SQUARE ROOT LAWS FOR FIRE ENGINE RESPONSE DISTANCES

Peter Kolesar

Edward H. Blum

October 1973
Any views expressed in this paper are those of the author. They should not be interpreted as reflecting the views of The New York City-Rand Institute or the official opinion or policy of the City of New York. Papers are reproduced by The Rand Corporation as a courtesy to members of its staff.
SQUARE ROOT LAWS FOR FIRE ENGINE RESPONSE DISTANCES*†

PETER KOLESAR AND EDWARD H. BLUM

The New York City-RAND Institute

An inverse square-root function is developed for the relation between average response distance and the number of locations at which response units are stationed in a region. Analysis of theoretical models, simulation data, and empirical measurements are used to confirm the square-root model.

The square-root response distance model is combined with response distance—response time relations to resolve decision problems important to the management of urban fire departments. The results can be used to find optimal resource allocations given resource constraints and response time standards, or to describe the response time consequences of proposed allocation plans.

I. Introduction

The time elapsed between a call for emergency service and the arrival of the responding unit or units is an important measure of the effectiveness of emergency service systems. Whether the service in question involves a fire department, police, ambulance, or emergency repair crews, when the call is important, speed is essential and response time often serves as a surrogate for more basic performance measures as lives saved, criminals apprehended, or property damage avoided [11], [12].

This paper develops a simple relationship between spatial average response distance and its key determinants that we call the “square root law” for response distances. It states that the average response distance in a region is inversely proportional to the square root of the number of locations from which emergency service units are available to respond. The principal result, when combined with a response time-response distance function, enables one to predict expected (average) response times in a region, given the following readily measurable parameters for the region: the number of active locations for emergency units, the geographical area being covered, the rate at which calls for service are generated, the expected time required to service each call, and finally, a constant of proportionality depending upon the detailed geometry—the street patterns, relative locations of calls and emergency units, etc. As with all simple models, the predictions so produced are approximations, which give results useful in narrowing the range of policies to be considered. Where necessary, the model’s results

* Received May 1972; revised November 1972.
† This work stems from part of the research of The New York City-RAND Institute’s Fire Project—in particular, part of the Project’s research to develop mathematical models for improved deployment of fire-fighting resources.

With such closely knit team projects, it is extremely difficult to separate responsibility or to give complete credit to all who contribute. Although it bears our names, this work is very much a joint product of many members of the Fire Project. We would like to note especially Grace Carter, who contributed the simulation models and ran the necessary experiments. Discussions with Warren Walker and Edward Ignall of The New York City-RAND Institute, and with Mike Florian and Pierre Robillard of the Université de Montréal were helpful in many ways.

The work of Peter Kolesar was partially supported by the National Research Council of Canada while he was visiting professor at the Université de Montréal during 1971.
can later be refined or checked in detail by more elaborate or tedious methods such as simulation. The following scenarios illustrate some of the model’s uses.\footnote{The examples are not hypothetical; our results have been applied in the manner indicated by the New York City Fire Department since 1970, and there is evidence of their applicability to other services as well [17], [22].}

\textit{Application 1}

Budgetary limitations set by higher levels in the city government imposed a constraint on the number of fire units available. Department management sought to determine the effect of alternative budget plans on the levels of service provided.

\textit{Application 2}

The lead time for acquisition of new facilities and other capital resources is quite significant. Hence we generated, using predictive models, forecasts of demand (total alarm rates, percentage of false alarms, percentage of alarms for structural fires by region) for several years into the future. Management is using the square-root model to determine the number of fire units needed at various times in the future to meet the anticipated demands while maintaining desired levels of response time. The model indicates clearly which areas may need more or fewer units; more detailed analyses must be carried out to make final decisions and to choose prospective new sites.

\textit{Application 3}

A change from uniform deployment of units and men around the clock to variable deployment with additional units and staffing during hours of peak demand has been suggested. The square root model has been used together with predictions of alarm rates by time of day to determine the time-variable staffing levels required to achieve desired response times for all regions of the City.

In focusing here on response distance and response time, and in dealing with simple analytical models, we necessarily over-simplify many elements of the real situation. For further discussion of the real, complex environment of urban fire deployment, see [2], [3], and [4]. The discussion which follows focuses on fire engines, but applicability of our results to other services should be clear. For some details, see [17].

\textbf{II. Square Root Law for Response Distance—Some Theoretical Models}

Inherently, the square root law follows from dimensional analysis: distance is the square root of area, and the area served from each fire house is inversely proportional to the density of fire stations. Thus it seems plausible that the expected fire company response distance should be inversely proportional to the square root of the density (number per square mile) of firehouses.

To be specific, consider a given region, and let \( N \) be the number of firehouses having units available to respond at a given time. Let \( D \) be the expected distance between points where fires could occur and the occupied firehouse nearest those points (i.e., the expected distance to the closest available unit). Then, the square root law predicts that \( D = K/(N/A)^{1/2} \) where \( K \) is a constant of proportionality and \( A \) is the area of the region being considered.

The analysis which follows shows the robustness of the square-root relationship and examines some interesting details. Formally, of course, the results discussed are stand-
ard fare in probability theory, and have been used by astronomers, geographers, chemists, and others. (See, for example, [1], [8], [9], [13], and [14].)

Simple Static Models

Consider first a probabilistic but static situation. Suppose that emergency units are located (one to a location) in a large region completely at random, i.e., according to the two-dimensional Poisson process with density \( \rho \), and that the incidence of fires is spatially homogeneous throughout the region. We focus attention on the random variable \( R_k \), the Euclidean distance from an arbitrary point to the \( k \)th nearest emergency unit. Consider first \( R_1 \). \( R_1 \) is greater than \( r \) if, and only if, there are no emergency units within a circle of radius \( r \) centered at the point in question. Thus,

\[
P[R_1 > r] = \exp \{-\rho \pi r^2 \}, \quad r \geq 0.
\]

It is easy to show that the expected response distance, \( ER_1 = \frac{1}{2} \rho^{-1/2} \). As the parameter \( \rho \) is the density of emergency units, in the limit for an extremely large area \( A \), \( \rho = N/A \) with probability one. Thus, if \( A \) is large, \( ER_1 = (2(N/A)^{1/2})^{-1} \).

More generally,

\[
P[R_k > r] = \sum_{j=0}^{k-1} \frac{\exp \{-\rho \pi r^2 \} (\rho \pi r^2)^j}{j!}, \quad r \geq 0, \quad k = 0, 1, 2, \cdots,
\]

and the expectation

\[
ER_k = \frac{1}{2 \rho^{1/2}} \sum_{j=0}^{k-1} t_j, \quad k = 1, 2, \cdots
\]

where \( t_0 = 1 \) and \( t_j = ((2j - 1)/2j)! t_{j-1}, \quad j = 1, 2, \cdots \).

To obtain the variance of \( R_k \), observe that \( 2\pi \rho r^2 \) has a chi-square distribution with \( 2k \) degrees of freedom, hence \( ER_k = k/\pi \rho \). Thus, \( \text{Var} R_1 = (4 - \pi)/4 \pi \rho \), and \( \text{Var} R_2 = (32 - 9 \pi)/16 \pi \rho \). As dimensional analysis suggests, the variances are inversely proportional to the density of units.

Summarizing, the average response distance varies inversely with the square root of the number of units. Tail probabilities decrease much faster. The probability that a given response distance, say \( r \), is exceeded, is \( \exp \{-\pi \rho r^2 \} \) which decreases exponentially with the number of companies.

As the dimensional argument suggests, the validity of these observations does not depend upon the particular model we have analyzed. Similar analyses of other models give the same structural relation between expected response distance and the number of available units. Indeed, straightforward analysis shows that, as \( k \) increases, \( ER_k \) converges from below to \( (k/\alpha \rho)^{1/2} \), where \( \alpha = \pi \) for the Euclidean metric and \( \alpha = 2 \) for the right-angle metric.

Such derivations assume that the region is infinite, that units and alarms are distributed homogeneously, and that distances are measured according to a "continuous" metric. In the real world, of course, none of these assumptions is strictly true, and homogeneity is rarely seen at all. We must see, therefore, to what extent the square-root relation still holds when any or all of these assumptions are violated.

The case when \( N \) is small and the region is finite has been studied by Leamer [19], who considered some problems in homogeneous finite regions (squares, triangles, and circles), with the number of facilities varying from 1 to 16, evenly spread throughout the region. Derived for Euclidean distances, his results indicate that even with a small number of units the response regions are quite similar to the optimal hexagonal shape.
associated with infinite regions and that the relationship between \( ER_1 \) and \( N \) is, indeed, very close to an inverse square root. We will examine the effects of inhomogeneity and real-world metrics in §III where we consider empirical results.

**Simple Dynamic Models**

The models considered in the preceding analysis are probabilistic but static, since they implicitly assume that all of the units assigned to a region are in quarters and available to be dispatched to an alarm. In reality, this is often not the case: at a given time some (perhaps many) of the units stationed in a region will be unavailable, busy responding to alarms or working at fires. In this situation, we must be concerned with the relation between the regional mean response time, averaged over some long time interval, and the number of companies assigned to the region. The following heuristic argument is meant to suggest why, in situations in which there is a low probability that all companies will be busy, the square-root law approximates well the relation between long-run average response time and the number of units assigned to the region.

Our model is the following: suppose that \( n \) units are stationed (one to a location) in a region, and we are interested in the relation between \( n \) and the long-run average response distance. We describe the state of the system at epoch (time) \( t \) by a \( n \)-dimensional vector \( X(t) \) whose \( i \)th component describes the state of unit \( i \) (in quarters, responding to an alarm, working at a fire, etc.). An important characteristic of the state of the system is \( N[X(t)] \), the number of units available to respond to an alarm at epoch \( t \). The analyses of the preceding sections suggest that the average response distance at epoch \( t \) will vary inversely as square root of \( N[X(t)] \), which clearly will change through time. Suppose also that alarms are geographically homogeneous and occur in time according to a Poisson process with alarm rate \( \lambda \). Suppose that as long as there are any units available one unit is dispatched to each alarm and that the total service time for each alarm is an exponentially distributed random variable with mean \( 1/\mu \). Service times are assumed to be mutually independent, and independent of the state of the system. Alarms which occur with all \( n \) units busy are handled by special procedures such as calling in units from outside the region, and will be regarded in this simple treatment as "lost calls." The units in the region do not respond to alarms outside the region.

The system described is the birth-and-death process usually called the \( M/M/n \) queuing system with losses. For this system, the stationary probability \( P_0 \), that exactly \( j \) units are busy is well known [9].

Now suppose, building on the static models, that as long as all the \( n \) units are not busy, the spatial average response distance of the closest available unit follows the square root law. That is, if \( j \) units are busy, the expected response distance is \( K/(n-j)^{1/2} \) and when all units are busy, the expected response distance is \( K_0 \). The long-run expected distance is then (writing \( \gamma = \lambda/\mu \))

\[
ER(n) = \left( \sum_{j=0}^{n-1} \frac{\gamma^j}{j!} \right)^{-1} \left( \frac{K_0 \gamma^n}{n!} + K \sum_{j=0}^{n-1} \frac{\gamma^j}{j!(n-j)^{1/2}} \right).
\]

Equation (1) does not facilitate the applications we desire. However, calculations reveal that if the probability that all units are busy is small, \( ER(n) \) can be approximated quite well by

\[
ER(n) \cong K/(n - \gamma)^{1/2}.
\]
This approximation is motivated by the observation that, when the probability that units are busy is "small", the expected number of units busy can be approximated by $\gamma$, and by the hope that $ER$ can be approximated by $K/E^{1/2}$ (where $E$ is the expected number of companies available). It turns out that (2) is an underestimate, a fact we will discuss later. To evaluate the accuracy of this approximation, (1) and (2) were calculated for values of $\gamma$ and $n$ in the range experienced in our work with the New York Fire Department. The errors involved in using (2) were typically less than 1% and never more than 5%. We decided that these errors were quite tolerable in light of the other approximations incorporated in this simple model and the accuracy of the parameter estimates.

In actual applications, we have made checks of the square-root approximation using more complicated queuing models, which explicitly account for real practices—i.e., that several units may work at a fire for different lengths of time, etc. [5].

III. Empirical Results

The models discussed in the previous section were derived under restrictive conditions. Actual circumstances are much more complex: calls for service are neither geographically homogeneous nor stationary in time; service times are state-dependent; response routes must follow street patterns, etc. An important question, therefore, is the extent to which inverse square root relationships hold for realistically complex situations. We examine this question now, using both empirical and simulation data.

Two relationships need to be examined; to be precise in their specification we use the notation of the last section. Let $N(t)$ denote the random variable—the number of units available in the region under consideration at time (epoch) $t$. Let $R(t)$ denote the random variable, the response distance at epoch $t$, and let $EV$ and $ER$ denote the long run averages of $N(t)$ and $R(t)$. We are concerned with:

(1) The relation between expected response distance $ER(t)$ and $N(t)$, the number of units available at the instants of the calls for service, that is, the function $ER[N(t)]$.

(2) The relation between long-run expected response distance $ER$ and $n$, the number of emergency units assigned to the region. Since $n$ is a major policy variable largely under managerial control this relation is of more general management interest than (1).

Collecting empirical data to examine these relationships is difficult. Experiments on operating systems to measure the appropriate variables for Case 1 involve formidable information gathering tasks. Studying Case 2 requires varying $n$. But operating fire departments would, since the experiments could be very costly or risky, sensibly not permit them. We have thus had to work with less direct, semi-empirical measurements based on simulations and historical data.

One of the tools our research group developed to analyze the consequences of various deployment strategies for the New York Fire Department is a large scale detailed simulation model of fire-fighting operations. This model generates alarms for incidents of various types and severity according to projections of historical patterns, under which alarm rates vary markedly throughout the region. The simulation uses complicated decision rules for the dispatching and dynamic relocation of fire engines and measures response distances that reflect actual street patterns [4]. This simulation model has been extensively tested to verify its correspondence with the real environment; it thus provides a means to test realistically the wider validity of the inverse square root law.

We have used this simulation to generate realistic data relating response distances to numbers of available units. Eight separate experiments were run using parameters appropriate to the borough of the Bronx in New York City; in each a different number
of ladder units was assigned to the Bronx. The number of ladders assigned varied from 12 to 31, with the home locations being selected on the basis of other analysis. (There were actually 24 ladder units in the Bronx at the time these experiments were initiated.) During each of the experiments, the simulation recorded response distance together with the number of companies available at the instants when alarms occurred. Data were collected separately for engines and ladders, and further broken down by region. These simulation data provide a means of examining in detail both the relationships noted above.

**Simulation Results**

We consider first the relation between $ER(t)$ and $N(t)$. Figure 1 displays simulated average response distances (in generalized distance units) for closest engines versus the number of companies actually available when the alarms occurred.

To test the square-root law hypothesis with these semi-empirical data, we fit to them the following relations, using least squares regression:

$$
\hat{R}(t) = \alpha N(t)^\beta,
$$

$$
\hat{R}(t) = \alpha / (N(t))^{1/2}.
$$

The inverse square-root law (4) is a special case of the exponential relation (3). The fit of (3) was made using a nonlinear regression program without transforming the data.

If square-root relations indeed hold, relation (4) should fit well and estimates of $\beta$ should be "close" to $-\frac{1}{2}$. Measuring how close $\beta$ is to $-0.5$ is not straightforward, since the simulation data do not satisfy the conditions requisite for classical statistical analysis. For example, the observations are not independent, and the square-root relation itself implies unequal variances. Examination of the sum of squared errors

![Figure 1: Average response distance of first engine vs. number of companies available at the instant of the alarm.](image-url)
indicates, of course, that (3) fits better than (4), but the difference between (3) and (4) is small. And an “eyeball” check of the graphs reveals little difference between (3) and (4).

Temporarily leaving aside our reservations about standard statistical tests, we calculate approximate 95 per cent confidence contours for $\alpha$ and $\beta$. In each case, these include $\beta = -\frac{1}{2}$, indicating that an inverse square root law gives a good fit. These data are typical of our other results. They indicate that the inverse square-root relationship between the average (and standard deviation of) response time and $N(t)$ holds well, even in realistic, complex situations.

We now turn to relation (2). The data just examined indicate that the square-root model describes the relationship between average response distance and the number of units available when an alarm occurs. But this does not assure that a square-root law describes the relationship between long-run average response distance and the average number of companies available to respond to an alarm. On the contrary, if the square-root law holds for the former, it cannot hold exactly for the latter, since the inverse square-root function is convex, and for a convex function $f(\cdot)$ of a random variable $X$, $Ef(X) > f(EX)$ (Jensen’s inequality). We find, however, that it gives a good approximation.

Figure 2 displays simulated average response distances for closest ladders versus predicted average numbers of ladder units available. (Predictions are made using the simple approximation that the average number of companies busy should be equal to the product of the predicted alarm rate and the predicted average total service time per alarm.) Simulations were run at different alarm rates, with from twelve to thirty-one units assigned to the region. The results again indicate that an inverse square-root function for both the averages and standard deviations fits well. Regressions of response distance against actual average number of companies available also support the relation and give an even better fit. We have emphasized the results using predicted availability, since, in practice, one would have predictions of availability rather than actual values.

Space limitations do not permit us to display here data generated for other areas,
for second closest companies, for variances of response distances, etc., all of which also confirm the validity of the square-root model [15]. In addition to the aforementioned simulation experiments run with the detailed model of fire-fighting operations in the Bronx, we also constructed and experimented with a "mini-simulation". In this more idealized model, alarms were generated according to a homogeneous Poisson process in a unit square. Fire companies could be located as desired, and distances could be calculated according to any metric. The results of simulations carried out with several metrics and various arrangements of companies were analyzed as above, yielding similar results confirming the robustness of the square-root relations even for situations with as few as three companies in the region.

We conclude this section by examining some simulation results for fire-station locations in Bristol, England, developed by J. Hogg [10]. Her basic data consisted of the locations of 6813 fires which occurred in Bristol from 1958 to 1964, and 15 sites at which fire stations could be located. Neither the site locations nor the fire incidence were evenly distributed spatially, both being more dense in the center of the rectangular region. In the analysis, nonstationarities of demand and travel velocity in time were considered, but possible unavailability of fire companies was ignored. Response times were calculated from knowledge of the distances involved and estimates of travel speed, which recognized variations by region and time of day.

Some of Hogg's results—replotted in Figure 3—give average travel times as a function of the number of firehouse locations occupied. These results were fit by least squares to models (3) and (4) and to the exponential function

\[ T = \alpha e^{\beta n}. \]

The sums of squared errors indicate that (3) fits best, (4) second-best, and (5) worst, with the difference between (3) and (4) being small. Simply looking at the graph also reveals little difference between (3) and (4). Moreover, an approximate 95 per cent confidence contour for \( \alpha \) and \( \beta \) includes \( \beta = -\frac{1}{2} \).

![Figure 3. Average response time vs. number of fire company locations (Bristol, England).](image-url)


**Limitations**

Now we examine in more detail the constant terms and residual errors from the least squares fits. Figure 2 shows systematic differences between average response distances generated by the Bronx simulations at different alarm rates: all the points obtained at a simulated rate of 21.3 alarms per hour lie above the fitted regression line representing the square root law, while all but one of the points obtained at 13.3 alarms per hour lie below it. The differences are small when 17 or more of the units are available in the region, which is most of the time, so that the “law” could be used in practical situations with considerable confidence. At low availability, however, the fits are less satisfactory, and the actual alarm rate does matter in a way not accounted for by the simplified theory used to derive the square root law. We have some understanding of the sources of the deviations. For example, we know that at high alarm rates the spatial distribution of units is affected by the deployment policies used, so the assumption of constant distribution of units is not strictly valid.

One can take care of such deviations in several ways: for example, by fitting separate regression lines for each alarm rate, or, as we have usually done, by supplementing the square-root law with analysis using appropriate queuing and simulation models, such as those described in [3] through [7].

**IV. Applications to Allocation Problems**

In this section, we illustrate how the square-root law can be applied to the analyses of resource allocation problems:

Consider the following allocation situation: We divide the City into \( m \) disjoint regions or neighborhoods, each of which is (roughly) homogeneous with respect to alarm rate, velocity of responding vehicles, types of fires, etc. Each region is to be assigned a number of fire companies which, except in special situations, will serve only that region. In region \( i \) (\( i = 1, 2, \ldots, m \)), we define

\[
A_i = \text{the geographic area (square miles)},
\]

\[
\lambda_i = \text{expected alarm rate in the period of interest (alarms per hour)},
\]

\[
1/\mu_i = \text{expected total time spent servicing an alarm by all the units employed (hours)},
\]

\[
n_i = \text{the number of units assigned to the region}.
\]

Assume that for district \( i \) the expected number of companies busy can be approximated by \( \lambda_i/\mu_i \) and that an inverse square root relationship holds for expected response distance,

\[
ED_i = c_i [A_i/(n_i - \lambda_i/\mu_i)]^{1/2},
\]

where \( c_i \) is a constant of proportionality depending on the street configuration, the location of house, etc. (Note that this function makes sense only if \( n_i > \lambda_i/\mu_i \))

Let us transform expected response distance to expected response time\(^2\) using the function \( ER_i = \beta_0 + \beta_1 ED_i \). Then overall expected response time in the City becomes

\[
ER = \sum_{i=1}^{m} \lambda_i ER_i / \sum_{j=1}^{m} \lambda_j.
\]

\(^2\) Space does not permit presentation of our empirical data on response time-response distance functions here, but it is important to remark that we have found this relationship to be surprisingly insensitive to time of day and to have a simple linear or square root form depending on the region of the City [15].
Note that we have ignored time of day variation for alarm rates and response velocities. If we wish to examine the relation between ER and \( n \) through time, we do so by posing a sequence of allocation problems for various times of day, using the appropriate parameters.

Optimization formulations we have found useful include:

1. Minimum Average Response Time Allocation. Problem: Find integers \( n_1, n_2, \cdots n_m \) which minimize \( ER \) subject to \( n_i > \lambda_i/\mu_i, i = 1, 2, \cdots, m \), and to \( \sum_{i=1}^{m} n_i \leq n \).

   In this formulation, we find that allocation of the total of \( n \) available companies which minimizes the city-wide expected response time per alarm. The restrictions \( n_i > \lambda_i/\mu_i \) require that we assign to each region at least as many units as will be busy on the average, which keeps our mathematical formulation sensible and workable. We note that with realistic (or real) response time-response distance functions the objective function is convex decreasing and separable in \( n_i \) so that a simple iterative procedure of examining marginal gains will determine the optimal allocation:

   **Step 1.** Set \( n_i = \lceil \lambda_i/\mu_i \rceil + 1 \), \( i = 1, 2, \cdots, m \). If \( \sum n_i > n \) the problem is infeasible; if \( \sum n_i = n \) the allocation is optimal; otherwise go to Step 2.

   **Step 2.** Calculate \( \Delta_i = R_i(n_i + 1) - R_i(n_i) \), \( i = 1, 2, \cdots, m \). Set \( n_j = n_j + 1 \) for \( j \) such that \( \Delta_j = \max \Delta_i \). If \( \sum n_i = n \) the allocation is optimal; if not, repeat Step 2.

2. Minimum Complement of Companies (Average Response Time Constraint). Problem: Find integers \( n_1, n_2, \cdots, n_m \) which minimize \( \sum_{i=1}^{m} \alpha_i \) subject to \( ER_i \leq \alpha_i, i = 1, 2, \cdots, m \).

   In this formulation, management or the public specifies the standard of protection (in terms of average response time) to be provided in each region, and the minimum number of companies necessary to achieve this protection is determined. The solution to this problem is to set \( n_i \) to the smallest integer larger than

   \[
   A_i \left[ \frac{\beta_i c_i^2}{\beta_i - \beta_i c_i} \right]^2 + \frac{\lambda_i}{\mu_i}.
   \]

   Clearly, other optimization problems can be formulated and solved. For example, one might seek to minimize the probability of long response times or formulate problems involving decisions about the overall number of companies as well as their assignment to regions. We have experimented with these models, though formulations (1) and (2) have been the workhorses of our analyses.

   In applications, we use the model more often to evaluate proposals generated by other analysis or by management than to prescribe "optimal solutions". An on-line computer program has been written which solves problems (1) or (2) and compares the values of \( ER_i \) for any set of allocations specified by the user. Such use of the program has proven attractive in applications. With the computer model, management can take into account a diversity of factors in creating allocation proposals and then use the model to evaluate them numerically for their impact on response time. The utility of these procedures and the validity of the calculations depend on how closely the original model assumptions are met and on how well the allocation regions are chosen. Regions should be neither so big as to be grossly inhomogeneous nor so small that the values of \( n_i \) obtained are meaningless.

**References**


