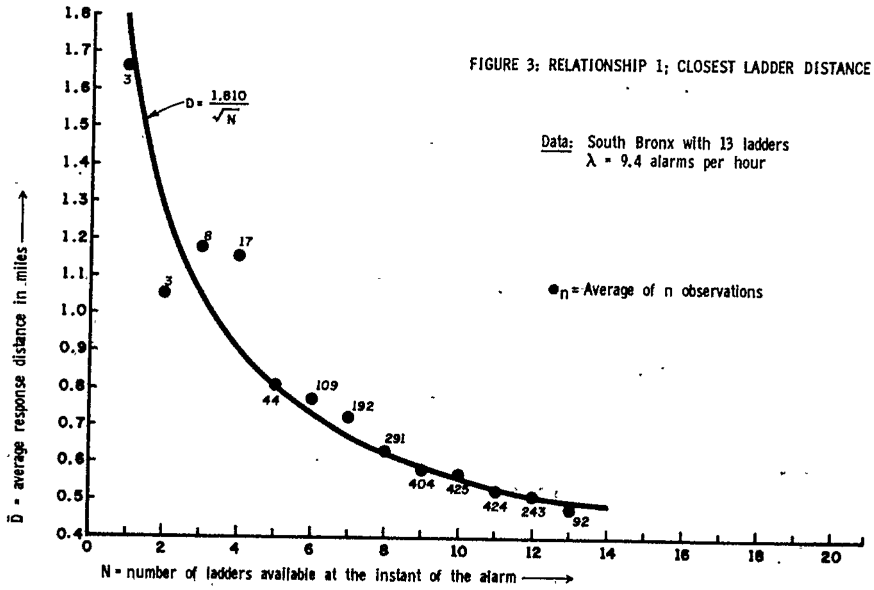

FIGURE 3: RELATIONSHIP 1; CLOSEST LADDER DISTANCE

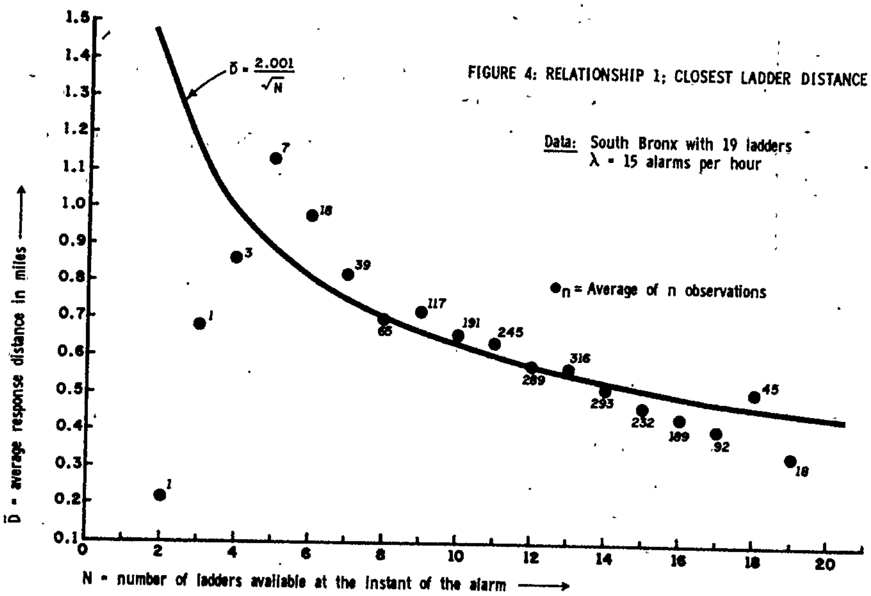

FIGURE 4: RELATIONSHIP 1; CLOSEST LADDER DISTANCE

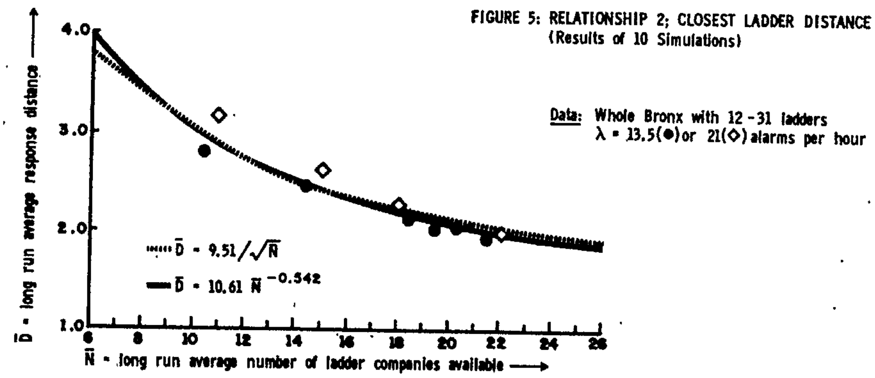

FIGURE 5: RELATIONSHIP 2; CLOSEST LADDER DISTANCE
(Results of 10 Simulations)

TABLE 1

Derivation Of New York 2 Relationship

	Region 1			Region 2	
Observed Availability	Observed and predicted (from availability) percent of NY 2 alarms receiving the indicated number of ladders (observed/predicted) One ladder	Two ladders	Observed Availability	Observed and predicted (from availability) percent of NY 2 alarms receiving the indicated number of ladders (observed/predicted) One ladder	Two ladders
.952	13.2/9.3	86.8/90.7	.898	19.3/19.4	80.7/80.6
.951	21.8/9.6	78.2/90.4	.882	24.7/22.7	75.3/77.9
.877	24.3/23.1	75.7/76.9	.769	35.1/40.9	64.9/59.1
.875	25.0/23.3	75.0/76.7	.752	41.9/43.4	58.1/56.6
.871	26.5/24.1	73.5/75.9	.739	45.8/45.4	54.3/54.6
.869	27.5/24.4	72.5/75.6	.718	52.5/48.5	47.5/51.6
.776	47.3/39.7	52.7/60.3	.555	72.3/69.2	27.7/30.8
.666	58.7/55.7	41.3/44.3	.466	72.7/78.3	27.3/21.7
.623	66.2/61.2	33.8/38.8	.379	85.1/85.6	14.9/14.4

TABLE 2

Verification of New York 2 Relationship

	Region 1			Region 2	
Observed Availability	Observed and predicted (from availability) percent of NY 2 alarms receiving the indicated number of ladders (observed/predicted) One ladder	Two ladders	Observed Availability	Observed and predicted (from availability) percent of NY2 alarms receiving the indicated number of ladders (observed/predicted) One ladder	Two ladders
.945	16.2/10.8	83.8/89.2	.849	27.8/27.9	72.2/72.1
.882	29.8/22.2	70.2/77.8	.705	49.8/50.3	30.2/49.7
.948	16.2/10.2	83.8/89.8	.847	28.1/28.3	71.9/71.7
.888	30.7/21.1	69.3/78.9	.699	50.7/51.1	49.3/48.9
.738	57.5/45.6	42.5/54.4	.494	75.5/75.6	24.5/24.4
.745	55.0/44.5	45.0/55.5	.485	75.8/76.4	24.2/23.6
.689	61.9/52.6	38.1/47.4	.469	86.1/78.1	13.9/21.9

REFERENCES

1. Carter, G., and E. Ignall, "A Simulation Model of Fire Department Operations," IEEE-System Science and Cybernetics, Vol. 6, No. 4, October 1970.

2. Carter, G., E. Ignall, and W. E. Walker, "A Simulation Model of the New York City Fire Department: Its Use in Deployment Analysis," Proceedings of the 1973 Winter Simulation Conference.

3. Cobham, A., "Priority Assignment in Waiting Time Problems," J. Operations Research Society of America, Vol. 3, P. 547, 1955.

4. Kolesar, P., and E. H. Blum, "Square Root Laws for Fire Engine Response Distances, "Management Science,"Vol. 19, No. 12, August 1973.

5. Kolesar, P., and W. E. Walker, "A Description of the New York City Police Department Patrol Simulation Model," The New York City-Rand Institute, WN-8178-NYC, April 1973.

6. Larson, R. C., Urban Police Patrol Analysis, The MIT Press, 1972.