AGRICULTURAL
ECONOMICS
RESEARCH UNIT

Lincoln College

THE AMBULANCE FACILITY LOCATION PROBLEM — A SURVEY OF METHODS AND A SIMPLE SOLUTION

by

JANET GOUGH AND OWEN McCARTHY

Research Report No. 73
December 1975

THE AGRICULTURAL ECONOMICS RESEARCH UNIT

THE UNIT was established in 1962 at Lincoln College, University of Canterbury. Its major sources of funding have been annual grants from the Department of Scientific and Industrial Research and the College. These grants have been supplemented by others from commercial and other organisations for specific research projects within New Zealand and overseas.

The Unit has on hand a programme of research in the fields of agricultural economics and management, including production, marketing and policy, resource economics, and the economics of location and transportation. The results of these research studies are published as Research Reports as projects are completed. In addition, technical papers, discussion papers and reprints of papers published or delivered elsewhere are available on request. For list of previous publications see inside back cover.

The Unit and the Department of Agricultural Economics and Marketing and the Department of Farm Management and Rural Valuation maintain a close working relationship in research and associated matters. The combined academic staff of the Departments is around 25.

The Unit also sponsors periodic conferences and seminars on appropriate topics, sometimes in conjunction with other organisations.

The overall policy of the Unit is set by a Policy Committee consisting of the Director, Deputy Director and appropriate Professors.

THE AMBULANCE FACILITY LOCATION PROBLEM - A SURVEY OF METHODS AND A SIMPLE APPLICATION

by

Janet Gough and W. O. McCarthy

Research Report No. 73
December 1975

PREFACE

As our list of publications indicates, an area of substantial involvement for the Unit in recent years has been in location economics. In these studies we have mainly been concerned with the location of facilities for processing, servicing and storage. As is appropriate the focus has been on primary industry applications.

Nevertheless there are certain problems in methodology to be investigated if we are to widen the scope of our research in primary industry. Central issues include moving from a static to a dynamic framework and allowing for stochastic supply and demand.

The purpose of this study is to explore this latter area. Although removed from agriculture the ambulance problem was considered appropriate because of its allied nature, the amount of previous work published on this topic and the availability of local empirical data.

Owen McCarthy
Director

December 1975

C O N T E N T S

			Page
PREFACE			
1.0	THE NATURE OF THE PROBLEM		1
	1.1	General	1
	1.2	The Ambulance System	1
2.0	THE SYSTEM		6
3.0	THE MODELS		8
	3.1	Fitzsimmons	8
	3.2	Toregas, Swain, ReVelle & Bergman	10
	3.3	Swoveland, Uyeno, Vertinsky & Vickson	11
	3.4	Savas	13
	3.5	Hall	14
	3.6	Volz	14
	3.7	Schneider & Symons	16
	3.8	Larson	17
	3.9	Kuehn & Hamburger	19
4.0	COMPARISON OF THE MODELS		22
	4.1	Assumptions Made	23
		4.1.1 Queuing Models	23
		4.1.2 Analytic Models	24
	4.2	Calculation of the Mean Response Time	25
	4.3	Amendment of the Feasible Location Set so as to Improve the Mean Response Time	26
	4.4	Evaluation of the Solution	27
5.0	CHOICE OF A MODEL		
	5.1	Advantages of the Models	29
	5.2	Disadvantages of the Models	30
	5.3	Selection of the Model	33
6.0	THE CHRISTCHURCH STUDY		36
	6.1	The Christchurch System	36
	6.2	Preparation of the Data	37
		6.2.1 Call Times	39
		6.2.2 Call Origins	41
		6.2.3 System Data	47
	6.3	Results	48
		6.3.1 Fitzsimmons Model : CALL	48
		6.3.2 Results of the Simulation Model	81
7.0	CONCLUSIONS		84
ACKNOWLEDGEMENTS			86
REFERENCES			87

APPENDIX I	:	Fitzsimmons Model : CALL	89
APPENDIX II	:	Swoveland, Uyeno, Vertinsky, Vickson Model	109
APPENDIX III	:	Possible Alternative Approaches	139
APPENDIX IV	:	Adaptions to the Simulation Program for the use with the Hooke & Jeeves Pattern Search Subroutine	142

LIST OF FIGURES

FIGURE		Page
1	Hospital & Depot Locations	38
2	Location of 359 Call Areas	43
3	Location of 82 Call Areas	44
4	Location of 60 Call Areas	45
5	Final Locations for Run 1 & Run 3	54
6	Final Locations for Run 1, Run 2, Run 3, Run 4 & Run 5	57
7	Final Locations for Run 3, Run 6 & Run 7	60
8	Final Locations for Run 3, Run 8, Run 9, Run 10, Run 11	63
9	Final Locations for Run 3, Run 12, Run 13, Run 14 & Run 15	66
10	Final Locations for Run 3, Run 9 & Run 16	69
11	Final Locations for Run 3, Run 17 & Run 18	72
12	Final Locations for Run 3, Run 19, Run 20 & Run 21	75
13	Final Locations for Run 3, Run 23, Run 24 & Run 25	80

LIST OF TABLES

TABLE		Page
1	Results of CALL	51
2	To Test (a) The effect of changing the number of Areas	53
3	To Test (b) The effect of changing the initial positions of the ambulances	56
4	To Test (c) The effect of Changing the Number (and Positions) of Hospitals	59
5	(d) To Test the effect of changing the Number of Calls LAMDA	62
6	(e) The Effect of Restricting the Final Positions of Ambulances	65
7	(f) The Effect of Changing the Number of Ambulances	68
8	(a) The Effect of Changing the Number of Monte Carlo Samples Taken	71
9	(b) The effect of Changing the Seed of the Random Number Generator	74
10	(c) The Effect of Changing the Number of Times RBAR is iteratively Calculated, to Remove the Effect of the Initial Value	76
11	(b) The Effect of Changing the Number of Calls, on the Model Assumptions	79
12	Simulation Model Results	82

1.0 THE NATURE OF THE PROBLEM

1.1 General

A number of service facility location problems are concerned with providing urban emergency services such as ambulances, fire and police. These problems differ from other service location problems such as the location of personnel (e.g. farm advisers) in that the time and location of demand for the service is uncertain.

Accordingly methods of solution are more complex.

The purpose of this study is to review various techniques which have been used for locating ambulance systems, to comment on their effectiveness, to suggest improvements and to work through a simple problem for Christchurch city.

1.2 The ambulance system

An ambulance system is concerned with several different functions. It must perform emergency services (consisting of road accidents, home accidents and sporting accidents); transfer functions (concerning transfer of patients between hospitals, or from hospital to home) and normal functions (consisting of priority calls which are not necessarily emergency calls).

Savas (4) considers the emergency medical care system (not to be confused with the emergency ambulance service) as being composed of three sub-systems:

(a) Communication

(b) Transportation

(c) Medical Treatment

of which the first two comprise the ambulance system. They are concerned with the locations of ambulances and hospitals, the dispatching of ambulances and the boundaries under which ambulances operate, as well as the design of ambulances, speed of travel and use of siren and similar factors.

The Medical Treatment sub-system is concerned with the equipment carried by ambulances, qualification of personnel, and availability of casualty stations.

The problem here considered is to attempt to improve the communications and the transportation aspects of ambulance recovery by means of locating ambulances in optimal positions. The constraints which are inherent in the general ambulance communications system are that all emergency calls must be answered, and that non-emergency and transfer functions must be performed according to any restrictions which may be placed on them.

The most important part of the ambulance service is the emergency service, and the 'optimality' of the ambulance locations is here considered primarily in terms of the emergency service.

The parameter usually used to measure the efficiency of the emergency service is the 'response time', which is defined as the time taken from receiving a call for an ambulance, to the arrival of the selected ambulance at the scene of the accident or emergency. The optimising parameter is therefore the mean response time for the system. Other parameters suggested which may be used as additional criteria for the evaluation of final solutions are: the mean retrieval time to the hospital, the average round trip time, the percentage of calls exceeding a certain 'maximum' response time, the number of calls exceeding this 'maximum' response time, the balance of the work load between ambulances as the percentage of calls handled by each ambulance, the standard deviation of the mean response time, the probability of a response time greater than a set of given times, the number of busy ambulances, with the associated probability for each occurrence, and the probability distribution of the response time.

In some of the studies considered (Toregas et al. (7), Volz (13)), an upper limit has been placed on mean response time from each area, or on the distance of any area from an ambulance station. This is used in an attempt to eliminate the problem which may arise of a low average response time overall, but a very high response time for one or two areas which have infrequent calls, that is, a low probability of an emergency arising from them.

The response time which may be considered as the most important measure of efficiency can be divided into three parts:

(a) Despatch delay - the delay in the decision making process as to which ambulance to send.

(b) Waiting time - the time waiting for the selected ambulance to become free.

(c) Travel time - the time taken for the ambulance to reach the scene of the emergency.

The despatch delay is usually fixed and constant for all cases, and an initial assumption in most cases is that the probability of all ambulances being busy is zero, thus making the waiting time zero. The main reason for this assumption is that Parzen (15) has shown that for a queuing system in which no waiting occurs, the number of busy servers has a Poisson distribution, with mean equal to the ratio of mean arrival rate to mean service rate. Savas (4) and Gordon and Zelin (10) have shown that for the assumption to be valid, the ambulance utilisation must be as low as 30% in the areas studied. However, in most systems where this has been tested (Fitzsimmons (1), (2), Volz (13)), the assumption has been shown to be quite valid. Therefore the variable response time which it will be required to minimise, may be considered as the time taken for the ambulance to travel to the accident scene. (Note that if a complete simulation is performed the possibility of a queue forming may be dealt with quite satisfactorily.)

Considering the major objective to be the reduction and minimisation of the variable average response, there are several courses of action available (Savas (4)). These are:

(a) To relocate ambulances away from a central point.
(b) To increase the number of ambulances.
(c) To both increase and relocate ambulances.

The first of these three courses of action will be considered, with a later extension considering the effects of increasing the number of ambulances to be worked on the solution or solutions obtained. This will be the form of the sensitivity analysis to be performed on the solution.

There are also several ways in which ambulances may be located.

(a) All ambulances may be located at a central point (trivial).

(b) Ambulances may be located uniformly, according to geographical boundaries.

(c) Ambulances may be located uniformly according to sub-regions of heaviest demand.

On investigation of the eight systems studied (1), (3), (4), (7), (9), (10), (12), (13), it was found that three different approaches to the selection of 'demand areas' have been used.

(a) 'Areas' have been delineated by placing a grid system over a map of the city or region (1), (3), (13), (7).

(b) 'Areas' have been selected by some population criteria, for example, each area contains approximately 10,000 people (9).

(c) 'Areas' have been selected according to some political boundary system (12).

The initial placement of ambulances within these 'areas' is generally according to distribution of demand. Swoveland et al. (9) use (b) in conjunction with a further grouping of 'areas' into 'regions' of greatest concentration of demand, to each of which an ambulance is assigned. The problem then becomes the selection of the optimal 'area' within each 'region' for the ambulance to be located. (Regional Response.)

Fitzsimmons (1), (2), (Lazarus (3), a further application of Fitzsimmon's method) and Volz (13) use the grid system and placement of ambulances is performed by an optimising algorithm. Hall (12) uses political boundaries and Savas (4), and Gordon & Zelin (10) use a direct placement according to distribution of demand. (These are different kinds of studies as they are concerned primarily with investigating the system and testing the hypothesis that locating ambulances away from the hospital site will improve the performance of the system.) Toregas et al. (7) are concerned with a network, for which distances are calculated according to a grid reference;

similarly Schneider & Symons (17) select sites on a radial network basis.

2.0 THE SYSTEM

All the specific studies have included descriptions of the operation of the system. The paper by Gordon and Zelin (10) is concerned with a preliminary investigation of a part of the New York ambulance system. Savas (4) later used their approach and deductions to produce a much more comprehensive report. This was not really an optimising study, but from investigation of the system, Savas was able to make a number of recommendations, and reach a number of conclusions relevant to other similar systems. Most of the later authors of ambulance location papers use Savas as a base reference.

Another part of the system has been considered by Shonick and Jackson (11). They are concerned with the hospital admittance problem, which they approach by means of a two line queuing model. This study is not concerned with the Medical Treatment sub-system, and it is assumed initially that an ambulance may always take a patient to the hospital closest to the scene of the emergency. ('Hospital' here means an emergency accepting hospital.) This is compatible with New Zealand policy.

Hall (12) describes the Detroit ambulance system, and Volz (4) describes the system in Washtenaw County, Michigan. This latter area is semi-rural, and possibly comparable to the typical New Zealand situation.

Fitzsimmons (1) outlines the system parameters as follows:

1. <u>Design Parameters</u>
   - (a) number and location of ambulances
   - (b) number and location of hospitals.

2. <u>Operating Policies</u>
   - (a) despatching policies
   - (b) retrieval policies
   - (c) speed of retrieval
   - (d) adaptive deployment policy.

3. Demand for Service

(a) system arrival rate

(b) incident location distribution.

3.0 THE MODELS

Three initial assumptions will be made about the system.

(a) That the probability of a call arriving at a time when all ambulances are busy is zero.

(b) That any patient may be taken to the nearest hospital, and that there is no limitation on hospital numbers as far as emergency admittances are concerned.

(c) That all calls may be considered as coming from the centroid of the 'area' defined.

The following sub-sections take each of the models considered in turn, and describe the way in which they are used to illustrate the ambulance systems they represent.

3.1 Fitzsimmons

The model developed by Fitzsimmons (Los Angeles) (1), (2) and later used also by Lazarus (Melbourne) (3) is very simple and flexible, and provides for several options, including multiple hospitals, and multiple ambulance types. It converges to a local optimum, and is a partial simulation model. The demand pattern used to estimate the mean response time is derived from actual demand data for a certain period in the form of the probability of a call coming from each area, where 'areas' are defined by a grid system. This is a common approach to defining the probability of a call. In the Fitzsimmons model, the same set of random numbers is used with the probability data each time a new set of locations is tested by establishing the expected mean response time, in order to allow comparative evaluation of the location sets. The Hooke and Jeeves technique (5) is used to determine a new 'better' location set. Fitzsimmons approaches the problem by noting from Savas (4) that the dispersal of ambulances throughout the service area considerably reduced the average response time. Previous studies have also produced evidence that reduction of the time taken to get the accident victim to hospital can reduce the fatality rate, thus reinforcing the assumption that response time is a useful criteria for the evaluation of an ambulance service. He constructs a queuing model where the 'waiting time' is considered as the despatch delay plus the wait for an ambulance, and the 'service time' is the total trip time, or the travel time to the scene, plus on-scene care time, travel time to hospital and transfer delay at the hospital. From Parzen (15) he concludes that the arrival times follow a Poisson distribution. The system then becomes one with exponentially distributed inter-arrival times, a general service time distribution and an infinite number of servers (due to the first basic assumption). The state of the system is considered to be the number of busy servers or ambulances. Monte Carlo methods are used to estimate mean response times for states where more than one ambulance is busy.

Additional assumptions made by Fitzsimmons includes one that is difficult to avoid in an analytical approach, that is, that all ambulances respond to calls either from their depot or from a hospital, also, that travel time may be computed as rectangular displacement between cartesian co-ordinates. Service time must be independent of the state of the system; however, as service time includes travel time to the accident this is not strictly true. Since total service is usually much greater than travel time to the scene, it may be considered approximately to hold.

This model established by Fitzsimmons computes the system mean response time for a given set of locations. It is then combined with the pattern search routine developed by Hooke and Jeeves (5), (14). Fitzsimmons considers that mean response time might not be sufficient criteria for selecting an 'optimum' set of locations, and suggests that a constraint might be added so that mean response from any area should not exceed a certain maximum. A number of addition parameters are presented for subjective evaluation. The program CALL (2) includes a facility allowing some ambulances to be fixed with certain boundaries before optimising for the remaining 'free' ambulances.[1]

Fitzsimmons (1) does give some consideration to the question of the minimum number of ambulances required for a particular system.

---

1 Note that the fixing of ambulances within these boundaries refers only to the final depot location of the ambulance and not to the area within which it operates.

This by definition is the number of ambulances such that the expected waiting time for an ambulance is zero. Bell (16) shows how this can be calculated if all ambulances are to be stationed at a single central station.

Fitzsimmons gives an approximate calculation of the minimum number of ambulances required for the dispersed station case, but only for the situation where all ambulances are considered able to provide service at any time.

Appendix I describes the 'Computerized Ambulance Location Logic' model CALL more fully.

3.2 Toregas, Swain, ReVelle & Bergman

Toregas et al. (7) formulate a model which can be solved by linear programming. The approach is to place an upper limit on the response time or the distance from the nearest facility for all areas, and then, assuming costs to be identical for all possible facility locations, to minimise the number of service facilities required to meet the response time or distance restrictions set.

This approach will give both number and location of facilities. However, it must be assumed here, that each facility location is able to respond immediately, at all times, to a call. This is essentially the same as restricting each ambulance to a fixed area (regionalised response), and to allow all calls to be answered immediately may require a number of ambulances to be positioned at each station.

A second study could be considered in which, assuming the minimum locations established by the linear programming minimisation above, the problem would be to determine the number of servicing units or ambulances which would be required at each depot, to meet the assumption that a unit should always be available. However, this form of solution would mean that a lot of unnecessary idle ambulance time would be incurred.

The other assumptions are general to the problem, that is, first that the demands may be assumed as occurring at central nodes of the defined areas, and that the service facility location may be either a subset or the whole of these nodes, and also, that the minimum distances or response times between the service location node and demand node pairs are known.

In order to perform a linear programming minimisation it is necessary to produce sets of possible locations which satisfy the restrictions that every demand node should be within the feasible distance/time from the service nodes. These are produced by a program which is linked to the MPS LP program. The solution of the LP program produces fractional results rather than 0-1 solutions and therefore consideration is given to the elimination of these, which is achieved by the addition of a single cut constraint.

3.3 Swoveland, Uveno, Vertinsky & Vickson

A different approach to the ambulance problem is made by Swoveland et al. (9) in which they apply a probabilistic Branch and Bound procedure. Simulation output is used to construct an analytical approximation to mean response time and the combined optimisation problem is solved by a probabilistic Branch and Bound algorithm.

The initial partitioning of the area under study is into 'areas'. The 'areas' are then grouped into 'regions' to which a particular ambulance is assigned. This assists in the balancing of workload between ambulances and also achieves a reduction in problem size.

A program, separate from the main program is used to generate an input call stream with nodes of origin, time of occurrence, priority, load-unload time and travel time from the node of origin to destination. Because it is a complete simulation the approach also has the facility that some calls may be cancelled on route or at the scene of the emergency. Queues may develop and different hospitals may be specified by means of a probability distribution for each 'area'. The two despatch rules, closest ambulance response and regionalised ambulance response are both used, and accordingly two separate solutions may be generated. It is interesting to note that although the second despatch rule is termed a regionalised response, the regions may actually overlap, so that a node may be served by more than one ambulance.

One of the interesting things about this approach is that a call may be given a priority. That is, service and transfer calls may be included in the study as well as emergency calls. This part of the model is more flexible than the Fitzsimmons approach.

Having generated the call stream, the next step is to obtain an estimate of the mean response time. The need for an approximation is established by restrictions on computing time, making it infeasible to simulate each possible series of assignments. The expected response time is approximated using a 'stability' hypothesis (which was tested and found acceptable). This hypothesis conjectures that the fraction of total calls per day arising at node i and serviced by the qth nearest ambulance will not be significantly different when produced by different 'assignments' where the 'assignments' refer to the different placements of ambulances in 'areas' within the 'regions'.

It is assumed that all responses to calls originate from depots (whereas Fitzsimmons allows for response from hospitals as well). Simulation runs designed to test the estimated response time with the computed response time showed that the approximation is satisfactory.

The minimisation or response time as a function of allocation is performed by a Branch and Bound procedure where the best assignment ('area') within the 'region' is determined. So, given a set of regions where each 'region' is given an ambulance, the best location or 'area' within that 'region' will be determined.

Data pruning is an essential part of the process as certain 'areas' may be omitted from the study (by amalgamation), and others may be 'fixed' by equating a 'region' with an 'area'. Considerable testing of the solution of the Vancouver study was performed, for the purpose of verifying the procedure. Also, the current situation was simulated for comparative purposes and proved that the exercise solution was marginally better. Parameters such as minimum and maximum expected response times, and priority ratings obtained by simulation added extra weight to the hypothesis that the calculated solution was superior to the currently operating system.

It has not been possible to obtain further information about the Branch and Bound optimising section of this procedure for which the estimate of the mean response time is calculated.

The full simulation part of the procedure has been tested and found to be very comprehensive, to the extent of reinforcing the necessity of some simplification in the procedure for calculating the mean response time for use with an optimising algorithm.

Appendix II describes the model more fully.

3.4 Savas

The simulation study performed by Savas (4) took as its general objective "to improve ambulance service", and used as measures of effectiveness

(a) the response time and
(b) the round trip time.

The specific objective then selected was to reduce response time (thus also reducing round trip time), and the alternatives available were to redistribute the existing ambulances, increase the number of ambulances, or combine the two. Cost and effectiveness on average response time, and percentage of calls exceeding a certain limit set on response time were considered. A good flow diagram of the simulation of a case is presented with the events following until the ambulance returns to its depot. The study was amenable to complete simulation and would also have been suitable for analytical solution because of the small size of the problem in which only one depot and one hospital were considered (see Fitzsimmons' thesis). Its main usefulness is in showing the effectiveness in reduction of average response time that will occur with the use of dispersed locations for ambulances.

Recommendations produced were that for a city, ambulances should be assigned to depots by a central unit, distributed geographically according to demand, and redistributed periodically as the pattern of demand changes, and also, that very rarely should two or more ambulances be stationed at one location.

Other results recommended that the number of ambulances should be sufficient to prevent the formation of significant queues, and that large service areas with no district restrictions on ambulance travel were the most efficient.

3.5 Hall

Another study was performed for Detroit, Michigan, by William K. Hall (12). The system is smaller than the New York and Vancouver systems and concerned with the use of police vehicles and ambulances (some of the police vehicles doubling as ambulances). The locations were fixed as police precincts, and three policies were examined for four centres using either one, two, three or four ambulances assigned to each precinct. Ambulances were not able to cross the precinct boundaries.

The three policies were:

(a) Assigning the vehicles to the sub-regions of heaviest demand for ambulances.

(b) Assigning the vehicles uniformly throughout the precinct on a geographical basis.

(c) Assigning the recovery vehicles to a single fixed station within the precinct (the police stations were the obvious choice here).

Substantial improvements were obtained, with the criterion for improvement being the probability of an ambulance being available.

3.6 Volz

Volz (13) in a study performed for Washtenaw County, sought to minimise the average response time to emergency calls, but he also used the restriction that the average response time to any point within the service area should be less than a given maximum. The despatch rule is that every time a call is received the nearest ambulance is sent. The remaining ambulances then relocate themselves in an optimal manner for the number of ambulances currently available.

This policy has been used in Washtenaw County for some time. The assumption that the probability of all ambulances being busy is zero is considered valid, since this event occurred only twice in the year's recorded data prior to the study.

One interesting feature of this model is that it may not necessarily be concerned only with emergency demand since the response time is calculated as a function of time of day, weather conditions and other related variables. Roads are divided into four categories. The average response time is computed analytically.

The minimisation procedure is carried out by amending solutions from an initial 'guessed' arrangement by means of a 'steepest descent' (14) method. Local minima exist and therefore it is necessary to restart the procedure from a number of different points. The validity of the model for Washtenaw County was checked by computing the average response time for the current system, and then checking this with the actual average response time. Subject to qualification due to differences in the current operation it proved satisfactory. The addition of the constraint to restrict the average response time from any area to a selected maximum caused a slightly different solution to be generated.

The computation does require a considerable amount of computer storage and computing time. Using an IBM 360 model computer Volz considers that he was very close to the limits of feasibility with a 30 x 24 mile area with a one mile square grid pattern and six ambulances. The study was performed for three, four, five and six ambulances, and it can be seen that the addition of one ambulance increases the computation and time factors exponentially rather than linearly. (From later experience, however, it may be noted that 720 grid areas constitute a large study.)

3.7 Schneider & Symons

The Schneider and Symons study (17) is part of a planned series of experiments designed to investigate different possible approaches to urban location problem solving. This particular paper discusses a man-computer interactive study of the location of ambulances in an urban area.

The assumptions made of the system are quite general except that a regional response rule with no overlapping is imposed. The regions are delineated by the rule that each ambulance serves only those accidents that are closer to it than to any other despatch centre. The disadvantages of regionalised response models such as Toregas and Hall, therefore apply at least in part, although here, the regions are determined internally to the problem and not externally as, for example, in the case of political boundaries.

Two objectives are stated:

(a) To minimise total travel time for each ambulance to reach all of the accident locations in their service area.

(b) To restrict the trip time from any ambulance centre to any accident location to less than X minutes.

The essential hardware components of the study were an IBM 1130 computer with 16K memory and a half million word disk with on line card reader, and one Advanced Remote Display Station (ARDS), incorporated with a data keyboard and 'joystick'. The programs are set up to present the analyst with a graphic display of the problem as nodes and intersections on the ARDS scope. The analyst then indicates the number of ambulances he wishes to work with and selects locations for these, using the 'joystick'. The computer calculates the required parameters and presents them to the analyst who proceeds to attempt to improve upon his previous solution.

The conclusions of the study are that the ADLOC (Ambulance Dispatch Centre Locator) system is a useful tool for finding good solutions to a simplified location-allocation problem, for helping people to understand the problem and for assisting the development of computer based heuristic techniques. Two heuristics are developed in this study which attempted to improve upon initial locations set by the analyst. Further work is intended to consider larger and more complex problems.

This type of approach to ambulance and general service facility location problems is certainly different. Its main use may be in the development and testing of heuristics suitable for use in the optimising phase of the problem. Also, the approach may be generalised to deal with a number of types of service facility rather more conveniently than the other models discussed. However, a considerable emphasis must be placed on the skill and experience of the analyst, and it is questionable whether the criteria for 'best solution' being used at this stage in the study are really sufficient.

3.8 Larson

Larson's paper (18) entitled "A hypercube queuing model for facility location and re-districting in Urban Emergency Services" is a very interesting one.

The abstract states "This paper develops computationally efficient algorithms for studying the analytical behaviour of a multi-serve queuing system with distinguishable servers. The model is intended for analysing problems of vehicle location and response district design in urban emergency services, includes interdistrict as well as intradistrict responses and allows computation of several point specific as well as area specific performance measures."

The model is later described as a "finite state continuous time Markov process" and thus is relevant to the discussion in Appendix III on the Markov process approach. The mathematics involved in the process, however, are very complex and it is considered that a comprehensive study of the method would involve some months. It does appear to have far more general application than the other models studied, which means it might be used for the investigation of fire station location, police vehicle location and other similar urban and rural services for which there is some form of stochastic demand.

Larson's interest is in the area of the general problem of allocating urban emergency units as can be seen from his earlier paper with J.M. Chaiken, "Methods for Allocating Urban Emergency Units : A Survey" (19). This paper discusses the general problems involved in urban emergency services which are considered as systems having the following properties:

(a) Incidents occur throughout the city which give rise to requests or calls for service; the times and places at which these incidents occur cannot be specifically predicted in advance.

(b) In response to each call one or more emergency service units are despatched to the scene of the incident.

(c) The rapidity with which the units arrive at the scene has some bearing on the actual or perceived quality of the service.

(d) In addition to such examples as fire engine and ladder trucks, police patrol cars and ambulances, emergency service units include certain tow trucks, bomb disposal units, and emergency repair trucks for gas, electric and water services.

It is noted that 'although all urban emergency service systems share the above characteristics, they may differ in certain significant details'.

The most important differences are that some emergency units have fixed locations (e.g. fire engines) whilst others are mobile (e.g. police patrol cars) and that some services require the despatch of one unit to a call (e.g. ambulances) whilst others require as a matter of course, the immediate despatch of a number of units (e.g. fire engines).

The section of this paper entitled 'Modelling Methods' is unsatisfactory to our purposes, as it is concerned most specifically with police patrol cars. However, it does differentiate between different types of nodes and different mathematical techniques which have been used. The problems of region delineation (more specifically for mobile units) and location, including pre-positioning and re-positioning are considered, although no satisfactory conclusions are reached.

In conclusion, the paper stresses the potential benefits accruing from analysis of such systems and looks forward to increased sophistication and application of models of the systems.

3.9 Kuehn and Hamburger

Problem evaluation is the area in which a solution is calculated in analytical terms and an evaluation procedure is used to determine whether or not a local optimum has been reached, and if not, how the present location pattern may be changed so as to improve the solution parameters.

Fitzsimmons uses a Hooke and Jeeves pattern search technique (5) to perform this process and Volz uses a method of steepest descents. Swoveland uses a Branch and Bound procedure.

Keuhn and Hamburger (8) discuss heuristic approaches to locating warehouses using three main heuristics which may be applicable to the ambulance system.

(a) Most geographical locations are not promising sites for a regional warehouse. Locations of promise will be at or near concentrations of demand.

Already it has been seen that initial location sets can be chosen with facilities located according to areas of demand. This does suggest an initial screening of potential locations. However, with the ambulance location problem this should be approached cautiously as a promising location may be disregarded at an early stage because of lack of information which becomes available later in the study. Also, a site which is not close to a demand area may be a useful part of a solution because it is able to service two or more demand areas at points around it.

(b) Near optimum warehouse systems can be developed by locating warehouses one at a time, adding at each stage of the analysis that warehouse which produces the greatest cost savings for the entire system.

For ambulance depot locations it is likely that the above may not be the case. Most of the methods considered work with a fixed number of locations, and attempt to locate these by initial placement, and shifting of these initial placements. Variation in number can only be achieved by 'running' the model several times with different numbers of facilities specified, and then comparisons may be made between the various parameters. The exception to this is the Toregas (7) model which seeks to minimise the number of ambulance facilities required.

One problem concerned with this second heuristic might be that the method would probably lead to a small concentration of depots around the central city area where the greatest concentration of demand occurs. A constraint would be needed to restrict the maximum response time allowable.

(c) Only a small subset of all possible warehouse locations need be evaluated in detail at each stage of the analysis to determine the next warehouse site to be added.

The reason for this proposal is connected with local demand, and only those areas where local demand would indicate a useful location are considered. This would seem satisfactory again in a situation where the facility deals only with demand from its particular area, but with the 'nearest ambulance' idea of despatch, it may prove risky.

The method followed by Keuhn and Hamburger selects the number of locations on the basis of adding the locations which give the greatest reduction in cost, until the addition of an extra facility would increase the total cost. Then it proceeds to a 'bump and shift' routine which modifies the solution by eliminating any warehouse which becomes uneconomical, and shifting warehouses from the currently assigned locations to any other potential sites from the original list. This part of the procedure might overcome the problems involved with the method of ambulance assignment, that is, the number of facilities could be determined by assuming each ambulance serves only a specific area, then in the second part, this could be extended to allow boundaries to be crossed.

Several reasons why the classic warehouse location problem is not equivalent to the ambulance problem are summarised by Fitzsimmons (1) with the primary reason being that ambulances are mobile, and that if the nearest ambulance is not available, then another, more distant one, must be sent to answer the call. This idea of a 'closest ambulance' despatch rule means that the response time to a call is dependent on the state of the system at the time of the call.

4.0 COMPARISON OF THE MODELS

In order to determine the set of the best possible locations, the problem may be approached as an iterative process, formulated as follows:

(a) Determine an initial set of feasible locations, according to a given demand pattern (pre-positioning).

(b) Obtain the mean response time for this set of locations, according to a given demand pattern.

(c) Test this solution for optimality (attainment of a local optimum) and if it is optimal, terminate the procedure - otherwise, amend the feasible location set and return to (b), (re-positioning).

Therefore, to specify the model it is necessary to specify:

(a) The assumptions to be made about the system.

(b) A method of obtaining the expected response time according to the current set of locations. This includes specifying a rule for the dispatch of ambulances (nearest ambulance, ambulance 'tied' to that area or 'regionalised response'), and a method of computing the response time either analytically, or through simulation.

(c) A method of amending the feasible location set so as to improve the average response time (the choice of the initial set of feasible locations will be dependent on the method used here).

(d) An evaluation criterion for determining the optimal (solution) set of locations. This will be required at two levels. Firstly, a means of determining the local optimum for a particular 'run' is required. Then, it may be considered necessary because of the nature of the solution surface which will probably contain a large number of local optima (Volz (13)) to specify some means of selection of a 'best' solution from a number of local optima obtained from several 'runs' of the problem using different initial conditions. That is, a number of additional criteria of optimality are required.

This section considers each of these above requirements and describes the manner in which the models specify the method.

The models are grouped into queuing models, which employ queuing techniques to approximate the expected response time, and analytical models using linear programming and semi-Markov process approaches.

The queuing models can be split further into those using simulation to obtain the expected response time, and those using analytical approximation methods to determine the mean response time.

4.1 Assumptions Made

The assumptions made depend on the type of model chosen. In most cases the assumptions have been tested either by testing the data or by verifying the model at the end of the analysis, using comparisons with actual data.

The three assumptions that the probability of a call arriving when all ambulances are busy, that all calls from an area originate from the centre of that area and that all demand points are potential supply points, mentioned in Section 3, are common to most models.

4.1.1 Queuing Models

(a) Fitzsimmons (1), (2), (3), Savas (4)

The main assumption is that the system can be modelled as a queuing process, for which it is required that the waiting time for an ambulance to become available is essentially zero (Parzen (15)). Fitzsimmons uses a nearest ambulance dispatch rule, and Savas similarly uses nearest ambulance in a descriptive study. Fitzsimmons assumes that distance between 'areas' can be calculated as a rectangular displacement. His model is a partial simulation model.

(b) Swoveland (9), Volz (13)

Swoveland et al. assume that all calls are answered from depots, and uses two dispatch rules, nearest ambulance and regionalised response. They also propose a 'stability hypothesis' in which it is assumed that the fraction of total calls arising per day at node i and serviced by the qth closest ambulance is independent of which 'area' in the 'region' the ambulance is stationed for all i and k. (This is used for the optimising section only.) The model is a complete simulation model.

Volz uses a nearest ambulance dispatch rule, and assumes an instant relocation of ambulances corresponding to each change in the system state. That is, the moment an ambulance is dispatched, the remaining ambulances instantly relocate themselves optimally. This is justified by determining that the driving times for relocation are small relative to the average time between calls. He also assumes that the factors affecting driving time, the location from which the call comes, and the number of ambulances in service at that time are independent of each other. The model is an analytical queuing model.

4.1.2 Analytic Models

Toregas et al. (7), Hall (12)

Toregas et al. (7) use linear programming to minimise the number of facilities in the form of ambulance depots are required. They make only one specific assumption and that is, that the minimum distance, minimum times between every node and service facility pair are known, and this is easily calculated. This model does not fall into the general iterative procedure and nor is it strictly relevant to the problem of increasing the effectiveness of the ambulance service. A linear programming cutting plane algorithm is used to minimise the number of facilities required subject to an upper limit constraint on response time. It is a set covering problem which determines the minimum number of facilities required, provided it can be assumed that each depot is able to provide an ambulance, at all times. This is an unrealistic assumption in the context of our study, therefore, the model will not be considered further for the present.

Hall (12) assumes vehicles operate only within their fixed boundaries (precincts) and by means of a semi Markovian determination of steady state probabilities obtains the number of vehicles required to serve that precinct, from a central point. It is considered that an extension of a Markovian process to the complete system under study would not be satisfactory, since one of the basic assumptions about a Markov process is that the current state is independent of the previous state for all states, and this is not practical when considering ambulances which are able to cross boundaries. Therefore, this study will also be ignored for the remainder of this section.

The model used by Schneider and Symons (17) is excluded from further investigation, as the hardware requirements make it impractical for our use and also because it is an exploratory model and intended for experimental use.

The Larson model (18) is excluded because of its generality. Nevertheless it is discussed later and recommended for further study.

4.2 Calculation of the Mean Response Time

Queuing Models

The arrival of calls is considered as a Poisson process, and the state of the system is given as the number of ambulances busy at that particular time.

(a) Fitzsimmons (1), (2), (3), calculates the mean response time for each system state greater than one, using simulation. For system states equal to one and zero the mean response time is determined by use of the probabilities of a call arising from each area in turn using a priority list. The system mean response time is calculated from the state mean response times as a weighted sum, and then iterated three times, in order to stabilise it.

Savas (4) simulates a series of calls from the demand patterns given and obtains a mean response time from this simulation. Because he is considering only one ambulance depot it is a simple matter to calculate the mean response time in this manner, as there is little computation involved in the decision as to which ambulance to send. (Note that several ambulances may be stationed at the depot.)

(b) Swoveland et al. (9) use simulation to compute the mean response time. For the optimisation they use an analytical approximation to the simulated expected response time as a function of the locations of the ambulances, which is justified by use of the 'stability hypothesis'.

Volz (13) also uses an analytical approximation to the expected response time, but in his model this is complicated by the introduction of a complex driving time model designed for use in an area where roads may vary considerably. It takes into account weather conditions, traffic conditions and other factors (but in its present form requires the use of a rectangular area, for ease of computation).

4.3 Amendment of the Feasible Location Set so as to Improve the Mean Response Time

Queuing Models

(a) Fitzsimmons uses the Hooke and Jeeves pattern search technique (5) to determine a 'better direction' of movement for all locations. Notice that the set of the number of locations is fixed. Locations are not 'dropped' or 'picked up'. The only way of obtaining solutions with alternative numbers of locations is by running the data several times with different numbers specified for each run.

Savas only considers the principle of location in general terms, and the situation of one depot and one hospital. It was determined by simulation of 'all possible locations' that the best placing for the depot was at the centre of demand.

(b) Swoveland again uses a technique involving a fixed number of locations. He uses a Branch and Bound procedure to determine in which 'area' of each 'region' the station should be placed. A lot of preliminary work is required to determine the optimal groupings of 'areas' into 'regions' and several different solutions might be obtained using different such groupings.

Volz minimises the mean response time using a discrete version of 'steepest descent' (14). There are a large number of possible local optima which may be obtained by this method and a number of runs using different starting points should be made.

4.4 Evaluation of the Solution

The method of evaluation of the solution is directly dependent on the method used to optimise, (minimise the mean response time). Because of the nature of the problem's objective function there will be a large number of local optima available, and both the Hooke and Jeeves pattern search technique used by Fitzsimmons and the method of steepest descent used by Volz will terminate on a local optimum. Of these two methods it seems that the former may avoid some of the more trivial local optima which will be found by the method of steepest descent, such as the clustering around intersections which was noted by Volz. However, it is likely that both methods produce different results when different starting points are used, and when parameters are varied slightly. Therefore, it is necessary to select a further set of parameters which may be studied in order to determine the 'best' solution. The parameters which may be used for this form of subjective evaluation of alternative solutions have already been mentioned in Section 1 and are:

(a) average round trip

(b) standard deviation of the mean response time

(c) number and percentage of responses to calls exceeding a certain fixed 'maximum'

(d) percentage of calls handled by each ambulance

(e) probability of a response time greater than a set of given times, e.g. 10, 15, 20 mins.

(f) probability associated with each state, or number of busy ambulances.

5.0 CHOICE OF A MODEL

Of the models studied thus far, the four which are relevant are those constructed by Fitzsimmons, Swoveland et al., Volz and Toregas et al. The Savas, Gordon & Zelin models may be omitted as the studies were principally concerned with gaining information about hospital and ambulance satellite stations, and with whether or not the use of satellite stations did reduce the mean response time, rather than with the optimisation of the actual location of the satellite stations. Hall may be omitted as his study is concerned with a dual function system, police vehicle and ambulance, and also a system where ambulances are restricted to certain areas and not able to cross boundaries. Schneider and Symons, and Lawson are excluded for reasons given at the end of 4.1.2.

Therefore, in evaluating the four models, two system characteristics will be required:

(a) that the mean response time be the primary measure of the behaviour of the system, or that the major objective will be to minimise mean response time.

(b) that the dispatch rule used will be that of nearest ambulance, which is the most realistic.

5.1 Advantages of the Models

Fitzsimmons: the model is simple, practically convenient as far as computing time and core space required are concerned, and available. It has a number of special features which may be used for experimentation allowing for such things as helicopter recovery, two types of retrieval policy (emergency and subemergency), relocation of ambulances at optimal locations corresponding to each change in the system state, (as in Volz's model) and the use of mixed types of ambulances. It is very tightly specified, and uses a good optimisation technique, recognised as a swiftly terminating method (14).

Swoveland et al.: two dispatch rules are available and a priority rating may be given to calls so that normal, transfer and emergency functions may be performed. Calls may be cancelled on the way, or at the scene of the call, and the analytical approximation to the mean response time is related to the complete simulation. Times and distances between 'areas' are entered for computation of the distances between nodes and units are able to respond at any time after they release their patient at a hospital or destination. Each may have up to 20 (different) possible destinations to which a probability is attached. If required each unit may be routed by an organised path, different from the shortest distance which would be the default. Distributions for loading time, unloading time and cancellation time are entered, allowing greater flexibility than the mean times used by Fitzsimmons.

<u>Volz</u>: the analytical approximation to the mean response time is again a function of a large number of variables which are relevant including weather conditions and time of day. This is because the model is a sophisticated point to point driving time one (as against the simple rectangular displacement used by Fitzsimmons), and a constraint is added so that the mean response time to any point in the service area will be less than a given maximum.

<u>Toregas et al.</u>: the model is simple and easily computed.

5.2 Disadvantages of the Models

<u>Fitzsimmons</u>: the dynamic aspects of the problem are not explicitly covered, in that a mean number of calls per day is taken from a year's data and used as the parameter for the Poisson call arrival distribution. This may be investigated by the use of different parameters calculated for different times of the year, and the comparison of results.

This is a disadvantage common to all models, due primarily to the cost in computing time, and core space involved in the programming of a possible change in parameters into the system. Volz considered the possibility of extension but was unable to attempt it because of the small size of the computer being used.

Fitzsimmons has programmed two ambulance speeds into the CALL system allowing a linear interpolation between them which proved suitable for the systems with which he was working, but which may not be generally suitable. One disadvantage which all models using a grid system and a rectangular displacement as an approximation to distance travelled have, is that this form of distance calculation simply may be not applicable. The geographical 'shape' of the Auckland area for example, is not suited to rectangular displacement, and a model which works with times between demand areas, is far more suitable. The Fitzsimmons model has been adapted to accept as input the time taken to get from each node to all other nodes. This is more satisfactory in some ways, but it means that the definition of the model is reduced, as placement must now be always at the centre of the 'area' defined. Using rectangular displacement, placement at different points within the 'area' is obtained. Also, the step size of PATS (the search routine) must be altered, as movement within an 'area' will not have any effect on the calculation of the system mean response, and at this stage, it is uncertain how this should be done.

A major disadvantage of Fitzsimmons model is that it is not possible to adapt the model to deal with transfer data and normal calls, mainly because of the necessary assumption that there is always one ambulance available to answer a call. Although the present study is primarily involved with emergency calls, a model which is able to be extended to represent the complete system is at least desirable.

Swoveland et al.: the main disadvantage with the Swoveland et al. complete simulation model is the amount of time required to compute the response time for each set of given locations.

Two forms of the model are available. One is simply a simulation model, which calculates the mean response time for a given set of locations, and for a given number of simulated calls. Further runs using slightly ammended 'temporary' positions of the ambulances may be made, but these positions must be entered by card. There is no linked optimisation. The second form of the model links the computation of the mean response time to a probabilistic Branch and Bound (integer programming) optimisation technique. This form of the model was not made available for study but it is known that an assumption is made concerning the simulated response time for different allocations within the regions. This substantially reduces the size of the problem, however, it may also reduce its accuracy, under different circumstances than that for which it was tested. Also, this form of the model must be used in conjunction with a regionalised response rule. Regions are able to overlap, which is usual, but this does not substantially reduce the disadvantage of regionalised type response.

The critical disadvantage with this model is therefore that the computer time required to perform the complete simulation (B6700 15 seconds CPU time) is too long to be practical for use with an optimisation technique with a possible number of iterations between 100 and 200. Thus, an analytical approximation has had to be used.

At present, the model does not contain the facility of a constraint on the maximum expected response time for each area, or a fixed maximum response time for the system. Swoveland does consider the possibility of adding such a constraint, however, and it would be a fairly simple matter to program this into the model. It is probably easier to think of the maximum response time as an evaluation parameter.

<u>Volz</u>: it is assumed that when an ambulance goes into service, the remaining ambulances are instantly relocated at the optimum points for that particular state of the system. This type of relocation is used in practice, therefore, it may be considered an advantage or a disadvantage, depending on the attitude of the authorities in the area under study.

The method of steepest descent will probably find a larger number of trivial local optima than the Hooke and Jeeves pattern search, and for a unimodal function it has been shown that a larger number of iterations (considerably larger) is required to find the optimum (Wilde (14)). Volz has pointed out that with an IBM 360 computer any number of ambulances beyond six becomes infeasible because of the storage limitations. It is unlikely that the Burroughs B6700 would be any more satisfactory, without changes to the programs, though Volz does suggest that newer, faster computers might get around these problems. He is willing to make his programs available if required, but pointed out that they are written partly in IBM 360 Assembler, which would have to be rewritten. Again, the driving time model is computed using a rectangular grid (for simplicity). The model is amenable to alteration, though it would probably increase the computations considerably.

Toregas et al.: the problem formulated is that of minimising the number of facilities required subject to all calls being answered, within a fixed maximum time period. It is not concerned with minimising the mean response time and therefore should really be excluded from the models under consideration. It does not take into account the possibility of an ambulance being busy, that is, it assumes that all facilities are able to supply ambulances whenever called. However, it is interesting as it is one of the few studies to consider the problem of specifying the number of facilities, rather than locating a given number. (See also Fitzsimmons thesis (16).)

It should be stressed that one of the main disadvantages of all the models is that they do not consider specifically the dynamic nature of the problem. Fitzsimmons uses average number of calls per day as his input parameter, Swoveland uses a rate per hour computed similarly from aggregated data. Toregas does not use demand data to generate an input call stream, and Volz also uses a daily average figure, though he recognises that it might be preferable to consider smaller time periods. (e.g. peak time demand only, from data taken over several weeks or months.) Thus, all the examples have assumed that the distribution of demand will be static with no seasonal or other cyclic variation, an assumption which does not seem intuitively realistic.

5.3 Selection of the Model

It is immediately apparent that three of the models which have been discussed are suitable for implementation in a New Zealand situation for the purpose of determining the optimal location of ambulances. They are those derived by Fitzsimmons, Swoveland et al., and Volz. They were all approached, in order to obtain further information and to determine possible access to the programs.

Volz's model was discarded initially, for two reasons. Firstly, although it is the only model of the three using a purely analytical calculation for the mean response time, the amount of computer space and time required for a system with more than six ambulances becomes excessive with an exponential increase for each additional unit. The B6700 may in fact have been able to cope with this, however, the second reason for rejecting this model was that a number of the routines are written in IBM 360 Assembler language which would have to be translated. This could be expected to take some considerable time (possibly two to three months) and detailed expertise which was not available.

Also, the method of steepest descent used in the optimising section has been shown to be inferior to the Hooke and Jeeves pattern search (at least under some circumstances) (14). However, had Volz's method been pursued, then it is probable that the analytical response time calculation could have been linked to the Hooke & Jeeves pattern search.

For all practical purposes, the Fitzsimmons model appeared to be the best. It is clearly and tightly defined, and uses an efficient means of optimisation. The assumptions are realistic, and the computing time involved for reasonable sized problems is small enough to allow for a considerable amount of experimentation with variation of the parameter. Therefore, a copy of the program was obtained and it was adapted for use on the B6700.

Despite the obvious usefulness of the Fitzsimmons model, it was decided to obtain further information about the Swoveland et al. model. In the paper in which it was presented (9), it was not particularly well specified. A number of features were mentioned, but it was not made clear whether these were actually incorporated in the model, or whether they were simply proposed extensions. Three features, which do exist are that queues are allowed to develop, patients may go to different hospitals and ambulances do not necessarily respond from a base or hospital, but may respond whilst returning to their base.

In the event, the simulation part of the model proved to cover a very wide range of possible situations, and produce a good full report. It also releases some of the previously required assumptions. It was therefore decided to use the simulation model to test the solutions obtained using the Fitzsimmons. The second part of the model (the optimisation) was not available and therefore it was also decided to attempt to link the simulation to PATS the Hooke and Jeeves pattern search subroutine. The problems attached to this are discussed in Appendix IV.

Both the Fitzsimmons model and the Swoveland et al. simulation model suffer from the same defect which is that the dynamic nature of the situation is ignored. Nevertheless by performing different runs for different rates and numbers of calls representing different times of the day, and different times of the year, some testing of the solution under dynamic conditions may be performed.

6.0 THE CHRISTCHURCH STUDY

6.1 The Christchurch System

The Christchurch St. John Ambulance service covers the whole of the Canterbury area from Kaikoura to Rakaia, and inland to Arthur's Pass. The Hospital Boards concerned virtually 'let the contract' for the operation of the ambulance service to the St. John Ambulance Association.

The service is maintained by means of fifteen ambulances, of which eight are based in Christchurch City, two in Rangiora, one in Cheviot, one in Kaikoura, one in Amuiri and two in Lyttelton. The regional ambulances all have primary responsibility in their own areas, and continuous radio contact is kept between all ambulances and the Christchurch control room.

The Christchurch City region may be defined by boundaries reaching from Tai Tapu to Christchurch Airport to Kaiapoi to Sumner. Details of road accidents occurring in this area were obtained from the Traffic Engineering Section of the Ministry of Transport in Christchurch.

There are certain specific features of the service in Christchurch which should be mentioned. This information was obtained from Mr G. Whitacker, the Station Manager in Christchurch.

At present the eight Christchurch ambulances operate from a central depot in Peterborough Street. All of these ambulances are used for both general service work and accident and emergency work. During the daytime, most of the ambulances are continuously involved in transporting patients from one hospital to another, or from hospital to home, and other general service work. Since the hospitals, public and private in Christchurch are well spread around the service area, (see Figure 1) a form of depot location is already thus established. It is a common occurrence for an ambulance to be diverted from a service function to answer an emergency call in the vicinity of its projected travel path.

Another feature of the Christchurch service which is apparently unique at present, is a provision for immediate skilled medical aid at the scene of the accident. A number of Christchurch doctors are involved in a scheme which sends a doctor to the scene of every major emergency. The doctor to be sent is selected from a list in a similar manner to the ambulance (i.e. for his proximity to the accident). It is therefore not uncommon for the doctor to arrive before the ambulance.

This means that the average on scene care time in Christchurch is longer than in other centres. In the latter the patient is simply transferred directly to the ambulance. In Christchurch quite considerable aid may be given before the patient is sent to hospital with a note from the doctor stating what medication has been administered.

Sirens are seldom used for retrieval in Christchurch because of the nervous system damage which may be effected causing movement of the patient, and compounding of the injuries. Also sirens may cause inattention among other drivers, and further incidents.

One other feature of the Christchurch service is that 98% of accident victims are taken to hospital. They may then be immediately treated and discharged, but in all cases of suspected concussion (very common with road accident victims) the patient is kept under observation for a time.

There is only one accident and emergency hospital, namely the Christchurch Central Hospital. However, it is likely that certain severe emergency cases near Princess Margaret Hospital would be taken there. However, all the other hospitals should be taken into account, because of their involvement with service runs.

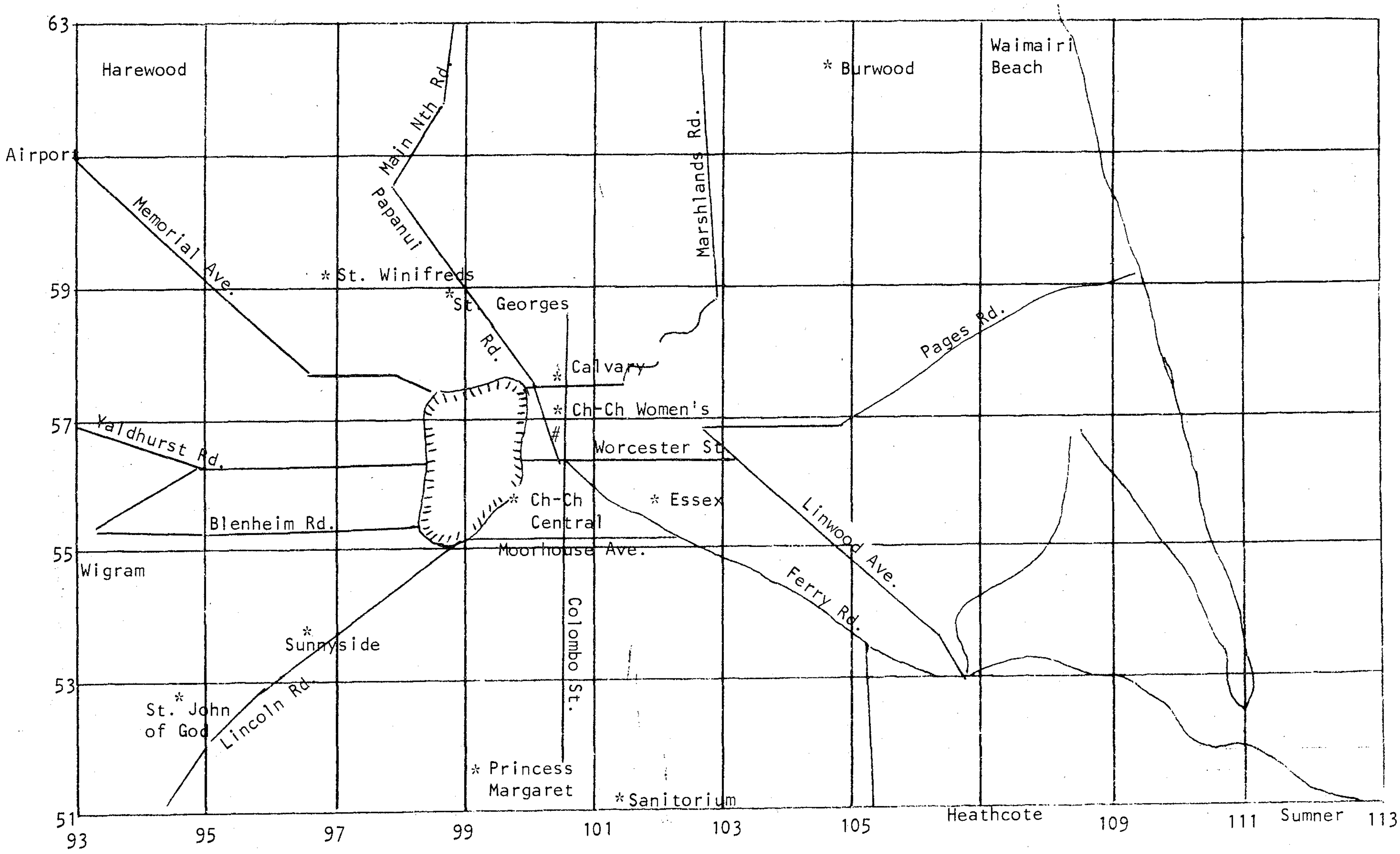
The total number of patients carried in 1973 was in the order of 30,000. Of this, 3,500 were the result of accidents and emergencies, including sporting accidents.

6.2 Preparation of the Data

A number of general assumptions are made about the ambulance system in both the models discussed. Some tests of these assumptions

FIGURE 1 : Hospital & Depot Locations

Hospital Locations *
Peterborough St. Depot #



should be made. A number of tests are suggested here, and examples of their application to the Christchurch study are given. The data requirements for both models (Fitzsimmons CALL, Swoveland et al.) are discussed.

6.2.1 Call Times

(a) The first piece of information required about the call times, is the mean number of calls. This has been purposely left as a vague statement because although both models require a single figure for this, either the mean calls per day (Fitzsimmons) or the rate of calls per hour (Swoveland et al.) it is unlikely that these parameters will provide sufficient information about the system.

For the purposes of predicting call times, both models assume that the intercall times will follow a Poisson distribution. This may be shown to be a valid assumption. However, it is extremely unlikely that the mean number of calls parameter will remain constant. It will be affected by time of day, weather conditions, time of year and various institutional factors such as Public Holidays.

For the Christchurch study, it was not possible to obtain information about calls on an hourly or daily basis. Therefore, the mean number of calls per day was deduced from the 1973 year total, as was the mean rate of calls per hour. Because of the scarcity of information, this value was varied, to test the sensitivity of the model. In most cases, LAMDA was set to 82, and RATE was set to 3.5.

However, in order to test the dynamic sensitivity of the solution, if possible, data should be obtained on an hourly basis, and the mean number of calls should be obtained for:

(i) peak times
(ii) off-peak times
(iii) summer

(iv) winter
(v) hours of darkness
(vi) hours of daylight
and any other period which appears likely to have specific significance.

(b) The assumptions made for the Fitzsimmons model are stronger than those of the Swoveland et al. model. One of these is that the probability of a call having to wait for an available ambulance is zero. That is, that no queue forms. The validity of this assumption may be determined from the output of CALL where in the System Performance table the probability of all ambulances being busy is printed out. This should be a suitably low number. It has been suggested that in order for the probability to be zero (or very close to it) the utilisation of ambulances should be approximately 30%. This can also be checked in the output of CALL and several runs were made, investigating this effect.

(c) For the CALL assumption that the service time is independent of the state of the system to be acceptable, the response time must be much smaller than the total service time. The output of CALL in all cases proved satisfactory in this circumstance, mainly because of the long mean time on the scene of accident (CARE = 20 minutes) and the mean transference time at hospital (TRSFR = 10 minutes), compared with an average system mean response time of around 5 minutes.

Notice that for CALL, the time at the scene of the accident and time transfer time at hospital are both constants. For the simulation model, these two parameters are treated as variables, and distributions of loading and unloading time must be entered, as well as a cancellation time distribution which is not considered in CALL.

Data to compile these distributions was not available for the Christchurch study, and the data used were contrived distributions with means equal to those in CALL. The cancellation time distribution was not used and the probability of a cancelled call was put to zero. The probability of an emergency call was put to 1.0, as CALL is able only to deal with emergency calls and it was attempted to use the simulation to represent the same system.

6.2.2 Call Origins

(a) In order to investigate the call origins, it is first necessary to divide the city, or area under consideration by means of a grid. It is not necessary in either case for the grid area obtained to be identical in size. The availability of accident or incident data will have a good deal to do with the grid pattern used. The main point to be concerned with in the selections of a grid area is that the centroid (for CALL) or the selected intersection or node (for the simulation) should be able to be considered representative of the whole grid area. The reason for this is that the probability of an accident occurring in the area will be applicable to the whole area.

For the Christchurch study, the availability of data was the main problem. It was not possible to obtain total work data, which includes transfer calls, normal calls and emergency calls and in any case CALL is not designed for use with transfer calls. Normal calls and emergency calls may be split in the further categories of road accidents, home accidents, sporting accidents plus some other types of calls which might also be considered transfer calls.

There may be some overlap here. However, the only data which was available with origin information for Christchurch was road accident data, which is held by the Ministry of Transport. With this even, there is no guarantee that an ambulance was actually sent to the accident, and in some of the minor cases, it is unlikely that one was sent. However, each of these accidents may be treated as a potential accident call. The total number of these 'accidents' was in the vicinity of 3,000. Because of the absence of further information it was decided to treat the area distribution of the accidents as representative of the area distribution of the total calls.

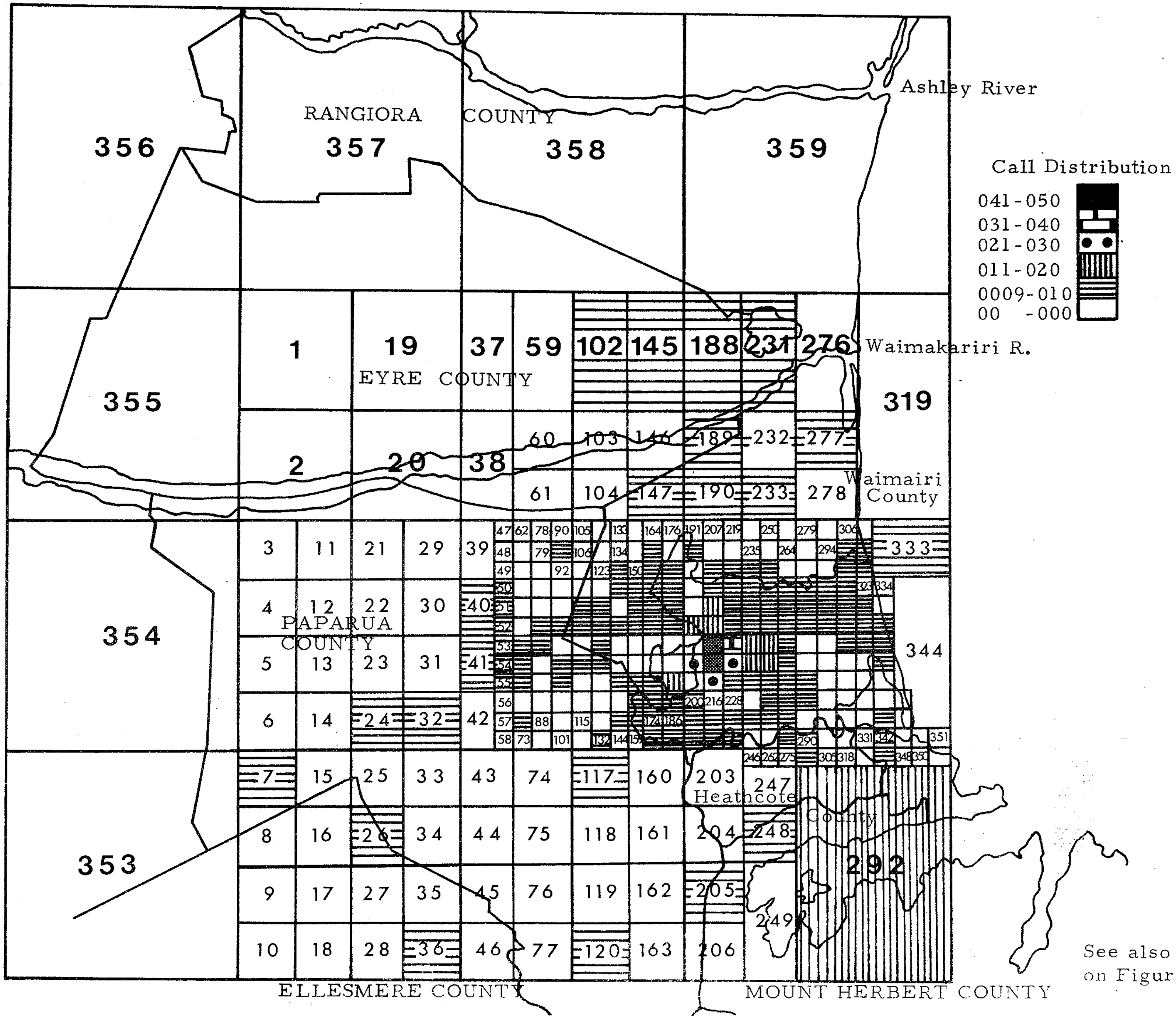
The total number of calls answered in 1973 was approximately 30,000. This concerned the whole Christchurch area. Therefore, as mentioned above LAMDA was set to 30,000/365 = 82 calls per day. This may be compared with the actual mean number of accidents per day, which is closer to 8, (consider too that there are eight ambulances in Christchurch).

The quality of the data therefore is poor, but as may be seen in section 6.3 (Results) it served as a vehicle for useful investigation of the properties of the models.

Some time was taken in selecting the grid and the total area to be considered. Originally, the calls were placed in a grid which contained 359 areas of varying sizes. This proved very unwieldy, and also, the probabilities associated with each area were very small. A second attempt reduced the area covered by removing Ellesmere County, most of Paparua and Eyre and Rangiora Counties. Amalgamation of the small city 'areas' reduced the number to 82. Some of these runs are shown in the results. A further reduction brought the number of areas down to 60, all of regular shape. Figures 2, 3 and 4 show these different areas.

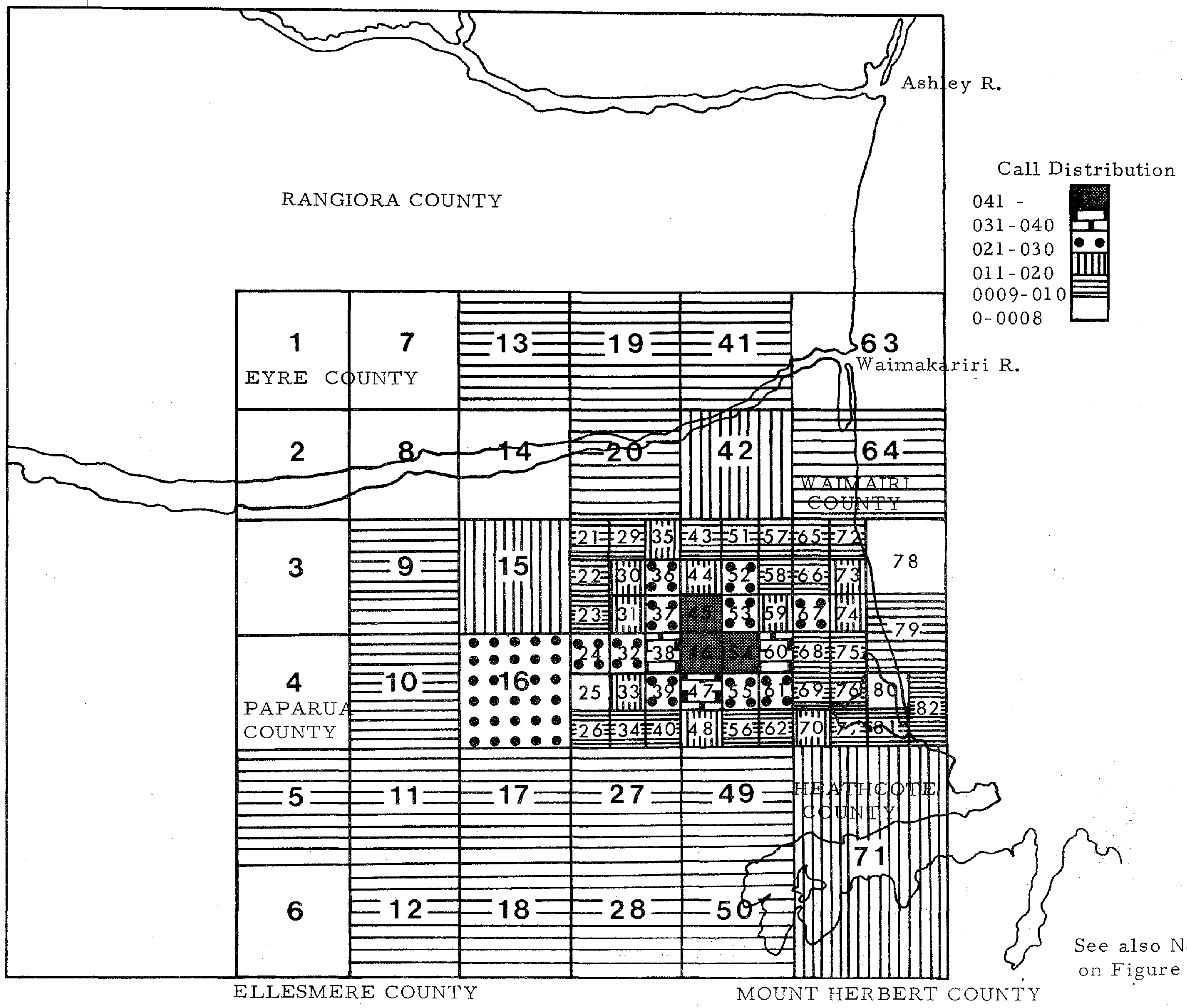
It is likely that the distribution of accidents will vary with time. Some attempt was made to test this for the Christchurch data, but the small number of accidents made the results meaningless.

(b) The second piece of information required concerning the call origins, is the distance travelled or time taken to travel between areas. For CALL, the



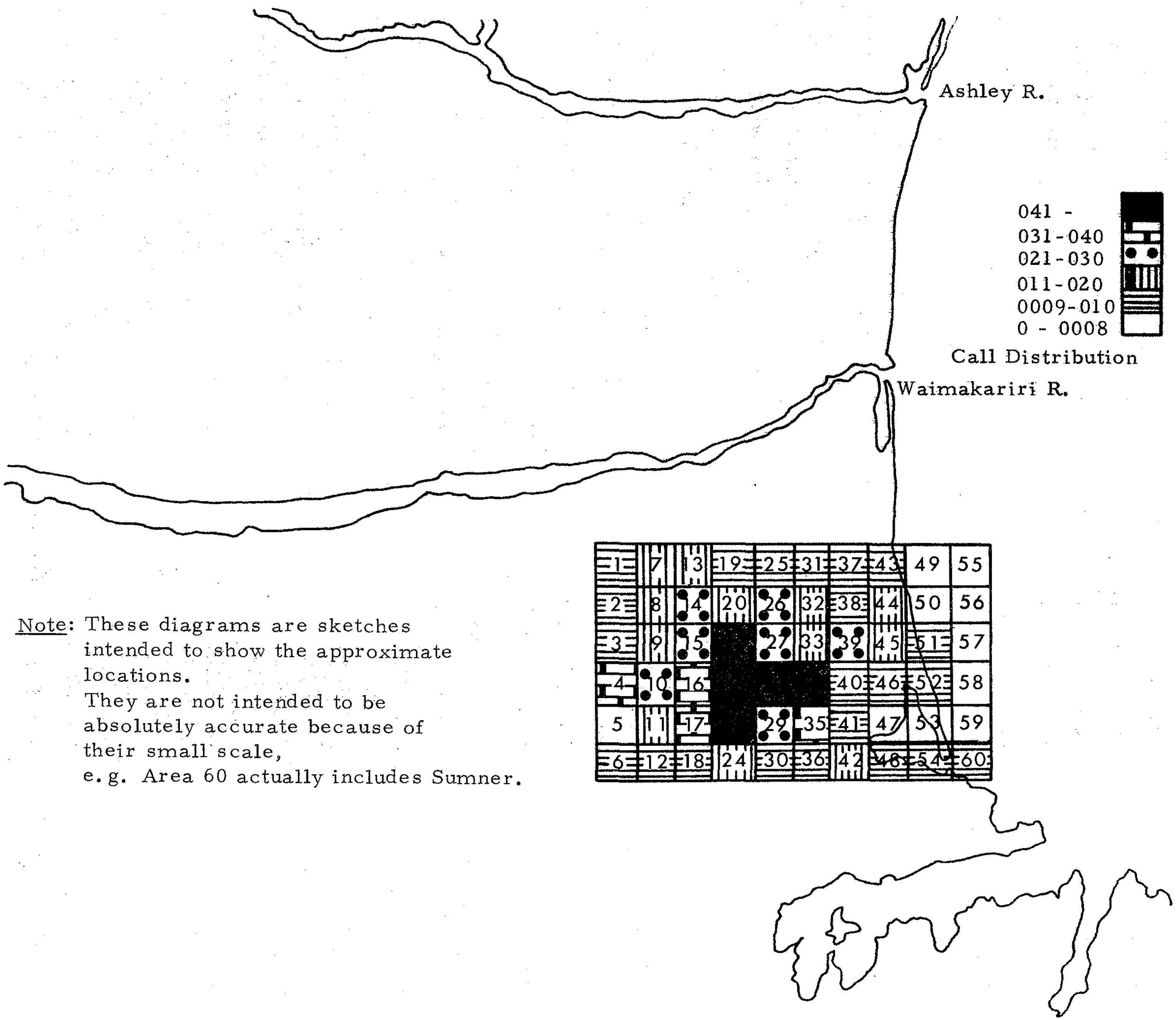
See also Note on Figure 4.

FIGURE 2 - Location of 359 Call Areas



See also Note on Figure 4.

FIGURE 3 - Location of 82 Call Areas



<u>Note</u>: These diagrams are sketches intended to show the approximate locations.
They are not intended to be absolutely accurate because of their small scale,
e.g. Area 60 actually includes Sumner.

FIGURE 4 - <u>Location of 60 Call Areas</u>

time is deduced from the rectangular displacement as a measure of distance.

Therefore, the required input consists of the city speed, the freeway speed, a maximum distance for the former and a minimum distance for the latter. The speed for the area between the two is a linear function of the distances and speeds. For Christchurch, speeds were set fairly arbitrarily as 25 m.p.h. for city speed CSPD and 45 m.p.h. for highway speed HSPD. These speeds were the same for all ambulances, as they are identical. The maximum city distance CITY was set as 27 and the minimum highway distance FWY to 40. From these values the time taken from one point to another is calculated. Because of this form of calculation, it is possible to locate the ambulances at any point with an area, and not necessarily at the centroid. Similarly, hospitals may be placed at any point within an area. A call is always considered to originate from the centroid of the area. As was noted in Appendix I this form of location is very convenient for use with the pattern search optimisation.

For the simulation model the times between each node are required. The definition of 'areas' does not need to be as rigid as for CALL, and therefore it is the points between which travel times are measured that may be considered as 'nodes' rather than centroids of areas. The boundaries of the areas do not need to be defined except for the calculation of the probabilities of calls. It may be necessary to use an approximate calculation based on distance to compute the travel times, but it is preferable to obtain these times by experimentation. Again, it is likely that there will be some variation of travel times according to peak hour traffic and off-peak hour traffic.

For the Christchurch study, as the simulation model was being used for comparison purposes with CALL, travel times were computed in exactly the same way as for CALL. No attempt to check the variation in travel times was made.

(c) In the simulation model it is possible to route ambulances back to their base from a hospital, node by node, so that when a call is received the current location of these ambulances may be checked to see if they are closer to the call than any of the 'idle' ambulances. Because CALL allows only response from bases or hospitals this facility was not used.

(d) Another facility of the simulation model is that each node may have a different number of allowable destinations, each of which is assigned a probability. This is a very useful facility particularly in a situation where different hospitals maintain different emergency units and functions. It is of particular use for the simulation of transfer functions in a situation such as Christchurch where ambulances perform transfer functions for 10 or more public and private hospitals and yet use only one hospital for emergency purposes.

In this study, since CALL is solely concerned with emergency use, one destination is allowed to each node, with a probability of 1.0.

6.2.3 System Data

Other system information is required for use with both programs.

For CALL this includes:

TRANS, the fraction of calls transported, and this was set to 0.98, meaning that 98% of all calls are transported to hospital. This is a large percentage, but consistent with the Christchurch policies.

LIMIT, the maximum number of search evaluations was set to 200. In most cases this was satisfactory, in one or two instances it was extended.

NAMB, the number of ambulances was set to 4, for all but a few runs where it was set to 5 or 8. There were two reasons for this. Firstly, although there are eight ambulances in the Christchurch fleet, the area under study is only a small part of the total area serviced, even though most of the calls do originate from it. Transfer calls were not being considered directly either. Secondly, the two runs in which NAMB was set to eight used such a large amount of computing time that it did not seem reasonable to perform many such runs. As this was intended as a test of the model, it was considered desirable to perform a large number of runs.

NHOSP, the number of hospitals was set to one in most circumstances and to two for some runs.

NFRAC, the number of one minute class intervals for the response time distribution was set to 40. This was an arbitrary selection.

For the Swoveland et al. simulation model all the data requirements have been mentioned.

6.3 Results

6.3.1 Fitzsimmons Model : CALL

A number of runs were made to test the parameters of the model, and also to test some of the assumptions of the model. A full table of these results is presented as Table 1.

In order to evaluate the results, each test is presented, together with its results, compared to a standard result. Run 1 and Run 3 are used as standard results. Comments are made on each test.

Tables 2 to 11 and Figures 5 to 13 show the results of these tests which are grouped into those performed on:

System parameters
Program parameters

Model assumptions

The tests are detailed as follows:

On System Parameters:

(a) A test on the effect of changing the number of areas (Table 2, Figure 5).

(b) A test on the effect of changing the initial positions of the ambulances (Table 3, Figure 6).

(c) A test on the effect of changing the number of hospitals (Table 4, Figure 7).

(d) A test on the effect of changing the number of calls LAMDA (Table 5, Figure 8).

(e) A test on the effect of restricting the final positions of ambulances (Table 6, Figure 9).

(f) A test on the effect of changing the number of ambulances (Table 7, Figure 10).

On Program Parameters:

(a) A test on the effect of changing the number of simulations runs made for the calculation of mean response time for states greater than 1. (Table 8, Figure 11.)

(b) A test on the effect of changing the seed of the random number generator (Table 9, Figure 12).

(c) A test on the effect of changing the number of times RBAR, the system mean response time is calculated in order to wash out the effect of the arbitrary initial setting of RBAR (Table 10).

On Model Assumptions:

(a) A test on the effect of changing the number of ambulances used. This test is related to the assumption that queues do not form, and the related assumption that an average ambulance utilisation of 30 per cent will prevent queues forming in most cases (Table 7, Figure 10).

(b) A test on the effect of changing the number of calls. This time LAMDA is changed in order to evaluate the effect upon the ambulance utilisation (Table 11, Figure 13).

(c) A test on the step reduction feature of the Hooke and Jeeves Pattern Search.

Some runs were originally made with 359 areas. However, none of them terminated because of the length of computer time required. Also, as can be seen in Figure 2, most of the outer areas had a very low probability of an accident.

Run 1 was made with 82 areas (see Figure 3) and Run 3 was made with 60 areas (see Figure 4). The other parameter for these two runs was identical (Table 1).

Table 2 shows the compared results of runs 1 and 3, where the initial positions of the ambulances were both at the corners of the system. There is an expected reduction in the mean response time at all levels. However, from Figure 5 it can be seen that the final locations for the ambulances are virtually identical. Therefore it was decided to use 60 areas, for most of the runs, thus restricting the problem to the city area.

TABLE 1

RESULTS of CALL

Initial Placements of the Ambulances at Corners of the Service Area (unless otherwise indicated)		LAMDA no. calls/ day	NAMB no. ambulances	NHOSP no. hospitals	NDIST no. districts	NEVAL	% of cases taken	Mean response final	x co-ord. final location	y co-ord. final location	% calls taken final	% utilisation final	Prob. of each state	Mean response each state	System mean Response	System mean time to Hospital	System mean Retrieval	System mean Service	Mean no. in System	Maximum Response Time
	1	82	4	1	82	157	100.00	8.67	100.0	56.0	21.09	51.88	.0951	5.361						
	2							9.81	97.7	56.3	28.02	70.73	.2238	5.917	6.805	41.715	9.910	41.316	2.353	36.46
Run 1:	3							9.50	102.0	56.0	22.97	57.57	.2632	6.208	(16.269)	(51.179)		(50.771)	(2.891)	(42.42)
	4							9.65	101.0	58.0	27.92	70.19	.2064	8.042						
	5												.2114	7.928						
	1	82	4	1	82	149	100.00	10.08	102.5	55.5	24.87	63.15	.0938	5.659						
Run 2:	2							9.74	97.7	56.1	26.46	66.67	.2220	6.267	7.042	41.952	9.910	41.553	2.366	36.73
Initial positions	3							9.25	102.2	57.2	28.46	70.92	.2627	6.563	(9.480)	(44.390)		(43.990)	(2.505)	(37.20)
Ch-Ch, P.M.H., Burwood	4							9.10	100.2	56.0	20.21	50.20	.2072	8.191						
Hospitals & Depot	5												.2142	7.928						
	1	82	4	1	60	140	100.00	6.58	100.0	56.0	22.76	50.55	.1175	3.694						
	2							8.20	97.7	56.1	26.98	62.41	.2516	4.237	5.183	37.958	7.776	37.601	2.141	27.54
Run 3:	3							7.20	102.0	56.0	23.56	53.17	.2694	5.054	(15.990)	(48.765)		(48.402)	(2.756)	(35.97)
	4							7.71	101.0	58.0	26.67	60.95	.1923	6.597						
	5												.1692	6.219						
		82	4	1	60	17	100.00	6.57	100.0	56.0	22.07	49.02	.1149	3.808						
Run 4:								8.23	97.7	56.1	27.15	62.84	.2486	4.457	5.579	38.355	7.776	37.998	2.164	27.55
Initial locations -								6.95	102.0	56.0	22.06	49.46	.2690	5.576	(5.109)	(37.885)		(37.528)	(2.137)	(27.55)
final locations for								9.08	101.0	58.0	28.70	67.83	.1940	7.499						
Run 3													.1736	6.219						
		82	4	1	60	131	100.00	7.59	102.3	56.2	22.53	51.33	.1161	3.956						
Run 5:								7.14	100.6	55.4	22.64	51.01	.2500	4.494	5.396	38.172	7.776	37.816	2.153	28.05
Initial locations:								8.28	97.8	55.6	25.64	59.42	.2692	5.283	(6.177)	(38.952)		(38.597)	(2.198)	(22.91)
depot								7.96	100.2	58.4	29.17	67.07	.1932	6.857						
													.1716	6.219						
		82	4	2	60	140	93.87	6.57	100.0	56.0	22.92	50.47	.1194	3.694						
							6.11	8.17	97.7	56.1	26.83	61.53	.2538	4.237	5.221	37.668	7.447	37.318	2.125	27.54
Run 6:								7.18	102.0	56.0	23.64	52.88	.2697	5.054	(16.102)	(48.549)		(48.191)	(2.774)	(35.97)
								7.71	101.0	58.0	26.59	60.27	.1910	6.597						
													.1661	6.511						
		82	4	2	60	127	B'wood	7.93	102.3	56.2	25.57	63.57	.0886	4.265						
							30.20	6.54	101.6	56.2	22.81	54.91	.2148	4.789	6.834	42.991	11.157	42.551	2.423	26.90
Run 7:							P.M.H.	8.19	97.8	56.5	32.00	80.05	.2602	5.199	(7.510)	(43.666)		(43.238)	(2.462)	(22.91)
							69.78	7.44	100.2	56.3	19.60	48.19	.2102	6.547						
													.2261	11.934						
		100	4	2	60	140	93.87	6.93	100.0	56.0	18.26	49.51	.0734	3.694						
							6.11	8.76	97.7	56.1	31.31	88.83	.1916	4.237	5.523	37.970	7.447	37.619	2.612	27.54
Run 8:								7.97	102.0	56.0	21.39	59.52	.2503	5.054	(14.323)	(47.323)		(46.998)	(3.2624)	(35.97)
								7.83	101.0	58.0	29.01	80.45	.2180	6.597						
													.2667	6.511						
		82	5	2	60	196	93.87	8.02	97.7	57.9	19.63	44.84	.1229	3.416						
Run 9:							6.11	6.51	99.3	56.0	20.98	46.12	.2557	3.833	4.712	37.159	7.447	36.809	2.096	27.99
initial locations:								6.66	102.0	56.0	20.94	46.22	.2701	4.300	(15.911)	(48.358)		(48.009)	(2.734)	(35.97)
corners & Sunnyside								7.24	102.0	58.0	19.89	44.55	.1887	5.270						
													.0939	7.554						
								7.77	100.1	53.9	18.54	42.10	.0617	6.511						
		100	5	2	60	173	93.87	7.92	97.7	56.2	18.69	51.95	.0765	3.443						
Run 10:							6.11	6.92	100.3	56.0	20.21	54.77	.1950	3.939	5.118	37.565	7.447	37.213	2.584	28.72
Initial locations:								7.13	102.0	56.0	18.54	50.51	.2519	4.469	(15.612)	(48.059)		(47.696)	(3.312)	(35.97)
corners & Sunnyside								7.48	102.0	58.0	20.35	55.92	.2170	5.285						
													.1402	7.369						
								9.40	100.1	53.9	22.19	63.94	.1204	6.511						
		200	5	2	60	217	93.87	8.54	100.3	55.8	6.27	29.73	.0050	4.151						
						(opt.not	6.11	7.84	101.7	57.8	32.28	179.04	.0263	4.484	6.112	38.559	7.447	38.209	5.307	26.53
Run 11:						found)		9.13	101.5	56.3	16.45	94.21	.0698	4.709	(10.747)	(43.194)		(42.962)	(5.967)	(35.97)
Initial locations:								8.77	98.5	56.0	39.65	225.05	.1235	5.130						
corners & Sunnyside													.1638	6.283						
								6.86	100.0	56.5	6.34	34.28	.6116	6.511						
		82	4	1	60	99	100.00	7.15	100.0	55.7	21.88	49.31	.1164	3.833						
								7.10	100.0	58.0	25.41	57.19	.2503	4.549	5.349	38.125	7.776	37.768	2.151	24.95
Run 12:								7.13	102.0	56.0	24.69	55.59	.2692	5.106	(15.990)	(48.765)		(48.402)	(2.756)	(35.97)
								8.43	102.0	57.8	28.00	65.19	.1930	6.870						
													.1710	6.219						

Run 13:	82	4	1	60	140	100.00	13.90 8.53 6.99 7.08	107.0 97.7 101.7 99.7	59.0 55.8 56.0 56.0	32.58 25.10 21.48 26.82	85.92 58.54 48.21 46.84	.1114 .2445 .2683 .1962 .1795	3.728 4.379 5.656 10.180 6.219	6.118 (15.990)	38.894 (48.765)	7.776	38.534 (48.402)	2.194 (2.756)	27.95 (35.97)
Run 14:	82	4	1	60	196	100.00	10.07 7.48 7.09 8.15	101.4 102.3 99.2 99.3	52.7 56.5 56.0 58.0	30.18 22.19 21.31 26.31	73.03 50.41 47.93 60.78	.1141 .2477 .2688 .1945 .1749	3.997 4.653 5.566 7.750 6.219	5.700 (15.990)	38.476 (48.765)	7.776	38.119 (48.402)	2.171 (2.756)	27.13 (35.97)
Run 15:	82	4	1	60	145	100.00	9.98 14.83 6.98 7.36	101.0 107.0 100.0 99.8	53.0 59.5 56.4 57.9	25.36 33.80 19.04 21.78	61.24 90.94 42.72 49.34	.1074 .2396 .2673 .1988 .1869	4.118 4.783 6.396 11.631 6.219	6.773 (15.184)	39.549 (47.960)	7.776	39.189 (47.627)	2.232 (2.712)	28.78 (35.97)
Run 16: initial locations: hospitals	82	8	1	60	358	100.00	5.42 6.36 6.45 6.80 5.83 5.56 6.31 6.10	100.2 99.8 97.6 106.1 103.5 102.0 96.0 100.0	55.9 53.6 59.3 58.7 58.1 55.9 55.9 57.8	20.20 13.87 5.74 5.45 9.23 18.61 8.66 18.23	43.52 30.62 12.71 12.17 20.11 40.24 19.10 39.98			2.891 (14.109) 205 iteration 3.087	35.666 (46.885)	7.776	35.310 (44.652)	2.011 (2.543) 205 iteration 2.022	
Run 17:	82	4	1	60	137	100.00	7.18 8.30 6.94 7.91	99.4 98.0 102.0 102.0	56.2 57.8 56.0 56.0	23.06 27.79 22.36 26.76	52.00 64.45 50.12 61.46	.1176 .2517 .2694 .1922 .1691	4.068 4.470 4.631 6.625 6.219	5.176 (15.908)	37.952 (48.684)	7.776	37.596 (48.752)	2.141 (2.552)	27.00 (35.97)
Run 18:	82	4	1	60	171	100.00	6.90 7.27 7.18 8.00	100.0 99.7 102.0 102.0	56.0 58.0 56.0 58.0	21.91 25.81 24.29 27.97	49.06 58.33 54.77 64.38	.1171 .2511 .2693 .1926 .1700	3.759 4.473 5.199 6.389 6.219	5.251 (16.025)	38.027 (48.801)	7.776	37.670 (48.437)	2.145 (2.758)	24.91 (35.97)
Run 19:	82	4	1	60	158	100.00	6.53 7.00 7.11 7.73	100.0 100.3 102.0 101.0	56.0 55.2 56.0 58.0	22.34 25.78 24.69 27.17	49.55 57.87 55.57 62.13	.1185 .2528 .2695 .1916 .1676	4.179 4.604 4.382 6.003 6.219	5.033 (15.685)	37.808 (48.461)	7.776	37.452 (48.103)	2.133 (2.739)	25.27 (35.97)
Run 20:	82	4	1	60	170	100.00	6.58 7.72 7.12 7.81	100.0 98.3 102.0 101.0	56.0 55.7 56.0 58.0	22.76 26.09 24.69 26.45	50.54 59.65 55.58 60.59	.1174 .2515 .2694 .1923 .1693	3.850 4.327 4.986 6.536 6.219	5.194 (15.809)	37.970 (48.585)	7.776	37.613 (48.225)	2.142 (2.746)	27.24 (35.97)
Run 21:	82	4	1	60	194	100.00	7.64 6.79 7.12 7.87	98.3 100.0 102.0 102.0	56.2 56.2 56.0 57.8	25.87 21.81 24.69 27.61	59.02 48.70 55.59 63.35	.1172 .2513 .2693 .1925 .1697	3.745 4.350 4.832 6.955 6.219	5.227 (15.929)	38.003 (48.705)	7.776	37.648 (38.343)	3.144 (2.753)	26.56 (35.97)
Run 22:	82	4	1	60	140	100.00	6.58 8.20 7.20 7.71	100.0 97.7 102.0 101.0	56.0 56.1 56.0 58.0	22.76 26.99 23.56 26.67	50.55 62.41 53.17 60.96	.1175 .2516 .2694 .1923 .1692	3.694 4.237 5.054 6.597 6.219	5.183 (15.989)	37.958 (48.764)	7.776	37.603 (48.409)	2.141 (2.757)	27.54 (35.97)
Run 23:	8	4	1	60	108	100.00	3.18 5.26 4.75 3.59	100.3 104.0 98.0 102.0	56.0 56.0 56.0 56.0	32.63 17.89 28.01 21.45	6.45 3.74 5.78 4.29	.8181 .1643 .0165 .0011 .0001	3.583 4.250 5.119 7.150 6.219	3.722 (14.823)	36.498 (47.598)	7.776	36.142 (47.243)	0.201 (0.262)	27.22 (35.97)
Run 24:	24	4	1	60	136	100.00	5.97 4.45 4.74 5.47	100.2 100.2 104.0 95.8	57.9 56.1 55.9 56.1	23.62 31.20 29.00 16.16	15.11 19.17 17.96 10.21	.5436 .3314 .1010 .0205 .0036	3.758 4.303 5.072 7.318 6.219	4.153 (16.135)	36.929 (48.911)	7.776	36.574 (48.555)	0.610 (0.809)	30.13 (35.97)
Run 25:	40	4	1	60	187	100.00	5.43 7.24 5.55 6.89	100.0 96.2 102.2 102.0	55.8 57.4 56.0 57.9	32.22 16.61 28.46 22.69	33.87 18.30 30.02 24.78	.3586 .3678 .1886 .0645 .0206	3.551 4.261 5.375 7.997 6.219	4.498 (17.091)	37.274 (49.867)	7.776	36.908 (49.510)	1.025 (1.375)	28.84 (35.97)

Note: Figures in Brackets refer to initial locations.

System Parameters.

TABLE 2 : To Test (a) The effect of changing the number of Areas

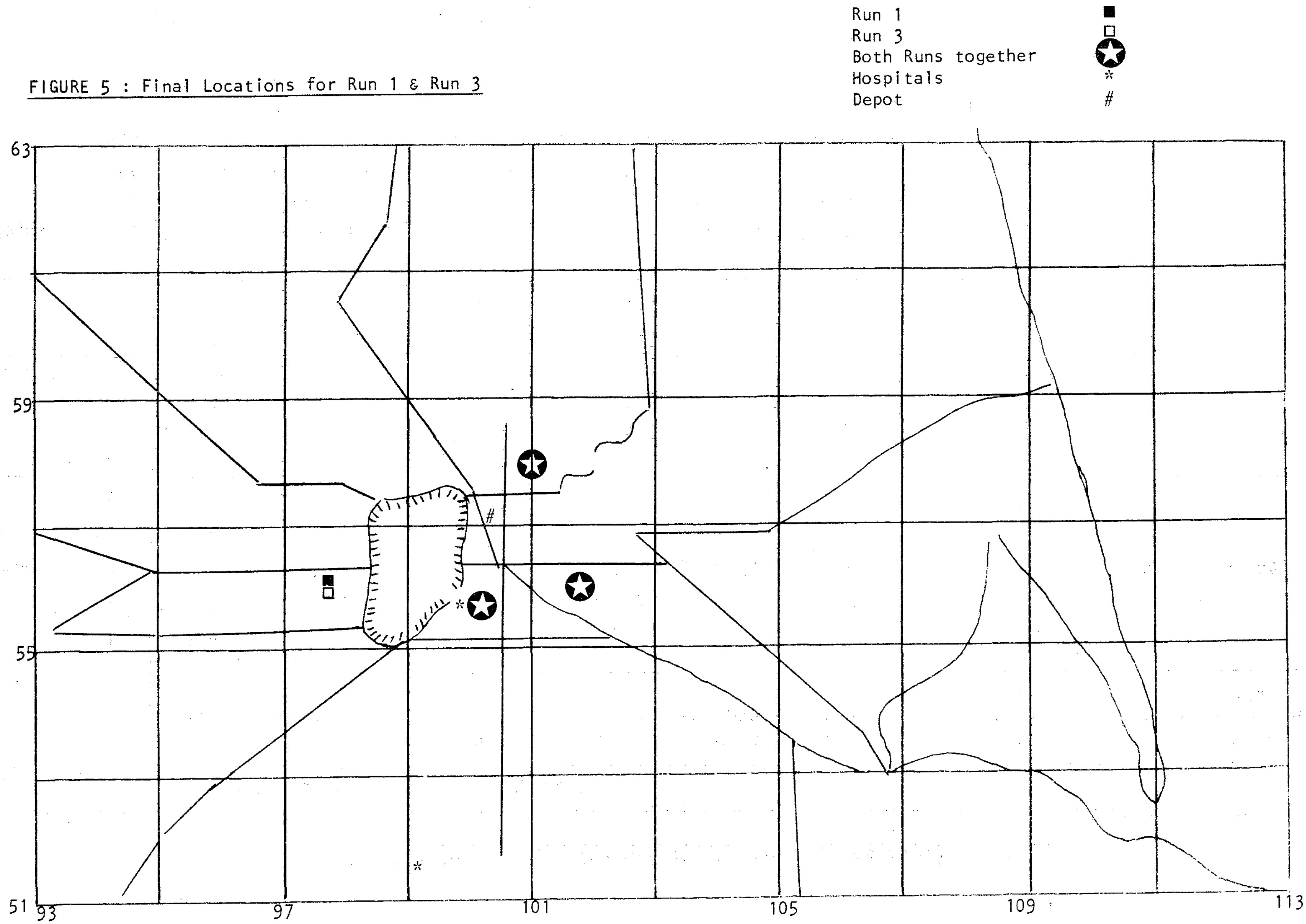
		By Ambulance		By State		For the System				
		Mean Response	% Utilisation	Probability	Mean Response	Mean Response	Mean Time to Hospital	Mean Retrieval Time	Mean No. in System	Maximum Response Time
Run 1	1	8.67	51.88	.0951	5.361	6.81	41.72	9.91	2.35	36.46
	2	9.81	70.73	.2238	5.917					
	3	9.50	57.57	.2632	6.208					
	4	9.65	70.19	.2064	8.042					
	5			.2114	7.928					
Run 3	1	6.58	50.55	.1175	3.694	5.18	37.96	7.78	2.14	27.54
	2	8.20	62.41	.2516	4.237					
	3	7.20	53.17	.2694	5.054					
	4	7.71	60.95	.1923	6.597					
	5			.1692	6.219					

Note: All the values given relate to the final locations of the ambulances.

Mean Time to Hospital refers to the mean length of time taken from when the call is received to transference at hospital.

Mean Retrieval Time is mean time from the accident scene to the hospital.

FIGURE 5 : Final Locations for Run 1 & Run 3



Note: Figure 5 relates to Table 2

Runs 1 and 2 were both made with 82 areas. For Run 1 the initial positions of the 4 ambulances were one at each of the four corners of the service area. For Run 2, the initial positions were Christchurch Hospital, Princess Margaret Hospital, Burwood Hospital and the Peterborough Street Depot. These are all marked in on Figure 6 as #. The final location for Run 1 and Run 2 are the same for 2 of the ambulances and different for the other 2. Looking at Table 3, the mean response times for the ambulances are 'balanced' better and the system mean response time is lower for Run 1 than for Run 2. In both cases however the percentage utilisation is not balanced well between the ambulances.

Runs 3, 4 and 5 were made with 60 areas. For Run 3 the initial positions were the four corners of the service area, for Run 4 the initial positions were the final positions for Run 3, and for Run 5 all ambulances were initially placed at the depot. The final locations for Run 4 were the same as the initial locations, as expected. The final locations for Run 3 were different to the final locations for Run 5 in all cases, though a general spatial relationship may be noted.

The mean response time for Run 3 is less than the mean response time for Run 5, as is the mean time to hospital.

Because of the slightly better results obtained in both cases for initial locations at the corners of the service area and also, because of the intuitive idea that less biasing will be involved if all ambulances are moved from a distant point, the remaining runs all use initial locations at the corners of the service area, unless otherwise stated.

Thus, Run 3 becomes the standard base point.

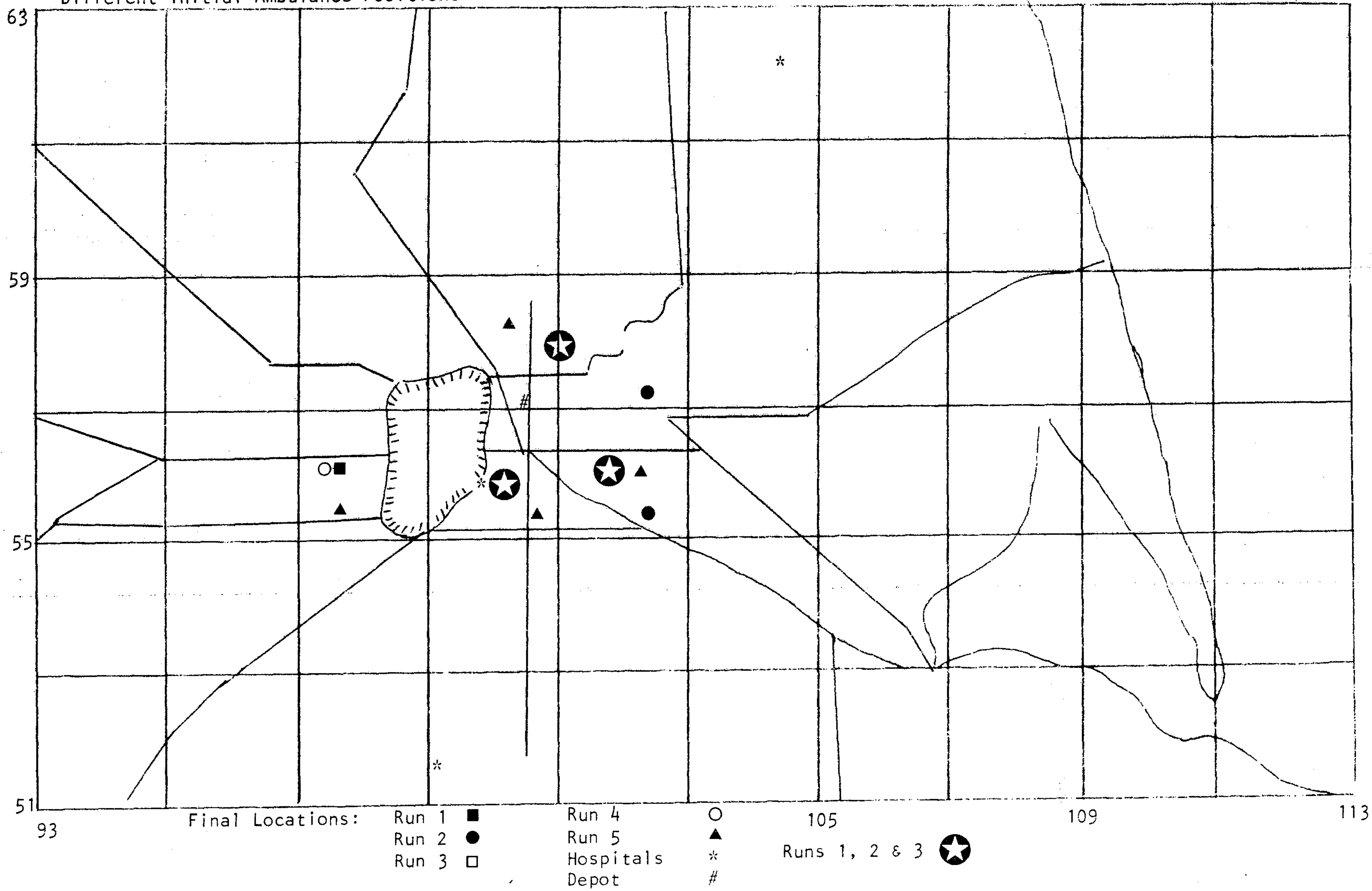
TABLE 3 : To Test (b) The effect of changing the initial positions of the ambulances

		By Ambulance		By State		For the System				
		Mean Response	% Utilisation	Probability	Mean Response	Mean Response	Mean Time to Hospital	Mean Retrieval Time	Mean No. in System	Maximum Response Time
	1	8.67	51.88	.0951	5.361					
	2	9.81	70.73	.2238	5.917					
Run 1	3	9.50	57.57	.2632	6.208	6.81	41.72	9.91	2.35	36.46
	4	9.65	70.19	.2064	8.042					
	5			.2114	7.928					
	1	10.08	63.15	.0938	5.659					
	2	9.74	66.67	.2220	6.267					
Run 2	3	9.25	70.92	.2627	6.563	7.04	41.95	9.91	2.37	36.73
	4	9.10	50.20	.2072	8.191					
	5			.2142	7.928					
	1	6.58	50.55	.1175	3.694					
	2	8.20	62.41	.2516	4.237					
Run 3	3	7.20	53.17	.2694	5.054	5.18	37.96	7.78	2.14	27.54
	4	7.71	60.95	.1923	6.597					
	5			.1692	6.219					
	1	6.57	49.02	.1149	3.808					
	2	8.23	62.84	.2486	4.457					
Run 4	3	6.95	49.46	.2690	5.576	5.58	38.36	7.78	2.16	27.55
	4	9.08	67.83	.1940	7.499					
	5			.1736	6.219					
	1	7.59	51.33	.1161	3.956					
	2	7.14	51.01	.2500	4.494					
Run 5	3	8.28	59.42	.2692	5.283	5.40	38.17	7.78	2.15	28.05
	4	7.96	67.07	.1932	6.857					
	5			.1716	6.219					

Also see Notes on Table 2.

FIGURE 6 : Final Locations for Run 1, Run 2, Run 3, Run 4 & Run 5

Different Initial Ambulance Positions



Run 6 was identical to Run 3 except that Princess Margaret Hospital was included as a possible destination. The final locations of the ambulance were the same in both cases (Figure 7). The work-load was split so that Christchurch Hospital took 93.87% of the load and Princess Margaret took 6.11%. This did not vary with the number of calls (see Run 8, Table 1 where final locations are again identical to Run 3). The response times for each ambulance were within .02 minutes and the mean response time for Run 6 was only slightly longer than for Run 3. There were some differences in the probabilities for the occurrence of each state but they were very slight.

Therefore, it was decided to use only Christchurch Hospital as an emergency centre.

Run 7 was made as an additional experiment. The initial locations were the Peterborough Street depot, and the two hospitals used as destinations were Princess Margaret Hospital and Burwood Hospital. Comparing the final locations with Run 5 (Figure 6 initial locations depot) there is little relationship. Looking at Figure 7, however, the straight line effect of the final locations can be seen to be due to the spread between the hospitals, where Burwood takes 30.20% of the load and Princess Margaret takes 69.78%.

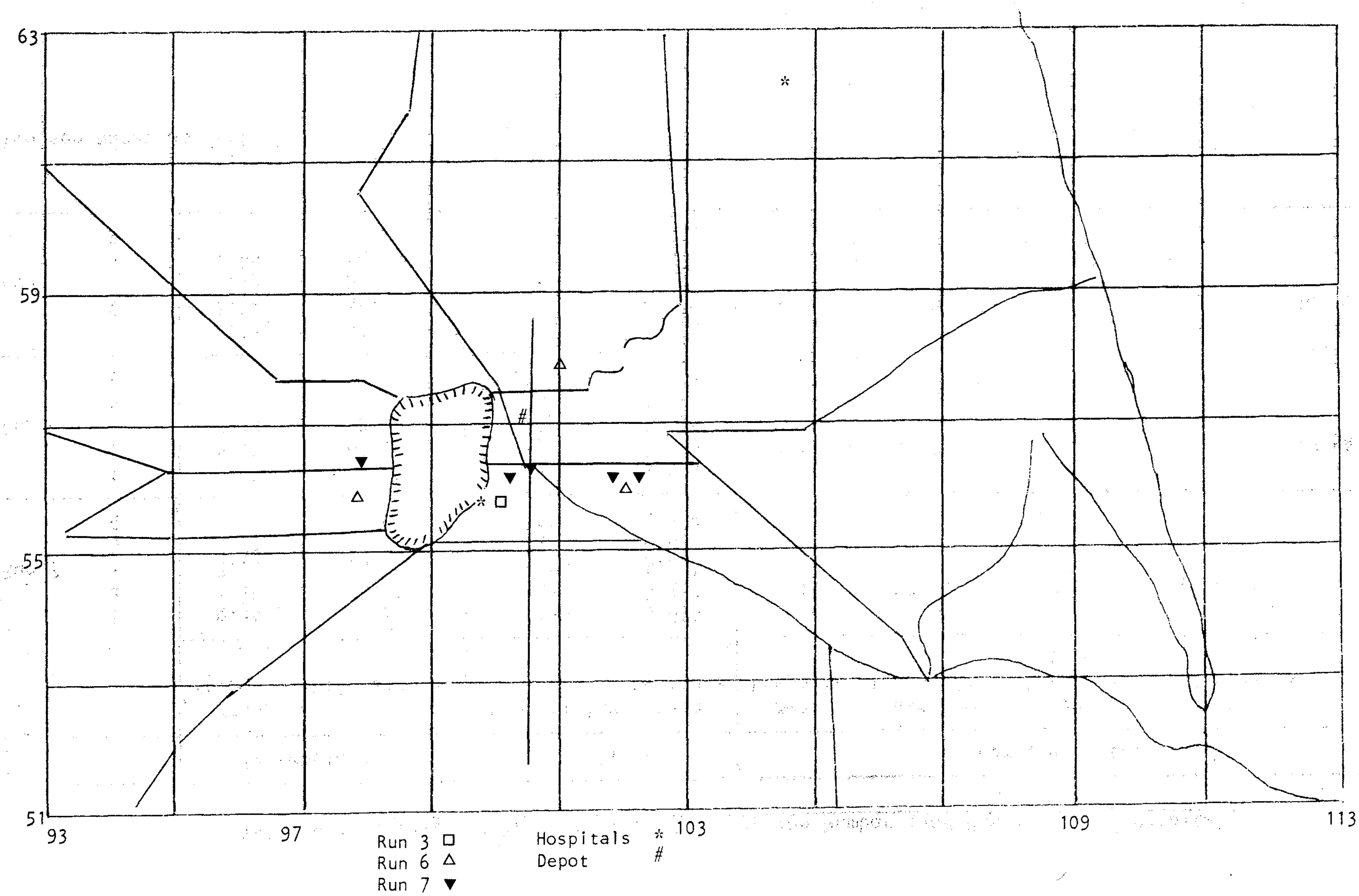
Another result of this test is that it becomes obvious that central placement of hospitals is desirable. This can be seen from comparison of the results of Run 6 and Run 7 in Table 4. Notice particularly here the bad balance in workload between the ambulances in Run 7.

TABLE 4 : To Test (c) The effect of Changing the Number (and Positions) of Hospitals

		By Ambulance		By State		For the System				
		Mean Response	% Utilisation	Probability	Mean Response	Mean Response	Mean Time to Hospital	Mean Retrieval Time	Mean No. in System	Maximum Response Time
	1	6.58	50.55	.1175	3.694					
	2	8.20	62.41	.2516	4.237	5.18	37.96	7.78	2.14	27.54
Run 3	3	7.20	53.17	.2694	5.054					
	4	7.71	60.95	.1923	6.597					
	5			.1692	6.219					
	1	6.57	50.47	.1194	3.694					
	2	8.17	61.53	.2538	4.237	5.22	37.67	7.45	2.13	27.54
Run 6	3	7.18	52.88	.2697	5.054					
	4	7.71	60.27	.1910	6.597					
	5			.1661	6.511					
	1	7.93	63.57	.0886	4.265					
	2	6.54	54.91	.2148	4.789	6.83	42.99	11.16	2.42	26.90
Run 7	3	8.19	80.05	.2602	5.199					
	4	7.44	48.19	.2102	6.548					
	5			.2261	11.934					

Also see Notes on Table 2.

FIGURE 7 : Final Locations for Run 3, Run 6 & Run 7



From Table 5 and Figure 8 it can be seen that the number of calls does not have a significant effect upon the final locations of the ambulances. In Run 8, LAMDA has been increased to 100. Otherwise the data were not altered. The final locations are identical. Otherwise, the mean responses have increased as expected. The probability of State 5 occurring, that is, the state when all ambulances are busy, has increased from .17 to .27 with an equivalent increase in the percentage utilisation of the ambulances. Clearly, the workload is too great for four ambulances (since it is now implied that there is a 27% chance of a patient having to wait).

Therefore in Runs 9, 10, 11 the number of ambulances has been increased to 5. In Run 9 LAMDA is equal to 82, in Run 10, LAMDA = 100 and in Run 11 LAMDA = 200. The initial positions of the ambulances are one at each corner of the area, plus one at Sunnyside Hospital. This is noted in Figure 8.

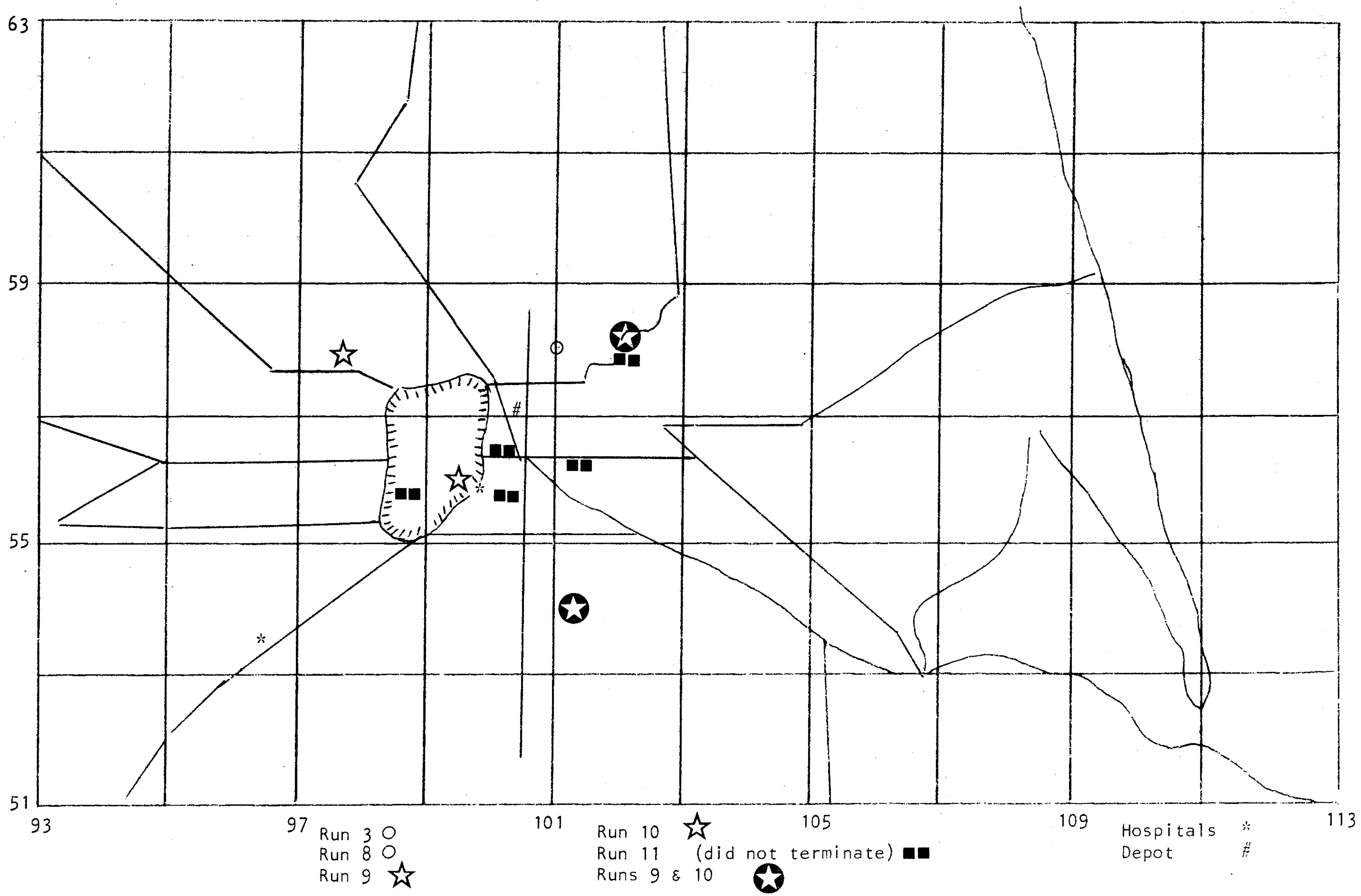
Again it can be seen from Figure 8 that the increase from 82 to 100 calls per day does not significantly affect the final locations of the ambulances. Only one position is substantially changed. With the increase to 200 calls per day, however, the problem increases in complexity to the extent that a solution was not obtained after 217 iterations. The results shown in Table 5 illustrate this. Ambulances 2 and 4 are used beyond their capacity (179% and 225% utilisation). It is likely that a solution is impossible for this situation. The ambulance utilisations thus become a guide to the feasibility of the system parameters. Table 12 (page 82) also shows the effect of changing the number of calls per day.

TABLE 5 : (d) To Test the effect of changing the number of calls LAMDA

		By Ambulance		By State		For the System				
		Mean Response	% Utilisation	Probability	Mean Response	Mean Response	Mean Time to Hospital	Mean Retrieval Time	Mean No. in System	Maximum Response Time
	1	6.58	50.55	.1175	3.694					
	2	8.20	62.41	.2516	4.237					
Run 3	3	7.20	53.17	.2694	5.054	5.18	37.96	7.78	2.14	27.54
	4	7.71	60.95	.1923	6.597					
	5			.1692	6.219					
	1	6.93	49.51	.0734	3.694					
	2	8.76	88.83	.1916	4.237					
Run 8	3	7.97	59.52	.2503	5.054	5.52	37.97	7.45	2.61	27.54
	*4	7.83	80.45	.2180	6.597					
	5			.2667	6.511					
	1	8.02	44.84	.1229	3.416					
	2	6.51	46.12	.2557	3.833					
Run 9	3	6.66	46.22	.2701	4.300	4.71	37.16	7.45	2.10	27.99
	4	7.24	44.55	.1887	5.270					
	5	7.77	42.10	.0989	7.554					
	6			.0617	6.511					
	1	7.92	51.95	.0755	3.443					
	2	6.92	54.77	.1950	3.939					
Run 10	3	7.13	50.51	.2519	4.469	5.12	37.47	7.45	2.58	28.72
	4	7.48	55.92	.2170	5.285					
	5	9.40	63.94	.1402	7.369					
	6			.1204	6.511					
	1	8.54	29.73	.0050	4.151					
	2	7.84	179.04	.0263	4.484					
Run 11	3	9.13	94.21	.0698	4.709	6.11	38.56	7.45	5.31	26.53
	4	8.77	225.05	.1235	5.130					
	5	6.86	34.28	.1638	6.283					
	6			.6116	6.511					

* 2 hospitals were used for Run 8.
Also see notes on Table 2.

FIGURE 8 : Final Locations for Run 3, Run 8, Run 9, Run 10 & Run 11



One of the facilities of CALL is that it enables restriction of the final positions of the ambulances within given boundaries. Table 6 and Figure 9 demonstrate this facility. Run 12 restricts an ambulance to the hospital area ($99 \leqslant x \leqslant 101$, $55 \leqslant y \leqslant 57$), Run 13 restricts an ambulance to the Marshlands area ($107 \leqslant x \leqslant 109$, $59 \leqslant xy \leqslant 63$), Run 14 restricts an ambulance to the Cashmere area (x unrestricted, $51 \leqslant y \leqslant 53$) and Run 15 restricts one ambulance to the hosptial area, one to Marshlands and one to Cashmere, and leaves one free.

From Figure 9 it can be seen that the results of restrictions to areas well away from the centre of demand are usually placements on the boundaries.

The mean response times do vary quite significantly with a difference of 1.29 minutes between the unrestricted case and the situation where three ambulances are restricted.

One of the interesting results is seen in Run 13 and Run 14 where the restricted ambulances carry 33% and 30% of the workload, as against unrestricted ambulance 1 which carries only 23% (Table 1). The reason for this is uncertain. Certainly the facility may be useful when it is required to obtain an even distribution of response times. Noted that the maximum response time is less in Runs 12 and 14 than in Run 3. Strategic restrictions could significantly improve upon this.

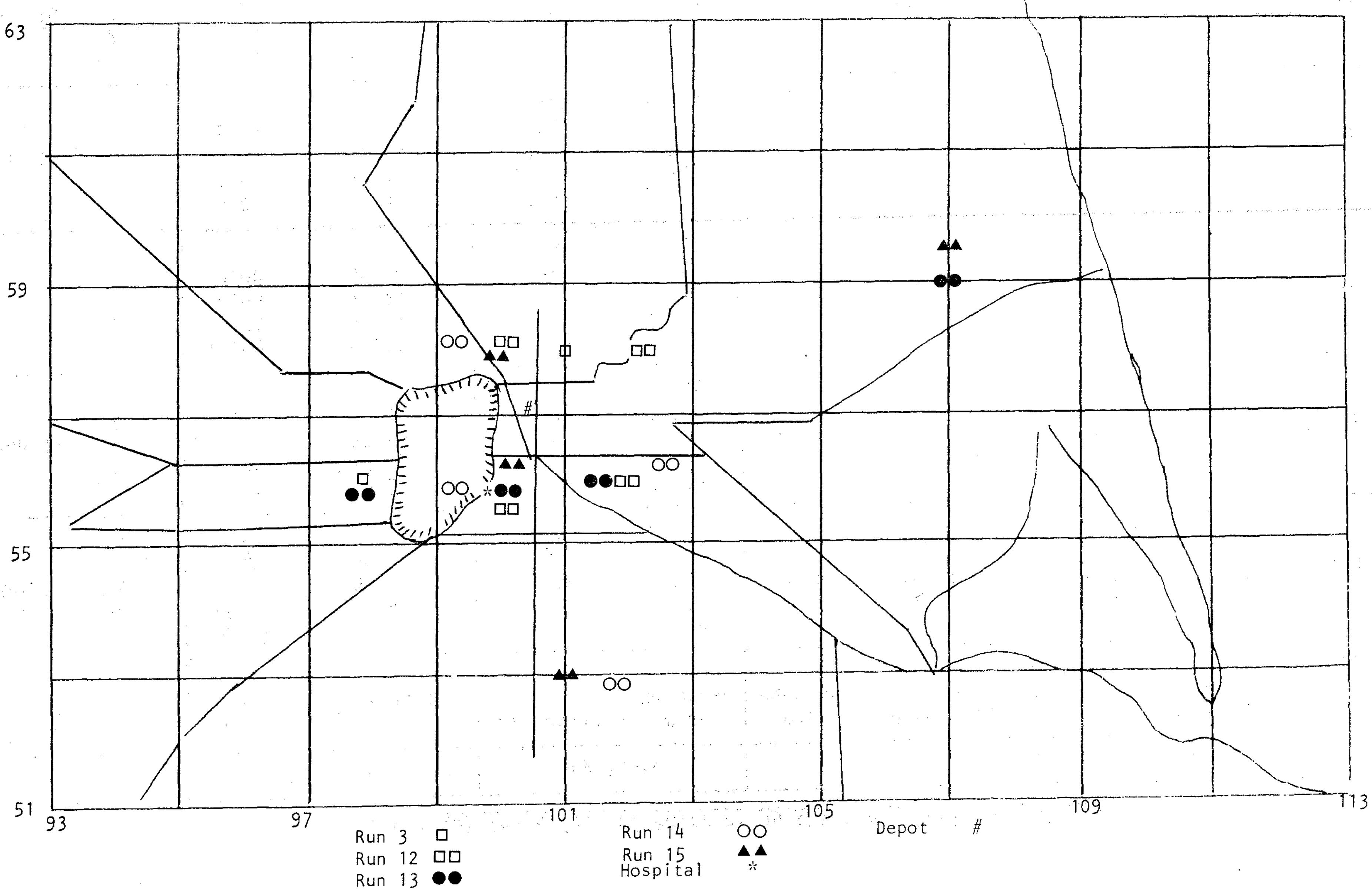
TABLE 6 : (e) The Effect of Restricting the Final Positions of Ambulances

		By Ambulance		By State		For the System				
		Mean Response	% Utilisation	Probability	Mean Response	Mean Response	Mean Time to Hospital	Mean Retrieval Time	Mean No. in System	Maximum Response Time
Run 3	1	6.58	50.55	.1175	3.694	5.18	37.96	7.78	2.14	27.54
	2	8.20	62.41	.2516	4.237					
	3	7.20	53.17	.2694	5.054					
	4	7.71	60.95	.1923	6.597					
	5			.1692	6.219					
Run 12	*1	7.15	49.31	.1164	3.833	5.35	38.13	7.78	2.15	24.95
	2	7.10	57.19	.2503	4.549					
	3	7.13	55.59	.2692	5.106					
	4	8.43	65.19	.1930	6.870					
	5			.1710	6.219					
Run 13	*1	13.90	85.92	.1114	3.728	6.12	38.89	7.78	2.19	27.95
	2	8.53	58.54	.2445	4.379					
	3	6.99	48.21	.2683	5.656					
	4	7.08	46.84	.1962	10.180					
	5			.1795	6.219					
Run 14	*1	10.07	73.03	.1141	3.997	5.70	38.48	7.78	2.17	27.13
	2	7.48	50.41	.2477	4.653					
	3	7.09	47.93	.2688	5.566					
	4	8.15	60.78	.1945	7.750					
	5			.1749	6.219					
Run 15	*1	9.98	61.24	.1074	4.118	6.77	39.55	7.78	2.23	29.78
	*2	14.83	90.94	.2396	4.783					
	*3	6.98	42.72	.2673	6.396					
	4	7.36	49.34	.1988	11.631					
	5			.1869	6.219					

* The ambulance under the restrictions are marked.

Also see notes on Table 2.

FIGURE 9 : Final Locations for Run 3, Run 12, Run 13, Run 14 & Run 15



The effect of increasing the number of ambulances is to decrease the mean response time. However, the computing time involved increases greatly, and in Run 16, 1000 seconds GPO time was used, without a solution being reached.

Figure 10 shows the final locations of ambulances for Run 3, for Run 9 where 5 ambulances are placed and Run 16 where 8 ambulances were tried. The initial locations for Run 16 were some of the hospitals shown on Figure 1. After 358 iterations, no solution was reached, though a very small mean response time is obtained.

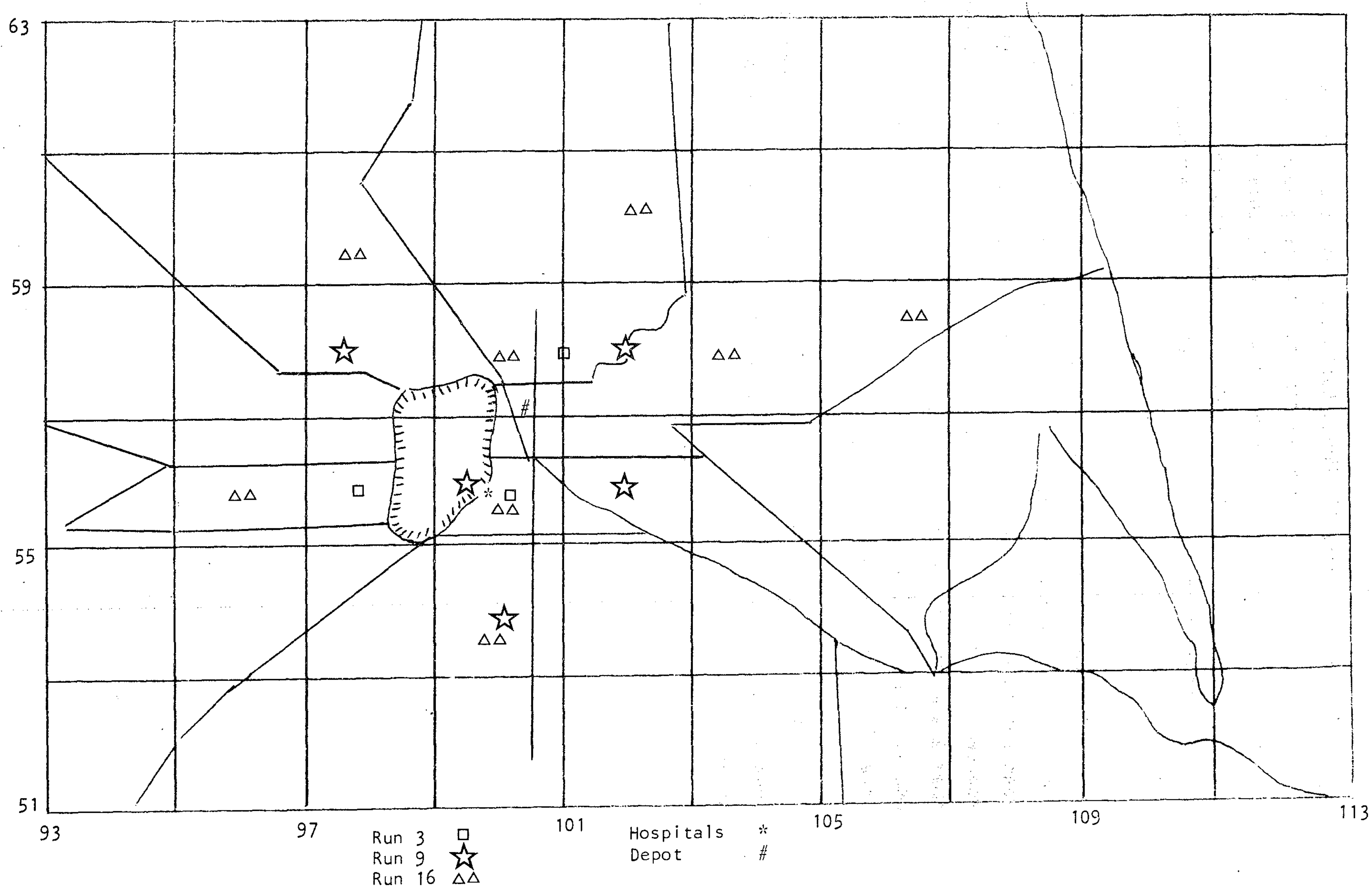
It should be noticed from Table 7 that the probability of all ambulances being busy for Run 9 is equal to 0.0617, which is considerably closer to zero than 0.1692 which is the case for Run 3.

TABLE 7 : (f) The effect of changing the Number of Ambulances

		By Ambulance		By State		For the System				
		Mean Response	% Utilisation	Probability	Mean Response	Mean Response	Mean Time to Hospital	Mean Retrieval Time	Mean No. in System	Maximum Response Time
Run 3	1	6.58	50.55	.1175	3.694					
	2	8.20	62.41	.2516	4.237	5.18	37.96	7.78	2.14	27.54
	3	7.20	53.17	.2694	5.054					
	4	7.71	60.95	.1923	6.597					
	5			.1692	6.219					
Run 9	1	8.02	44.84	.1229	3.416					
	2	6.51	46.12	.2557	3.833					
	3	6.66	46.22	.2701	4.300	4.71	37.16	7.45	2.10	27.99
	4	7.24	44.55	.1887	5.270					
	5	7.77	42.10	.0989	7.554					
	6			.0617	6.511					
Run 16	1	5.42	43.52							
	2	6.36	30.62							
	3	6.45	12.71							
	4	6.80	12.17			2.89	35.67	7.78	2.01	
	5	5.83	20.11							
	6	5.56	40.24							
	7	6.31	19.10							
	8	6.10	39.98							

Also see notes on Table 2.

FIGURE 10 : Final Locations for Run 3, Run 9 & Run 16



For Run 17, only 100 Monte Carlo samples were taken for the calculation of the mean response time for states greater than 1, instead of the usual 200. As can be seen from Table 8, this does not greatly affect the system mean response time nor the maximum response time. It does, however, affect the final locations of the ambulances (Figure 11).

For Run 18, 300 samples were taken and this did increase the computation time and the number of iterations required. The system mean response time becomes greater, while the maximum response time is reduced. The final ambulance positions are affected, to the extent that 2 ambulances are positioned together. The location at which this occurs is a very stable position (x = 102, y = 56) and is found as a final location for most runs.

The reason for the choice of 200 samples is unknown; however, it does appear that the model is too sensitive to changes in the number of samples.

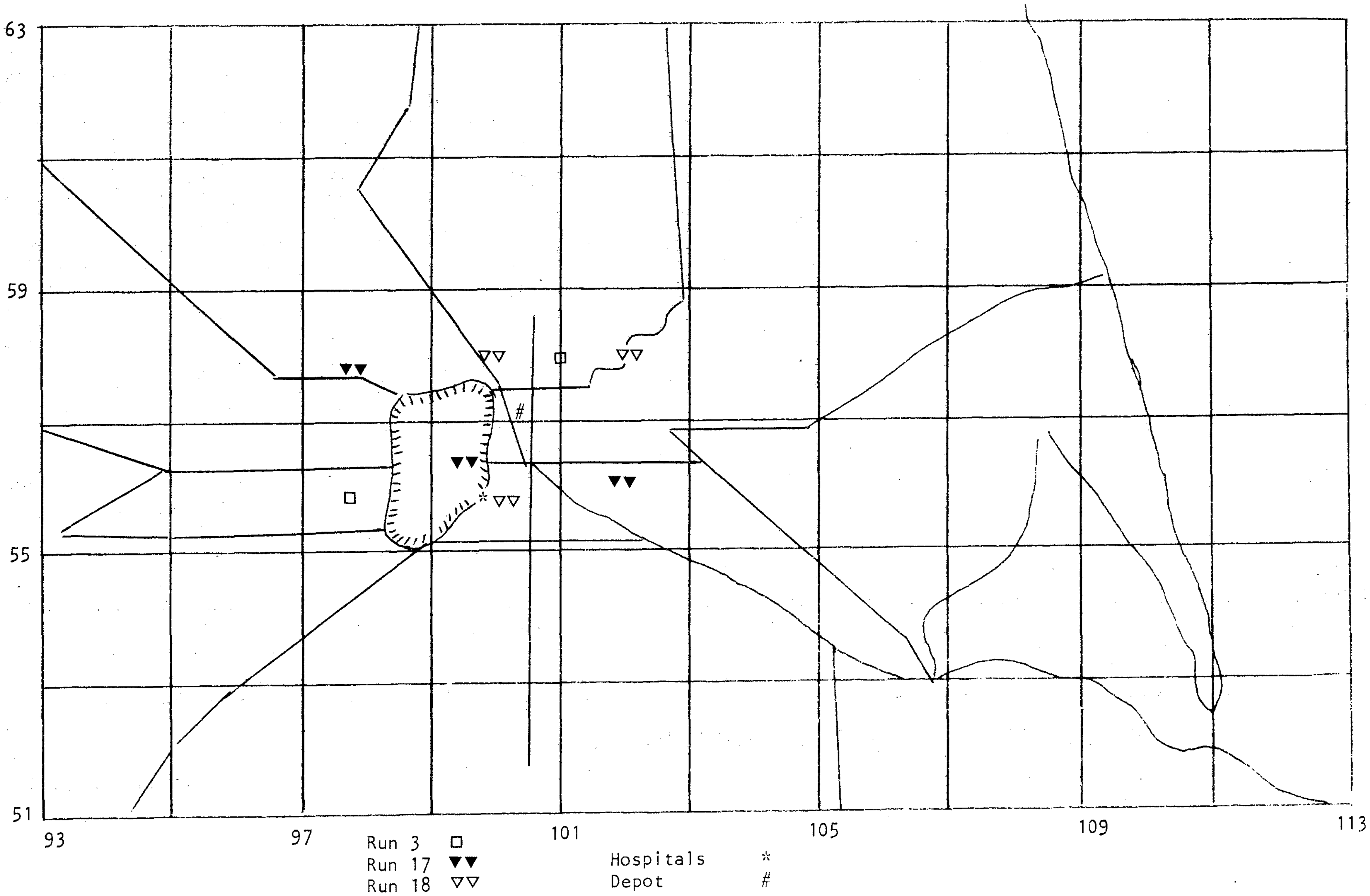
Program Parameters.

TABLE 8 : (a) The effect of changing the number of Monte Carlo Samples Taken

		By Ambulance		By State		For the System				
		Mean Response	% Utilisation	Probability	Mean Response	Mean Response	Mean Time to Hospital	Mean Retrieval Time	Mean No. in System	Maximum Response Time
	1	6.58	50.55	.1175	3.694					
	2	8.20	62.41	.2516	4.237					
Run 3	3	7.20	53.17	.2694	5.054	5.183	37.96	7.78	2.14	27.54
	4	7.71	60.95	.1923	6.597					
	5			.1692	6.219					
	1	7.18	52.00	.1176	4.068					
	2	8.30	64.45	.2517	4.470					
Run 17	3	6.94	50.12	.2694	4.631	5.176	37.95	7.78	2.14	27.00
	4	7.91	61.46	.1922	6.625					
	5			.1691	6.219					
	1	6.90	49.06	.1171	3.759					
	2	7.27	58.33	.2511	4.473	5.251	38.03	7.78	2.15	24.91
Run 18	3	7.18	54.77	.2693	5.199					
	4	8.00	64.38	.1926	6.389					
	5			.1700	6.219					

Also see notes on Table 2.

FIGURE 11 : Final Locations for Run 3, Run 17 & Run 18



Run 19 was a RNG seed = .5, Run 20 uses seed = 4.0 and Run 21 uses seed = .1554. (The base run, Run 3 has a seed = .170998.)

From Figure 12, it can be seen that for Run 3, Run 19, Run 20, the final locations of the ambulances are very similar. Ambulance 2 is the only one which takes up a different position. For Run 3 and Run 20 it is on the west side of the Park (the statistics for these two runs are very close) and for Run 19 it has shifted to a position on Moorhouse Avenue, close to Colombo Street. Run 21 produces only one ambulance position, the same as for Run 3, and that is at x = 102, y = 56, the most stable position. The other three final locations are, however, quite close to the standard, Run 3.

From Table 9, it can be seen that the system mean response times, the mean times to hospital and the probabilities of each state occurring are very similar. The mean response times for each ambulance are similar, and reflect only the differences in the final locations. There is some difference in the maximum response times.

Overall, the model does not appear to be too sensitive to changes in the seed of the RNG.

TABLE 9 : (b) The effect of changing the seed of the random number generator

		By Ambulance		By State		For the System				
		Mean Response	% Utilisation	Probability	Mean Response	Mean Response	Mean Time to Hospital	Mean Retrieval Time	Mean No. in System	Maximum Response Time
	1	6.58	50.55	.1175	3.694					
	2	8.20	62.41	.2516	4.237					
Run 3	3	7.20	53.17	.2694	5.054	5.18	37.96	7.78	2.14	27.54
	4	7.71	60.95	.1923	6.597					
	5			.1692	6.219					
	1	6.53	49.55	.1185	4.179					
	2	7.00	57.87	.2528	4.604					
Run 19	3	7.11	55.57	.2695	4.382	5.03	37.81	7.78	2.13	25.27
	4	7.73	62.13	.1916	6.003					
	5			.1676	6.219					
	1	6.58	50.54	.1174	3.850					
	2	7.72	59.65	.2515	4.327					
Run 20	3	7.12	55.58	.2694	4.986	5.19	37.97	7.78	2.14	27.24
	4	7.81	60.59	.1923	6.536					
	5			.1693	6.219					
	1	7.64	59.02	.1172	3.745					
	2	6.79	48.70	.2513	4.350					
Run 21	3	7.12	55.59	.2693	4.832	5.23	38.00	7.78	2.14	26.56
	4	7.87	63.36	.1925	6.955					
	5			.1697	6.219					

Also see notes on Table 2.

FIGURE 12 : Final Locations for Run 3, Run 19, Run 20 & Run 21

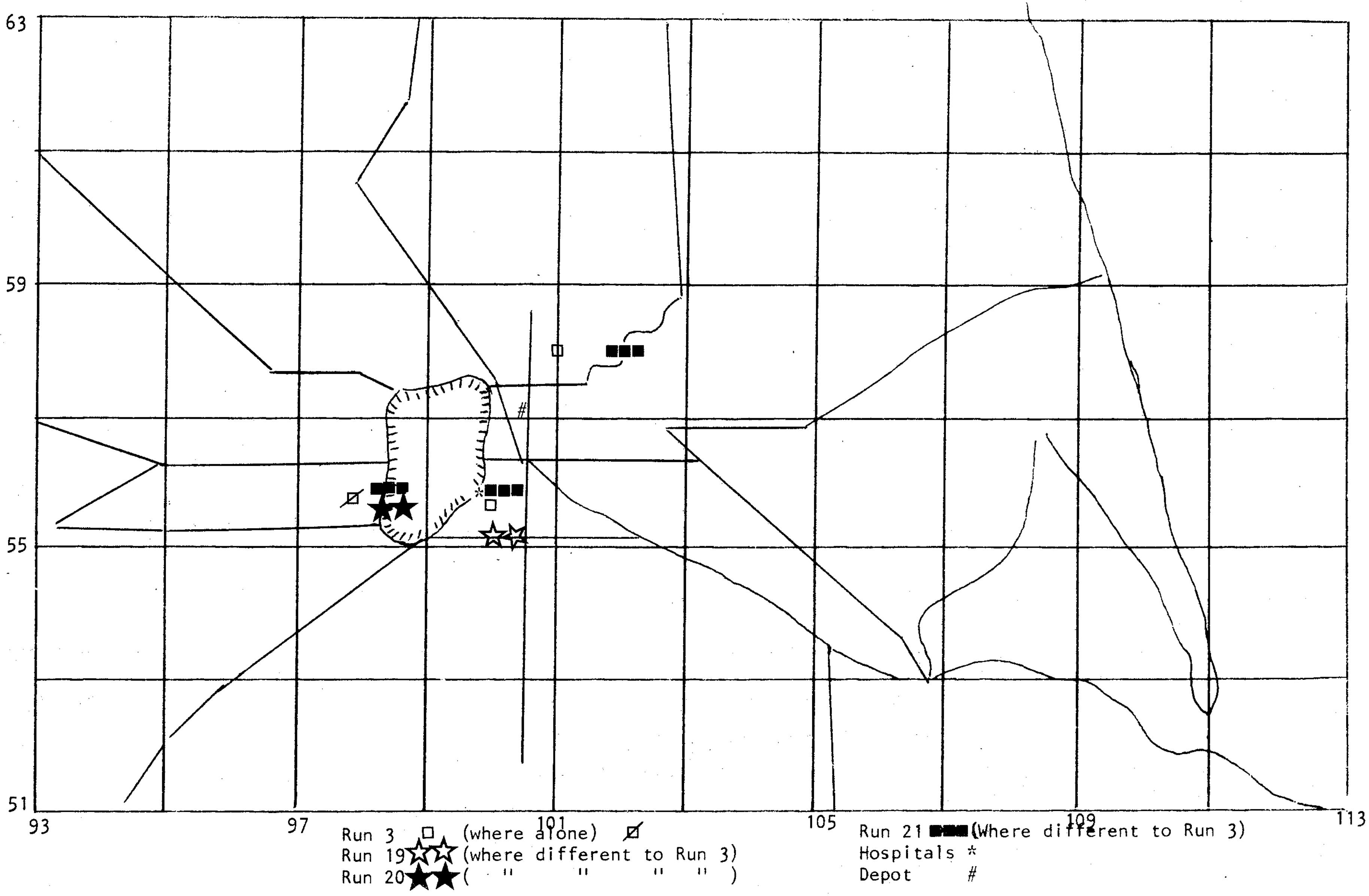


TABLE 10 : (c) The effect of changing the number of times RBAR is iteratively calculated, to remove the effect of the initial value.

		By Ambulance		By State		For the System				
		Mean Response	% Utilisation	Probability	Mean Response	Mean Response	Mean Time to Hospital	Mean Retrieval Time	Mean No. in System	Maximum Response Time
	1	6.58								
	2	8.20								
Run 3	3	7.20				5.183	37.96	7.78	2.14	27.54
	4	7.71								
	1	6.58								
	2	8.20								
Run 22	3	7.20				5.183	37.96	7.78	2.14	27.54
	4	7.71								

For the base run Run 3, RBAR was calculated 3 times. For Run 22, RBAR was calculated 6 times. The results for Run 22 were identical to the results for Run 3. Therefore it is sufficient to calculate RBAR three times only.

Also see notes on Table 2.

Model Assumptions

(a) The effect of changing the number of ambulances used on the model assumptions.

Refer back to Table 7 and Figure 10. CALL assumes that no queues will form, an assumption which is required, since only under these circumstances will the state of the system have a Poisson distribution (used in OBJECT). Therefore, it is required that the probability of state N + 1 occurring, where N is the total number of ambulances, should be close to zero. From Table 7 it can be seen that with 82 calls, for four ambulances, the probability of state 5 occurring (i.e. all ambulances busy) is .1692.

i.e. P (queue) = .1692.

This is not close to zero, and therefore the assumptions of the model break down.

Considering Run 9 with 5 ambulances,

P (queue) = .0617 and this is probably close enough to zero to be satisfactory.

The next section considers this assumption from a different point of view.

(b) The effect of changing the number of calls.

In order to satisfy the assumption that no queues should form, it has been suggested (2) that typically a 30% ambulance utilisation should be achieved. Therefore in Runs 23, 24, 25 LAMDA was varied, so as to ascertain the effect upon the utilisation.

In Run 23 LAMDA was set to 8. This is the actual expected number of accidents per day. The utilisation from Table 11 averaged about 15% and the probability of all ambulances being busy was .0001. For the final locations (Figure 13) it may be noticed that three are approximately the same as for Run 3, the standard. The maximum response time is slightly longer than for Run 4 with which there are two final locations in common.

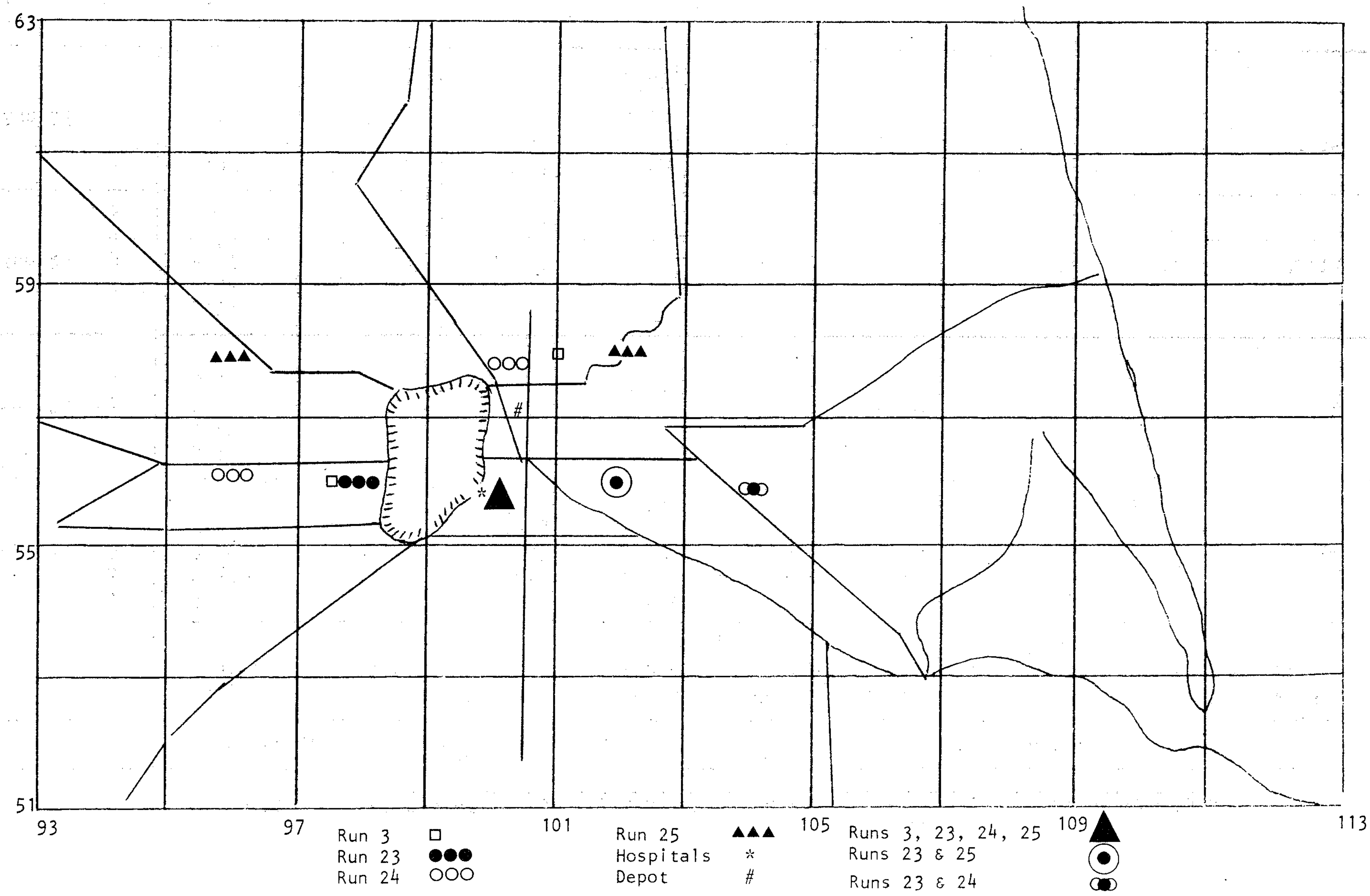
It appears therefore that 40 accidents per day could be handled satisfactorily by a system of four ambulances.

TABLE 11 : <u>(b) The effect of changing the number of calls, on the model assumptions</u>

		By Ambulance		By State		For the System				
		Mean Response	% Utilisation	Probability	Mean Response	Mean Response	Mean Time to Hospital	Mean Retrieval Time	Mean No. in System	Maximum Response Time
	1	6.58	50.55	.1175	3.694					
	2	8.20	62.41	.2516	4.237					
Run 3	3	7.20	53.17	.2694	5.054	5.18	37.96	7.78	2.14	27.54
	4	7.71	60.95	.1923	6.597					
	5			.1692	6.219					
	1	3.18	6.45	.8181	3.583					
	2	5.26	3.74	.1643	4.250					
Run 23	3	4.75	5.78	.0165	5.119	3.72	36.50	7.78	0.20	27.22
	4	3.59	4.29	.0011	7.150					
	5			.0001	6.219					
	1	5.97	15.11	.5436	3.758					
	2	4.45	19.17	.3314	4.303					
Run 24	3	4.47	17.96	.1010	5.072	4.15	36.93	7.78	0.61	30.13
	4	5.47	10.21	.0205	7.318					
	5			.0036	6.219					
	1	5.43	33.87	.3586	3.551					
	2	7.24	18.30	.3678	4.261	4.50	37.27	7.78	1.03	28.84
Run 25	3	5.55	30.02	.1886	5.375					
	4	6.89	24.78	.0645	7.997					
	5			.0206	6.219					

Also see notes on Table 2.

FIGURE 13 : Final Locations for Run 3, Run 23, Run 24 and Run 25



(c) The step reduction feature of the Hooke and Jeeves pattern search.

As a check on the search technique, Run 4 was made using as initial ambulance placements, the final placements of Run 3. This proved interesting, as although the final positions of the ambulances were the same as the initial positions in Run 4, the parameters were fairly significantly different. One would expect, all other things being equal, that they should be the same. The only explanation available at this time is that possibly the search routine does not satisfactorily establish a base point under one complete shift, and therefore returns the current RBAR as system mean response, rather than the base RBAR. (In this the initial one.) It is not particularly satisfactory.

6.3.2 Results of the Simulation Model

Two sets of data files were set up for the Swoveland et al. simulation, and a small series of runs with 60 areas were made. The results of these runs are presented in Table 12. When the original data were collected it was not realised that distributions of loading and unloading times would be required; therefore the data used for this is not particularly satisfactory.

The results are fairly self explanatory, and demonstrate as expected the suitability of the CALL results. Because the location of the ambulance is designated simply by a node number, it may be better to use an increased number of smaller 'areas' in the region of concentrated demand. This would provide a finer definition of the location which is available in CALL because of the co-ordinate system.

The overall mean response for the third data set of 6.1 mins is suitably similar to 5.2 for CALL Run 3, to which it corresponds. The start up time should be deducted from this. (Changing the seed of the RNG to 4.0 - not shown here - reduces this

TABLE 12 : Simulation Model Results

Constant parameters, start up time - 2 mins., number of calls simulated - 200.

	Ambulance locations	Rate of calls	Overall Mean Response (mins)	Average Utilisation %	node	Ambulance % of calls	% of calls with response GT 20 mins
Estimated loading and unloading time distribution	Corners of Area	3.5/hour = 84/day	11.4	22.41		33.81 36.19 11.43 18.57	9.52
	Burwood Sunnyside P.M.H. Ch-Ch		6.4	16.31		20.48 25.71 20.48 33.33	0.00
	Final positions for CALL Run 3		6.1	16.01	16 22 28 27	25.71 35.71 22.38 16.19	0.00
Constant loading time = 20 unloading time = 10	Corners of area		11.4	22.41		33.81 36.19 11.43 18.57	9.52
	Corners of area	1.7/hour = 42/day	11.9	11.24		40.00 39.05 10.48 10.48	4.76
	Final positions for CALL Run 3		5.9	7.63		28.57 35.71 21.90 13.81	0.00

time to 5.5 mins.)

The difference in the percentage of calls taken by each ambulance is probably due to the fact that the fourth ambulance is really on the border of areas 21 and 27 rather than in 27 as used here.

The average utilisation is considerably less, probably due to the loading and unloading time distributions.

In general, if sufficient data can be obtained for the simulation model, then it appears that it is likely to be useful in providing further reports of the system along with checks of the assumptions required for CALL, for example, zero waiting time.

7.0 CONCLUSIONS

(a) Concerning the Models

CALL is a useful means of obtaining some idea of the optimal locations of ambulances. It can also be used to ascertain the optimality of certain specified locations. However, the results do show that there is a large amount of variability in the final results. That is, the models appear to be fairly sensitive to program and system changes.

The simulation model is more precise than CALL and consequently requires not only more data but more accurate data than CALL. By removing some of the assumptions needed for CALL, it provides more information about the system.

However, as it is not possible at present to optimise with the simulation model, it is suggested that CALL should be used to provide feasible location sets for emergency calls (for which it is most useful). The simulation model should then be used to test out these location sets for all calls. The feasibility of this process will depend largely on the availability of data.

(b) Concerning the System

Christchurch is not a particularly suitable area for the application of ambulance location models. It is of a regular geographical shape and the demand for ambulances is concentrated on the centre of this area, which means that a central placement is necessary for all ambulances. In fact, the present situation of all ambulances based at the depot, yet spending most of their time moving in cyclic patterns between hospitals may be a better solution than any of the results obtained using CALL. However, it has served as a useful test for the models.

On the other hand, Auckland has an irregular geographical shape, and from preliminary investigation of the road accident data, it does appear that the spread of accidents is quite large. This means that a rule of thumb solution to the location problem will not be easy to obtain, and therefore the two models might be usefully applied. This may be dependent on the availability of data.

(c) Concerning Similar Urban Facility Location Problems

The models used to study the ambulance location problem are generally not applicable to other urban facility problems. The main reason for this is the concept of the nearest ambulance dispatch rule. However the hypercube model of Larson (18) is worthy of further investigation, in view of it being, potentially at least, generally applicable.

ACKNOWLEDGEMENTS

Thanks are due to Professor James A. Fitzsimmons, University of Texas at Austin and to Dr Dean Uyeno, University of British Colombia for providing their programs and advice. Also to Professor Richard A. Volz, University of Michigan for advice. The Traffic Engineering Section of the Ministry of Transport in Christchurch provided all the accident data and were very helpful with advice on interpretation of it. Mr G.F. Whitacker, Station Manager, the St. John Ambulance Association assisted with information about the Christchurch Service.

REFERENCES

(1) FITZSIMMONS, J.A. 1970. Emergency Medical Systems: A Simulation Study and Computerized Method for Deployment of Ambulances. Ph.D. thesis, University of California, Los Angeles.

(2) FITZSIMMONS, J.A. 1973. A Methodology for Emergency Ambulance Deployment. Management Science 19 : 627-636.

(3) LAZARUS, S.S. 1973. Locating Ambulances in Metropolitan Melbourne: Proceedings, National Operations Research Conference, Melbourne.

(4) SAVAS, E.S. 1969. Simulation and Cost Effectiveness Analysis of New York's Emergency Ambulance Service. Management Science 15 : B608-627.

(5) HOOKE, R. & JEEVES, J.A. 1961. Direct Search Solution of Numerical and Statistical Problems. Journal of Association for Computing Machinery 8 : 212-229.

(6) SNYDER, R.D. 1971. A Note on the Location of Depots'. Management Science 18 : 97.

(7) TOREGAS, C., SWAIN, R., ReVELLE, C., & BERGMAN, L. 1971. The Location of Emergency Service Facilities. Operations Research 19 : 1363-1971.

(8) KUEHN, A.A. & HAMBURGER, M.J. 1963. An Heuristic Program for Locating Warehouses. Management Science 9 : 643-666.

(9) SWOVELAND, C., UYENO, D., VERTINSKY, I., & VICKSON, R. 1973. Ambulance Location: A Probabilistic Enumeration Approach. Management Science 20 : 686-698.

(10) GORDON, G. & ZELIN, K. 1970. A Simulation Study of Emergency Ambulance Service in New York City. Tans. N.Y. Academy of Science Series II, 32.

(11) SHONICK & JACKSON. 1973. An Improved Stochastic Model for Occupancy Related Random Variables in General-Acute Hospitals. Operations Research 21 : 952-965.

(12) HALL, W.K. 1972. The Application of Multifunction Stochastic Service Systems in Allocating Ambulances to an Urban Area. Operations Research 20 : 558-570.

(13) VOLZ, R.A. 1971. Optimum Ambulance Locations in Semi Rural Areas. Transportation Science 5 : 193-203.

(14) WILDE, D.J. 1964. Optimum Seeking Techniques. Prentice Hall, New Jersey.

(15) PARZEN, E. 1962. Stochastic Processes. Hoden Day.

(16) BELL, C.E. & ALLEN, D. 1969. Optimal Planning of an Emergency Medical Service. Unpublished paper. 36th National Meeting Operations Research Society.

(17) SCHNEIDER, J.B., & SYMONS, J.G. 1971. Locating Ambulance Dispatch Centres in an Urban Region: A Man-Computer Interactive Problem Solving Approach. Regional Science Research Institute Discussion Paper No. 49.

(18) LARSON, R.C. 1974. A Hypercube Opening Model for Facility Location and Redistricting in Urban Emergency Services. Comput

(19) CHAIKEN, J.M., & LARSON, R.C. 1972. Methods for allocating Urban Emergency Units: A Survey. Management Science <u>19</u> : 110-130.

APPENDIX I

FITZSIMMON'S MODEL : CALL

CALL - a part simulation program used to investigate the properties of emergency ambulance facilities : James A Fitzsimmons.

1. Program Description and Methodology

The main program initially reads all the required information about the system, the co-ordinates of the centroids of each 'area' and the probability of a call coming from that area, the positions of the hospitals, the initial positions of the ambulances, and their allowable 'ranges' and speeds.

The cumulative distribution over all areas, from area 1 to area n is calculated, and a quick look up 'key' used to obtain the area related to a certain call probability is calculated.

Although entry is allowed for vehicles of different speeds, the actual ambulance speeds are averaged out over all the ambulances for use in calculating times. This is done, just before the calculation of the mean response time for each hospital, which is calculated as the sum of the probability of a call from each district, multiplied by the time taken to get from each hospital to each district. The mean retrieval time to each hospital is calculated as the sum over all hospitals of the time to the closest hospital times the probability of a call for each district, multiplied by 1.25. The mean number of cases delivered to each hospital is calculated.

The main program then moves into the pattern search routine, by calling subroutine PATS, which calls subroutine OBJECT, in order to evaluate the system parameters for the initial locations given. PATS calls OBJECT initially, in order to evaluate the system parameters for the initial positions of the ambulances. (A description of PATS is given at the end of this section in order not to detract from the main logic of the program.) Briefly, PATS uses a Hooke and Jeeves pattern search technique (5) to improve the locations of the ambulances. The parameter tested in all cases is RBAR, the system mean response time which is calculated in OBJECT for each new set of locations and returned to PATS until an optimum is reached.

Subroutine OBJECT

OBJECT begins by constructing a travel time matrix, with the times for each ambulance to each district, for the given locations of the ambulances. Thus, if there are four ambulances and one hundred districts, then a 4 * 100 matrix will be calculated. As for the travel times used for calculation of the hospital retrieval times there are three rates of travel. Firstly, city speed, where the distance between ambulance and district is less than the given 'city distance' (CITY), secondly freeway speed, where the distance is greater than the given freeway speed (FWY), and thirdly, if the distance is greater than the city distance and less than the freeway distance, the speed (rate of travel) is calculated on a linear interpolation from the city distance to the freeway distance.

For each district, the travel time matrix is used to create a table giving the ordering (numerical labels 1, 2 ... N where N is the number of ambulances) of the ambulances, from closest (in time) to furthest away. (The ambulance priority table is printed out for the final locations obtained.) A 'bubble sort' routine is used to calculate the ambulance priority table.

The next parameter for the given locations calculated, is the maximum response time (RMAX) and it is calculated from the priority table, according to the last ambulance on the list. RMAX is selected as the longest time for all of these.

Response Time for State 1 : no busy ambulances

Mean responses for all states are now calculated. For state 1, where all ambulances are idle, it is obvious that the ambulance which will be sent will be the ambulance stationed closest to that district. The mean response is therefore calculated as the sum of the minimum times to get to each district (first-in ambulance on the priority table) times the probability of a call coming from that district, for all districts.

i.e. $$RES(1) = \sum_{I=1}^{NDIST} TRAVL(I,K) * PROB(I)$$

RES (1) is the mean response time for state 1.

NDIST is the total number of districts.

K is the number of the closest ambulance to district I.

TRAVL (I,K) is the travel time to district I for ambulance K.

PROB (I) is the probability of a call occurring in district I.

Response Time for State 2 : one busy ambulance

The mean response time for state 2 is equal to the sum over all districts of the time for the closest ambulance (first-in ambulance on the priority table) to get to that district times the probability of the closest ambulance being available if one ambulance is busy, plus the time for the second closest ambulance times the probability of the closest ambulance not being available if one ambulance is busy.

K   is the closest ambulance to district I

L   is the second closest ambulance to district I.

K and L are read from the ambulance priority table.

TRAVL (I,K) is the travel time to district I for the closest ambulance.

TRAVL (I,L) is the travel time to district I for the second closest ambulance.

Therefore,

$$RES(2) = \sum_{I=1}^{NDIST} (TRAVL(I,K) * PRB_I) + (TRAVL(I,L) * (1-PRB_I))$$

Where PRB_I = the probability of the first in ambulance for district I being available given one busy ambulance, and

$1-PRB_I$ = the probability of the first in ambulance for district I not being available given one busy ambulance.

PRB_I is calculated as

$$PRB_I = 1 - \sum_{\substack{i \text{ s.t.} \\ k_i = K}} PROB(i)$$

where k_i is the closest ambulance to district i.

That is, PRB_I = 1 - the sum of the probabilities of a call occurring in those districts for which K (the closest ambulance to district I) is the closest ambulance.

Let I_k represent the set of districts for which K (the closest ambulance to district I) is the closest ambulance, that is, all i (i = 1,2 ... NDIST) such that $k_i = K$. Also, let $P(I_k)$ equal the sum of the probabilities of a call occurring in I_k.

That is, we wish to show that $PRB_I = 1 - P(I_k)$.

Now, if one ambulance is busy, then it will be the closest ambulance to the district in which the first call has arisen.

The probability of ambulance K being the ambulance sent to the first call is therefore equal to $P(I_k)$.

Thus, the probability of ambulance K not being busy equals $1 - P(I_k)$.

i.e. $PRB_I = 1 - P(I_k)$.

Response time for States 3 ... (N + 1) : 2 or more busy ambulances

For states greater than 2, the form of calculation used for one busy ambulance would become extremely complex. Therefore a Monte Carlo simulation procedure is used to estimate the mean response time.

For each state 3 to N corresponding to 2 ... (N - 1) busy ambulances, 200 possible assignments are simulated, the response time calculated, and summed, and finally the total is divided by 200. The seed of the random number generator is set so that each time the subroutine OBJECT is called, the same stream of random numbers will be generated.

State 3 : 2 busy ambulances

A random number between 0 and 1 is generated and using the quick look up 'key', the area at which the cumulative probability is closest to the random number is selected as the call origin. All ambulances having initially been set to 'idle' status, the closest ambulance to that district is now set to 'busy'. Another district is selected, in the same fashion. If the closest ambulance is busy, then the second closest ambulance is also set to 'busy'. Thus two busy ambulances have been obtained. Select a third district, and calculate the response time (RSP_1) as the time for the closest 'idle' ambulance to reach that district. Notice that this may be the first, second or third ambulance on the ambulance priority table for that district.

This process is repeated 200 times and RSP is accumulated so that

$$RSP = \sum_{i-1}^{200} RSP_i$$

and thus RES (3) = RSP/200.

For state 4, it can be seen that three ambulances must be set to 'busy' and then RSP_i will be the time for the closest 'idle' ambulance to reach the district which the call arises in. That is, four districts are selected and the ambulance which answers the call may be the first, second, third or fourth ambulance on the priority table.

When the mean response times for states 0,1,2, ... N have been calculated the process is completed.

System Mean Response Time

A short iterative procedure is then used to calculate RBAR which is the mean response of the system. This uses the fact that if no queue forms, then the distribution of the states of the system is Poisson, with mean equal to RHO (defined below).

An initial RBAR is calculated as

RBAR = .25 * RES (1) + .75 * RES (2)

The calculation from here onwards is repeated three times to remove the initial value set for RBAR.

XMU which is the mean service time is then calculated as

XMU = RBAR + CARE + (RET + TRSFR) * TRANS

where

CARE is the mean on-scene care time

RET is the mean retrieval time

TRSFR is the mean transfer time at hospital

TRANS is the proportion of cases which are transported.

This is consistent with the definition of service time discussed in the general summary of methods.

RHO, the mean arrival rate / mean service rate = $\frac{\lambda}{\mu}$ is calculated as LAMDA x XMU / 1440

where

LAMDA is the mean number of calls per day (mean arrival rate).

$$\text{RHO} = \frac{\text{mean arrival rate}}{\text{mean service rate}} = \frac{\text{no. calls per minute}}{\text{no. services per minute}} = \frac{\text{LAMDA/1440}}{\text{1/XMU}} =$$

$$\text{LAMDA} * \frac{\text{XMU}}{1440}$$

$$\text{since no. services} = \frac{\text{total service time}}{\text{mean service time}}$$

$$\text{and no. services/min.} = \frac{1}{\text{mean service time}}$$

the probability of each state occurring is required. This is calculated using the assumption that calls arrivals are Poisson distributed. FREQ (J) is the probability of state J occurring and

$$\text{PREQ (1)} = e^{-\frac{\lambda}{\mu}} = \frac{1}{e^{\frac{\lambda}{\mu}}}$$

and RBAR the mean system response time is now set to FREQ (1) * RES (1) which is a function of the previous value of RBAR

$$\text{since RBAR} = \text{FREQ (1)} * \text{RES (1)}$$

$$= \frac{1}{e^{\frac{\lambda}{\mu}}} * \text{RES (1)}$$

$$= \frac{1}{e}\ \text{LAMDA} * \frac{\text{XMU}}{1440} * \text{RES (1)}$$

where

$$XMU = RBAR + CARE + (RET + TRSFR) * TRANS.$$

FREQ (J) for J = 2,3 ... N is then calculated as

$$FREQ\ (J) = FREQ\ (J-1) * \frac{RHO}{J-1}$$

that is

$$FREQ\ (2) = \frac{1}{e^{\frac{\lambda}{\mu}}} \cdot \frac{(\frac{\lambda}{\mu})}{1}$$

and

$$FREQ\ (3) = \frac{1}{e^{\frac{\lambda}{\mu}}} \ \frac{(\frac{\lambda}{\mu})}{1} \ \frac{(\frac{\lambda}{\mu})}{2}$$

$$= \frac{1}{e^{\frac{\lambda}{\mu}}} \ \frac{(\frac{\lambda}{\mu})^2}{2}$$

or

$$FREQ\ (J) = \frac{1}{e^{\frac{\lambda}{\mu}}} \ \frac{(\frac{\lambda}{\mu})^{J-1}}{(J-1)!}$$

Finally

FREQ (N+1) is calculated

$$FREQ\ (N+1) = 1 - \sum_{J-1}^{N} FREQ\ (J)$$

and RBAR is calculated as

$$RBAR = \sum_{J=1,N} FREQ\ (J) * RES\ (J)$$

When this has been repeated three times, the value of RBAR is considered to be sufficiently stabilised.

The first time OBJECT is called, the initial value parameters are isolated and kept in storage to be printed out with the final position parameters. At the termination point of PATS when further moves of the ambulances cannot improve the value of RBAR, PATS returns to the main program, where subroutine LOAD is called.

Subroutine LOAD

LOAD calculates further parameters of ambulance percentage loads and mean responses for the final ambulance locations obtained from PATS. The proportion of calls taken by each ambulance is calculated as the sum over all districts having that ambulance as the first in ambulance of the probability of a call from that district, times the probability of the state (i=1,N), where the probability of state N = the probability of stage N + the probability of state N + 1 (all ambulances busy). Actual numbers carried by each ambulance are calculated as the proportion taken by each ambulance times the total number of calls transported.

The mean response time for each ambulance is calculated as the sum over all districts having that ambulance as first-in ambulance of the travel time for the first-in ambulance for the district times the proportion of cases transported by that ambulance.

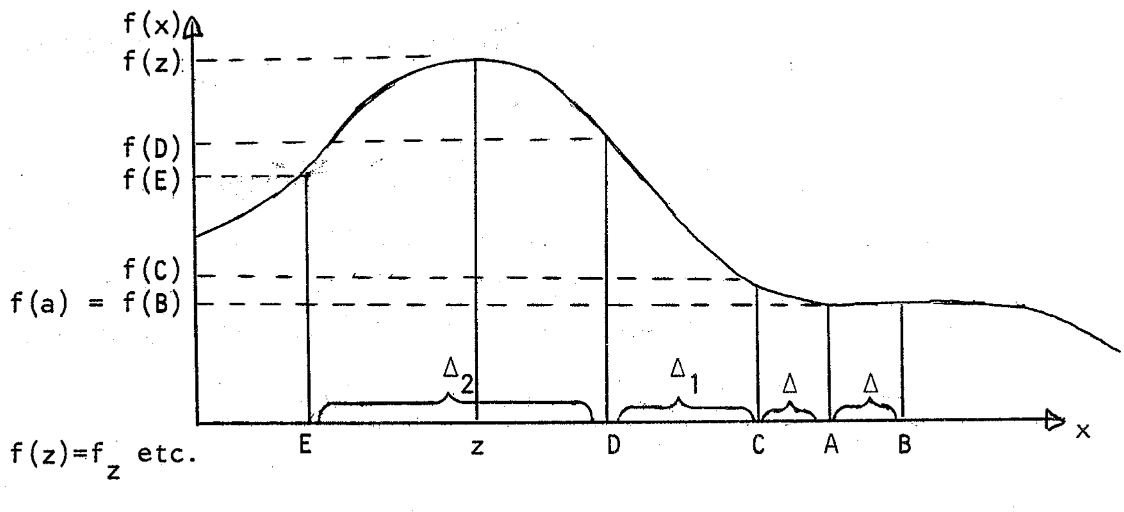
Load returns to the mainline program where percentage utilisation of each ambulance is calculated, and the parameter tables are printed out.

PATS

PATS is a subroutine which uses a Hooke and Jeeves Pattern Search technique (5) to obtain a local optimum. In order to obtain a good estimate of the global optimum, it may be necessary to use different sets of initial conditions, to evaluate the results.

Simple Example

As a simple illustration of the Hooke and Jeeves Pattern Search technique, consider the situation as shown in the diagram below.



We have a function f(x), and we wish to determine the optimum value of f(x), or the value of x for which f(x) is optimum. (We are maximising here, whereas in the ambulance example we are minimising.)

This example has a single global optimum.

Begin the search by evaluating the function at $x = A$, a point on the flat section of f(x). The value of the function is f_A.

Consider a move of Δ to the right, where Δ is a certain fixed increment, the size of which is set initially. Evaluate the function at this new point $x = B$, as f_B.

Test f_A against f_B, and since $f_A = F_B$ the flag is set as a failure, and we return to A.

Consider then a move of Δ to the left, evaluating f_C at $x = C$. Here f_C is greater than f_A which implies a success. Therefore set $\Delta_1 = 2\Delta$ and consider point D, at a distance Δ_1 from C, to the left of C.

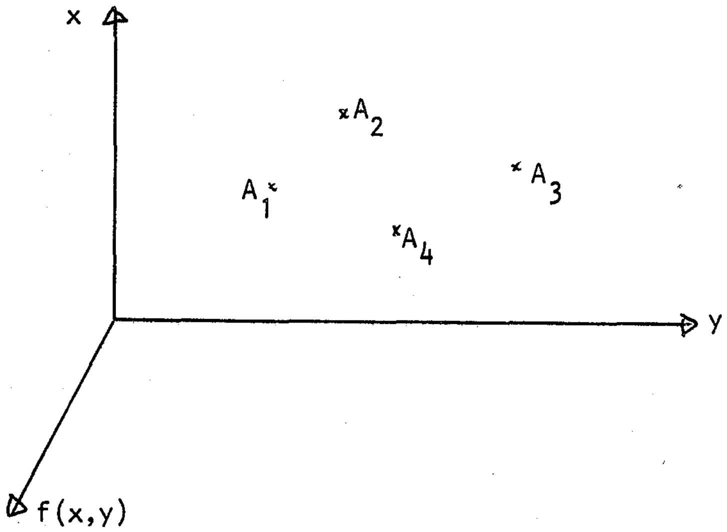
Another success implies that Δ_2 is set to $2 * \Delta_1$.

The next point E, at a distance Δ_2 from D will be a failure, and therefore the step size is reduced, and moves are made forwards and backwards (to the left and to the right) until another successful 'direction' is found. Eventually, the situation will arise where the step size has been reduced to its minimum allowable level Δ_n and f_Z will be obtained at $x = Z$, such that no move in either direction of Δ_n will improve the value. Therefore the maximum is reached.

Problems become immediately apparent. For example, if Δ is such that $f_C = f_A$ also, then the process will terminate in this vicinity. Also, if the function rises slightly so the f_B is just greater than f_A but then falls again, then the local optimum around f_B will be the termination point. Careful choice of Δ is therefore required, and a number of 'runs' using different starting conditions will reduce the effects of these problems.

The Ambulance Situation

In the ambulance situation we have n (say 4) variable units at A1, A2, A3, A4 as shown below.



In applying the Hooke and Jeeves technique, each ambulance is test moved in turn, first in the x direction, and then in the y direction, (that is, we have 8 possible moves to be tested sequentially, $A1_x$, $A1_y$, $A2_x$, $A2_y$, $A3_y$, $A4_x$, $A4_y$) corresponding to eight variables in a single dimensional case), and at each 'move' the objective function (system response time) is tested until a potential successful direction is established. After the 8 test moves, the ambulance is actually moved, and the Δ of which it is likely that there will be a different one for each of the eight 'moves' are adjusted. It can be seen that as the number of ambulances increases, the computation involved increases rapidly.

Detailed Procedure

A flow diagram of the subroutine has been drawn up, and the following explanation of PATS was obtained by following through the various possibilities. A list of the parameters involved is found at the end of the section.

The subroutine PATS uses as input the x and y co-ordinates of the ambulances. It does not differentiate between the x and y directions, nor the ambulances, except in terms of checking the boundaries of the system. Therefore, it could also be simply used for a number of one dimensional elements, or even three dimensional elements. The only required changes would be in the boundary checks and calculation of the D(I) increments, which are initially set as a function of the boundaries for each particular ambulance.

Initialisation: At the beginning of the subroutine, all the parameters are set to their initial values. Some of these parameters, for example LT2, LT3, LT4, LT5, LT6, LT7, ID1, ID2, ID3, ID4 simply show the brakdown of the routes the search proceeds through, and as such are not shown in the logic of the subroutine. However, their values are explained in the table of parameters at the end of this section.

The number of moveable 'elements', in this case, 2 times the number of ambulances, is entered as a formal parameter to the subroutine: (N = 2 * NAMB). The subroutine then works through, setting K = K + 1 through K = 1, 2 ... N, replacing K = N + 1 by K = 1.

After initialisation, subroutine OBJECT is called, and this returns a value of RBAR. SN, OLDSN, SP and SC are then all set to RBAR, initial value. SN will in future be the value of RBAR obtained from the previous 'move', where 'move' refers to the movement of a single 'element' (SN is tested against SP for 'success' or 'failure') and SC is the previous 'best position'. SC is only changed after all 'elements' have been moved and tested, during the adaptive logic sequence.

Boundary check: Element K is increased by D(K), set as a function of the boundaries of the element K. A boundary check is then made to ensure that the element lies within its own boundaries. If at any stage an element is found to have passed outside of its boundaries, then if LA = 2, 5 the move is reversed, LA is set to 3,6 and the control returns to the boundary check. In the case of LA = 3,6 then a failure is considered to have occurred and the step size for that element is reduced, before returning to the main line of the logic with the element restored, to consider that same element again.

Test SP: After the boundary check OBJECT is called again, and SN is set equal to RBAR, SP is tested against SN to see if the new value of the objective function is <u>less</u> than the old value. (Note that for a maximisation, the test must be reversed.) If successful then D(K) is multiplied by ALP = 2 if LA = 2,5 so that the next time that 'element' is incremented, the amount moved will be 2 times the previous shift. Or D(K) is set to -D(K) if LA = 3,6 and LA will be changed to 2 or 5 before the sequence returns again. That is, a successful direction for that element has been established (whether it be forward or backward).

Failure: If a failure occurs, then if LA = 2,5 (that is, a successful direction was established for the previous element), then LA is set to 3 or 6, the element is restored to its previous value, D(K) is unaltered and the logic returns to the boundary check, still working with the same element. If a failure occurs when LA = 3,6 then the logic enters the <u>reduction of step size</u> routine.

Note that before an element is quitted a successful direction must be obtained. That is, LA can only be equal to 3,6 when OBJECT is called, if that same element has just failed a boundary test, or if a 'failure' has occurred when LA was equal to 2 or 5. Before the next element is considered a direction must be established.

Once D(K) has been reduced by the reduction routine, then the control returns to the main line of the logic and considers the next element K + 1.

Base Point Test: For K + 1,N the procedure continues. KK is incremented at the same time as K. However, KK is allowed to reach N + 1. When KK = N + 1, then a <u>base point test</u> is made. SP, the current RBAR is tested against SC the previous 'base point', that is, the RBAR current when KK was last equal to N + 1.

Adaptive Logic: If the base point test is successful, and L4 = 1 then LA is set to 7, M1 is set to 1 and the procedure returns to CALL OBJECT with K = 1 and LA = 7. This also occurs if L4 = 2 and NPF less than 5. This means that after a new SN has been returned, the procedure moves into the adaptive logic sequence. This occurs when each element has been tested and a successful direction has been found, such that the value of the objective function for the new position of all the elements (SN) is less than the value of the objective function for the previous position of all the elements (OLDSN).

V is calculated as (OLDSN - SN) * 100 divided by OLDSN times the number of times OBJECT has been called since the OLDSN was established. That is, it is the percentage change in OLDSN. Depending on the size of V, GR is calculated. That is, if V is small GR is increased, and vice versa. OLDSN is set to SN, SP is set to SN, SC is set to SP, LA is set to 4 and X(I) is replaced by a function of GR, X(I) and Q(I), the value of X(I) the previous time the adaptive logic was entered (previous base point), for all I = 1, 2 ... N.

X(i) is replaced by Q(I) + GR * (X(I) - Q(I))

where

Q(I) is the previous base point value of X(I) and X(I) from the RHS is the value of X(I) obtained during the routine establishing a successful direction.

GR is greater than or equal to 2.2.

That is, the previous base point is incremented by GR times the move made establishing the direction (all previous successful moves are incremented).

X(I) boundaries are checked, and if an element has passed outside of its boundaries, then it is placed on the boundary.

Return: The logic returns to CALL OBJECT with LA = 4 so that SP = SN (the new RBAR), LA is set to 5, and we return to the boundary check, ready to begin a new sequence with K = 1 (KK also set to 1 during the adaptive logic).

If the base point test is not successful, L4 = 1 implies that L4 is set to 2, and if M1 is greater than N, we terminate. However, if L4 = 2, SP is set to SC, X(I) to Z(I), NPF = 0, and continue as for M1 less than N, M2 is set to 1, M1 = 1, KK = 1, L4 is set to 1, LA is set to 2 and we begin again with K = 1 at the boundary check.

Termination: Therefore, there are four conditions for termination.

(1) If the number of 'moves' becomes greater than the number set as a limit. A 'move' is made each time OBJECT is called to return a new value of SN. The test is made during the adaptive logic sequence, before the new values of the parameters are set.

(2) If the base point test, occurring when KK is greater than N is failed, L4 = 1, and M1 is greater than N. (M1 is incremented during the sequence reducing the step size, and restored to 1 at the completion of the base point test, therefore, M1 greater than N implies that all the elements have been through the step size reduction since the last base point test.)

(3) If an element is outside of its boundaries, and LA = 1,4,7,8 at this point.

(4) If LA = 8 after OBJECT has been called.

In fact, only two termination conditions apply, since it is impossible for LA to equal 1,4,5,8 at a boundary test, and since (4) is the result of (2), that is, if the conditions for (2) occur, then LA is set to 8, and we return to the point where OBJECT is called.

Thus (2) may be considered a normal exit.

PATS parameters

NEVAL = KEVAL	is incremented by 1 each time OBJECT is called.
KOUNT	is incremented by 1 (and written out with NEVAL, SN, V, GR) each time the adaptive logic loop is entered.
LT	is incremented each time OBJECT is called (after the first time) such that LT5 = LT5 + 1, if LA = 5,

the total of LT2 + LT3 + LT4 + LT5 + LT6 + LT7 = NEVAL - 1. Note also that LT7 = KOUNT.

LA = 2,5	implies that a successful move has been made and a 'direction' established.
LA = 3,6	implies that that element has just failed a boundary test, or has just failed a normal test, and has been restored. A direction has yet to be established. If it has just failed a normal test then the step size will be reduced for that element.
LA = 2,3	implies L4 = 1 implies that we are looking for a direction, and that we do not have one yet.
LA = 5,6	implies L4 = 2 implies that we have just been through the adaptive logic, and we have a successful direction, KK = 1, K = 1, and we begin again on the first element (which has been incremented by GR times the direction and test again).
LA = 8	implies termination.
NPF	keeps track of the number of times the step size is reduced within each sequence K = 1, ... N.
K	indicates the ambulance and direction of the current movement. K = N + 1 implies, K = 1, and that all elements have been tested and that a base point test should be made.
KK	is incremented as K, and is set back to 1 after the base point test or adaptive logic.
ID	is incremented during the adaptive logic routine, depending on the size of V and OLDV ID1 + ID2 + ID3 + ID4 = KOUNT.
SN	is the value of RBAR for the current position.
SP	is the value of RBAR for the previous move.
SC	is the value of RBAR for the previous base position.
LT, LSN	are zero throughout this use of PATS, and obviously relate to some other use of the subroutine.

The parameters printed out below the KOUNT table are:

NEVAL, KOUNT, LT3, LT4, LT5, LT6, LT7, K, KK, M1, M2, NPF, LA, L4, ID1, ID2, ID3, ID4.

2. Description of Input

The initial input parameters of the system are:

LAMDA	The average number of calls per day
CARE	The average on scene time, in minutes
TRSFR	The average transfer time at hospital, in minutes
TRANS	The proportion (in decimal form) of calls transported
CITY	The maximum travel distance for city speed
FWY	The minimum travel distance for freeway speed (note that FWY must be greater than CITY)

These parameters are written on to the first data card in FORMAT (10 F 8.3).

LIMIT	is the maximum number of search evaluations, before termination
NAMB	is the number of ambulances in the fleet
NHOSP	is the number of hospitals
NDIST	is the number of districts
NFRAC	is the number of one minute class intervals beginning at zero, for theresponse time distribution. (This can be determined by setting an initial arbitrary value, say 40, and checking the results.)

These parameters are written on to the second card in FORMAT (1015).

District Data

After the two parameter cards, information about the districts to be used is required. The co-ordinates of the districts are read in 10 F 8.3 format. Firstly all the x co-ordinates of the centroids of districts are read in, then all the y co-ordinates of the centroids. Therefore, for 82 districts, 9 + 9 = 18 cards will be required.

Notice that it is possible to have different sized 'areas', and that the numerical ordering of the districts is also unimportant, though the data must always be read for the same areas in the same order.

Then, the probability of a call arising from each district is entered in 10 F 8.6 format, so that the probabilities all sum to 1.0. The selection and size of the districts is largely dependent on these probabilities, and it must be remembered that the probability assigned to an area is used as being applicable to every point within that area (for the Hooke and Jeeves Pattern Search), and not just at the centroid.

Hospital Data

For each hospital a separate card is required in FORMAT (3A4, 2 F 8.3). The first 12 columns of the card provide an alphameric description of the site of the hospital, and this is followed by the x and y co-ordinates of the hospital. Notice that these co-ordinates are precise, and not necessarily the centroids of an area.

Ambulance Data

A card is also required for each ambulance in FORMAT (215, 8 F 8.3). This contains AMBNO, the ambulance designation number, usually 1,2 ... N then TYPE the vehicle type, which 1 denotes a land vehicle, and 2 denotes a helicopter, LOCX the initial location x co-ordinate, LOCY the initial location y co-ordinate, CSPD the average city speed for that ambulance, HSPD the average freeway speed for that ambulance, XMAX the maximum x co-ordinate, YMAX the maximum y co-ordinate, XMIN the minimum x co-ordinate, YMIN the minimum y co-ordinate.

3. Extension of CALL

As mentioned earlier in Section 3.1, it is often not convenient to use rectangular displacement between nodes as a basis for the calculation of the time taken to travel between these nodes. Therefore, CALL has been amended so as to accept a travel time matrix between nodes, instead of the parameters CITY & FWY. Therefore, if CITY is entered as zero on the first parameter card additional card input is required as follows.

(a) After the district data, x and y co-ordinates of the centroids of the areas, and before the probabilities by area, enter the maximum and minimum co-ordinate of each area as lower limit x, upper limit x, lower limit y, upper limit y in 12 F 6.3 format. That is, 3 areas per card.

This information is required to determine the node number given the co-ordinates of a point.

(b) After the ambulance data, that is, at the end of the card deck, enter the travel time matrix.

Data for each node begins on a new card.

TRAVEL (1,1) is set to 0, so no cards are required for node 1.

For node 2 enter TRAVEL (2,1) in F 4.2 format (minutes)
For node 3 enter TRAVEL (3,1) and TRAVEL (3,2)
so that for node r enter TRAVEL (R,1), TRAVEL (R,2) --- TRAVEL (R,R-1) and the number of cards required will be $\frac{R-1}{20}$ (+ 1 if R-1 not exactly divisible by 20).

TRAVEL (R,R) is set to 0.

When all the cards have been read in, TRAVEL (I,J) for those values not entered (the upper triangle of the matrix) is set to TRAVEL (J,I).

The use of the travel time matrix in determining an optimum has not been tested. The obvious problem in the case of the Fitzsimmons model is that now all travel times are being measured between one point in node A and one point in node B, for all A and B. The disadvantage comes from the use of DEL, used to calculate the D(I) step, size in the pattern search routine. The problem is, that if D(I) is less than one half of the distance across the area, then a shift of D(I) will have no effect upon RBAR, as the travel times from that station will be the same. Therefore, DEL must be adjusted so that the D(I) are greater than are half the distance across the areas.

D(I) = DEL * (XMAX(I) - XMIN(I))

In our example XMAX(I) - XMIN(I) is usually constant for all I at 20. Therefore D(I) = .1 x 20 = 2.0 which is equal to the distance across an area. Therefore, DEL is probably suitable.

4. Output

General Description

The first part of the output consists of a rewrite of the input data. This is very useful as a check, as it is written clearly and explicitly (e.g. LAMDA = etc.) and decimal places are printed in the centroid co-ordinates and the probabilities.

Next is written the KEY used for quick reference selection of the 'area' for the random number generator. This is followed by a list of the parameters involved in the Hooke and Jeeves Pattern Search. See the note at the end of this section for a full explanation of these parameters.

Then, the system parameters are reported. The first table is headed 'SYSTEM PARAMETERS' and gives the mean on scene care time, the mean transfer time at hospital, the probability of transport, the mean number of incidents per day, and the number of search points evaluated. All of these except the last are input parameters, listed in report form.

The second report table is headed 'HOSPITAL' and lists the hospitals, their locations, the number of cases handled per day, and the percentage of cases handled by each hospital.

The third table is headed 'AMBULANCES' with sub headings 'INITIAL LOCATION' and 'FINAL LOCATION'. Under each subheading is given the x and y co-ordinates, the mean response time corresponding to those co-ordinates, the number and percentage of calls per day for each ambulance and the percentage utilisation, which has been calculated as the total expected round trip time times the percentage load.

The fourth table is headed 'SYSTEM PERFORMANCE', and for each system state 1,2 ... N + 1, corresponding to 0,1 ... N busy ambulances is given the probability of that state occurring and the mean response time in minutes for that state, for initial location and final location.

Also for initial and final location is given the mean system response time (minutes), the mean time to hospital (minutes), the mean retrieval time (minutes) and the mean number in the system, which is RHO, the mean arrival rate/mean service rate.

The fifth table is headed 'DISTRIBUTION OF RESPONSE', and gives the distribution over one minute intervals from 0 to NFRAC. The values for this table are calculated at the same time as the response times for each system state and it is printed out for both initial and final conditions.

The sixth table is headed 'RESPONSE TIME BY DISTRICT'. This shows for each district the minimum response time and the maximum response time determined from the priority listing of ambulances for each district.

APPENDIX II

SWOVELAND, YUENO, VERTINSKY, VICKSUN MODEL

1. Program Description and Methodology

The main program begins by setting all the variable parameters to the default values, and then reads the NAMELIST card and adjusts the variables accordingly.

Next it reads INTERC, the alphameric names of the nodes from the file stored on unit 1, and sets them into a COMMON/INTER/block, for later use. The names are stored in numerical node order.

If NUM is greater than zero, that is, the input call stream has not already been created, it calls CALPRO, the subroutine which sets up the input files by generating time of call, location, destination, loading and unloading time, and cancellation time if the call is cancelled. This information is stored on unit 7 as file INPUT. File 4 is rewound (it contains the travel time information) and MSIM is called to perform the required simulation according to location information read off cards. MSIM calls SETUP, READS and/SIM which itself calls UPDATE and DPATCH. File 8 is created and the simulation output information about the calls is stored on it. File 8 is then given an ENDFILE and rewound in the main program.

Finally STATIS is called. This subroutine prints out the final reports and uses HIST to print out the required histograms. The computations for the output options are performed here too.

Because of the complexity of the simulation an explanation of the more important variables is given here, in the order in which they are defined in the main program. Whether they appear in NAMELIST or COMMON is also specified. This list should be particularly valuable for reference when studying the flow diagrams, at the end of the section.

Main Program Variables

D(NODES,NODES)	travel time matrix
PATH(NODES,NODES)	routing matrix
NAMB(NODES,NODES)	
PAR(6,j),j=1,NZ	input stream information computed in CALPRO and stored on unit 7
PAR(1,j)	time of call
PAR(2,j)	selected node for incident
PAR(3,j)	destination (hospital), = -1 if the call is cancelled
PAR(4,j)	travel time plus loading time, (cancellation time if PAR(3,j) = -1
PAR(5,j)	unloading time
PAR(6,j)	1 if a transfer call, 2 if a regular call, 3 if an emergency call
OUT(7,j),j=1,NZ	simulation output information about call, computed in SIM and stored on unit 8
OUT(1,j)	+ 1 if call is dispatch, 0 otherwise
OUT(2,j)	response time for call j
OUT(3,j)	number of the ambulance servicing the call
OUT(4,j)	node at which the ambulance is located when call j dispatched
OUT(5,j)	region at which call takes place
OUT(6,j)	waiting time for dispatch for call j
OUT(7,j)	1 if ANU, 2 if call cancelled (PAR(3,j) = -1)

Terminology

ANU call	a call cancelled after the ambulance has arrived at the scene
DISPATCH	action of assigning a particular ambulance the task of responding to a call
NORMAL call	regular priority call which is not a transfer call or emergency call
TRANSFER call	low priority call involving transfer of a patient, e.g. hospital to hospital
CANCEL call	call resulting in the dispatch of an ambulance which is cancelled before the ambulance arrives at the scene

EMERGENCY call	high priority call requiring immediate service (often implies a faster travel time)
USE(K)K=1,NZ	
.TRUE.	when call j (USE(j)) is assigned an ambulance
STAT(j,k)	not used

Main Common Statement Variables

* ALLOW	allowable delay for a transfer call before dispatch required (minutes x 10)
REL(j)j-1,NREG	time at which next ambulance update for ambulance j takes place, e.g. when j has delivered a patient to hospital and will next become available
TEMP	delay occurring when a queue has formed
KSTAT(j),j=1,NREG	0 if ambulance j at home, 1 if free, (but not at home), 2 if on way to a call, 3 if with patient
NA(i,j)	jth region in which node i is located
DD	response time for dispatched ambulance
NLOC(j)j=1,NREG	next location for ambulance j
NREG	number of ambulances (= number of regions)
NBAS(j)j=NREG	base location for ambulance j
NAMS(j)j=1,25	the regions to which node j belongs
LOC(j)j=1,NREG	node at which ambulance j is currently located
ASGN(j)j=1,NREG	the number of the last call assigned to ambulance j
* WIND	0 if same input stream is to be used for each run (i.e. unit 7 is to be rewound), otherwise (default = 0)
* LOOK	number of calls to read in addition to NCALLS (default = 10)
* IQ	1 if dispatch rule is nearest ambulance, 0 if dispatch is regional (default = 1)
* STUP	start up time for ambulance leaving base - minutes times 10 (default = 20, that is 2 minutes)

* FAST	fractional emergency call speedup (1 by default)
FREE	no. of free ambulances
NEXT	time at which the next ambulance update takes place

Other Variables

II(j)j=1,25	data on unit 9 used for regionalised response consists of the regions to which the node concerned belongs
NPBAS(j)j=1,NREG	new base position for ambulance j (read from cards as P)
* NUM	number of calls to be created for input stream by CALPRO
FH	number of free ambulances at their bases
FNH	number of free ambulances not at home
BUSY	number of busy ambulances
INTERC(i,j)j=1,9	alphameric description of node j
NZZ	maximum possible number of calls = 3000
NAX	maximum number of ambulances = 30
* NODES	number of nodes (default = 82)
* RATE	number of calls per hour (default = 0)
* CUTOFF	number of hours to wait before collection statistics (default = 0)
* SEED	seed for the random number generator (default = 0)

* Those variables marked with an asterisk are NAMELIST variables and values for these are read in by card (otherwise default value is as given).

Subroutine CALPRO

CALPRO sets up the input call stream for the main simulation and reads the travel time matrix off unit 4 and the call probabilities off unit 2. Then, it reads the probabilities of emergency calls, transfer calls and cancelled calls off unit 2 also.

A random number is selected, R1, and the functions IINOD, JJNOD, IAFLD, IAFUN, IACAN are initialised using this random number. Notice that R1 is a convenient notation not used in the program. Then for each call, 1,2 ... NUM the following procedure is worked through:

A random number R2 is generated and the time of call is calculated as:

$$T_L = T_{L-1} + \frac{e^{-R2}}{\text{no. calls/6 secs}}$$

where

$T_0 = 0$

so that the time of the Lth call is

$$T_L = \frac{\Sigma}{I=1,L} \quad \frac{e^{-R2}}{\text{no. calls/6 secs}}$$

and set PAR(1,L) equal to this. (L is the number of the incident.)

R3 is generated and a node for the location of the call is selected directly using INODE(R3). This is PAR(2,L). R4 is generated and used to determine the type of call, (PAR(6,L)). If R4 is less than the probability of a regular call (P_2) then it is designated a regular call, (where the probability of a regular call = 1 - prob. emergency - prob. transfer).

If P4 is greater than P_2 but less than P_4 (where P_4 = prob. transfer + prob. regular) then it is a regular call, and if R_4 is greater than P_2 and greater than P_3 then it is an emergency call.

R5 is generated so that if R5 is less than or equal to PCAN (where PCAN is the probability of a call being cancelled) then the call is cancelled, PAR(3,L) is set to -1 and PAR(4,L) is equal to the time of cancellation, determined by generating R6 and returning the function ICAN(R6).

If R5 is greater than PCAN then R7 is generated and PAR(3,L) represents the destination of the call (that is, the hospital) returned from JNODE(R7,PAR(2,L)). That is, it is dependent on the location of the call.

For a non cancelled (that is, PAR(3,L) L = -1), R8 and R9 are generated and used to calculate PAR(4,L), in this case the travel time plus the loading time (IFLD(R8)) plus the unloading time (IFUN(R9)).

All this information in the array PAR is written onto unit 7 to be used later for the main simulation. Disk files 4,2,7, are rewound.

CALPRO makes use of five functions IINOD, JJNOD, IAFLO, IAFLD, IACAN, which are later entered (after initialisation and summing of the required distributions) at ENTRY as INODE, JNODE, IFLO, IFUN, and ICAN.

They all use a cumulative distribution calculation to return a value for a random number.

IINOD(INODE)	returns a node number using the probability distribution of events
JJNOD(JNODE)	returns a destination given the location of the event, using the distribution of the possible destinations for that node
IAFLD(IFLO)	returns a loading time given the distribution of loading times
IAFUN(IFUN)	returns an unloading time given the distribution of unloading times
IACAN(ICAN)	returns a cancellation time given the distribution of cancellation times
CALPRO variables	(only those variables which do not appear in the main program)
I(j)	node number
M(j)	number of possible destinations from node j
P(j)	probability an incident will occur at node j
N(j,k)	kth destination from node j
AP(j,k)	probability an incident in node j will go to kth destination
II(j)	= j if all nodes have possible destinations, = jth node with a possible destination

III	number of nodes with possible destinations

(a node has a possible destination if N(j,k) for that node are not zero for all k)

FCAN	is the probability of a cancelled call
P3	is the probability of an emergency call
P1	is the probability of a transfer call
P2	(1-P1-P3) is the probability of a normal call
P4	(1-P3) is the probability of a non emergency call
IPR	(PAR(6,j)) type of call, 1 = transfer, 2 = regular, 3 = emergency

Subroutine MSIM

Subroutine MSIM calls three other subroutines, SETUP, READS and SIM. SETUP is used to read the travel time matrix D(NODES,NODES) and the path matrix PATH(NODES,NODES) from unit 4 and 3 respectively. If the response criteria is regionalised response it also reads REGDEF from unit 9 which contains the node number (nodes in numerical order) plus all the regions to which that area belongs.

READS reads the input cards for the ambulance locations for each run. Because of its importance for input, a full description of READS is given at the end of this section.

SIM performs the main simulation of calls and returns the system parameters to MSIM. A full description of this is also given.

MSIM therefore puts NCALLS, IN, LM to zero initially, then calls SETUP to return D and PATH. It then calls READS, which reads the parameter card V (which returns a value of NCALLS) and the ambulance location cards for the first run.

If NCALLS is zero then the routine terminates - a null operation. NZ is set to NCALLS + LOOK. LM is tested. If LM - 1, that is, the only card in the run which has been read contains an 'S' in column

1 then the routine terminates. This is the normal exit.

If LM = 0 and IN = 1, this implies the first run. LM = 0 and IN = 1 implies the second or subsequent runs, and if WIND = 1 (default) then unit 7 is rewound so as to use the same call stream as for the previous run. IN is set to 1 then (whether 0 or 1 previously) KSTAT(K) and LOC(K), status and current node for unit K are set to 0 for all K. FREE is set to NREG and NEXT the time at which the next ambulance update is due to take place is set to 1000000.

SIM is then called to perform the main simulation and to compute OUT, then NCALLS, NZ, IQ, STUP, ALLOW, NREG, FAST are written on to unit 8.

For each call j=1, NZ if OUT = 1, implying dispatch, OUT(k,j), k=2, 7 is set to zero and OUT(k,j) k=1, 7, the output from SIM for each call is written on to 8.

Then the logic returns to CALL READS again, and terminates only when LM is returned as 1, implying that all the runs have been completed.

MSIM variables

NA(IZ,J),J=1	NAMS(IZ) is the th region to which node IZ belongs
JNC is priority	(PAR(6,j))
KT	current call being dispatched
NCALLS	number of calls to be simulated for the run
NPT	call currently to be assigned
NF(j)	jth free ambulance
NZ	total number of calls to be simulated (NCALLS*LOOK)
IN	run counter

Subroutine READS

Each card is read as L1, L2, L3, REL(j),j=1,17 in format A1, I4, I4, 17A4.

In column 1 of each card there must be a letter read as A1. If this letter is V, then NCALLS is set to L2, and the logic returns to read another card. Therefore every data set should contain a card with V at the head.

If L1 is equal to P this signifies a permanent ambulance location for ambulance number L2 at node L3. For the nearest ambulance response rule NREG is put to zero in SETUP, therefore for the first card with P on it L2 will be greater than NREG and so NREG is put equal to L2. NPBAS(L2) = L3.

For example, card P 1 55 puts ambulance 1 at node 55 and another card is read. NREG is put equal to the largest value of L2 appearing on a P card. For the first run, all the ambulance locations should be specified on P cards.

If (in a subsequent run) L2 is equal to T, this indicates that the ambulance L2 should be repositioned at L3 <u>for this run only</u>. On following runs, ambulance L2 will be put back to the position it last occupied on a P card. Notice that on all runs after the first, those ambulances not specified on P or T cards will hold the position they last held on a P card.

If L1 is equal to R, this indicates the end of a run. Any further descriptive information in columns 10-67 will be written on to unit 8 (to be written out with the printed report for that run later as a heading), and the program returns from the subroutine to MSIM. That is, only one set of run cards is read each time READS is called.

If L1 is not equal to P, T, V, R, that is, it is equal to S (or an error has occurred), then LM is set to 1 and the program returns to MSIM. This implies termination of the program. (The last run has been made.)

Subroutine SIM

The simulation of calls in SIM proceeds as follows. Each call 1, NZ is considered sequentially.

(a) Consider a call to be dispatched. KT is incremented KT=KT+1, where KT is the call currently being dispatched.

(b) If the call currently to be assigned (NPT) has been assigned, that is USE(NPT) is .TRUE. then NPT is incremented NPT=NPT+1.

Then KTIME is set to PAR(1,NPT), that is, the time at which the call NPT occurs.

(c) If NEXT, the time at which the next ambulance update is due to take place is less than or equal to KTIME, then the update is performed by calling UPDATE, and NEXT is set equal to the value for which the next ambulance update is expected. Then return to (c) and text NEXT again.

If NEXT is greater than KTIME then the call NPT is considered.

(d) If no queue exists, that is, FREE is greater than zero then KNODE is set to PAR(2,NOT), USE(NPT) to .TRUE. and DPATCH is called to determine DD the response time for that call. (NB: nearest ambulance response means that the ambulance having the lowest response time of all the ambulances that are free at that time is sent.) Thus DPATCH returns DD the response time, and JDISP the number of the ambulance that is sent. KSTAT(JDISP), NLOC(JDISP,REL(JDISP) are then adjusted and if REL(JDISP) is less than NEXT, then NEXT is set equal to REL(JDISP).

OUT(1,NPT) = 1, and depending on the states of the ambulances the number of free and busy ambulances are adjusted.

OUT(2,NPT) = DD, OUT(3,NPT) = JDISP and OUT(4,NPT) = LOC(JDISP), OUT(6,NPT) = 0 and OUT(7,NPT) = 2 if a normal call and equals 1 if the call is cancelled and ANU. If KT less than NZ, that is, more calls are remaining to be simulated then the process returns to (a) and increments KT.

If at (d) a queue is found to exist then NEXT is adjusted to the next event due to occur and UPDATE is used (note that in the provisions of update if the time waited for a transfer call exceeds ALLOW then the call is cancelled). If a queue still exists (that is the update does not result in an ambulance becoming free) then branch back to (d) and repeat. Once a free ambulance is obtained (e) JNC is set to 0 and if a call remains that has not had an ambulance assigned to it find the first call with emergency priority. (If all calls occur at a later time than NEXT then return to (b) * also and retest. USE.)

Assign the ambulance which has become free to the emergency call (or one with highest priority) and set all OUT values plus REL, NLOC and KSTAT depending on whether or not the call is later cancelled. If KT is greater than or equal to NZ then return. Otherwise if there is still a queue return to (a) and increment KT or if the queue has been dispersed return to (e) and put JNC to zero. This may continue to revert to JNC = 0 until PAR(1,j) greater than NEXT for all j = NPT ... NZ in which case it returns to b (see *).

Subroutine STATIS

IRUN is initialised to zero and headings are written out for RATE, SEED, CUTOFF before IRUN is incremented as IRUN=IRUN + 1 ALLOW/10 and STUP/10 are written out (now in minutes) and if IQ = 1, which we have assumed throughout this description the ambulance positions as node number are alphameric description are written out. If IQ = 0 then a small routine works on the region and area number numbers and then also prints out ambulance positions (see flow diagram). In both cases, the dispatch rule is then written out as a reminder.

ZERO is called and this sets a number of the summation parameters to zero.

For each call i = 1,NZ, PAR(J,i), j = 1,6 is read off unit 7 as IT, N, JN, IT1, IT2, IPR and OUT (j, i), j = 1,7 is read off unit 8 as ID, IRES, JAMB, JRES, JREGN, IWAIT, NOTU. Statistics for each type of call are then calculated. There are performed as conditional summations. Statistics are printed out by call type and by ambulance. The first part of the standard output is a histogram of response time in minutes for all calls.

Then, for the first run the optional output cards are read. The options are HITR, HINM, HIEM (histograms for transfer, normal or emergency calls) CROS (cross over matrix for regional response) REGN (regional response only) SUBS (statistics related to specified nodes on the card for nearest ambulance response) and FINI (the termination card).

The options are read into an array which is used to determine the options for subsequent runs. Therefore the same options are used throughout the experiment.

FINI must always be the last options card, at which point unit 7 will be rewound (since PAR is the same for all runs), the end of run messages will be printed out and the logic returns to increment IRUN again.

The normal exit from the subroutine and thence from the program occurs when no further run card information is available on unit 8. That is, an end of file message is received (see main program ENDFILE statement).

Notice that if x runs are made, and the same call stream is used (NZ calls), then there will be NZ records of information read as PAR from unit 7 (though in fact there will be NUM which must be greater than NZ calls stored for use), but there will be x times NZ record of ca-l information read from unit 8 as OUT, which contains different information for all x runs.

Subroutine HIST

This is general subroutine which prints out a bar histogram. The x axis labels in the original subroutine were printed out using a variable format statement, where the format was built up by manipulation of the bits of variable FMT used to represent the format. On the B6700 it is not easy to duplicate this type of manipulation and so the labels are simply printed across the bottom of the histogram and must be positioned correctly one each below each of the vertical lines) by the analyst.

Subroutine MEAN

This subroutine returns the mean and sample standard deviation of the x (i) as AV and SD, given an input of

$$AV = \sum_{i=1}^{F} x(i)$$

the sum of the variables

$$SD = \sum_{i=1}^{F} x(i)^2$$

the sum of the squares of the variables, and F, the number of variables.

2. Input

Tape Files

These notes should be read in conjunction with the Simulation Model Users Guide. They have been rewritten here to clarify the relation to the B6700 requirements, for which they have been set up.

In order to run the ambulance simulation, a number of input tape files relating to the performance of the system must be set up. These are accessed by a series of file declarations statements of the form.

FILEƀƀ1 = (filename), UNIT=DISK, RECORD = i, BLOCKING = j [1.]

at the beginning of the program, after the files AFIL, BFIL, CFIL, DFIL have been read off tape onto disk (see workflow). These statements have been patched into the original program. One statement is required for each file.

AFIL

The first of these files is called AFIL for the B6700 system and it contains the alphameric names of each node (in nodal numeric order) in 9A4 order, thus allowing 36 columns for each identification. The internal name of this file in the program is INTERX and it is

---

1. ƀ denotes a blank space

accessed as unit 1. The form of the file declaration statement is

FILEɃɃ = AFIL, UNIT = DISK, RECORD = 14, BLOCKING = 15

RECORD = 14 implies that the file is stored in card image form with each record having 14 words or 6*14 = 84 characters (card has 80 characters).

Storing AFIL

In order to store this file a small program has been written to read cards containing the alphameric identification (one per card) in columns 1-36 and to create a disk file. The disk file is locked by a statement ENDFILE 1 followed by LOCK 1, and then File AFILE is copied to tape. The corresponding file declaration at the beginning of this program must contain an extra attribute, AREA = i*j where i is equal to the number of nodes + rounding to a whole number divisible by the BLOCKING parameter (here 15), and j = 1, that is, if there are 82 nodes, AREA = 90*1. Notice that the maximum number of nodes possible is 150 therefore j will always be equal to 1.

BFIL

The next file is BFIL and the information on this file is used to route the ambulance back to its base after a call. It is used primarily for regional type ambulance response. The data is entered as a matrix of dimension (NODES*NODES), row-wise in 26I3 format.

That is, if there are 82 nodes, then the first 82 entries R(1,j) j = 1, 82 will require $\frac{82}{26}$ = 3 plus 1 record, making a total of 4 records (or file rows).

The next row of the matrix R (2,j), j = 1, 82 will begin on a new record or file row thus meaning that a total of 4*82 = 328 records will be required.

That is, the form of the file declaration statement in the simulation program will be

FILEɃɃ3 = BFIL, UNIT = DISK, RECORD = 14, BLOCKING = 15

Since again the records are stored in card image form.

Calculation of BFIL

Consider the information stored on BFIL, (R(i,j), j = 1,n) i = 1,n) R (I,1) is the next adjacent node to I on the route to node 1 from I and this will take a time $T_{I,1}$ to travel (see later travel time matrix CFIL) R(I,j) is therefore the next adjacent node to I that an ambulance will go to when travelling from node I to node j.

In fact the nodes need not necessarily be adjacent, as the time characteristic will take care of the total travel time which will be sum of the total steps made.

For example:

An ambulance has dropped a patient at node 56 and now wishes to return to its home base at node 62. (Route is 56, 57, 59, 62) - Find

R (56, 62) = 57 time $T_{56, 57}$

R (57, 62) = 59 time $T_{57, 59}$

R (59, 62) = 62 time $T_{59, 62}$

R (62, 62) = 0 time $T_{62, 62} = 0$

and total time take = $T_{56, 57} + T_{57, 59} = T_{59, 62} + T_{62, 62}$.

For the Vancouver test study it was assumed that an ambulance replied to a call only from a hospital or from its home base. That is, the ambulances are all returned home in a single step, rather than node by node, and therefore each row of ROUTE (the internal name of BFIL) was identical and equal to

1, 2,..................26,
27, 28,................52,
53, 54,................78,
79, 80, 81, 82

For example, consider a unit at node 8 wishing to return to node 25, Row 8 will be 25, and the return journey is made in one step.

Storing BFIL

A small program to enter the values of ROUTE into BFIL has been written, so as to read the values off card and store them as for AFIL. For the storing program the file declaration will be of the form

FILEƀƀ3 = BFIL, UNIT = DISK, RECORD = 14, BLOCKING = 15, AREA = i*j

i = NODES* number of cards required per row, that is the total number of file rows.

If NODES = 82 then i = 82* 4 = 328, however, this must be rounded to be divisible by 15 (BLOCKING) and so put i = 330 and j is equal to 1.

Notice that if $i * \frac{\text{RECORD}}{\text{BLOCKING}}$ is greater than 500, it is desirable to change i, so as to make this quantity which will be called ii less than 500.

That is, if NODES = 150 say, then i = 150 * 6 = 900

now,

$$900 \frac{*14}{15} = 840$$

so put,

i = 450 and j = 2 (thus i*j = original i).

CFIL

The third file is CFIL and it contains the travel time matrix which has the internal name of TRAVEL (or D). It is entered row-wise in 16I5 format (again in card image) and therefore requires NODES/16 file rows plus 1 if NODES is not completely divisible by 16 for each row of the matrix. The unit of time is .1 minutes, so that one minute is entered as 10.

TRAVEL is stored as unit 4 and the file declaration card in the main program has the form

FILEƀƀ = 4, UNIT = DISK, RECORD = 14, BLOCKING = 15

The values may be calculated in any fashion required, and if accurate information is not available, it may be necessary to compute the travel time matrix as a function of the rectangular displacement between nodes. Note that in this case, the whole matrix must be entered.

TRAVEL (i,i) will be equal to 0 (unless some sort of internal median travel time is required) but TRAVEL (i, j) need not necessarily be equal to TRAVEL (j, i) (consider the idea of a one way street section where traffic in one direction may need to travel further and therefore take longer than traffic in the other direction).

Storing CFIL

A small program to read off card the matrix and store it as CFIL on tape has been written, and it requires a file declaration of the form

FILEȻȻ3 = CFIL, UNIT = DISK, RECORD 14, BLOCKING = 15, AREA = i*j

i = nodes*times the number of cards required per row of the matrix, that is, the total number of file rows.

If NODES = 82, then i = 82*6 = 492 plus rounding to a suitable number makes i to 510. Put j equal to 1.

$ii = i * \frac{\text{RECORD}}{\text{BLOCKING}} = 476$ and this is less than 500, so is satisfactory.

However, if NODES were set to 150, then i would be 150 * 10 = 1500 and $1500 * \frac{14}{15} = 1400$, which is about three times 500.

Therefore set i to 510. 1500/3 plus rounding to make it divisible by 15 and put j = 3 thus i*j = 1530 which allows sufficient storage for all the file rows.

DFIL

The fourth file contains all the probability distribution information and is stored as DFIL on unit 2. There are several sections of this file, which are stored in different formats, and it is not stored in card image form.

First Section

The first section of the file is the probability distribution for each of the nodes. There is one record per node, and this is in format 2I2, F7.5, 20I2, 20F7.5.

Format	Characters	Contents
I2	1 - 2	node number (numeric ordering 1,2,.. .. NODES)
I2	3 - 4	number of allowable destinations (hospitals) for that node
F7.5	5 -11	probability of a call arising from that node
I2	12 -13	node number of 1st possible destination
I2	14 -15	node number of 2nd possible destination
etc.		
I2	50 -51	node number of 20th possible destination
F7.5	52 -58	fraction of calls going to 1st possible destination
F7.5	185 -191	fraction of calls going to 20th possible destination

For example:

21ƀ5ƀƀ01500ƀ2ƀ9152975 ƀƀ50000ƀƀ30000
ƀƀ25000ƀƀ04000ƀƀ01000 (7*15 = 105 blank spaces)
making a total of 191 characters.

This would indicate that there is a 1.5 chance of a call arising at node 21, that there are 5 possible destinations to which a call from node 21 could be taken, these are nodes, 2, 9, 15, 29, 75, and the respective probabilities of going to these nodes are .5, .3, .15, .04 and .01.

Second Section

The next section of DFIL contains three file rows each of format E20.5. They contain information about the types of calls.

a) probability of a cancelled call PCAN
b) probability of an emergency call P1
c) probability of a transfer call P3

Notice that PCAN + P1 + P3 is not equal to 1.0 as might be thought initially. In fact P4, the probability of a normal call, must be introduced and this is such that P1 + P3 + P4 = 1.0.

Third Section

The third section of DFIL contains the distributions of the loading times, the unloading times and the cancellation times respectively.

a)	maximum possible loading time (= LT)	I2
b)	fraction of calls completing that activity for 1,2,3, ... 10 minutes	10F10.5
	fraction of calls completing that activity for 11, 12, ... 20 minutes	10F10.5
	until LT minutes is reached.	

The number of file rows required for a) plus b) will therefore be 1 + (LT/10) if LT is an integral multiple of 10, otherwise

1 + (LT/10) + 1

Similarly sections a) and b) are repeated with the distributions of the time taken to unload patients at hospital, and for the times at which cancellations of calls may occur.

Notice also, that the distributions of the times must sum to 1.0.

The file declaration for the main program is

FILEbb2 = DFIL, UNIT = DISK, RECORD = 37, BLOCKING = 30

RECORD = 37 as the largest number of characters per file row is equal to 191 from the first section of the file and 37 * 6 = 222 (this value of 37 was discovered by experimentation with the file blocking). Notice that RECORD * BLOCKING must always be divisible by 30. When RECORD was equal to 14, BLOCKING was put to 15, but now that RECORD is 37, BLOCKING must be equal to 30 for this to hold.

Storing DFIL

For the storing program which reads the required information off cards, adjusts the row sizes and stored in on tape as DFIL, the file declaration requires an extra attribute of AREA = i*j.

This is a little more complex to work out than in the previous instances.

Consider an example: NODES = 82, maximum loading time = 30 minutes, maximum unloading time = 25 minutes, and maximum time for cancellation = 15 minutes.

There will be 82 + 3 + (1+3) + (1+3) + (1+2) equal to 96 file rows. (Round to 120 as 120/30 = 4) or in general terms NODES + 3 + (LT/10 + 2) + (UT/10+2) + (CT/10+2) (allowing possible extras for LT greater than (LT/10) * 10).

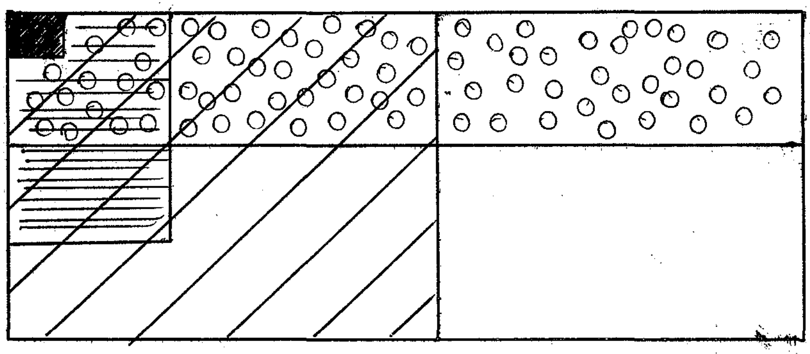
Put i - 120 in the specific example: Now, ii-i*37/30 = 148 and this less than 500, so we may use i = 120 and j = 1.

If ii were greater than 500, then adjustments as seen for the earlier files would have to be made.

If the regionalised response rule is to be used, then another file with internal name REGDEF is required to be stored on tape and accessed as unit 9. It has format 26I3 (card image) and contains:

I3 cols 1 - 3	Node number of the area
I3 cols 4 - 6	Code number of the 1st region to which that node belongs
I3 cols 7 - 9	Code number of the second region to which that node belongs
.	
.	
.	
I3 cols 76 - 79	Code number of the 25th region to which that node belongs.

This explains the possibility of overlap of regions. Here it is seen that any node may belong to up to 25 different regions at the same time.



Region 1

Region 2

Region 3

Node 1

Node 1 may therefore be serviced by ambulances from regions 1, 2& 3.

Internal Program Files

During the operation of the program, two output disk files are created. These are stored as INPUT on unit 7 and OUTPUT on unit 8.

INPUT

INPUT is created by CALPRO and contains the call stream for the main simulation. The number of calls stored is NUM and for each call, the time, location, destination, priority, type and loading, unloading and cancellation times (if relevant) are obtained and stored. The file declaration in the main program is

FILEƀƀ7 = INPUT,UNIT=DISK,RECORD=6,BLOCKING=25,AREA=i*j

i = NUM + rounding, j = 1

Consider NUM = 175; this implies that i = 175 (divisible exactly by 25) and ii = $175 * \frac{6}{25} = 42$.

Now the maximum number of calls that it is possible to simulate is 3000, in which case i would equal 3000, and ii would equal $3000 * \frac{6}{25} = 720$ and it might be considered desirable to amend j to 2 and put i to 1500.

However, for most circumstances it can be seen that it is likely that it would be satisfactory to put AREA = (NUM + rounding) * 1.

OUTPUT

OUTPUT is created by MSIM and is stored to be recalled by the report writing subroutine STATIS. It contains the output characteristics of the calls, after simulation.

There are several sections of information stored on unit 8. The first section is alphameric information which is read off the run card (R card) from columns 10 - 67. It is in the nature of a header card and is printed out at the head of each run.

The second section of information is the System parameter information of NCALLS, NZ, IQ, STUP, ALLOW, NREG, T, FAST, (NBAS (j), j = 1, NREG) which is stored with format 2I4, I1, I3, I4, I2, 2F4.2, 30I2 making a total character length of 66 characters.

This is followed by a file record for each call NZ containing OUT (i,j), j = 1,- which are the call characteristics for call i. These are stored with format I1, I4, 3I4, I4, I1 making a total of 22 characters. Thus the maximum record length comes from the first section which has 68 characters.

For each run made, all these three sections of file records are required. Therefore, for a total of R runs and NZ calls simulated in each, the number of file records will be

R * (2 + NZ)

The file declaration in the main program therefore has the form

FILEɃɃ8 = OUTPUT, UNIT = DISK, RECORD = 90, BLOCKING = 9, AREA = i*j
where,

i = R* (2 + NZ) plus rounding.

Consider,

R = 3, NZ = 150;

then

$$i = 3*\ (2 + 150) = 456,\ \text{plus rounding equals } 459$$

and

$$j = 1$$

now

$$ii = i* \frac{90}{9} = 4590.$$

Therefore consider $i = 54$ and $j = 10$, ii will now be equal to 540, but this is close enough to 500 not to be of concern. Notice that i (original) has been increased from 456 to 540. This has been done so as to allow the new i to be divisible by 9 ($9*6 = 54$).

File Attributes

There may be some confusion as to the reasoning behind some of the values determined for the file attributes. For the four input files, some difficulty was experienced in obtaining the correct form of the data on the test files supplied, and these attributes are related to the form in which the original test data was received. It is likely therefore that some better arrangement be made, however, for the moment, it is likely that the given characteristics will perform satisfactorily and reasonably efficiently. Likewise with the two internal files, the attributes might well be adapted, but it is considered that the given forms will be no worse than most people would obtain.

Changing the File Attributes

It will be necessary to change the file attributes in the main program for problems of different dimensions.

The given file declarations are

```
FILEƀƀ1 = AFIL, UNIT = DISK, RECORD = 14, BLOCKING = 15
FILEƀƀ2 = DFIL, UNIT = DISK, RECORD = 37, BLOCKING = 30
FILEƀƀ3 = BFIL, UNIT = DISK, RECORD = 14, BLOCKING = 15
FILEƀƀ4 = CFIL, UNIT = DISK, RECORD = 14, BLOCKING = 15
FILEƀƀ7 = INPUT, UNIT = DISK, RECORD = 6, BLOCKING = 25, AREA = 175*1
FILEƀƀ8 = OUTPUT, UNIT = DISK, RECORD = 90, BLOCKING = 9, AREA = 54*10
```

Files 1, 2, 3, 4 will not need to be changed, because they are already created (AREA = i*j must be changed for the individual storing programs). For files 7 and 8, the AREA attribute must be changed for each different experiment. In order to do this, it is not necessary to patch new file declarations in to the source deck, as these attributes may be changed in the workflow language.

Workflow

The workflow required to copy the files from tape and operate the simulation program using the object deck is as follows.

```
? JOB SIMULAMB  (or any other desired name)
? USER LINCO22/ <PASSWORD>                        <user password>
? BEGIN
? COPY D64/ = FROM D64
? CHANGE D64/AFIL TO AFIL
? CHANGE D64/BFIL TO BFIL
? CHANGE D64/CFIL TO CFIL
? CHANGE D64/DFIL TO DFIL
? RUN D64/OBJECTUYENO    -   run card
? DATA FILES
? REMOVE (LINCO22)
? END JOB
```

In order to change say the AREA = 175*1 in file declaration for file 7 to AREA = 200*1 put a card after the run card of the following form

```
? FILE FILE7 (TITLE = INPUT, AREA SIZE = 200, AREAS = 1)
```

The remaining characteristics will remain as in the original file declaration. This makes it little easier to understand what is occurring also, as it makes it clear that by putting AREA = 200*1 we are stating that we require 1 section of core only, and it will contain 200 records.

Card Input

The first card contains values for those variables which are required to take values other than the default options assigned.

That may be any of the following variables:

IQ	dispatch rule parameter default option IQ = 1 implies nearest ambulance response otherwise IQ set to 0 implies regionalised response. If IQ = 0 then a further input file must be set up on unit 9.
LOOK	default = 10 is the number of calls which will be simulated in addition to NCALLS, the number read on the V card.
STUP	default = 20 is the time (in minutes times 10) taken for the ambulance to leave its base.
FAST	default = 1.0 is the fractional speed up for emergency calls
WIND	default = 1 implies the input stream of calls is to be rewound for each run, otherwise the value of NUM must be sufficient to allow a different call stream for each run.
NUM	default = 0 implies that the input stream has already been calculated and stored on unit 7 otherwise, NUM must be set to the number of calls required to be generated and stored on unit 7.
RATE	default = 0, is the number of calls per hour
SEED	default = 0 is the seed for the random number generator
CUTOFF	default = 0 is the number of hours to wait before counting statistics
ALLOW	default = 100 is the allowable delay before a transfer call must either be dispatch or disgarded (minutes * 10)
NODES	default = 82 is the number of nodes or areas

Any of these parameters which require changing are entered on a NAMELIST card.

Card 1

Col 1	blank
Col 2	&
Cols 3 - 7	BEGIN
Cols 9 - 80	variable names and values ending with &END

For example:

b&BEGIN SEED=4, STUP= 24, NUM = 175, WIND = 1.5 &END

Card 2

This card contains the number of calls required for the simulation procedure, that is NCALLS.

Col 1	V
Cols 2 - 5	value of NCALLS (must be less than NUM-LOOK)

Card Set 3

The next cards set up the locations for the ambulances. These will be the permanent or base positions and there must be a card for each ambulance.

Col 1	P
Cols 2 - 5	the number of the ambulance (numerical ordering)
Cols 6 - 9	the node number of the base position of that ambulance

Card 4

The run card, indicating the end of the run.

Col 1	R
Cols 10 - 67	Header information to be printed out at the beginning of the run

If a second run is required then use:

Card Set 3a

This time it is only required to put in ambulance location cards for those ambulances which it is desired to place in positions other than the base positions specified in card set 3. If it is desired to shift an ambulance simply for one run, and then to have it return to its base position for further runs, then put a T in column 1. If, however, it is desired to create a new base position for the ambulance then put a P in column 1.

Card set 3a must be followed by a card 4, as must and further location sets such as card set 3b. There may be as many of these varying location sets as required, provided they are each followed by a card 4.

Card 5

This card signifies that all the required runs have been specified.

Col 1 S

These location cards are followed by the Options cards. The required options are specified by four letters in columns 1 - 4.

Options Cards

HIEM	indicates it is required to print a histogram of the mean response time for emergency calls
HITR	histogram for transfer calls
HINM	histogram for normal calls
SUBS	with node numbers in cols 5 - 7, 8 - 10, 11 - 13 ... 77 - 79 indicates that the means and standard deviations for the collection of nodes specified will be calculated.. It is possible to have more than one SUBS card.
REGN	with a collection of node numbers specified as in SUBS. This option produces the same results as SUBS for a regionalised dispatch rule.
CROS	this is used with regional dispatch only, also, and gives the fraction of calls which occur in each region.

As many or as few of these as is desired may be used, provided they are followed by the last card.

Last card

Cols 1 - 4 FINI

A typical deck might then be:

```
ƀ&BEGIN NUM=175, RATE=1.0, STUP=24, SEED = 4, NODES = 100  &END
Vƀ150
Pƀƀƀ1ƀƀ55
Pƀƀƀ2ƀƀƀ2
Pƀƀƀ3ƀƀ13
Pƀƀƀ4ƀƀ18
RƀƀƀƀƀƀƀƀFIRST RUN
Tƀƀƀ1ƀƀ27
Pƀƀƀ2ƀƀ30
RƀƀƀƀƀƀƀƀSECOND RUN
Tƀƀƀ3ƀƀ35
RƀƀƀƀƀƀƀƀTHIRD RUN
S
HIEM
SUBSƀ12ƀ13ƀ14ƀ15ƀ16
FINI
```

This provides for a first run with the four ambulances at nodes 55, 2, 13, 18 respectively. In the second run ambulances 1 and 2 are shifted to nodes 27 and 30, whilst 3 and 4 remain at 13 and 18.

In the third run ambulance 3 is shifted to 35, ambulance remains where it has been throughout at 18, ambulance 1 returns to its base position at 55, and ambulance 2 stays at its new base position at 30.

3. Amendments to the Program

In order to run the simulation program on the B6700 it was necessary to make a number of changes. Some of these have already been mentioned, but they are repeated here:

1. File declarations for internally created and external data files were inserted at the beginning of the main program.
2. Some dimensions were increased to allow for full generality. Maximum number of nodes in 150 and maximum number of calls is 3000.
3. NODES was added as a variable to the NAMELIST. Previously, if

any number of nodes other than 82 was required then several changes had to be made to the program. Now, nodes is entered as a variable (default = 82) and no program changes are required, unless NODES is greater than 150.

4. File 8 is given an ENDFILE, as it requires this for termination of the call simulation.

5. One of the major changes was concerned with the random number generator. For the B6700 it is not necessary to use a subroutine to call the random number generator. Therefore, the subroutine RANDOM (X,A) is deleted and the statements CALL RANDOM (X,0) are replaced by X = RANDOM (ASSED) where ASSED is initialised to SEED.

6. Similarly the subroutines TIME1(T) and TIME2(T) are deleted. It was not considered necessary to replace them.

7. The partial word statements which appeared in the orginal program in the form e.g. INTEGER * 2 TRAV (NODES, NODES) were replaced by INTEGER TRAV (NODES, NODES). It is assumed that the reason for these statements was a space saving one. The B6700 ignores the partial word form, printing out a syntax warning and because of the memory structure is not concerned with space saving. It should be noted also that both integer and real variables are stored in the same way.

8. The remaining changes occurred in the subroutine HIST where as it has been mentioned already a bit manipulation function HCREP(L) was used to provide the format for printing out the axis labels. HCREP (L) has been deleted and HIST modified so that the labels are simply printed in standard format. The labels correspond to each of the vertical lines of the histogram.

4. Proposed Extensions

It is considered that it would be a useful extension to the simulation model to link it to the Hooke and Jeeves pattern search routine PATS. In order to do this it is necessary to suppress the print out of statistics at each stage.

The required changes have been documented and one outlined in Appendix IV. They are stored on a stack of 'patch cards' for the program D64/SOURCEUYENO but have not been satisfactorily tested. Notice that the same reservations concerning DEL and the travel time matrix entry suggested for Fitzsimmons model also apply.

APPENDIX III

POSSIBLE ALTERNATIVE APPROACHES

From a preliminary look at the problem it seemed that it might be possible to use some form of Dynamic Programming or Markov Chain approach to the problem. This belief was further fostered by the discovery of the paper by Hall (12) who used a semi-Markov process to derive steady state probabilities for a study of Detroit, Michigan.

However, after some manipulation of the problem it was determined that there is really no simply way of 'fitting' the problem into a Markov process framework, nor is it really desirable to apprach a problem with the object of forcing it into a particular framework.

The inherent feature of a Markov process for these purposes is that the probability of a transition to state j during the next time interval, given that the system now occupies state i, is a function of i and j and not of any history of the system, before its arrival in state i and that these probabilities p_{ij} may be specified.

Therefore

$$\sum_{j=1}^{N} P_{ij} = 1, \text{ for all } i, 0 \leqslant p_{ij} \leqslant 1$$

where there are N system states.

Markov Process Application

A Markov Process steady state determination might be used to calculate the mean response time for the system. To do this it is necessary to compute the transition matrix.

The state variable is the number of busy ambulances at a point in time. N is the total number of ambulances. An event is defined as a change in the system state.

Originally the situation will be limited, so that two events may not occur simultaneously, that is, two ambulances leaving their station at the same time must be partitioned into two events.

P (i, j) is an element of the transition probability matrix P and represents the probability of the system moving from state i to state j when an event occurs.

Therefore

$$p(i, j) = 0 \qquad \text{for all } i = j$$

and for the first row of the matrix

$$p(1, j) = 0 \qquad \text{for } j = 3, 4 \ldots N$$

implies

$$p(1, 2) = 1$$

For all other rows the only non zero elements are

$$p(i, i-1) \text{ and } p(i, i+1)$$

(cf. random walk).

This will not provide any information about the frequency distribution.

Therefore, consider an event to be defined as occurring at a specific time interval. That is, the system is to be checked at minute intervals, and the probabilities to be defined according to the changes that have occurred in that period.

Consider the system in state i, and it is required to compute p(i, j) for j=1, N.

An essential feature of the Markov Process for steady state determination is that the p(i, j) are independent of the history of the system. But if at time t (events defined at t = 1,2 ... T) probabilities p(i, j) will depend on how long ago those i ambulances became busy. Thus, the transition probabilities are not independent of the history of the system.

A semi Markov process analysis with a two dimensional state variable (police and ambulance vehicle deployment) was used by Hall (12). Travel distance by vehicles were restricted and this enabled the service time distribution to be well defined and the techniques used to determine the transition probabilities were given by E.H. Moore and R. Pyke, Estimation of the Transition Distributions of a Markov Renewal Paper - Boeing Scientific Research Laboratories, 1964. An attempt is being made to obtain this paper.

Another Markov Process type application has been noted in Larson (18), a very mathematically complex paper.

Dynamic Programming Application

The principle of optimality used in dynamic programming is that an optimal policy has the property that whatever the initial state and initial decisions are, the remaining decisions must constitute an optimal policy with regard to the state resulting from the first decision.

Dynamic programming of location allocation system is discussed by A.J. Scott, Combinatorial Programming, Spatial Analysis and Planning where it is used for the successive optimal location of a given number of facilities. The stages are thus defined as the successive locations of each additional facility. Using this approach, it appears that the major problem involved is that ambulances are mobile and not always at their station, in which case, the same objections as those applied to Kuehn and Hamburger must be accepted.

Attempts were made to formulate the problem using different definitions of state variables, and stages, but it seems difficult to avoid the problem of ambulances being able and required to work outside their own 'area'. The problem may be conceptual only. A method of determining the mean response time for each state and stage is still required, and future development could be in the Markov process area.

Larson in a paper Models for the Allocation of Urban Police Patrol Forces, M.I.T.O.R. Centre Technical Report No. 44, 1969, apparently used a dynamic programming apprach.

APPENDIX IV

ADAPTIONS TO THE SIMULATION PROGRAM FOR USE WITH THE HOOKE & JEEVES PATTERN SEARCH SUBROUTINE.

PATS - nearest ambulance response

1. Amend Main Program

```
      COMMON/ONE/CUTOFF                                          00016001
      COMMON/TWO/AMAX,AMIN,BMAX,BMIN
      COMMON/THREE/XU(400),XL(400),YU(400),YL(400)
      DO 1010 J=1,NODES                                          00046001
      READ (5,212)XU(J),XL(J),YU(J),YL(J)
  212 FORMAT(4F10.0)
 1010 CONTINUE
      READ (5,212)AMAX,AMIN,BMAX,BMIN
C
C     ONE CARD PER NODE WITH UPPER X, LOWER X, UPPER Y
C   LOWER Y - AREAS MUST NOW BE RECTANGULAR
C   AMAX AND AMIN GIVE UPPER AND LOWER X VALUES
C   BMAX AND BMIN GIVE UPPER AND LOWER Y VALUES
C
```

Some conditional factor might be used to indicate that an optimisation is required.

2. Amend MSIM

```
      COMMON/ONE/CUTOFF                                          00269001
      COMMON/TWO/AMAX,AMIN,BMAX,BMIN
      COMMON/THREE/XU(400),XL(400),YU(400),YL(400)
      COMMON/FOUR/XMAX(40),XMIN(40)
      COMMON/FIVE/NREG
      DO 1010 J=1,20                                             00296001
      XMAX(J)=AMAX
      XMIN(J)=AMIN
      JJ=J+20
      XMAX(JJ)=BMAX
      XMIN(JJ)=BMIN
 1010 CONTINUE

       CALL PATS(NREG)                                           00320100
```

3. Create a new Subroutine STAT(TOTS) containing the essential parts of STATIS called by OBJECT. TOTS is the simulation program equivalent of RBAR.

```
      SUBROUTINE STAT(TOTS)                                        00337001
      COMMON/ONE/CUTOFF
      COMMON/FIVE/NREG
      ISTAT=600*CUTOFF
      READ(8,5,END=200) (DESCR(I),I=20)
    5 FØRMAT(20A4)
      READ(8,7)NCALLS,NZ,IQ,STUP,ALLOW,NREG,
               FAST
    1 (DESCR(J),J=1,NREG)
    7 FØRMAT(2I4,I1,I3,I4,I2,4X,F4.2,20I2)
      DO 20I=1,NZ
      READ(7,21)IT,N,JN,IT1,IT2,IDR
   21 FØRMAT(I6,2I2,2I4,I1)
      READ(8,25)ID,IRES,JAMB,JRES,JREGN,IWAIT,NOTU
   25 FORMAT(I1,I4,3I2,I4,I1)
      IF(JN.GE.0)NOTU=0
      IF(IT.LT.ISTAT)GØTØ20
      IF(ID.EQ.0) GO TO 20
      A1=IRES/20.
      A2-(IRES+IT1)/10.
      A3=A2+(IT2-IWAIT)/10.
      IF(JN.EQ.-1)GØTØ20
      IF(IPR.EQ.1)GØTØ26
      IF(IPR.EQ.2)GØTØ27
      NM(3)=NN(3)+1
      AV(3)=AV(3)+A1
      GO TO 20
   26 NM(1)=NM(1)+1
      AV(1)=AV(1)+A1
      GO TO 20
   27 NM(2)=NM(2)+1
      AV(2)=AV(2)+A1
   20 CØNTINUE
      TOTN=0
      DO 82 L=1,3
      F=NM(L)
      TOTN=TOTN+NM(L)
      IF(F.GT..5)TOTS=TOTS+AV(L)
   82 CONTINUE
      TOTS=TOTS/TOTN
      WRITE(6,1010)TOTS
 1010 FORMAT('MEAT RESPONSE IS', F8.2)
      CALL EXIT
      END
```

4. Create a new subroutine OBJECT to be called by PATS

```
      SUBROUTINE OBJECT                                    01488000
      COMMON/ONE/CUTOFF
      COMMON/THREE/XU(400),XL(400),YU(400),YL(400)
      COMMON/FOUR/XMAX(40),XMIN(40)
      COMMON/FIVE/NREG
      CALL AREANO(NREG,XX,YY)
      CALL SIM(D,PATH,PAR,OUT,USE,NODES,NZ,NCALLS,NZZ)
      REWIND7
      REWIND8
      CALL STAT(REAR)
      RETURN
      END
```

PATS . AREANO TO BE SEQUENCED SUBSEQUENT TO OBJECT

Notes

When the simulation model is linked to the Hooke and Jeeves Pattern Search a number of limitations are placed on the model. All 'nodes' are now defined in terms of 'areas' and these areas must all be rectangular. The areas must be ordered, that is, the numbering must be consecutive. Because times are given as time between nodes and not calculated as rectangular displacement PATS is used in the restricted form, and the additional subroutine AREANO must be used.

As well as the changes listed above, PATS and AREANO must have their COMMON statements changed.

PATS will require

```
      COMMON/ONE/CUTOFF
      COMMON/THREE/XU(400),XL(400),YU(400),YL(400)
      COMMON/FOUR/XMAX(40),XMIN(40)
      COMMON/FIVE/NREG
```

and AREANO will require

```
      COMMON/THREE/XU(400),XL(400),YU(400),YL(400)
```

None of the above recommended changes has been adequately tested.

RECENT PUBLICATIONS

RESEARCH REPORTS

48. *Proceedings of an N.Z. Seminar on Project Evaluation in Agriculture and Related Fields,* R. C. Jensen (ed.), 1968.
49. *Inter-Industry Structure of the New Zealand Economy, 1961-5,* B. J. Ross and B. P. Philpott, 1968.
50. *Fresh Vegetable Retailing in New Zealand,* G. W. Kitson, 1968.
51. *Livestock Targets in North Canterbury Hill Country: The Impact of Changing Prices,* J. L. Morris, H. J. Plunkett and R. W. M. Johnson, 1968.
52. *Sectoral Capital Formation in New Zealand, 1958-65,* T. W. Francis, 1968.
53. *Processing Peas: A Survey of Growers' Returns, 1967-8,* B. N. Hamilton and R. W. M. Johnson, 1968.
54. *Fertiliser Use in Southland,* R. W. M. Johnson, 1969.
55. *The Structure of Wool and Wool Textile Production, Trade and Consumption, 1948-68,* B. P. Philpott, G. A. Fletcher and W. G. Scott, 1969.
56. *Tower Silo Farming in New Zealand—Part 1: A Review,* D. McClatchy, 1969.
57. *Supply and Demand Projections of the United Kingdom Meat Market in 1975,* D. R. Edwards and B. P. Philpott, 1969.
58. *Tower Silo Farming in New Zealand—Part II: Economic Possibilities,* D. McClatchy, 1969.
59. *Productivity and Income of New Zealand Agriculture, 1921-67,* D. D. Hussey and B. P. Philpott.
60. *Current Trends in New Zealand Beef Production and Disposal,* D. McClatchy.
61. *Land Development by the State: An Economic Analysis of the Hindon Block, Otago,* E. D. Parkes.
62. *An Economic Analysis of Soil Conservation and Land Retirement on South Island High Country,* R. W. M. Johnson, 1970.
63. *A Regional Analysis of Future Sheep Production in New Zealand,* R. W. M. Johnson, 1970.
64. *An Economic Assessment of the Middle Class and Upper Middle Class Market in Malaya as a Potiential Outlet for New Zealand Meat and Dairy Products,* K. Y. Ho, 1970.
65. *Capital Formation in New Zealand Agriculture, 1947-67,* R. W. M. Johnson, 1970.
66. *Distribution Costs and Efficiency for Fresh Fruit and Vegetables,* G. W. Kitson, 1971.
67. *The Optimisation of a Sixteen Sector Model of the New Zealand Economy,* T. R. O'Malley, 1973.
68. *An Analysis of Lands and Survey Development Projects, 1945-69,* H. J. Plunket, 1972.
69. *Quantitative Techniques for Forecasting: A Review with Applications to New Zealand Wool Prices for 1974-5,* Joan Rodgers, 1974.
70. *A Practical Guide to Tax Planning using Procedures for Income Equalisation,* P. J. Charlton, 1975.
71. *Studies in Costs of Production: Process Peas and Beans, 1974-75,* W. O. McCarthy, R. G. Moffitt, P. W. Cosgriff and P. D. Chudleigh, 1975.
72. *Location of Farm Advisory Officers in New Zealand—an Application of Facility Location Analysis,* Joan R. Rodgers, Owen McCarthy and Vicki Mabin, 1975.
73. *The Ambulance Facility Location Problem—a Survey of Methods and a Simple Application,* Janet Gough and W. O. McCarthy, 1975.
74. *Studies in Costs of Production: Town Milk Supply Farms 1973-74,* R. J. Gillespie, 1976.
75. *Stabilising Post-Tax Incomes of New Zealand Sheep Farms,* P. D. Chudleigh, M. J. Blackie and J. B. Dent, 1976.

MARKET RESEARCH REPORTS

1. *Processing Plant Location Studies: I: Theory and a Simple Application to N.Z. Wool Selling Centres,* W. O. McCarthy, J. L. Rodgers and C. R. Higham, 1972.
2. *Processing Plant Location Studies: II: Policy Alternatives for N.Z. Wool Selling Centres,* C. R. Higham, J. L. Rodgers and W. O. McCarthy, 1972.
3. *Doing Business in Japan,* W. O. McCarthy (ed.), 1972.
4. *The Japanese Distribution System and Implications for New Zealand Traders,* G. W. Kitson, 1973.
5. *Prospects and Strategies in Promoting Tourism Between Japan and New Zealand,* G. W. Kitson, 1973.
6. *Market Assessment,* W. O. McCarthy (ed.), 1973.
7. *Optimum Site, Number and Location of Freezing Works in the South Island, New Zealand — A Spatial Analysis,* R. J. Brodie and W. O. McCarthy, 1974.
8. *The Japanese Food Market and Implications for New Zealand,* G. W. Kitson, 1975.
9. *Structure and Corporate Relationships in the Japanese Wool and Wool Textile Industries,* G. W. Kitson, 1976.

DISCUSSION PAPERS

3. *Economic Evaluation of Water Resources Development,* R. C. Jensen, A.N.Z.A.A.S., Christchurch, 1968.
4. *An Illustrative Example of Evaluation Procedures,* A. C. Norton and R. C. Jensen, N.Z. Assn. of Soil Conservators, May 1968.
5. *The Shape of the New Zealand Economy in 1980,* B. P. Philpott and B. J. Ross, N.Z. Assn. of Economists, August 1968.
6. *Economic Problems of New Zealand Agriculture,* R. W. M. Johnson, A.N.Z.A.A.S., Christchurch, 1968.
7. *Recent Trends in the Argentine Beef Cattle Situation,* R. W. M. Johnson, November, 1968.
8. *Price Formation in the Raw Wool Market,* C. J. McKenzie, B. P. Philpott and M. J. Woods, N.Z. Assn. of Economists, February 1969.
9. *Agricultural Production Functions,* A. C. Lewis, N.Z. Assn. of Economists, February 1969.
10. *Regional Economic Development in the Context of the Changing New Zealand Economy,* B. P. Philpott, Nelson Development Seminar, April 1969.
11. *Quarterly Estimates of New Zealand Meat Price, Consumption and Allied Data, 1946-65,* C. A. Yandle.
12. *Indicative Economic Planning with a Sixteen Sector Projection Model of the New Zealand Economy,* B. J. Ross and B. P. Philpott, A.N.Z.A.A.S., Adelaide, 1969.
13. *Recent Developments in the Meat Industry with particular reference to Otago and Southland,* R. G. Pilling, Otago Branch, N.Z. Econ. Soc., October 1969.
14. *The Future Profitability of Beef Production in New Zealand,* R. W. M. Johnson, N.Z. Inst. Agr. Sc., August 1970.
15. *Demand Prospects for Beef, B. P. Philpott,* N.Z. Inst. Agr. Sc., August 1970.
16. *The Structure of Wool and Wool Textile Production, Trade and Consumption, 1958-69,* B. P. Philpott and W. G. Scott, June 1970.
17. *Trends in the Terms of Exchange and Productivity in the New Zealand Dairy Industry,* R. W. M. Johnson, June 1970.
18. *Marketing Margins for New Zealand Lamb and for all Lamb and Mutton in the United Kingdom,* A. C. Lewis and S. M. C. Murray, July 1970.
19. *A Pilot Optimisation Model for the 1972-3 N.D.C. Plan,* B. P. Philpott and T. R. O'Malley, August 1970.
20. *Recent Trends in Capital Formation in New Zealand Agriculture, 1964-9,* R. W. M. Johnson and S. M. Hadfield, 1971.
21. *Productivity and Income of New Zealand Agriculture, 1921-67,* (Supplement to Research Report No. 59), S. M. Hadfield, 1971.
22. *Some Aspects of the Economics of Nitrogen Storage Farming in New Zealand,* B. P. Philpott, I. D. Greig and A. Wright, 1972.
23. *Economic Aspects of Stone Fruit Marketing in New Zealand,* A. W. Smith 1972.
24. *New Zealand, The Ten, and Future Market Strategies,* C.C.C. Bulletin, No. 559, W. O. McCarthy, 1972.
25. *The Wool Acquisition Controversy,* C.C.C. Bulletin, No. 577, W. O. McCarthy, 1974.
26. *Productivity,* C.C.C. Bulletin, No. 579, B. J. Ross, 1974.
27. *Investment on the Rural Scene,* paper presented to N.Z. Inst. of Valuers Seminar B. J. Ross, 1974.
28. *The Oil Crisis and International Economic Stability,* B. J. Ross, 1974.
29. *Christchurch Tomorrow—A discussion of the future development of Christchurch as a Regional Centre,* J. W. Wood, 1975.
30. *Use made of Transport by Farmers: A Pilot Survey with Findings Relating to Ashburton County, New Zealand,* T. I. Ambler, 1975.
31. *A Postal Sample Survey of Sheep Farmer Attitudes to Incentives and Obstacles to increasing Farm Output and other Agricultural Policy Issues,* J. G. Pryde, 1975.
32. *Proceedings of a Seminar on Costs Beyond the Farm Gate, 12th March 1976,* J. G. Pryde, W. O. McCarthy, D. L. Fyfe (eds.), 1976.
33. *A Postal Survey of the Opinions of a Group of Farm Management Society Members on Incentives and Obstacles to Increasing Farm Output,* J. G. Pryde, 1976.
34. *A Statistical Analysis of Sources of Variance of Income on Sheep Farms in New Zealand,* P. D. Chudleigh and S. J. Filan, 1976.

Additional copies of Research Reports and Market Research Reports, apart from complimentary copies, are available at $2.00 each. Discussion Papers are $1.00 (except No. 32 which is $3.00). Remittance should accompany order.