MANAGEMENT SCIENCE
Vol. 35, No. 2, February 1989
Printed in U.S.A.

TESTING THE VALIDITY OF A QUEUEING MODEL OF POLICE PATROL*

LINDA GREEN AND PETER KOLESAR
Graduate School of Business, Columbia University, New York, New York 10027

This paper describes efforts to validate a multiple car dispatch queueing (MCD) model of police patrol operations using New York City data. The MCD model was designed for use in a computer system that has been disseminated to many police departments in the U.S. to help planners allocate patrol cars among precincts. It has also been used to evaluate specific changes in patrol policy in New York. We define validation as a series of hierarchical procedures ranging from tests of mathematical correctness to evaluations of model robustness. We discuss the difficulties and limitations of assessing the validity of a model of a loosely managed system in which human behavior is central and in which controlled experiments cannot be performed. Focusing on specific uses, we conclude that the MCD model is a good although imperfect description of patrol operations in New York and is a potentially useful planning tool for many other urban police departments.
(QUEUEING; VALIDATION; POLICE PATROL)

1. Introduction

This paper reports on our efforts to test the validity of the multiple car dispatch (MCD) queueing model of police patrol car operations that was developed by Green (1984). The MCD model was used to test the feasibility of changing from two to one police officers per patrol car in New York City (Green and Kolesar 1984a), and has been implemented in a revised version of the widely used Patrol Car Allocation Model (PCAM) (see Chaiken and Dormont 1978).

The development and testing of this queueing model originated with a request from the New York City Police Department (NYCPD) for a review of its own queueing based patrol car allocation system which had been significantly underestimating delays. The NYCPD had routinely relied on its model in creating allocation plans, and the model played a role during the financial crisis of the 1970's when difficult decisions had to be made about cutbacks on patrol resources. The seriousness of the actual and potential applications of the MCD model led us to an extensive validation effort for which we have not found parallels in the queueing literature. Though there have been several papers in the computer performance literature on the application and validation of analytic queueing models (see e.g. Chandy and Reiser 1977), there is no treatment in the Management Science/Operations Research literature of the validity of queueing models of systems in which human behavior plays an important role. Such validations are clearly harder to perform and are less likely to be definitive than those of purely mechanistic systems such as computers, or automated production systems. While the question of what constitutes adequate validation of such models is ambiguous, we strongly agree with Gass (1983) that knowledgeable and confident use of these models requires more thorough validation efforts than have been reported to date.

Since there were few guides in the queueing literature itself, we took the several fine discussions of validation issues and approaches for *simulation* models as our starting point. We were particularly influenced by the papers of Van Horn (1971), Naylor and Finger (1968), Fishman (1973), and Sargent (1979).

* Accepted by Arnold Barnett; received July 1986. This paper has been with the authors 5 months for 2 revisions.

0025-1909/89/3502/0127$01.25

Our approach to model validation is through a hierarchical multistage process in which at each stage, tests of the model are performed which, if failed, obviate the need to continue. The stages are:

1. *Model Verification*: An extensive effort to assure that the equations defining system performance and their implementation via a computer program are correct.

2. *Model Significance*: A determination of whether the estimates produced by the model differ substantially from those produced by the model it is intended to replace—in this case, the Cobham M/M/s priority model (1954).

3. *Output Validity*: A determination of whether given reasonable estimates of system parameters, the model produces reasonable estimates of system performance. We do not question here *why or how* the model works—it is viewed as a black box and judged only on its performance.

4. *Structural Validity*: An examination of the validity of its key structural assumptions—asking, in effect, why it works.

5. *Sensitivity Testing*: Testing the sensitivity of the model to violations of its key assumptions. Of particular concern are those assumptions which were identified as being suspect in stage 4, and those assumptions that might be violated in future applications.

This paper addresses the validity issues of stages 3, 4 and 5. Stage 1 is a trivial yet time-consuming exercise. The issue raised in stage 2 was explored in Green and Kolesar (1984b). The results showed that the MCD model generally produces significantly higher estimates of delay than several versions of the Cobham (1954) model which it was a candidate to replace, and hence it results in more reasonable citywide allocations.

Stage 3 is crucial since it examines the model's ability to predict actual system performance on the precinct level. Indeed, this is what is usually thought of as the standard validation method. The ideal would be a controlled experiment in which each input parameter would be known precisely and could be varied to test the model's accuracy under broadly different conditions. For many reasons, detailed later, this was impossible. §3 describes an alternative which, though admittedly problematic, gave important insight about the model's validity.

To some extent, stages 4 and 5, presented in §§4 and 5, compensated for the incompleteness of our procedure in stage 3. These tests are, of course, important in their own right for the better understanding of the phenomena modeled and of the circumstances under which the model is useful.

Our validation effort was guided by and premised on the use of the MCD model as a tool for patrol car allocations as done by PCAM and for analyses of policy decisions such as in the feasibility study reported in Green and Kolesar (1984a). In these "macro" applications, the model determines the number of patrol cars required in each police command (precinct) for a specified service level, e.g. to keep expected delay less than five minutes. Since queueing delay is very sensitive to the number of servers in the system, a model may produce estimates of delay that are significantly off (in percentage error) yet accurately enough identify desirable patrol car allocations. Thus, our primary concern was whether the MCD model's estimates of delay are, in general, of the right order of magnitude to produce good estimates of adequate resource levels, and under what circumstances its estimates may have to be modified.

2. Background

2.1. *The NYCPD Patrol System*

The interest of the NYCPD in the MCD model and our proximity to and familiarity with its operations led to our use of New York City for our test. The City has 73 police commands called precincts, which on average have a population of nearly 100,000, are about four square miles in area, and receive about 35,000 police emergency calls per year. Each precinct's patrol cars operate independently—that is, cars neither patrol outside their precinct nor are typically dispatched across precinct boundaries.[1] Approximately 1,200 police patrol cars—each with two officers—are fielded each day. These cars, called radio motor patrols (RMPs), serve on three nonoverlapping tours of duty: Tour 1 from midnight to 8 AM, Tour 2 from 8 AM to 4 PM and Tour 3 from 4 PM to midnight.

RMPs respond to 911 calls, perform preventive patrol, and react to incidents they encounter. While *official* NYCPD policy dispatches one RMP per job, dispatchers often assign back-ups if deemed appropriate. Moreover, a "primary" car may itself request one or more "back-up" cars, and cars often "assign themselves" to back-up based on the radio transmission they hear between the dispatcher and the primary car. Thus, the number of cars that respond to a call may be greater than the number dispatched. Consequently, approximately 30% of calls get a multiple car response. To make matters more complex, when not patrolling or responding to jobs, RMPs often go "out-of-service" for activities such as precinct assignments, getting gas, car repairs, and meal breaks.

2.2. *The Multiple Car Dispatch Queueing Model*

The key aspects of the MCD queueing model (which is used here to depict a single precinct-tour) are:[2]

- There is a fixed number, S, of identical cars (servers) on patrol duty throughout the tour.
- Emergency calls arrive one at a time according to a stationary Poisson process at rate λ.
- There are K priority classes of calls, labeled $k = 1, 2, \ldots, K$, and P_k is the probability that a particular call has priority k. Priority 1 is highest and there is no preemption.
- Calls in priority class k require the services of j patrol cars with probability C_{kj}, $k = 1, 2, \ldots, K, j = 1, 2, \ldots, S$.
- If a call requiring j servers arrives to find sufficient cars free, service begins at once. All the required cars begin serving simultaneously.
- If not enough cars are available, the call enters the queue according to its priority except that it cannot take the first position in queue if the call in first position already has any cars assigned to it.
- A call in first position in queue has cars assigned to it as they free up and these cars are held for that call until the required number of cars are assigned. At this point the cars begin serving simultaneously.
- Service times of the various cars at each call are independent, identically distributed exponential random variables with mean $1/\mu$, regardless of the priority or of the number of cars needed by the call.

(In what follows we use the terms "server" and "car" interchangeably without implying any distinction.)

2.3. *Data Sources*

The NYCPD agreed to provide data to test the MCD model. Our initial plans were influenced by the validity testing reported by Crabill et al. (1975) in which observers rode in patrol cars and kept detailed logs of all activities while data were simultaneously extracted from the 911 computer system. However, the cost of such a detailed experiment

[1] Our analysis showed advantages to interprecinct dispatching and NYCPD is currently moving in that direction.

[2] Detailed derivation of such operating characteristics as expected delay, probability of delay, and expected number of cars available may be found in Green (1984) which also contains an extension of the MCD model in which the number of patrol cars dispatched may depend on how many are available at the moment that the call arrives.

was prohibitive. Thus, we used only data gathered via the police computer system (acronymed SPRINT), which provides the real-time support for patrol car dispatching.

SPRINT creates detailed records for every call for police service. From these we were able to obtain data on arrival times, service times (including travel) and "dispatch queueing delays"—the interval between the call's entry into the SPRINT system by a telephone operator and dispatch of the patrol cars. Dispatch queueing delay is the major performance measure used by the NYCPD in its planning.

Although SPRINT maintains the real-time status of each patrol car in the field, it does not create historical files on patrol car availability. Thus, we could not determine from these records when cars actually began and ended their tour of duty or when they were out-of-service. To extract data on the crucial issue of patrol car availability, we had to intervene manually and consult SPRINT "status reports" that were only maintained in real time. Each precinct was sampled every 15 minutes. Limited by personnel and one slow teletype terminal, we were able to extract the relevant data during only one tour of duty for two weeks (excluding weekends) for three precincts.

In addition to formal data collection, we spent many hours observing police telephone operators and dispatchers, and riding in RMPs. This first-hand experience was essential for assessing the importance of the various assumptions of the model and in understanding the computer-generated data.

2.4. *Selection of Test Precincts and Time Period*

The three test precincts, chosen in collaboration with the NYCPD, cover a broad range of conditions. Selection criteria included the amount and severity of 911 calls as well as the authors' familiarity with the areas. (Two are the home precincts of the authors and the third was studied extensively by one of the authors while working with the NYC-Rand Institute.)

- The 50th Precinct in the Riverdale section of the Bronx is a largely upper middle class residential neighborhood. It is 5.2 square miles in area, and has an emergency call rate of about 2.7 calls per hour of which about 19% are "high priority-crime related." This precinct was chosen as an example of what we loosely term a "light" precinct.
- The 26th Precinct is on the upper west side of Manhattan and includes several educational institutions (including Columbia University) surrounded by some elegant housing in its south and abandoned tenements in Harlem in its north. The precinct is 0.94 square miles in area, and has an emergency call rate of about 3 calls per hour, of which about 23% are high priority-crime related. This precinct was chosen as an example of a "moderate" precinct.
- The 77th Precinct in central Brooklyn fits the stereotype of a ghetto area with rundown housing and many abandoned buildings. The precinct is 1.7 square miles in area, has an emergency call rate of about 7 calls per hour of which about 28% are high priority-crime related. This precinct was chosen as an example of a "heavy" precinct.

3. A Comparison of MCD Model Estimates to Actual Precinct Performance

A cornerstone of the validation was to determine whether, given accurate estimates of the parameters of a precinct, the MCD model produces reasonably accurate estimates of observed dispatch queueing delays. Ideally, this would be done by running a series of experiments in several precincts during which all operations would be carefully controlled and monitored. However, this was impossible for several reasons: First, as mentioned previously, limited resources and access allowed for only one test period of ten days. Second, the scarcity of patrol officers and cars and the potentially dangerous consequences of manipulating them for test purposes ruled out significant deviations from current operations. (Political considerations, as well, made the NYCPD management very re-

TABLE 1

Counts of Calls for Police Service by Priority and Number of Patrol Cars Dispatched—Ten-Day Experiment

No. of Responding Cars		26th Precinct High Priority	26th Precinct Low Priority	50th Precinct High Priority	50th Precinct Low Priority	77th Precinct High Priority	77th Precinct Low Priority
1	Count	62	131	54	114	149	265
	Percent	63.9%	85.6%	75.0%	86.3%	68.7%	85.2%
2	Count	23	20	16	14	51	38
	Percent	23.7%	13.1%	22.2%	10.6%	23.5%	12.2%
3	Count	7	1	1	3	10	6
	Percent	7.2%	0.6%	1.4%	2.3%	4.6%	1.9%
4	Count	3		1		5	2
	Percent	3.1%		1.4%		2.3%	0.6%
5	Count	2	1		1	2	
	Percent	2.1%	0.6%		0.7%	0.9%	
Total		97	153	72	132	217	311

luctant to add or subtract even a single car.) Third, tighter management of patrol car operations could only be achieved with the cooperation and knowledge of the patrol officers themselves, and might possibly introduce reporting and behavioral bias.

We also ruled out using a detailed simulation as a "laboratory" in which we could control all operating conditions since: (1) the loose management of the system and the lack of precise data on major parameters made it impossible to define rules which such a simulation could follow; and (2) a simulation model itself would have to be validated, raising the very same problems we were addressing for the analytic model. Thus, we were limited to monitoring existing operations as carefully as possible without control. We attempted to compensate for this lack of control by selecting three precincts with different environments and different levels of congestion. As economists often do, we thereby substituted a "cross-sectional" analysis for a more satisfactory but infeasible "longitudinal" analysis.

Since patrol car allocations are typically done bi-monthly, the queueing model is used to predict long-run average behavior. Thus, we aggregated our data over the ten days to

TABLE 2

Input Parameters for MCD Model

	Precinct 26	Precinct 50	Precinct 77
Average No. Calls/ Hr. (λ)	3.125	2.550	6.600
Average Service Time (Mins.) ($1/\mu$)	28.04	37.33	30.06
Proportion of High Priority Calls p_H	0.388	0.353	0.411
Dispatch Probabilities			
$p(1\|H)$	0.639	0.750	0.687
$p(2\|H)$	0.237	0.222	0.235
$p(3\|H)$	0.124	0.028	0.078
$p(1\|L)$	0.856	0.864	0.852
$p(2\|L)$	0.144	0.136	0.148
Average No. of Effective Cars on Duty ($\bar{n}$)	3.30	4.21	5.95

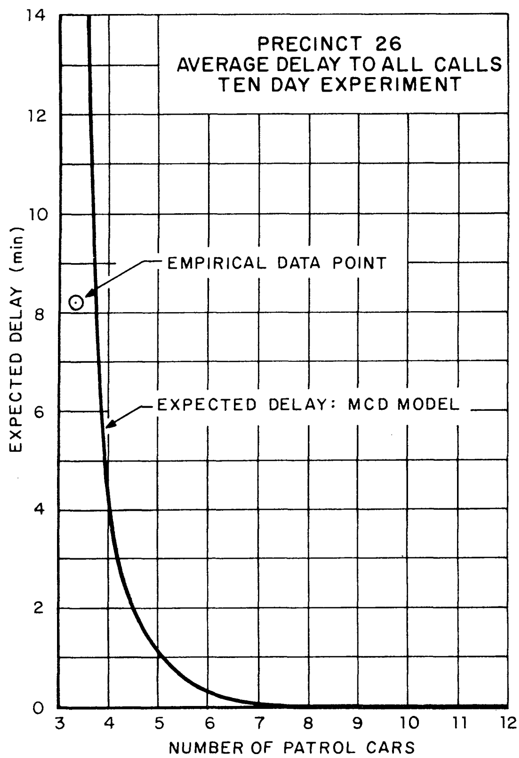

FIGURE 1

better approximate steady-state conditions. This yielded six data points—two for each precinct—consisting of statistical estimates of average delay of the first responding car to high priority calls and the average delay of the first responding car to all calls regardless of priority. Such statistics are central in NYCPD's planning of car allocations, and are key features of the PCAM system.

We now discuss how the parameters of the MCD model were estimated from the empirical data.

3.1. *Estimation of Call Rates and Proportions of Call Types*

Table 1 shows the number of calls for service and the numbers of responding cars between 8 AM and 4 PM by priority class and precinct. (Though the 144 "incident codes" used to characterize calls are nominally grouped into five priority classes, NYCPD planning and operations were primarily based on two priorities.) Although Table 1 shows that a small fraction of jobs actually received more cars, our test of the model assumed that high priority jobs never needed more than three cars and that low priority jobs never needed more than two cars. We had several reasons for this modeling assumption. From our field observations, we knew that when more cars are available, more are likely to *respond*, (through self-assignment) to a job, though they do not all actually *work*. (They are available for a subsequent call.) Moreover, the maximum number of cars per job in the model is bounded by the total number of cars. Since the ten-day average of the number of cars on duty (see Table 2) was less than four for one of the precincts, it was impossible for the model to assume that jobs "needed" more than three cars. Thus, the model presumed five call types (number of priorities × number of responding cars) and

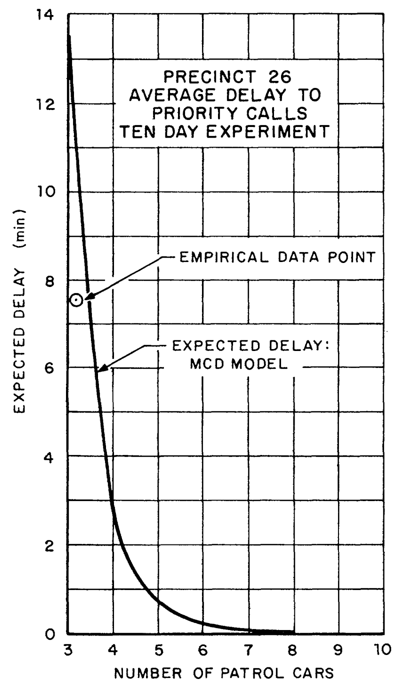

FIGURE 2

included in the high priority (H) 3-car category statistics, those H jobs that received three or more patrol cars, and in the low priority (L) 2-car category statistics, all L jobs that received two or more cars. Although, as can be seen in the table, the proportions of cars needed by job type did not vary significantly by precinct, we used the precinct specific proportions in our analysis.

3.2. *Estimating the Average Service Time per Patrol Car*

We computed for each precinct the average service time for all responses during the ten-day period. Service time, the interval between the dispatcher's assignment of the car and the time when the car radios in its final disposition, incorporates travel time and (possibly) time spent writing reports after "actual servicing" has ended. The average service times for each precinct are given in Table 2.

3.3. *Estimating the Number of Patrol Cars on Duty*

The number of patrol cars on duty, a parameter that one would suppose to be the simplest of the model's inputs to estimate, was in fact, quite difficult to obtain since:

1. Patrol cars come on and go off duty at different times. For example, although in the three precincts studied, the day tour runs from about 7:30 AM to 3:30 PM, some cars come on duty "around" 9 AM and end patrol "around" 2 PM for another type of assignment. Furthermore, actual "start" and "stop" times are not officially recorded, and vary from day to day.
2. As mentioned previously, a car is not available to patrol and respond to calls for service at all times during its tour of duty, but may go out-of-service for some intervals during its tour.
3. In New York City, police cars in the field are not all "ordinary" RMPs. There are a variety of "special" cars and supervisors' cars that may respond to certain types of 911

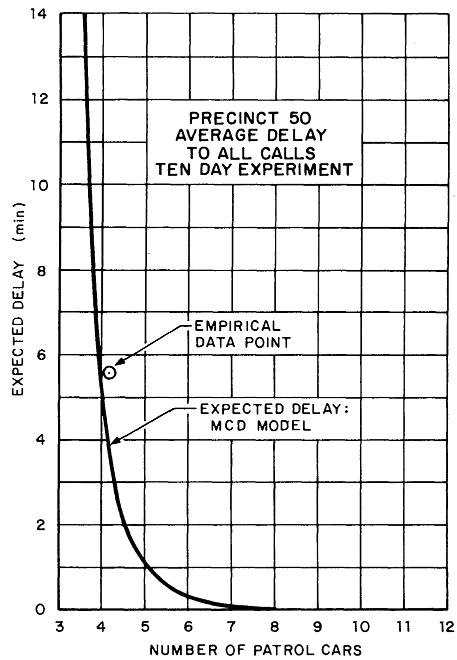

FIGURE 3

calls or even occasionally behave exactly as an RMP. No fixed rules govern these cars' behavior and there isn't a straightforward way to equate them to RMPs.

To deal with these problems, we defined the concept of the number of "effective" patrol cars on a tour—the average number of cars servicing or available to service calls. To compute this number, we first estimated the average number of RMPs out-of-service each day from our sampling of the SPRINT status reports. We next estimated by precinct, by day, an "RMP equivalency" for all "special" (including supervisors') cars from the ratio of the total number of jobs worked by these cars to the number of jobs worked by an average RMP that day. The out-of-service factor for the day (the average ratio of out-of-service time to total tour time) was then applied to obtain the number of "effective special cars" for the day. The *total number of effective cars* was thus the number of RMPs fielded minus the average number of RMPs out-of-service plus the number of effective special cars. This number varied between 2.65 and 4.01 in Precinct 26, 3.63 and 5.33 in Precinct 50, and 4.80 and 7.60 in Precinct 77. Averages are presented in Table 2.

3.4. *Results*

The estimates of the model's input parameters appear in Table 2. Notice that although the model requires an integer number of cars, the actual average numbers of effective cars are fractional. In fact, the number of effective cars was nonstationary and it is not clear how to estimate the resulting delays in such circumstances. We looked at two heuristic methods.

One method is to draw the expected delay vs. numbers of servers curves generated by the MCD model and read off the expected delay corresponding to the (fractional) average number of effective cars. These curves together with the observed delay-average number of effective cars point are shown in Figures 1 through 6. We believe that this estimation method should produce a lower bound of the actual average delay (all else being equal)

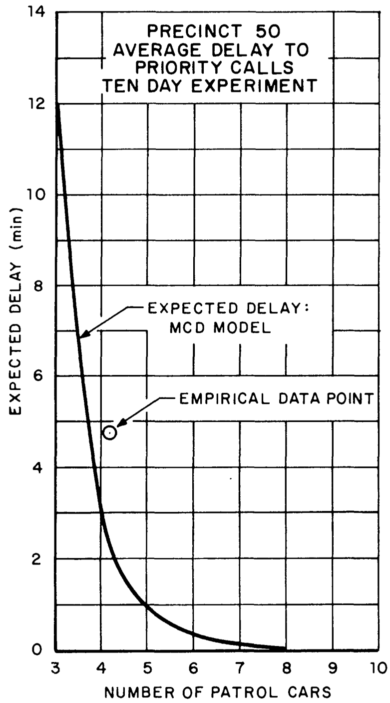

FIGURE 4

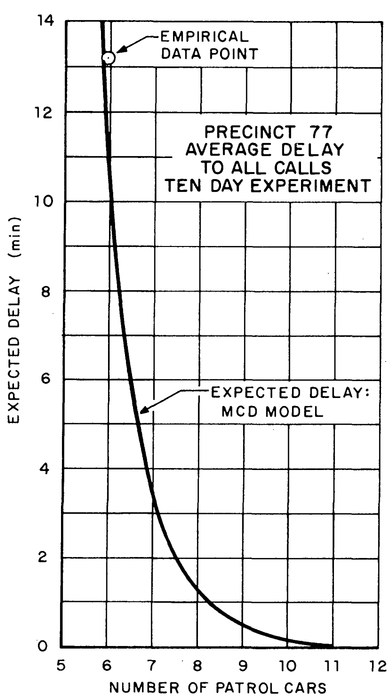

FIGURE 5

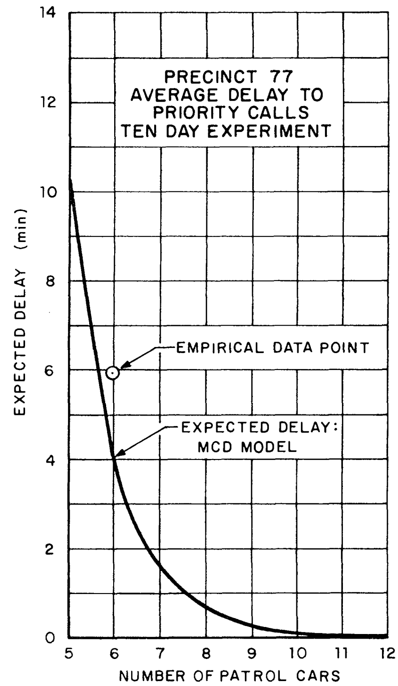

FIGURE 6

because of the analogy with results for nonstationary arrival processes (see Rolski 1981 and Green and Kolesar 1987) which indicate that stationary models using the average arrival rate produce a lower bound on the actual delays. These numbers are referred to as stationary estimates in Table 3. (The missing entry for Precinct 26 indicates a value that was off the plotted graph.)

Another estimation method is to interpolate linearly between the MCD delays generated by the integer numbers of cars that bracket the average. Were the actual system operated for long periods of time at each of the two integer points, this method would accurately estimate the actual delays. Unfortunately, the patrol system in each sample precinct was operated at many staffing levels which changed frequently. The delays at both the integer values and the interpolated estimates are given in Table 3.

It was striking that the numbers of effective patrol cars in the 26th and 77th precincts were so low relative to the precinct load that both the actual and theoretical queueing delays were much longer than the 1 minute standard desired by the NYCPD for high priority jobs. At such low staffing levels the model predicts very large changes in expected delays from the addition or deletion of even a single patrol car. Thus, the linear interpolation of the MCD delays produces high estimates.

The general indication is that the MCD estimates are "in the right ballpark," and are consistent with the usefulness of the model as a planning tool. However, it is clear that this experiment was not an adequate test for several reasons:

1. The limited amount of data resulted in high levels of variability in the performance measures (see the high standard errors in Table 3).

2. We were not able to test a full range of operating conditions in each precinct, and the cross-sectional approach didn't yield the degree of diversity we would have liked.

3. Due to marginally adequate actual levels of patrol car staffing, the traffic intensities in the sample precincts were in a range where delay is extremely sensitive to small changes in the number of effective cars.

TABLE 3
Comparisons of Observed and Estimated Delays (Minutes)

Precinct	Call Category	MCD Delay Estimates		Stationary** Estimate	Observed Delay (Standard Error)
26	All Calls	3 Cars	33.61		
		3.30 Cars*	24.89	NA	8.28
		4 Cars	4.54		(1.19)
	High Priority	3 Cars	13.63		
		3.30 Cars*	10.41	9.20	7.53
		4 Cars	2.89		(2.00)
50	All Calls	4 Cars	4.82		
		4.21 Cars*	4.05	3.50	5.60
		5 Cars	1.14		(0.51)
	High Priority	4 Cars	2.94		
		4.21 Cars*	2.49	2.20	4.78
		5 Cars	0.81		(0.76)
77	All Calls	5 Cars	50.33		
		5.95 Cars*	48.30	13.20	13.31
		6 Cars	9.73		(1.39)
	High Priority	5 Cars	10.22		
		5.95 Cars*	9.91	4.20	5.91
		6 Cars	3.99		(0.86)

* Obtained from linear interpolation.
** Obtained from graph.

4. The data on the number of effective cars were not highly reliable.

In addition, the nonstationarity in the number of effective cars raised the issue of how to predict delays from the model—a problem that we had not anticipated before the experiment. (Variability in the number of cars on patrol is not considered a problem for planners since it is assumed that precinct management keeps the number of effective cars virtually constant.)

In the concluding section of the paper, we discuss the insights we gained about conducting this kind of test in the future.

4. The Validity of the Structural Components of the MCD Model

We now consider in turn the validity of the key structural assumptions of the MCD model:

- *The call arrival process* is a mixture of stationary Poisson processes.
- *Service times* of individual cars are independent exponentially distributed random variables.
- *The queueing discipline* is that all required cars start service together, that calls are served on a first-come-first-served basis within priority class without preemption and that cars are assigned to waiting calls as soon as they become available.

4.1. *The Call Arrival Process*

The assumption of a Poisson input process has often been used to model the occurrence of emergency events (see, for example, Larson 1972, Chaiken and Larson 1972, Chelst 1981 and Kolesar 1981) but its validity has not been extensively examined in the literature. (For analyses supporting the Poisson nature of fire alarms see Finnerty 1977 and Carter and Rolph 1979.) We knew that call rates differed by season, and between weekends and weekdays, so we focused our analysis of this question on weekdays in a single season. Our data base was calls for police service during the 39 weekdays (excluding also the 4th of July) for June and July of 1980.

We also knew that the arrival process was highly nonstationary. Plots of the number of calls received in 15-minute intervals (as shown in Figure 7 for Precinct 77 with an SAS Cubic Spline smoothing 1981) clearly supports the existence of a strong temporal pattern. The effect of this nonstationarity on the model's utility, which is discussed in the next section, was a major concern to us.

Though the arrival process was clearly nonstationary, it might still be Poisson. The nonstationary Poisson process is a counting process $\{N(t), t \geqq 0\}$, obeying the standard Poisson process assumptions of "independent increments" and "orderliness" but with arrival rate $\lambda(t)$ an explicit function of time, that is

$$\lambda(t) \equiv \lim_{\Delta t \to 0} \frac{P\{N(t+\Delta t) - N(t) > 0\}}{\Delta t}.$$

For such a process the integral of $\lambda(t)$

$$\Lambda(t_1, t_2) = \int_{t_1}^{t_2} \lambda(t)dt$$

is the expected number of events in the interval $t_1 \leqq t \leqq t_2$ and the number of arrivals in such intervals has a Poisson distribution with parameter $\Lambda(t_1, t_2)$. Moreover, the time transformation $\tau(t) = \Lambda(0, t)$ converts the original process into a stationary Poisson process (see Heyman and Sobel 1982, or Cox and Lewis 1968).

These properties suggest two approaches to testing the validity of the nonstationary Poisson process model: (1) observe call counts in certain fixed intervals during the day and determine if they follow patterns consistent with the Poisson distribution; (2) estimate $\Lambda(t)$, transform time and observe if interarrival times on the rescaled axis are independent and exponentially distributed. Both procedures have major heuristic aspects. The first, since the theory states that counts in *any* interval are Poisson but we analyze counts for

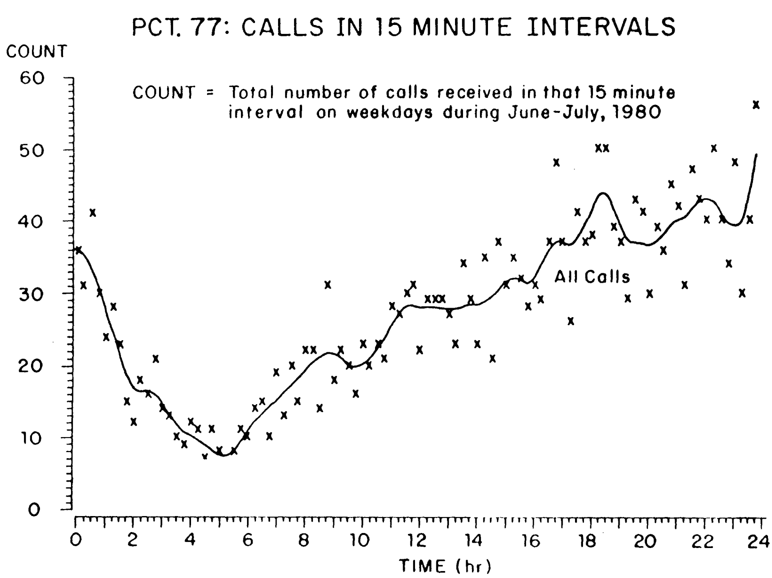

FIGURE 7

only *certain* intervals, and the second, since in our application no explicit function is specified, $\Lambda(t)$ itself must be estimated heuristically.

Analysis of Call Counts. We examined call counts in the 96 15-minute intervals, and in the 24 1-hour intervals that comprise the day. (In each case the first interval starts at midnight.) Rather than attempt to analyze 96 or 24 histograms, each containing 39 observations we focused on the relationship between the mean and variance of each of the data sets by doing linear regression. The summary of the regression results given in Table 4A shows a good correspondence to the mean/variance relation predicted by the Poisson model: the regression estimates of the parameters of the model $s^2 = a + b\bar{x}$ ($\bar{x}$ is the sample mean and s^2 the sample variance) consistently show that $\hat{a}$ is nearly zero and that $\hat{b}$ is nearly one; and that the model fit measured by r^2, the square of the correlation coefficient, is higher than 0.925. We refit a model $s^2 = c\bar{x}$ and again $\hat{c}$ was nearly one in all cases. Plots of residuals supported the hypothesis of linearity. We must remark, however, that the estimates of b and c were consistently slightly larger than one and formal tests of the hypotheses that $b = 1$ or $c = 1$ sometimes failed.

Analyses of Call Interarrival Times. $\hat{\Lambda}(0, t)$, the average cumulative count of calls received up to time t, was used to estimate $\Lambda(0, t)$ and to rescale time. Figure 8 is a histogram of the transformed interarrival times from Precinct 26. As shown, the correspondence between actual and predicted values is very close—certainly close enough to justify employing a nonstationary Poisson model—yet with 6256 observations the Chi-

TABLE 4

Test of the Poisson Model

A. *Call Counts—Summary of Regressions of Sample Variances vs. Sample Means*

	Precinct 26	Precinct 50	Precinct 77
Interval = 15 *min.*			
a*	0.049	−0.017	−0.0105
b*	1.048	1.116	1.106
c†	1.083	1.096	1.100
Std. Error†	0.026	0.032	0.024
r^2†	0.949	0.926	0.958
Interval = 1 *hour*			
a*	−0.031	0.109	0.095
b*	1.122	1.005	1.013
c†	1.160	1.039	1.131
Std. Error†	0.048	0.045	0.061
r^2†	0.958	0.958	0.936

B. *Transformed Interarrival Times—Summary Statistics*

	Precinct 26	Precinct 50	Precinct 77
Coef. of Variation $s/\bar{x}$	1.050	1.025	1.033
Sample Mode	0.026	0.026	0.026
Mean/Median Ratio $\bar{x}/\tilde{x}$	0.648	0.678	0.619
Auto Correlations (Lag of)			
1	0.008	0.004	0.021
2	0.015	0.004	−0.003
3	−0.012	−0.004	0.002
Sample Size n	6256	3438	9039

* Parameters of the model $s^2 = a + b\bar{x}$.
† Parameters of the model $s^2 = c\bar{x}$.

square test has such high power that the hypothesis of exponentiality would be rejected. Plots for the other two precincts showed similar good fits.

We also examined other statistics on the adjusted interarrival times for correspondence to the exponential model. From Table 4B we observe that for each precinct

• the coefficient of variation is nearly one, consistent with an exponential model,

• the sample mode is $0.026 = \frac{1}{39}$, essentially the smallest possible interarrival time save zero,

• the ratio of the sample mean to the sample median is close to the theoretical value 0.693 predicted by an exponential model,

• consistent with the assumption of independence of successive interarrivals, the lag auto-correlations are nearly zero—and do not differ significantly from this value at the usual 5% level.

We therefore conclude that a nonstationary Poisson model gives an adequate description of the overall statistical pattern of calls for service.

4.2. *Service Times*

Overall, service times (per car) were well approximated by the exponential distribution. Figure 9, a histogram of service times to all calls in Precinct 50, shows a very close correspondence to frequencies predicted by an exponential model. This plot is typical of our results for other precincts.

Other analyses further supported exponentially by showing that for each precinct-tour:

• the sample standard deviation divided by the sample mean was close to one (the range was 0.98 to 1.30 with an average of about 1.10),

• the sample mode in all cases but two was 0 or 1 (the data as recorded by the SPRINT program were truncated to the nearest minute).

• the sample median divided by the sample mean was usually close to the 0.69 predicted by the exponential probability law.

We next examined the assumption that service times are identically (and exponentially) distributed by priority class. Histograms of service times within priority classes (not

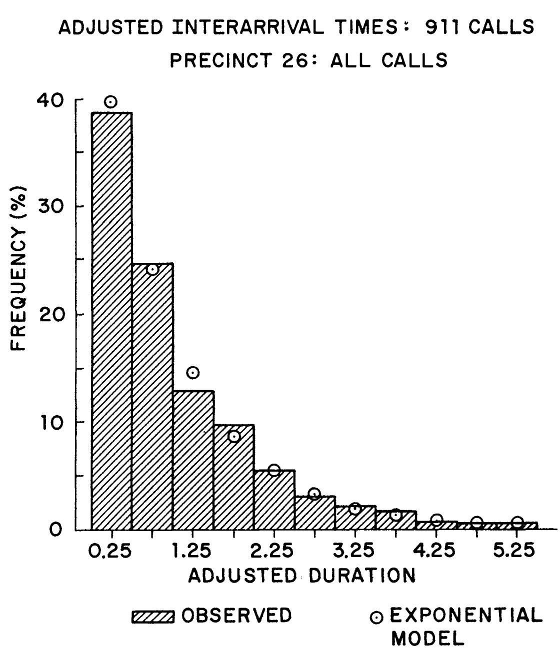

FIGURE 8

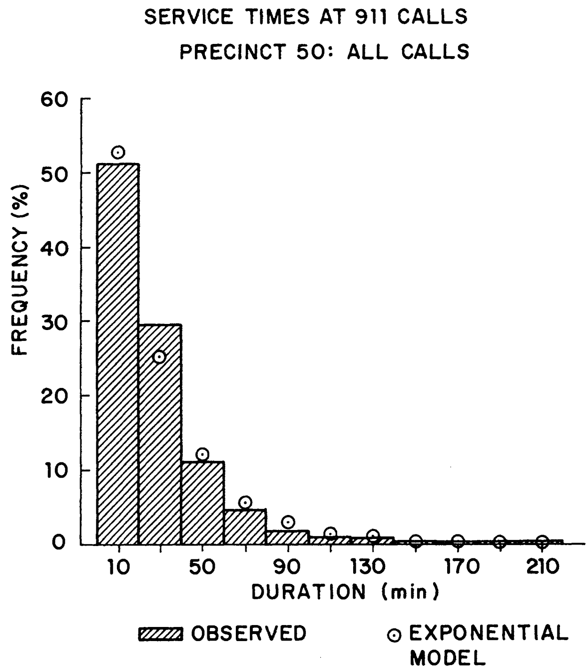

FIGURE 9

shown here) always displayed the characteristic exponential shape so it sufficed to examine the differences in mean service times by priority class. The results across precincts were quite similar and showed that service times for high priority calls which go to the head of the queue (both in practice and in the model) were only 6.7% lower on average than service times to low priority calls. Since the model uses the overall average, actual expected delays assuming all else the same will be somewhat lower than predicted by the model. To this extent the error introduced by assuming that average service times are equal across priority classes is conservative.

Next, we examined whether service times are independent and identically distributed regardless of the number of patrol cars working. Statistics in Table 5 show that although there are precinct to precinct differences that may require individual service time models, broad patterns apply across the precincts. We focus attention on the rightmost column containing the aggregate pattern for all three precincts and observe that:

1. Mean service times per car vary with the number of cars working at a job: the mean service time of cars working alone is about 10% higher than the overall mean service time while the mean service times of cars working in pairs are about 15% lower than the overall mean. (About 60% of responses are made alone and about 75% of jobs are one-car jobs while about 25% of responses are made in pairs and 16% of jobs are two-car jobs.) The mean service times at three-car jobs (about 10% of responses and 4% of jobs) are about 20% lower than the overall mean service time.

2. For two-car jobs, the joint distribution of service times follows a pattern similar to that predicted by an independent exponential model: The average service time of the shorter working car $\bar{U}$ is 59% of the average of both cars $\bar{X}$. The average service time of the longest working car $\bar{V}$ is 146% of the average of both cars working together. (Under the model of independent identically distributed exponential service times, the theoretical value of these ratios are 50% and 150% respectively for $\bar{U}$ and $\bar{V}$.) The correlation between X_1 and X_2 is 0.30 implying about a 10% improvement in prediction of say X_1 given X_2 or vice versa—not very strong.

TABLE 5

Statistical Summary of the Analysis of the Independence of Service Times of Patrol Cars Working at the Same Jobs

		Pct. 06	Pct. 50	Pct. 77	All Pcts
All Jobs, All Cars	$\bar{X}$	21.32	27.53	25.53	24.49
One-Car Jobs					
Mean Time	$\bar{X}$	23.11	29.52	29.17	27.22
Two-Car Jobs					
Mean Time (both cars)	$\bar{X}$	19.01	22.56	20.95	20.58
Mean Shorter Time	$\bar{U}$	10.73	13.53	12.71	12.18
Mean Longer Time	$\bar{V}$	27.29	31.59	29.20	28.98
Correlation X_1, X_2	ρ	0.26	0.34	0.31	0.30
Three-Car Jobs					
Mean Time (all cars)	$\bar{X}$	17.51	23.42	19.40	19.29
Mean Shortest Time	$\bar{U}$	7.69	11.61	9.15	8.97
Mean Median Time	$\bar{W}$	15.65	21.92	17.26	17.34
Mean Longest Time	$\bar{V}$	29.19	36.73	31.78	31.55
Average Correlation X_i, X_j	ρ	0.29	0.63	0.39	0.41

3. For three-car jobs the joint distribution of service times follows a pattern similar to that predicted by an independent exponential model: The ratio of the average service time of the shortest working car, $\bar{U}$, the median working car, $\bar{W}$, and longest working car $\bar{V}$ to the average of all three cars $\bar{X}$ are respectively 47%, 90%, and 163% as compared to theoretical values of 33%, 83%, and 183%.

Overall, we conclude that the i.i.d. exponential model is quite good for two-car jobs and reasonable for three-car jobs.

4.3. *The Queueing Discipline*

Start Times. Is it true that, as assumed by the model, patrol cars assigned to the same job start work together? To explore this we determined whether the times at which cars are assigned are close, relative to their service times and end times.

Consider two-car jobs and define variables Δs and Δf as the difference between the start times and the finish times of the two cars, respectively. If both cars start together and if their service times are independent and identically distributed exponential random variables with mean θ, we should have $\Delta s = 0$, and $E(\Delta f) = \theta$. The observed data aggregated over all precincts show that the sample averages are $\bar{\Delta} s = 6.40$ minutes, $\bar{\Delta} f = 16.86$ minutes, and $\bar{\theta} = 20.58$ minutes. Thus $\bar{\Delta} f$ is of the same order of magnitude as $\bar{\theta}$ as predicted by the model, and $\bar{\Delta} s$ is much smaller than $\bar{\theta}$ or $\bar{\Delta} f$, but is clearly not zero.

What are the reasons for and implications of $\bar{\Delta} s$ being six minutes? A frequency histogram of Δs shows that over 50% of the time Δs is less than one minute. Yet, the same histogram shows some values of Δs are as large as three hours and that about 10% of Δs values are more than 20 minutes. Scatterplots of Δs vs. Δf show that when Δs is large, Δf is also usually large. This implies that, on such jobs, very long service is given by cars that *do not* work together, but on the contrary, that work *consecutively*. Conversations with NYCPD managers confirm that these jobs are actually very, very long one-car jobs instead of long two-car jobs, as would be implied by the average service time data alone.

To be consistent with the model's assumptions, Δs should not be included in the service time of the first car. To the extent that it is part of the empirical service time data, the model will tend to overestimate delays. This effect will be most pronounced at high traffic intensity.

Priority Behavior. Discussions with dispatchers and detailed examination of SPRINT records confirmed that, in general, calls are served on a first-come first-served basis within priority class. Exceptions are that certain "super-H calls," which are crimes in progress with potential for personal injury, are usually handled before other less urgent jobs in the *H* priority class.

The model assumes no preemption, and in fact, preemption is very rare for many reasons. One is that leaving a job in progress would create a confusing situation. Another is that when patrol officers are working at a job, they frequently cannot listen to their radios and therefore cannot be reached for reassignment.

Other Sources of Delay. The MCD model assumes that cars are assigned to incoming or waiting jobs as soon as they are available. In reality, some delay may occur in the dispatch process itself, mostly since each dispatcher is responsible for keeping track of RMP status and assigning cars to incoming calls in some two to four adjacent precincts. To estimate the extent of such dispatching delays, we simultaneously examined the status reports and the SPRINT records for our two-week test period for each precinct, and identified "super-H" jobs that appeared to arrive when at least one RMP was available. These jobs should not have been delayed. (RMP availability was estimated by examining the status report that immediately preceded the job's arrival and checking the assignment of other jobs that arrived in the interim.) Of these jobs, we counted those that were delayed for at least two minutes as suffering a dispatch processing delay. (We used two minutes as our benchmark since SPRINT truncates clock time to minutes, and delays recorded as one minute could actually have been only a few seconds long.) Table 6, containing counts and average delays for these jobs, shows that almost half of all super-H jobs that arrived when a car is available had dispatch processing delays and the average delay for calls delayed more than our benchmark was over four minutes. Thus, dispatcher processing delay was a real factor in the operation of these precincts.

5. Sensitivity of the MCD Model

The next stage of our validation deals with the model's robustness and generalizability. We ask "Which of the model's major assumptions would, if significantly violated, render it inappropriate as a planning tool?" Here, we examine only those components of the patrol system that can be explicitly modeled, since our objective is to evaluate the MCD model relative to possible alternative models, including simulation. Thus, even though we know that both the "out-of-service" phenomenon and the existence of special cars can significantly affect system performance, the lack of any well-defined rules that govern car behavior makes these characteristics impossible to treat explicitly by any alternative model.

TABLE 6

Dispatch Processing Delays to Super-H Calls—Ten-Day Experiment

	Pct. 26	Pct. 50	Pct. 77
No. of Super-H Calls	44	37	118
No. Arrived When RMP Available	20	24	40
No. with Dispatch Delay $\geqq$ 2 Mins.	12	9	18
Avg. Dispatch Processing Delay for Calls with Dispatch Delay $\geqq$ 2 Mins.	4.58	3.11	4.67

5.1. *Exponentiality of Service Times*

Although our data show that the exponential distribution gives a close fit to the service times in New York, Larson (1972) points out that it is only a very crude fit in some other cities. An approximate (but accurate) version of the MCD model without priorities but with general service time distribution (see Federgruen and Green 1983) was used to compare expected delays for exponential and nonexponential service times. We set the coefficient of variation of the service time to values lower or higher than 1 and using the mean service time per car from the SPRINT data, we fitted either an Erlang (k) (for coefficient of variation less than 1) or a hyperexponential distribution (for coefficient of variation greater than 1). The coefficients of variation chosen represent extremes reported from actual patrol car service times. We used the SPRINT data to obtain the arrival rate, and the distribution of dispatched cars per job. Expected delays for two precinct tours with the lowest and highest proportion of multiple car dispatches at several levels of congestion are shown in Table 7.

Note from §4 that the actual coefficients of variation in our sample precincts were usually slightly higher than 1, so that the exponential model might slightly underestimate delays (excluding other factors). But observe that for both tours the exponential model correctly shows that with three servers delays are excessive, with four servers they are in the marginally tolerable range, and with five servers delays are acceptable. Since the range of variation in delays between the hyperexponential and Erlang 2 models is 40% or more and the delays with the exponential model fall approximately halfway between the extremes, individual percentage errors may be large. Yet, we believe that the key observation is that the exponential MCD model gives the right order of magnitude for expected delays and would thus lead to making proper patrol car allocations.

5.2. *Stationarity of the Call Arrival Process*

To test the effect of nonstationarity of the call process, we developed a simulation model and simulated each of the three sample precincts using the actual arrival times, numbers of responding cars, and service times as recorded on the SPRINT tapes for the 39 days of our sample. The system was simulated continuously around the clock with the number of cars on duty held constant throughout a tour, but changed from tour to tour. In a series of experiments to represent varying levels of congestion, we selected staffing levels for each tour that would keep the magnitude of expected delay (as predicted by the queueing model) relatively balanced throughout the day. The resulting delays for each tour were compared to those generated by the MCD model with input parameters equal to tour averages.

TABLE 7

Comparison of Expected Delays for Several Service Time Distributions in Precinct 26

# Servers	$cv = 0.71$ (Erlang 2)	$cv = 1.00$ (Exponential)	$cv = 1.25$ (Hyperexponential)
		Tour 1	
3	16.0	21.7	29.3
4	2.7	3.4	4.3
5	0.7	0.9	1.0
		Tour 2	
3	53.4	69.1	88.8
4	6.2	7.9	10.0
5	1.7	2.0	2.4

cv = coefficient of variation.

TABLE 8
Comparison of Expected Delays for Simulation of 39 Test Days and MCD Model

	Tour 1			Tour 2			Tour 3		
Precinct	No. of Servers	Sim	MCD	No. of Servers	Sim	MCD	No. of Servers	Sim	MCD
26									
Experiment 1	3	51.7 (19.3)	21.7	3	31.4 (16.9)	69.1	4	42.2 (15.0)	30.3
2	4	6.9 (3.0)	3.4	5	2.2 (1.4)	2.0	6	2.2 (1.2)	1.8
3	5	2.3 (1.3)	0.9	6	0.6 (0.5)	0.6	7	0.9 (0.7)	0.6
50									
Experiment 1	3	5.8 (5.2)	3.4	4	2.1 (1.2)	3.7	5	2.6 (1.8)	3.1
2	4	0.9 (1.0)	0.6	5	0.6 (0.6)	0.2	6	0.6 (0.3)	0.9
3	5	0.2 (0.3)	0.1	6	0.1 (0.2)	0.2	7	0.1 (0.1)	0.3
77									
Experiment 1	4	47.5 (14.3)	16.2	6	8.0 (2.8)	13.0	7	15.5 (7.1)	15.5
2	5	10.8 (4.8)	3.9	7	3.0 (1.3)	4.2	8	6.7 (2.7)	5.2
3	6	3.3 (1.7)	1.2	8	1.3 (0.6)	1.6	9	2.9 (1.2)	2.1
4	7	1.1 (0.9)	0.4	9	0.6 (0.3)	0.6	10	1.5 (0.7)	0.9

Note: Figures in brackets are half-widths of 90% Confidence intervals; thus, for example, 51.7 (19.3) means that a 90% Confidence interval for the expected delay in the 26th Precinct on Tour 1 with 3 servers is 32.4 to 71.0.

Table 8 shows the results for several levels of system congestion for each of the three precincts. (The notation 6.9 (3.0) means that the 90% confidence interval of the simulated value is 3.9–9.9.) (Confidence levels were generated using batch means, see Law and Kelton 1982.) The confidence levels for the simulated delays are quite large due to the variability of delays. In spite of this, the results for Tours 2 and 3 indicate a correspondence between the expected delays for the nonstationary case and those generated by the (stationary) MCD model. Of course, the correspondence is better for higher staffing levels. The worst case is Tour 1 where the simulated expected delay is significantly larger than that predicted by the queueing model. This occurs because in each precinct, the call arrival rate has its maximum near the end of Tour 3 (see Figure 1) and if Tour 3 is staffed to handle the average load, large queues occur at the beginning of Tour 1. Worse, at the start of Tour 1, the call arrival rate is still near the daily maximum while only four hours later, it's at the daily minimum. A staffing level appropriate to the Tour 1 average call rate is therefore inadequate for the first several hours of the tour when substantial queues build up.[3]

In summary, the existence of nonstationarity may lead to model predictions which significantly underestimate true delays and hence to inadequate allocations of cars. This appears to be particularly true for a tour that begins during a peak arrival time.

[3] This is in part a reflection on the inappropriateness of the traditional (12 × 8), (8 × 4), (4 × 12) tour structure in the face of significant nonstationarity with a peak near the end of one of the tours.

6. Conclusions

In conducting this study, we not only learned about the validity of the MCD model for allocating patrol cars, but we also gained some important insights about how to evaluate the usefulness of analytic queueing models of man-machine systems. We first present our conclusions for the MCD model.

Overall, we conclude from this study that the MCD model is a valid tool for its intended purpose: the allocation of patrol cars across precincts and tours. The provisos to this conclusion are: First, since our testing was done in New York City, our results may be particular to that locale. Second, the goal of patrol car allocation models is to indicate staffing levels that will result in a speedy response to emergency calls, and the model works best as a predictor when the actual system is so staffed. Third, nonstationarity of the call arrival process may distort the predictions and therefore the MCD model should be used in a manner that better reflects the changing load on the system (see below). Last, but not least of our provisos is that the actual patrol system be reasonably tightly managed—that is, patrol cars and dispatchers will follow to a reasonable approximation the car assignment and work protocols being modeled. It is our strong feeling when these conditions are met, the MCD model will indicate patrol car allocations that achieve desired response performance and that appropriately balance forces among different precincts and tours.

From this validation effort we recommend that future use of the MCD model explicitly take into account two major factors. The first is the nonstationarity of the call arrivals. We have seen here and in our ongoing research on the issue of nonstationarity that delay estimates generated by a stationary queueing model using average arrival rates may differ sharply from the actual delays in a nonstationary system. However, our findings indicate that it is possible to correct for this to a great extent by using a series of stationary models on a partition of the time interval (tour) under consideration and estimating the overall delay from the average delay over such a partition. In fact, the current version of PCAM uses this sort of approach (Walker 1983).

The second factor is the dispatching process. Significant dispatcher processing delays ought not to occur! Changes in the way data are handled and displayed on the dispatcher's terminal screen and a change to increased digital reporting that reduces voice communications would drastically reduce or eliminate them. Moreover, at least from the dispatching delay point of view, having interprecinct dispatches of patrol cars that are handled by the same dispatcher would be advantageous. Such improvements are clearly cost effective relative to adding patrol cars. It is not necessary (or possible) we believe to build a single model that incorporates dispatcher processing and patrol car operations.

What about the validation process itself? We have several conclusions. First, we believe that it is essential to perform a multiple facet type of validation for models of systems of this sort. Though the "black box" approach of stage 3 is the usual kind of validation test performed for models of more mechanistic systems (see e.g. Lazowska 1984), it is not adequate in a system which has a significant human component. There are several reasons: (1) it's harder to perform a truly controlled experiment; (2) it's more difficult to test the model under all important conditions, e.g., in the patrol system case it would have been extremely difficult to gather the real-time data needed for the midnight to 8 AM tour; and (3) there are parts of the system which are poorly understood and/or cannot be modelled at all (even by a simulation).

So, even if the results of our stage 3 testing yielded virtually perfect model predictions, we would not be confident that they would be equally good in other precincts, for other tours, or for different seasons. We would also be at a loss to identify under what circumstances the model's predictions should be modified. Similarly, the testing of stage 4 or 5 alone would be inadequate since the system is more than the sum of its components.

Of course, if at all possible, even a partially controlled field experiment is important. Given the somewhat inconclusive outcome of our stage 3 test, we now realize that we should have tried harder to get the NYCPD to consent to a longer experiment in which the numbers of cars in the precincts was controlled. Another limitation to our findings was due to the high workload in 2 of the 3 test precincts which resulted in extreme sensitivity of the MCD estimates. Some future tests should be conducted under more moderate workload conditions.

We also observe that field experience is crucial for conducting any validation test in a complex environment. Without the first-hand knowledge we had gained about the *entire* police patrol system, we would have been at a loss as to how to gather or interpret the data. Furthermore, we would have been totally unaware of the existence of dispatcher processing delays.

Finally, the definition of model validity has to be modified when modelling systems in which human behavior plays an important role. The correspondence between the model and the system is likely to be significantly less tight than for mechanistic systems. This is tolerable if the model still captures the essential dynamics of the system and is used as a first step in the decision process. It is also important for the model users to understand that the more loosely managed and adaptive the system, the less accurate the model will be. Even in such situations, the model may still help identify poor management and perverse behavior from the performance benchmarks it establishes.

[4] This work was supported by a grant from the National Institute of Justice and by the Faculty Research Fund of the Graduate School of Business, Columbia University.

References

BRIGHAM, G., "On a Congestion Problem in an Aircraft Factory," *Oper. Res.*, 3 (1955), 412–428.

CARTER, G. AND J. ROLPH, "Analyzing the Demand for Fire Department Services," Chapter 8 in *Fire Department Deployment Analysis. I*, W. Walker et al. (Eds.), North Holland Publishing Company, New York, 1979.

CHAIKEN, J. AND P. DORMONT, "A Patrol Car Allocation Model: Background, Capabilities and Algorithms," *Management Sci.*, 24 (1978), 1280–1300.

—— AND R. LARSON, "Methods of Allocating Urban Emergency Units," *Management Sci.*, 19 (1972), 110–130.

CHANDY, K. M. AND M. REISER, *Computer Performance: Proc. Internat. Sympos. Computer Performance Modeling, Measurement and Evaluation*, North-Holland Publishing Co., New York, 1977.

CHELST, K., "Development of One- vs. Two-Officer Patrol Units: A Comparison of Travel Times," *Management Sci.*, 27 (1981), 213–230.

COBHAM, A., "Priority Assignment in Waiting Line Problems," *Oper. Res.*, 2 (1954), 70–76; correction *ibid.* 3, 547.

COX, D. AND P. LEWIS, *Statistical Analyses of Series of Events*, Chapman and Hall, London, 1968.

CRABILL, T., W. WALKER AND P. KOLESAR, "Validation of a Police Patrol Simulation Model," *Proc. 8th Annual Simulation Sympos.*, Tampa, FL, 1975, 229–247.

EDIE, L., "Traffic Delays at Toll Booths," *Oper. Res.*, 2 (1954), 107–138.

FEDERGRUEN, A. AND L. GREEN, "An $M/G/c$ Queue in Which the Number of Servers Required is Random," *J. Appl. Probab.*, 21 (1984), 583–601.

FINNERTY, J., "How Often Will the Firemen Get Their Sleep?" *Management Sci.*, 23 (1977), 1169–1173.

FISHMAN, G., "On Validation of Simulation Models," *Proc. AFIPS* 1973 *National Computer Conf.*, Washington, D.C., 1973, 51–57.

GASS, S., "Decision-Aiding Models: Validation, Assessment, and Related Issues for Policy Analysis," *Oper. Res.*, 31 (1983), 603–631.

GREEN, L., "A Multiple-Dispatch Queueing Model of Police Patrol Operations," *Management Sci.*, 30 (1984), 653–664.

—— AND P. KOLESAR, "The Feasibility of One-Officer Patrol in New York City," *Management Sci.*, 30 (1984a), 964–981.

—— AND ——, "A Comparison of the Multiple-Dispatch and $M/M/c$ Priority Queueing Models," *Management Sci.*, 30 (1984b), 665–670.

—— AND ——, "Effects of Nonstationarity in Markovian Queueing Systems," Columbia Univ. Graduate School of Business Working Paper, 1987.

HEYMAN, D. AND M. SOBEL, *Stochastic Models in Operations Research*. Vol. 1, McGraw-Hill, New York, 1982.

KOLESAR, P., "Ten Years of Research on the Logistics of Urban Emergency Services," *Operational Research* '81, North-Holland Publishing Company, 1981, 557–568.

——— AND W. WALKER, "A Simulation Model of Police Patrol Operations: Program Description," Report R-1625/2-HUD/NYC, The Rand Corporation, Santa Monica, CA, 1975.

LARSON, R., *Police Patrol Analysis*, MIT Press, Cambridge, MA, 1972.

——— AND M. MCKNEW, "Police Patrol Initiated Activities within a Systems Queueing Model," *Management Sci*., 28 (1982), 759–774.

LAW, A. AND W. D. KELTON, *Simulation Modeling and Analysis*, McGraw-Hill, New York, 1982.

LAZOWSKA, E. D., *Quantitative System Performance*, Prentice-Hall, Englewood Cliffs, NJ, 1984.

NAYLOR, T. AND J. M. FINGER, "Verification of Computer Simulation Models," *Management Sci*., 14 (1968), 13-92–13-101.

POLLOCK, S., J. BIRGE AND W. HOPP, "A Model for Wide Area Patrol Policy Evaluation," mimeo report, Department of Industrial and Operations Engineering, University of Michigan, Ann Arbor, 1982.

ROLSKI, T., "Queues with Nonstationary Input Stream: Ross's Conjecture," *Adv. in Appl. Probab*., 13 (1981), 603–618.

SARGENT, R., "Validation of Simulation Models," *Proc*. 1979 *Winter Simulation Conf*., San Diego, CA, 1979.

The SAS Institute, *SAS* 1979, *Users Guide*, Cary, NC, 1979.

———, *SAS/GRAPH Users Guide*, Cary, NC, 1981.

VAN HORN, R., "Validation of Simulation Results," *Management Sci*., 17 (1971), 247–258.

WALKER, W., "Handling Time-Varying Call Rates and Overload Conditions in PCAM," private communication, 1983.