EMS Management Journal
Volume 1, Number 1
January -March, 2004

Operations Research Models for the Deployment of Emergency Services Vehicles

Jeffrey B. Goldberg, MS, PhD

ABSTRACT

This paper is a review of the development and current state of the art in operations research for deployment and planning analysis pertaining to Emergency Medical Services and Fire Departments. These public safety systems have received a great deal of attention in the operations research community since they provide important services to people and the problems are amenable to mathematical modeling and solution. The concentration here is on both analytical work and applications. Modeling and problem assumptions are emphasized rather than clever solution procedures and mathematical derivations. This paper is organized both chronologically and by modeling approach / problem issue, so the interested reader can easily trace the chain of modeling improvements that are most closely applicable to their particular problem. The bibliography contains over 115 relevant books, articles, and web sites that represent over 35 years of work from all of the leading operations research analysts and practitioners in the field.

Keywords: deployment, ambulances, dispatch, emergency medical services
Citation: Goldberg J: Operations Research Models for the Deployment of Emergency Services Vehicles. 2004 Jan-Mar *EMS Mgmnt J* 1(1):20-39.

PROBLEM MOTIVATION AND BACKGROUND

The design and operation of emergency service vehicle systems has been a vibrant area for operations research (OR) professionals since the mid 1960's. The primary concern of operations research is to develop and solve mathematical models to help make decisions. In this applications area, there have been 100's of journal papers covering models for important decisions such as:

- the location of fixed position fire stations and possibly variable position ambulance bases;
- the dispatching of vehicles;
- the number of vehicles of different types, staffing and equipment being carried; and
- how and when to re-deploy resources under different system states.

The reason for the large amount of work is quite simple. The systems are important to the public and hence designing and operating them well leads to a clear sense of purpose for the researcher.

The goal of this paper is to share this body of knowledge with emergency medical services (EMS) managers, fire service (FS) mangers, and their physician medical directors; three groups that may not be familiar with this body of work from the operations research field. In the past, there have been many papers reviewing this literature; however these have generally been targeted at operations research professionals. The approach of this paper is different in that it is organized along tasks and problem areas that are necessary for designing and operating effective systems. The focus is on the problems and the underlying modeling assumptions as opposed to the specific mathematics of the models. Once the capabilities of the models are known, then the interested reader can get more information as needed. Also, the paper contains sections covering work on

---

Presented: Sympoisum on the Future of EMS and Fire Deployment and Operations at the 2003 EMS Performance Conference & 1st Annual NEMSMA Meeting, Tampa, FL, November 9, 2003
Author: Associate Professor, Department of Systems and Industrial Engineering, The University of Arizona, Tucson, Arizona; jeff@sie.arizona.edu; 520-621-6547; http://www.sie.arizona.edu/faculty/goldberg.html
Manuscript received: 09 November 2003; *Accepted for publication:* 14 November 2003
Online Access: http://www.emsmj.com/v1n1/deployment/default.htm

data collection, building valid models, successful applications, and past review papers. Potential gaps in the literature are discussed where appropriate. The discussion for each topic proceeds chronologically to give the reader a feel for the progressive improvements in the body of work. For simplicity, each paper is contained in only one section, however, there are instances where there could be alternatives and the most appropriate section has been chosen.

The discussion centers mainly on EMS, however much of the material is relevant to the FS as well. One section contains models for managing multi-equipment types and here both FS and EMS are mentioned. Police deployment models are often fundamentally different in that they must allow for a patrolling vehicle as opposed to the stationary vehicles normally seen in FS and EMS situations. The materials in this paper are generally applicable to any system where the vehicles sit at a location and wait to be called into service.

SYSTEM OPERATION AND PERFORMANCE CRITERIA

The scope of this paper is based on the standard emergency call process:

1. The call (demand) comes to the system via 911 or some other mechanism.
2. The severity of the call is estimated.
3. The dispatcher evaluates the system status and determines the appropriate vehicle(s) to send to the scene.
4. Upon arriving on scene, service is provided.
5. The vehicle(s) may or may not provide transport to a hospital.
6. After completion of service (and transport) the vehicle goes into an idle state and returns to a predetermined location to await another call.

The decisions of dispatching and vehicle location are critical factors in system success. If one cannot do both of these well, there will be inefficiencies in the system. Note that both types of decisions must be made in a dynamic environment. However, significant planning can be done a priori. For example, standard dispatch strategies (such as "send the closest idle vehicle") can be evaluated for situations where they might not be the best policy. When these situations arise, the dispatcher can be alerted to go off of the standard strategy and onto a contingent strategy. Vehicle base locations can also be evaluated using specifics of situations and again, the system manager can be alerted on cases where it is better to go off of the standard plan.

The primary objective of EMS and FS deployment is to get the appropriate equipment to calls in a safe and timely fashion. The issues of selecting the appropriate equipment are not the focus of this paper. Generally, the dispatcher has some tools to make these decisions, based on the phone triage process and the state of the system. Also, vehicle safety is not considered here because it is assumed that the vehicle will arrive on scene in a safe manner. The final issue of timeliness is the primary objective that is used in operations research models. All of the OR typically makes the following assumptions:

- There is a standard time, T, such that if the first vehicle arrives on scene within T minutes, then the call service is deemed a success. The specific value of T may vary with the type of call as more serious calls have lower T values.
- The area is partitioned into zones. These zones may take on any shape, but all calls from a zone originate in the population center. All travel to and from the zone is measured from the zone center point. Data is collected and aggregated at the zone level.

There are many ways that timeliness is measured. For example, one can operate to:

- Minimize the total or average time to serve all calls.
- Minimize the maximum travel time to any single call (ensures that no demand point is too far from equipment).
- Maximize Area Coverage - cover as many zones in the area as possible within T minutes of travel.
- Maximize Call Coverage - cover as many calls in the area as possible within T minutes of travel.

Note that the 3rd and 4th examples are not equivalent since some zones in the area may have markedly different call loads. In reality, each of these examples are surrogates for the true objective of reducing as much morbidity and mortality as possible. The assumption is that if calls are answered and serviced quickly, then this will lead to better clinical outcomes, patient satisfaction and compliance to regulatory standards for response time performance.

Besides timeliness, there are other objectives of EMS and FS deployment systems. These include:

- Minimize Cost – Cost is primarily a function of the amount of labor (man-hours) needed to staff the unit-hours used per year, the number of base stations that must be opened and maintained, and the number of vehicles that must be purchased, supported, and serviced. Labor is typically the largest cost (ReVelle [1989]).
- Maximize Coverage Equity – Here, the system manager must balance area performance against the performance in a smaller group of zones. For example, it may not be acceptable to have zones that are poorly served while having the area at a reasonable level and some zones that are extremely well served. By changing decisions, more equitable systems can be designed. Marsh and Schilling [1994] contains a broad review of coverage equity issues in EMS and FS systems.
- Maximizing Labor Equity – Here it is important for the system manager to balance the workload for all employees in the system. This reduces employee burnout and hard feelings.

MODELING ISSUES - GRANULARITY, DATA REQUIREMENTS, AND VALIDITY

It is often difficult and expensive to experiment with an actual system. Mistakes are costly both in money and in potential morbidity and mortality. Collecting data to verify a good system might take months of data collection. Instead of experimenting on the actual system, OR professionals generally build models of systems that can be implemented and experimented with on a computer. System errors can be found on the model before they are implemented on the actual system. It is generally well worth the cost of building the model, collecting data, and running the model as opposed to trying to experiment on the actual system.

When using a model to help make decisions, there is significant work that must be done before any analysis can begin. First one must structure the granularity of the model and define zones. Next, one gathers demand, service, and travel time data based on that structure. Dai, Wang, and Yang, [1994] discuss a decision support system and data base structure that can be used for planning and for operational decisions. Third, the model is implemented; usually in software. Finally, one validates the model to convince the decision maker that model output has some correlation with the output of the actual system. Each of these steps has been the topic of OR studies.

Granularity of the Zone Structure

The zone structure is often formed based on the convenience of the model builder or the data collection system. Since most urban and suburban EMS and FS systems have tens of thousands of calls per year, it is impossible to model down to the call level. Instead, all calls in a "small area" are aggregated to a single zone.

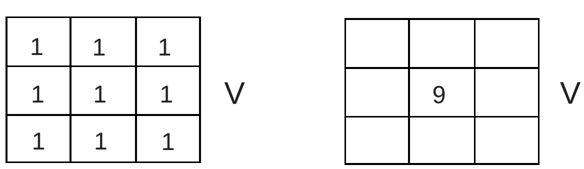

Figure 1 - Call Aggregation with a Vehicle Location

At first glance this looks like a minor issue in modeling, however, looks can be deceiving. Consider the aggregation in Figure 1. Here, we have 9 calls, one in each of the nine address blocks. Instead of using these individual locations, we aggregate all calls to the center of the blocks and this is our zone.

The problem here is that timeliness measured on the aggregated system may greatly overestimate the timeliness of the actual system. Consider the 8-minute call coverage criteria typically used in EMS systems and assume that a single vehicle is located exactly 7.5 minutes from the center of the blocks directly to the right. In the aggregated problem, all calls would be considered covered since the vehicle is 7.5 minutes away. In the actual problem, calls in the left column and calls in blocks directly above and below the center block may not be covered as these travel values may be larger than 7.5 minutes. Similar examples using travel time or vehicle utilization as the criteria can easily be constructed. As the zones become smaller, the inaccuracies due to aggregation become smaller as well. Currently, it is reasonable to solve problems with 100's of zones (depending on the model).

This problem was first realized in Hillsman and Rhoda [1978] and they defined 3 specific types of errors:

- A errors – errors in distance measurement for the call since the original call location is not the location of the aggregated calls.
- B errors – errors in distance measurement due to not knowing the true location when a vehicle or facility is located at an aggregated zone.
- C errors – errors in dispatching due to not knowing the correct distance from vehicles or bases to calls in aggregated zones.

When the criteria is total system travel distance, Current and Schilling [1987] show how to eliminate A and B errors, and Hodgson and Neuman [1993] use a GIS approach to eliminate C errors. Francis and Lowe [1992] derive bounds on how poorly the model approximates the actual objective. Erkut and Bozkaya [1999] discuss A and C errors and give practical strategies for reducing the errors and setting up models that have small errors. Note that as computing power increases and larger models can be formulated and solved, less aggregation is needed and this problem becomes less critical. At this time, aggregation can still cause problems in models that use coverage or travel time objectives.

Travel Time Modeling

Models require estimates of travel time to make decisions concerning dispatching order, determine coverage areas, and compute estimates of the criteria. Without accurate travel time estimates, most models would have little predictive value and the decisions that they suggest would be suspect at best. There has been little OR work in this area, especially given its importance (there is surely additional work in the transportation literature, but that is not the focus of this paper). Travel time is generally assumed to be known exogenously and occurs on a street network. When solving real problems, this simply is not sufficient. Often this data exists (at least in rough form) at a county government agency responsible for traffic management and planning.

Volz [1971] uses linear regression to determine speed coefficients on different road types (for example, freeway limited access roads, four or more lane roads with at least two lanes in each direction, three lane roads with a left turn lane, and local two lane streets) and then uses these coefficients with an estimate of the road types on the travel route. Goldberg et. al. [1990] uses this approach in predicting the mean and standard deviation of base to zone travel times in the Tucson EMS system. Kolesar [1973 and 1975] presents models based on regression studies in New York. For short trips, Kolesar [1975] suggests that, travel time is proportional to the square root of the distance traveled, while for longer trips, travel time is proportional to the distance traveled. Chelst and Jarvis [1979] estimate the probability distribution of resulting travel times (after the model is solved, these will be the travel trips one expects to realize) for urban emergency service systems. Their work is based on the results obtained from the Hypercube model of Larson [1974] used to predict the probability of different system busy states. Repede and Bernardo [1994] perform a detailed model of travel in Louisville using a 47,000 call database and the UNIFIT curve fitting package. Recently, Van Buer et. al. [1996] considered the problem of locating 1-way streets and cul-de-sacs so as to enable all emergency services reasonable access while reducing crime.

DEMAND DATA MODELING AND PREDICTION

The ability to predict demand is of paramount importance; however there has been little systemic OR study in this area. The typical approach is to tally past demand for each zone over some time period (a year or six months), and then assume that future demand will behave similarly to past demand. Similarly includes both quantity and spatial similarity. Even when the quantity of demand is changed, this is usually done in a proportional manner (for example, assume that demand in each zone will increase by 10%).

The initial study in the OR literature is Kamenetzky, Shuman, and Wolfe [1982]. They develop a regression model with four independent variables; area population, area employment rate, % of the population white and married, and housing units per area resident. They perform transformations on these variables and develop a model with an excellent R^2 value of .92. They also predict demand by category of call. Mabert [1985] models the demand for police calls in Indianapolis. Box-Jenkins time series models are used to predict the number of calls per day, however there is no attempt at predicting the spatial distribution.

Although not within the scope of this paper, there has been work in this area in the healthcare literature. Puig-Junoy, Saez, and Martinez-Garcia [1998] formulates a utility theory model to explain why people use the emergency room as a substitute for their primary care physician. This approach might be used to help explain why people use the 911 system for "less than emergency" calls. Also, growth and city planning models could be used to estimate demand for zones that have no call history.

MODEL VALIDITY

Model validity refers to the model's ability to predict output and to make decisions that will work as well as predicted in the actual system. This is a key step in the modeling process. Unless the model makes valid predictions then the model will have little to no value.

Almost all models have "face validity" where the model looks reasonable to the casual observer. The next level is "replication validity." Here, the analyst inputs data on past operation of the actual system and the model replicates the operation of the system including:

- Predicting coverage and travel time close to those realized in the actual system.
- Making the same dispatching decisions as the actual system made.
- Predicting vehicle utilization close to that realized in the actual system.

The final level is "prediction validity" where the analyst inputs data for a future system and the model predicts how the future system will behave. Often future validity cannot be fully placation determined until the system is implemented. Hence if the model has face and replication validity, then the decision maker is generally convinced of the quality of the model's output.

There have been many studies in model validity since the typical models used in planning require making significant assumptions (these will be discussed later). In Jarvis [1975], it is shown that a spatial queueing model can predict vehicle utilizations usually to within a few percent. Halpern [1977] shows on small systems that when you assume that all calls have the same service time then the vehicle busy probability estimates may not be accurate when the assumption is not satisfied. Goldberg et. al. [1990], Valenzuela et. al. [1990], and Repede and Bernardo [1994] write about high precision validity studies with actual data in Tucson and Louisville. ReVelle [1989] discusses modeling developments to improve the validity of models that use area coverage and call coverage criteria. Borras and Pastor [2002] compare the output of probabilistic covering models to the output of more detailed previously validated models. They show that the probabilistic models are highly accurate. Saydam and Aytug [2003] show how to improve accuracy when measuring expected coverage. Chiyoshi, Galvao, and Morabito [2003] show that when vehicles are busy a high percentage of the time, then popular covering models do not predict coverage well even when they include features specifically designed to remedy the accuracy problem.

RAND REPORTS

In the late 1960's and the 1970's, a powerful group of OR analysts worked at the RAND Institute in New York. The works that they did were strong both in theory and in application and covered a wide spectrum of uses in EMS, FS, and police planning and operations. The following models were distributed and used extensively throughout the country (Chaiken [1978]):

- PAM - Parametric Allocation Model - used for planning the number of fire bases or ambulances that were needed in an urban area.
- FHSEM - Firehouse Site Evaluation Model - covering model to determine firehouse locations
- FIRESIM - Simulation Model of Fire Department Operations - detailed simulation model to evaluate a single deployment of base locations
- Hypercube Queueing Model - a model that tracks the status of vehicles and predicts the fraction of time that the system is in each state (to be discussed later as this is one of the foundation OR models for EMS deployment)

There is a large body of work that came out of RAND during this period and some of it was also published in the general OR literature. The NFPA Fire Station Location Bibliography [1996] has a large list of these papers including technical reports, users manuals for software, and applications reports. We do not provide additional details here as these models have become a bit dated, however they form much of the basis of the initial thoughts on the problems of designing and operating EMS and FS systems.

COMPUTER SIMULATION MODELS

The term "computer simulation model" is used in the OR literature to describe models that have a high degree of detail and try to precisely mimic the operation of the actual system. These are generally tied to a detailed database and coded in a computer language that is tailored for building detailed models. They have a high degree of face validity and can obtain extremely accurate replication validation results.

These models are descriptive in that they take a system design and set of operating rules and then compute the performance of that single system on any desired criteria. To find high quality solutions, one must embed the simulation model in a search routine to test different designs and rule structures. The difficulty with these models is that they often take large amounts of computer time to obtain estimates of system parameters and hence this search routine is also a long process. If one has the computing resources, the detailed data set and the time to search for solutions, then simulation is an excellent tool for obtaining high quality designs and operating rules.

Savas [1969] and Fitzsimmons [1971] were early users of simulation modeling for EMS systems. Berlin and Liebman [1974] used simulation to evaluate solutions that they obtained from a simple covering model. This approach of using simulation on a small set of solutions found by other less precise modeling tools is a popular approach. Monarchi, Hendrick, and Plane [1977], Fitsimmons and Stikar [1982], and Repede and Bernardo [1994] all have an approach similar to Berlin and Liebman in that they use simulation to evaluate decisions made with the help of other models.

Trudeau, Rousseau, Ferland, and Choquette [1989] cover a solution process that includes data collection as well as simulation modeling for planning and operation of an EMS system. Erkut and Polat [1992] built a detailed simulation model that was used in Istanbul, Turkey to help minimize total travel time and the percentage of calls that are served late. Goldberg et. al. [1990a] discuss a similar model used in Tucson, AZ to help validate another faster less detailed model. Zaki, Cheng, and Parker [1997] describe a detailed model that was used in the Richmond, VA area. Also, this work has the ability to consider pre-empted calls and contains a validation study as well as estimates for travel time distributions.

ANALYTIC MODELS FOR PLANNING

The large majority of OR work is in the area of "models for planning." Here, the decisions are generally for determining the location of vehicle bases and the equipment assigned to each base. The problem is static in that a single set of demand and travel time data is used in the model. Mathematical programming is generally used to find optimal solutions to the models and when the models are difficult to solve, heuristic procedures* are used. The term "optimal solution" is in some sense misleading. The solution is only as good as the underlying assumptions of the model. If these assumptions do not produce a valid model, then the solution found may have little correlation to optimal solutions in the real problem. There is generally a modeling tradeoff in that high validity high precision models (e.g., simulation models) cannot be easily searched for optimal solutions. To guarantee optimality, usually one must have a rather stylized model with a significant assumption set. Larson and Stevenson [1972] suggest that "real problems" have rather flat objective functions and many solutions exhibit near optimal behavior.

Toregas et. al. [1971] first proposed the "set covering model." Their objective is to minimize the number of vehicles needed to cover all zones. In essence, they are minimizing cost and ensuring a fair coverage. Each potential vehicle location has a set of demands that it covers. All demand points are equally important, and a single static covering distance (or time) for each demand is used. If even one vehicle covers a demand, then that is sufficient, even though sometimes the vehicle may be busy. Hence this model generally overestimates coverage. This model typically looks like laying out "travel circles" to cover all demands. Neebe [1988] finds all solutions to this model as the covering distance standard varies. As expected, as you increase the covering standard, then fewer vehicles are needed.

Church and ReVelle [1974] take somewhat of a dual approach. They hold the number of vehicles fixed (and hence fix costs) and then locate the vehicles to cover as many calls as possible. The model is called the "maximal covering model," and this objective can result in some zones that are not covered. As in Toregas et. al. [1971], a zone is covered if even one vehicle is located within the travel standard. There is no allowance for busy vehicles.

In both the set covering model and the maximal covering model, it is assume that there is a single time period. Since location decisions often involve extensive capital costs, one might want to consider a multiple period problem over a long horizon (multiple years and decades), Schilling [1980] considers a model that is divided into time periods (years for example). The work extends the maximal covering model to consider a different location set for each time period. The model is multi-objective in that there is an objective to maximize total demand covered in each period. It includes constraints that limit the total number of vehicles placed in each time period and ensure that if you locate a vehicle at a site in time period t, then you will also use the site in time periods t+1, t+2 … T. Current, Tatick, and ReVelle, [1997] model the case of locating a set of vehicles over a long horizon when the total number of vehicles and facilities is uncertain. They concentrate on finding the locations for near-term decisions so that the system will be in a good situation when the next decision is to be made. The model objective includes construction and operations costs.

Both set covering and maximal covering are "integer programming models" since all decisions are required to be integer valued (either you locate at a site or you don't, there is no "half location"). Typically, these types of problems are mathematically difficult. ReVelle [1993] shows that even though the problems can be difficult, in most instances the solution can be obtained by readily available software for real sized problems (100's of demand points and 50 potential vehicle sites). Galvano and ReVelle [1996] discuss a Lagrangian heuristic for cases when the maximal covering model is difficult. A Lagrangian heuristic is a process where one relaxes restrictions (constraints) on the solution in order to get a starting point to generate alternatives that satisfy all restrictions. The technique is commonly use to solve problems where the decisions are constrained to be integer valued.

The key deficiencies in the set covering and maximal covering models are:

*Heuristic procedures may find good solutions, but are not necessarily the best and are often performed by trial-and-error

- The use of a single objective, when in fact, both cost and performance are key criteria.
- The inability to consider sometimes busy vehicles and therefore uncovered demands even though the model claims full coverage. This also leads to the assumption that the closest vehicle to each zone is the only vehicle that ever answers calls for that zone.
- Each call requires a single vehicle and only a single type of equipment is considered. The real problem has multiple equipment types that cooperate and share calls.
- All demand, travel time, and service time data are assumed to be deterministic. In the real problem, demand timing and location are both random (but somewhat predictable and modelable based on past data). Service times and travel times also are random and sometimes highly varying.
- Using a single set of data leads to a single set of locations and hence these models have no ability to analyze dynamic real-time decisions such as repositioning.

Works that try to remedy these deficiencies are discussed below.

MODELS THAT INCORPORATE MULTIPLE OBJECTIVES

Many of the models discussed in the remainder of section 4 include multiple objectives. Often these objectives are combined into a single objective using a weighting factor. For example, if f(x) and g(x) are objective functions, then if w is between 0 and 1, then:

$$w * f(x) + (1-w) * g(x)$$

is a combined objective (this is called the “weighting method”). Another approach with multiple objectives is to find the set of non-dominated solutions (called the “Pareto set”). A solution x is non-dominated if no other solution is better or equal to x on all objectives. The final approach used is called "goal programming." Here, each objective is a goal with a prescribed level for success. For example, if the goal is to minimize average travel time, then a specific level might be to have average travel time less than or equal to 8 minutes. The user then specifies the "cost" if the goal is not met and this cost may be related to the magnitude of the unmet amount. Usually, the total cost of unmet goals is minimized. The difficulty in goal programming is in specifying the cost of not meeting the goal.

Schilling, ReVelle, Cohen, and Elzinga [1980] extend the maximal covering model by dividing the demand in each zone into two call types, each with a different priority. They then formulate two objectives - maximize the covering of the highest priority calls and maximize the covering of next lower priority of calls. They then find the set of non-dominated solutions. They also consider two vehicle types as well as a budget constraint for purchasing new equipment. Daskin and Stern [1981] extend the set covering model to include an objective of maximizing the number of zones that are covered by more than one vehicle. They first solve a set covering problem to find the minimum number of vehicles needed for covering. Since there are usually multiple sets of locations that attain full coverage with this minimum number of vehicles, they then choose a particular solution by finding the location set that maximizes the number of zones with multiple coverage. One can also use the weighting method to combine the objectives of minimizing the number of vehicles and maximizing the number of multiple covered zones. Taylor, Baker, and Clayton [1989], develop a model for allocating ambulances to sectors in a county. The model is more of a macro location rather than specific site location since it has only 10 sectors. Goal programming is used with objectives of minimizing system response time, staying within a pre-specified budget, and keeping vehicle workloads low.

Revelle, Schweitzer, and Snyder [1996] extend the maximal covering model to the case where facility sites may not be used to cover their own zones. This model has application in disaster situations where if a zone has a call, then everything within the zone is hit, not just a single element within the zone. The multiple objectives considered are maximizing the demand covered and maximizing the number of vehicle sites that are at least double covered. Badri, Mortagy, and Alsayed [1998] develop an extensive goal programming model that includes minimizing fixed and recurring operating costs, maximizing demand coverage, minimizing system travel distance and maximum travel time, and minimizing the use of areas where water availability might be a problem (application in Dubai, UAE).

When appropriate, works that use multiple objectives will be noted in the remaining subsections.

MATH PROGRAMMING MODELS THAT INCORPORATE BACK-UP COVERAGE AND BUSY VEHICLES

Church and Weaver [1985] extend the maximal coverage model and develop the "vector assignment p-median model." Here, it is assumed that a zone's demand will be served x% of the time by its closest vehicle, y% of the time by the 2nd closest vehicle, …One must estimate these %'s using another model or past experience. The model locates a fixed number of vehicles so that the total travel distance is minimized. Clearly the approach is not valid when the %'s change based on the specific set of locations chosen.

Hogan and ReVelle [1986] extend the maximal covering model and the set covering model by adding a second objective to maximize the number calls covered by 2 or more vehicles. They combine the objectives using weighting factors and explore how the solution changes as the weight on each of the two objectives changes. Also, they add constraints to model a secondary coverage criteria so that all calls are covered within the secondary time limit (for example 15 minutes) while trying to maximize the number of calls covered within the shorter primary limit (for example 8 minutes). Gendreau, Laporte, and Semet, [1997], develop a search for a model that maximizes the number of double covered calls.

Pirkul and Schilling [1988] model the objective of maximizing covered calls subject to limits on the number of calls that each vehicle can answer (as measured by the number of calls it may serve in the study horizon) and a requirement for backup coverage. Narasimhan, Pirkul, and Schilling [1992] extend the model to include the objective of maximizing the number of covered calls of different service levels. They include constraints on the number of vehicles, the call capacity of each vehicle, and defining coverage for different levels of service. Here, a vehicle may provide, primary, secondary, or both levels of service and each level may have a different coverage time standard.

Daskin [1983] assumes that each vehicle is busy with probability equal to p. One can estimate p accurately by estimating the total workload (travel and service time) divided by the total work time available. Given p and the number of vehicles that cover a demand zone, one can estimate the probability that the zone is covered by computing the probability that at least one of the covering vehicles will be idle. This computation is simplified by assuming that the vehicles operate independently and this is generally called the "independence assumption." It is clear that the assumption is not entirely valid since as nearby vehicles are busy, it is more likely that a vehicle is busy since it now has a large primary coverage area. Daskin develops a heuristic procedure for finding the best set of locations assuming that the number of vehicles is fixed. The model maximizes the expected demand covered (the model is called the "maximal expected covering model") and hence is builds on Church and ReVelle [1974]. The paper also considers how the coverage changes as the number of vehicles changes.

ReVelle and Hogan [1989] extend the maximal expected covering model. First, all vehicles are assumed to have equal busy probability. Then a constraint is formed to ensure that each zone is covered with probability p. For example, assume that vehicles are busy 20% of the time and you want a zone covered 95% of the time. Then if only one vehicle covers the zone, the probability of coverage is only 80% since 20% of the time, the covering vehicle is busy. If you have 2 vehicles covering, then the zone is covered 96% of the time; 80% by the first vehicle and 16% (80% of the remaining 20%). So, to meet the 95% constraint, you need 2 or more covering vehicles for the zone. For any busy probability and any coverage level, it is simple to compute the number of covering vehicles required. The model is then extended to include the case where vehicles have unequal busy probabilities; however it is still assumed that for any zone, all vehicles serving that zone have equal busy probabilities (called the "zone based busy" assumption). Then one can form the identical constraints to the first model, but the number of covering vehicles required for each zone may be different (if the busy probability is lower, then fewer vehicles are required to get to p since the closest vehicles answer a higher fraction of calls). This assumption on equal busy probability for all vehicles covering a zone is problematic in that the same vehicle is assume to have two different busy probabilities depending on which zone it is covering. Ball and Lin [1993] formulate a model similar to ReVelle and Hogan [1989] as it includes a constraint on the number of covering vehicles to ensure that calls are covered with probability p. This model also uses the zone based busy assumption to derive the coverage reliability constraint. ReVelle and Hogan [1989a] use a similar approach to extend the set covering model to include coverage probability.

Marianov and ReVelle [1994] extend the set covering model to ensure that each demand has a vehicle actually available within the time standard, with probability p. The objective of minimizing the required number of servers remains the same. Queuing theory is applied to the development of the availability constraints so that one does not need to assume that vehicles are independent (as was done in Revelle and Hogan [1989] and Daskin [1983]). Instead, the model uses the zone based busy assumption. Silva and Serra [2003] extend the Marianov and Revelle [1994] work to include multiple call priorities. Marianov and ReVelle [1996] extend ReVelle and Hogan [1989] by using queuing theory to develop the required coverage constraints. The zone based busy assumption is still required. Marianov and Serra [1998] extend the work of Marianov and ReVelle [1996] to include decisions on base location and decisions on allocating customers to bases. The bases behave as queuing systems so calls may wait in line. Marianov and Serra [2002] use the techniques in their [1998] work and extend the set covering model to include minimizing the number of facilities and the number of vehicles. Constraints on queue length are also included. In all of the Marianov papers in this paragraph, all demands assigned to a base must go to that base, regardless of the busyness of the base. This limits the validity of these models in an EMS situation.

Models That Consider Multiple Vehicle Types And Multiple Vehicles Per Call

Schilling et. al. [1979] extend the maximal coverage model and consider both facility location and equipment allocation to facilities. Their model is called "FLEET" and considers two types of equipment (ALS and BLS for example). The objective is to maximize the amount of demand that is covered by both types of vehicles. There is a fixed number of each type of equipment and a fixed number of sites to open. Each equipment type has a different coverage standard. Facility location costs are assumed to be equal for all possible site locations, so once the number of sites is fixed the location cost is fixed.

Charnes and Storbeck [1980] consider a goal programming model. They locate a fixed number of ALS and BLS vehicles so that the goal for ALS coverage is met for each zone and the goal for BLS coverage is met for zones that are not covered by ALS vehicles. There is an implied cost of not meeting the different coverage goals and the model tries to minimize the overall cost of not meeting coverage goals. The model builder must determine the specific cost of not meeting goals. Also, Storbeck [1982] uses a goal programming model to evaluate the tradeoff in having some zones with multiple coverage and some zones with no coverage given the cost of having no coverage is known. Moore and Revelle [1982] considers the "hierarchical service location problem." This model is an adjustment of the FLEET model and considers a demand covered if it is covered by either type of vehicle as opposed to both types. The goal is to minimize the amount of demand that is not covered. Instead of fixing the number of vehicles of each type, a constraint is included to ensure that the budget is not violated. The model uses an assumption that the potential site locations are known. One can solve the model for a number of budget levels and see the tradeoff between coverage and expense.

Marianov and ReVelle [1992] model the simultaneous location of engines and ladder trucks and includes limits on the number of vehicles at each location. The approach is an extension of Hogan and ReVelle [1986] model for double coverage. Coverage in this model requires two truck companies and three engine companies within the standard response distance. When busy vehicles are included in the model, then the coverage requirement is modified using an approach similar to that used in ReVelle and Hogan [1989]. ReVelle and Snyder [1995] construct the "FAST model" to locate both fire and ambulance vehicles. The model is multiobjective in that it attempts to maximize call coverage for fire and call coverage for ambulances. Instead of finding one "optimal" solution using a weighting method, the approach is to generate all solutions that are "efficient" in that there is no other solution this is better on both objectives. This set of solutions (generally termed the "Pareto set" frontier) is given to the decision maker(s) who then make the trade-off and select a single alternative. Constraints include the number of vehicles of each type, a construction budget for stations, and the notion that each site can only be for ambulance or fire trucks. Jayaraman and Stinastava [1995] extend Daskin's maximal expected coverage model to include primary and secondary vehicles. The objective is to maximize the sum of calls that are primary and secondary covered. Primary coverage is defined to have "enough" primary within the primary time standard and secondary coverage is similarly defined. Constraints include limits on the number of primary and secondary vehicles and the number of facilities.

Serra [1996] defines the "coherent covering location model." Here, there are two types of vehicles, ALS and BLS for example. ALS vehicles can provide ALS and BLS service while BLS vehicles provide only BLS service. The model has two criteria, maximize call coverage by ALS vehicles and maximize call coverage by an ALS or BLS vehicle. There are limits on the number of ALS and BLS vehicles and a distance standard that ensures that BLS vehicles are located near ALS vehicles (this is the "coherent" part of the model). The notion of coverage depends on the service type as each has its own acceptable distance standard. Mandell [1998] considers a similar ALS/BLS system with the same coverage requirements. The model includes constraints similar to those in Revelle and Hogan [1989] to ensure reliable coverage in the presence of busy vehicles as well as a queuing model to estimate the precise probability of successful coverage. The work of Mandell represents the most advanced of the static planning models for ALS/BLS systems.

Amiri [1998] formulates a model to minimize the cost of setting up and operating facilities and includes primary and secondary vehicles and primary and secondary service levels. Each demand must be covered by both a primary and secondary vehicle. Marianov and Serra [2001] develop a model to locate high and low level servers and allocate zones to either type of server. The servers operate like a queuing system and calls may wait in line. The objective is to minimize the cost of opening and operating facilities. Both models have limited applicability to EMS systems as demands must go to the assigned facility and not the closest idle facility.

Batta and Mannur [1990] extend both the maximal covering model and the set covering model to the case where calls require multiple units of the same type. When multiple units are involved, coverage depends on the response order for the call. For example if a call requires 3 units, then the call can be considered covered if the first unit is within 5 minutes, the second unit within 7 minutes and the third unit within 10 minutes. The specific time standards can differ for any system and must be given for each call and each arriving vehicle.

MODELS THAT CONSIDER PARAMETERS THAT ARE RANDOM

Hypercube Approaches

Larson [1974] introduces the "Hypercube model" for evaluating the performance of a set of base locations. Like simulations, this model is descriptive and must be embedded in an optimization framework in order to search for good solutions. The model assumes that calls come to the system based on a Poisson process (as opposed to the deterministic assumptions of all previously discussed works in section 4) and that service time for each call is exponentially distributed with mean ì . (A Poisson process is a standard process used to model arrivals to a system. It is the result of having a large number of potential customers, *N*, where each has a small probability, *p*, of using the system in a short time interval. The product *N*p*, denoted by λ, is called the "intensity of the process" and is the average number of arrivals per unit time. Given λ, it is a simple matter to calculate the probability distribution on the number of arrivals in any time period, *t*, as this follows a Poission distribution with a mean of λ**t*.). Larson uses these assumptions to formulate a large model. with a state for every possible combination of idle and busy vehicles. For example, if there are 5 vehicles, then the state (1, 0, 1, 1, 0) corresponds to vehicles 1, 3, and 4 being busy and vehicles 2 and 5 idle. If a call arrives when the system is in this state, then it will be served by either vehicle 2 or 5, depending on the location of the call relative to the location of the two idle vehicles. In general, the model has 2^N states where N is the number of vehicles. When N is even 20, this is computationally difficult. The class of models is called “Markov Models” due to the assumption that the probability of going to any future combination of busy and idle vehicles is due only to the current combination and the probability that future events occur. The manner in which we arrived at the current combination is not relevant in predicting future states (this memoryless property in which the future does not depend on the past other than through the current state is called the “Markov Property”).

In Hypercube, each call is assumed to require one vehicle and it is assumed that each zone has a unique preference ordering of the available vehicles. This unique preference order simply implies that for any call, there is a dispatch preference order. The dispatcher will go down the order and dispatch the first idle vehicle on the list. Generally, the preference is distance based, but this is not required in the model. There could be a difficulty if two vehicles were equally preferred for a call and Burwell, Jarvis, and McKnew [1993] extend Hypercube to consider co-located servers and dispatch ties. The output of the Hypercubemodel is the probability of being in any state state (combination of busy and idle vehicles) and this irobability can then be used to compute the traditional criteria of system travel time, system coverage, vehicle workload and busy probabilities, and the number of non-first-choice dispatches in the system.

To remedy the computational difficulties of Hypercube, Larson [1975] develops the "A-Hypercube, the Hypercube approximation model." The model consists of equations to compute the busy probability of each vehicle. The key extension in this work is the development of factors (called "Q-factors") that can be included in models and used to relax the assumption that vehicle busy probabilities are independent. The model still requires that the mean service time be a single value for all calls, however one can do a calibration procedure to estimate the appropriate service time. Goldberg and Szidarovsky [1991] demonstrates the convergence of the computational methods used in A-Hypercube.

Batta, Dolan, and Krishnamurthy [1989] combine the ideas in A-Hypercube with those in Daskin's expected covering model. This new model remedies the independence assumption in Daskin's model by including Larson's Q-factors in the objective and the authors build a heuristic to find good sets of locations. Benveniste [1985] extends A-Hypercube by adding the objective of minimizing system travel distance and developing in an optimization strategy that selects server locations and service areas (what zones are first choice for each zone). Jarvis [1985] extends A-Hyprecube by including general call service times. This enables the modeling of systems where the call service times depend both on the serving vehicle and the demand zone. Validation studies on this model show that estimates of vehicle busy probabilities are extremely accurate (less than 5% error) when compared to busy probabilities of the actual system. Goldberg and Paz [1991] extend Jarvis' model by adding the objective of maximizing the expected number of calls covered (they estimate the probability that travel is less than the coverage time standard) and by embedding the new model in a location selection heuristic. The assumption of a fixed vehicle preference ordering for each zone remains.

Queueing Approaches

Berman and Larson [1982] consider the situation where demand occurs according to a Poisson process, service is random and follows a general distribution that is independent of vehicle location. The model locates a set of vehicles to minimize total expected travel distance to serve all demands. The model considers the possibility that more preferred vehicles are busy and hence a less preferred vehicle must be sent and the possibility that the system is completely busy and call must queue or be sent to a system operating in parallel. The model is similar in spirit to the model proposed in Jarvis [1985]. Berman, Larson, and Parkan, [1987] decompose the model of Berman and Larson into a set of single vehicle problems. They then locate each server optimally. Information on shared workloads are passed between the single vehicle problems until the method converges to a stable solution. The goal is to minimize system travel and waiting time.

Models That Consider The Dynamic Real-Time Nature Of The Problem

The models previously discussed in section 4 tend to be "single use" models. A user would solve the model for a single data set of demands, travel times, and service times, and obtain insight on good sets of locations for that data set. This is problematic in that the data is typically not stationary and has dramatic changes over the day, the week, and even the year.

One approach for dealing with the dynamic nature of the problem is to break the week into 168 hourly periods and solve the model for each hour. Here, a user will have to integrate solutions so that the system runs smoothly and is not jumpy with vehicles changing locations repeatedly. Also, one can use the models and do pre-planning for atypical situations. For example, if 25% of the vehicles are busy, one could solve a model with 25% less capacity and see how the system should be designed. This solution can now be used to help in deciding how to re-deploy. So, the typical strategy for dealing with the dynamic nature is to use the static models and do a great deal of experimentation to pre-plan for contingent situations. Unfortunately, one cannot anticipate every possible situation and one must still figure out how to integrate and implement the solutions from the different model runs.

There has been work in the OR literature on two dynamic problems; repositioning and dispatching. We close this section with a short discussion on each problem.

Dynamic Re-positioning

For the short run problem (hours or real time), Repede and Bernardo [1994] extend Daskin's maximal expected covering model to allow different location sets at different times of the week. A set of constraints to limit the number of vehicles during each hour is also included to control costs. There is no constraint on how many times a vehicle may changes bases during a day. Amiri [2001] uses a similar approach when designing general service systems. The model also includes an objective that considers costs such as labor time, travel time, and the waiting time of calls. The decisions are the locations of the facilities and the available capacity at each open facility for the day, and the zones assigned to each open facility. Once a facility is opened and given capacity, then that capacity is fixed for the entire day. The zone assignments can vary with each hour. Gendreau, Laporte, and Semet, [2001], develop a model that considers the objective of maximizing double coverage of demand. The constraints include: the number of vehicles at each site, the moving of the same vehicle repeatedly, long travel trips, and round trips between two sites. The model has been used on data generated from the Montreal, Canada EMS system and is designed to be solved quickly in real time so that it can help with relocation decisions in real time. It is embedded in a software system that also considers dispatching as well as vehicle reassignment. This final work seems to be the only work on real-time repositioning decisions.

Dispatching

The dispatching decision for EMS calls are generally quite simple in that the closest idle vehicle is usually dispatched to the call. There can be complications however and sometimes it may be better for the system if the second or more closest idle vehicle is sent. This event is most likely to occur when there are two (or more) vehicles that are approximately equal in travel time. Here, either vehicle can get to the call successfully, but there is little OR work done on dispatching of ambulances.

Fire calls are often a different matter Chelst and Barlach [1981] consider the decision when multiple vehicles must be dispatched. The work uses the ideas in the Hypercube and A-Hypercube models and the major concern is estimating the time of the first responder on scene. Swersey [1982] develops a decision model to deciding how many fire companies to dispatch to a call. Ignall, Carter, and Rider [1982] develop a model for determining how many and which fire companies to dispatch in a high workload urban system. Minimizing immediate response time and minimizing future losses are the criteria in both the Swersey and Ignall papers. Cuninghame-Greene, and Harries [1988] show that using the closest idle vehicle dispatch rule is optimal when the criterion is minimizing average response time. Weintraub, et . al. [1999] present a model and application for assigning and routing repair vehicles for the Emergency Services Division of the electricity utility in Santiago, Chile. This system operates similarly to an EMS system in that it has random calls and various traffic patterns during the day. The key feature of real time routing is primarily of interest to EMS and FS operations managers.

APPLICATIONS AND SOFTWARE SYSTEMS

There are numerous papers covering the actual application of planning and deployment models discussed in the previous sections. In this section, we include only those projects that are tied to specific cities or areas and papers that deal with strategies for implementation and have not been previously mentioned. We note that the usage of these models is international in scope and we have only sampled a subset of all of the actual applications.

When performing a successful study, it is important to know where problems might occur. Hogg [1968] first discusses a methodology for performing studies, defines the key problems that must be modeled, and discusses the approach in Bristol County, England. Chaiken [1978] provides key lessons for implementing OR studies in FS and EMS cases. Chaiken collected data that showed the extensive usage of RAND models for EMS and FS planning. Hypercube was extensively distributed and used. Issues such as extensive and accurate model documentation, the presence of an advocate in the organization, and data requirements enhance the model's usability and re-usability. Alternatively, Baker and Byrd [1980] write that if the study must be done quickly, sophisticated models may not be the appropriate approach as they can take substantial time. We organize this section by general modeling approach; math programming, queuing and simulation, and non-traditional approaches that have not been previously discussed.

Plane and Hendrick [1977] use maximal covering models to locate fire companies in Denver, CO. Schreuder [1981] uses mathematical programming to locate fire stations in Rotterdam. Eaton et. al. [1981], Eaton et. al. [1985], and Eaton et. al. [1986] use maximal covering models and the model of Daskin and Stern for locating ambulances in Columbia (South America), Austin, TX, and the Dominican Republic respectively. Badri, Mortagy, and Alsayed, [1998] describe a multi-criteria approach for locating fire stations in Dubai, UAE. The authors describe the entire solution process from setting the criteria to building a goal programming model, to assessing the current and suggested plans. Alsalloum and Rand [2003] use goal programming to model two criteria; cost as estimated by the number of vehicles, and the probability that a demand is actually covered. They use the approach in Riyadh Saudi Arabia and thoroughly discuss developing travel time models to estimate the probability that travel is within the coverage time limit.

Savas [1969] describes a simulation model that was used in New York to locate ambulances. The model suggested that locating ambulances near demand areas, rather than near hospitals, would increase performance. Fitzsimmons and Srikar [1982] use a location search routine and a simulation model to find based locations in Austin, TX. Brandeau and Larson [1986] describe applying the Hypercube model to locate ambulances in Boston, MA. Fujiwara, Makjamroen, and Gupta [1987] use Daskin's expected covering model on data from Bangkok, Thailand. They first find good locations with the model and then further evaluate these candidates using a detailed simulation model. Goldberg et. al. [1990] use Jarvis' spatial queueing model with search heuristic to find improved base locations in Tucson, AZ. Repede and Bernardo [1994] and Repede, Jeffries, and Hubbard [1993] use an adjustment to Daskin's expected covering model within a user-friendly software system to locate emergency medical vehicles in Louisville, KY. This is one of the first efforts to consider the "time" dimension explicitly in their modeling efforts as they adjusted their model for different times of the day and developed different sets of locations. Morabito and Mendonca [2001] use the Hypercube model to consider the deployment of ambulances in Brazil.

Carson and Batta [1990] use a scenario approach to locate a single vehicle on the Amherst Campus of the State University of New York at Buffalo. They consider 4 scenarios that model 4 periods during the day and try to find a location that minimizes area travel time for each scenario. McAleer and Naqvi [1994] use data analysis to help reduce area travel time in Belfast, Northern

Ireland. Swersey, Goldring, and Geyer, [1993] build a cost model to analyze the issues in merging fire and emergency medical units in New Haven, CT.

Two papers deserve mention even though they are not directly inline with the topic. Richard, Beguin and Peeters [1990] consider rural fire station location. This problem is different than urban location as the travel times are typically much longer. They consider three models for finding locations and one main result is that in these situations, having service equity can be expensive. Lane, Monefeldt, and Rosenhead, [2000] discuss a system dynamics model for a hospital emergency room. Even though this final model is not directly concerned with EMS and FS design and operation, the approach is an excellent example of model building and the system dynamics approach is applicable in many areas for medium and short term planning.

There are now many companies that provide significant EMS and FS modeling and analysis services and software systems. These companies work in all phases of EMS and FS planning and decision making including site and equipment location, dispatching and transport, and record keeping.

Isera uses its operations research scheduling technology to help model the vehicle loading problem and provide labor schedules for EMS and FS providers. Deccan International uses computer aided dispatch data to estimate necessary modeling parameters for their model - FIRE/EMS ADAM. EMS and FS personnel can use a drag-and-drop graphical-user-interface to experiment with different system designs (different equipment locations for example). Results are presented in easy to understand graphs, maps, and tables. Queues Enforth Development developed a graphical user interface for the Hypercube model and a computer aided dispatch program in the mid to late 1980's. They have expanded their product line to include training and products tailored to police applications. The Trapeze Software Group specializes in software for vehicle scheduling, dispatching, and reporting. They integrate intelligent vehicle systems with dispatch decision making to make patient pick-up and delivery systems more efficient.

PREVIOUS REVIEW PAPERS

We conclude this work with the numerous efforts at review papers in the OR literature. All of these are targeted towards the OR researcher and hence are mathematically more advanced than what we have presented here. The length of the list and the long time span again emphasize the importance of the problem and the richness of the potential application.

- Revelle, Marks, and Liebman, [1970] - considers private and public location models.
- Chaiken, and Larson, [1972] - reviews of all of the RAND reports and work in ES, EMS, and police.
- ReVelle, Bigman, Schilling, Cohon, and Church, [1979] - reviews all of the set covering and demand covering work of the 1970's. This paper is context free in that it simply talks of location as opposed to EMS or FS location.
- Ahituv and Berman [1988] - covers in detail a large part of the OR work on service systems from the 1970's and 1980's. Problem areas such as dispatching, repositioning, and base location are included. The text is geared towards the researcher and the practitioner. Results and applications are presented in the main chapters while appendices contain the detailed mathematical analyses.
- Brandeau and Chiu [1989] - covers all the models of the 1970's and 80's including the queueing work that came out of MIT. This paper is primarily for the OR researcher and contains a discussion of current open research problems.
- ReVelle [1989] - primarily covers the work from Johns Hopkins University on covering models including the probabilistic models and multiple coverage models of the late 1980's.
- Schilling, Jayaraman, and Barkhi, [1993] - contains a comprehensive review of all facility location problems that involve covering an area.
- Swersey [1994] - Comprehensive review in a handbook in operations research for public sector decision making. Spans police, FS, and EMS deployment.
- Marianov and Revelle [1995] - contains a broad general review of math programming models for siting EMS and FS bases. Includes probabilistic as well as deterministic models.
- Hesse-Owen, and Daskin [1998] - considers work relating to strategic decisions over an extended time horizon. This work covers some dynamic problems as well as probabilistic problems where some of the data is not known with certainty.

- Brotcorne, Laporte, and Semet [2003] - considers past work in deterministic and probabilistic covering models and optimization approaches. This paper has a special section on dynamic re-positioning of vehicles and this sets it apart from much of the other work.

DIRECTIONS FOR FUTURE WORK

Even though there has been a considerable amount of OR work applied in EMS and FS planning and operation, potentially fruitful areas for future research remain. As with the rest of the paper, the focus is on modeling approaches rather than model solution. It is clear that developing valid models with heuristic solution techniques is the correct approach rather than developing simplistic models with algorithms.

The major area of needed work is the real-time operation of EMS and FS systems. The decisions involved in vehicle relocation and dispatch are complex and simulation and analytic modeling can be used to help consider the impact of such decisions. For example, using simulation, one could "fast-forward" in time and evaluate the impact of relocation or dispatching decisions for the next 30 minutes or an hour. It may be possible to build a generic simulator that would accept various data and operational strategies as input. The key constraint on such an approach is computational time as the decision must be make quickly in real-time. Another off-shoot in this direction would be to build an "agent" that would look at system status and suggest relocation opportunities to the system manager. The agent could be running in background using the current system data as well as data from a traffic management system. Dispatching decisions would be automatically made along with any relocation needed.

Another area in daily planning that has had almost no attention is shift scheduling. It is not difficult to estimate the required number of vehicles needed per hour, however one must staff those hours with vehicles and crews. The goal is to meet the needed service level at minimum total cost. The difficulty of the problem is that each city has its own shift length and starting times and these must be used when doing the scheduling. In some sense, a problem has to be solved that fits shifts into each day to meet demand and then crews must be assigned to each shift so that the labor scheduling policies are viable. There has been considerable OR mathematical programming work in airline crew scheduling and the EMS scheduling problem has some similar attributes. Isera has worked extensively in this area, however their work is proprietary.

In the area of planning, there is little modeling work done in the design of ALS/BLS systems with the true cooperation and dispatching policy between the different types of vehicles. The mathematical programming work in this area has problems in that the dispatch policy of "send a BLS if an ALS is busy or further away" is not modeled since the busy and idle states of each vehicle are not modeled. This is difficult to do with analytical queuing models as well. Since computing power has increased significantly since 1974, it might be interesting to extend the ideas in Hypercube and A-Hypercube to the multi-vehicle type case as these might be computationally feasible.

The final area is data collection and modeling. If you cannot obtain data to run in the model, then there is little need for the model! Little work has been done on long term demand forecasting and hence long term capital investment models have little value. Most models use deterministic data or the average of a sample since there are few good estimating procedures to obtain distributions. Accurate travel and service time estimates are critical for building valid detailed models, but little OR work has been done in this area. Either this work must be brought from other areas (such as transportation engineering), or it has to be developed.

REFERENCES

Ahituv, N., and Berman, O., 1988, Operations Management of Distributed Service Networks, A Practical Quantitative Approach, Applications of Modern Technology in Business Series, Editor Eric Clemons, Plenum Press, New York. [Link to publisher]

Alsalloum, O., and Rand, G., 2003, "A goal-programming model applied to EMS system at Riyadh City, Saudi Arabia," Working paper 2003/035, Lancaster University Management School, Lancaster, UK. [Link to abstract]

Amiri, A., 1998, "The design of service systems with queueing time cost, workload capacities, and backup service," European Journal of Operations Research, vol. 104, pp. 201-217. [Link to abstract]

Amiri, A., 2001, "The multi-hour service system design," European Journal of Operations Research, vol. 128-3, pp. 625-638. [Link to abstract]

Badri, M., Mortagy, A., and Alsayed, C., 1998, "A multi-objective model for locating fire stations," European Journal of Operational Research, vol. 110-2, pp. 243-260. [Link to abstract]

Baker, D., and Byrd, J., 1980, "A lesson in timing: a nonemergency solution to an emergency service decision," Interfaces, vol. 10-3, pp. 30-33. [Link to journal]

Ball, M., and Lin, F., 1993, "A reliability model applied to emergency service vehicle location," Operations Research, vol. 41-1, pp.18-36. [Link to journal]

Batta, R., and Mannur, N., 1990, "Covering-location models for emergency situations that require multiple response units," Management Science, vol. 36-1, pp. 16-23. [Link to journal]

Batta, R., Dolan, J. and Krishnamurthy, N., 1989. The maximal expected covering location problem: Revisited. Transportation Science, vol. 23, pp. 277-287. [Link to journal]

Benveniste, R., 1985. Solving the combined zoning and location problem for several emergency units. Journal of Operational Research Society, vol. 36, pp. 433-450. [Link to journal]

Berlin, G., and Liebman, J., 1974, "Mathematical analysis of emergency ambulance locations," Socio-Economic Planning Sciences, vol. 8, pp. 323. [Link to abstract]

Berman, O. and Larson, R., 1982, "The median problem with congestion," Computers and Operations Research, vol. 9, pp. 119-126. [Link to abstract]

Berman, O., Larson, R. and Parkan, C., 1987. The stochastic queue p-median problem. Transportation Science, vol. 21, pp. 207-216. [Link to journal]

Borras, F., and Pastor, J., 2002, "The ex-post evaluation of the minimum local reliability level: an enhanced probabilistic location set covering model," Annals of Operations Research, vol. 111, pp. 51-74. [Link to journal]

Brandeau, M. and Chiu, S., 1989, "An overview of representative problems in location research," Management Science, vol. 35-6, pp. 645-674. [Link to journal]

Brandeau, M.L. and R.C. Larson, 1986, "Extending and applying the hypercube queueing model to deploy ambulances in Boston," in Management Science and the Delivery of Urban Service, edited by E. Ignall and A.J. Swersey, TIMS Studies in the Mangemnt Sciences Series, Vol. 22, pp. 121-154, North-Holland/Elsevier.

Brotcorne, L., Laporte, G., and Semet, F., 2003, "Invited review: ambulance location and relocation models," European Journal of Operational Research, vol. 147-3, pp.451-463. [Link to abstract]

Burwell, T., Jarvis, J., and McKnew, M., 1993, "Modeling co-located servers and dispatch ties in the hypercube model," Computers and Operations Research, vol. 20-2, pp. 113-119. [Link to abstract]

Carson, Y., and Batta, R., 1990, "Locating an ambulance on the Amherst Campus of the State University of New York at Buffalo," Interfaces, vol.20-5, pp.43-49. [Link to journal]

Chaiken, J., 1978, "Transfer of emergency service deployment models to operating agencies," Management Science, vol. 24-7, pp. 719-731. [Link to journal]

Chaiken, J., and Larson, R., 1972, "Models for allocating urban emergency units: a survey," Management Science Vol. 19-4, pp. 110-130. [Link to journal]

Charnes, A., and Storbeck, J., 1980, "A goal programming model for siting multi-level EMS Systems," Socio-Economic Planning Sciences, vol. 14, pp. 155-161. [Link to abstract]

Chelst, K., and Barlach, Z., 1981, "Multiple unit dispatches in emergency services: models to estimate system performance," Management Science, vol. 27-12, pp. 1390-1409. [Link to journal]

Chelst, K., and Jarvis, J., 1979, "Estimating the probability distribution of travel times for urban emergency service systems," Operations Research, vol. 27-1, pp. 199-204. [Link to journal]

Chiyoshi, F., Galvao, R., and Morabito, R., 2003, "A note on solutions to the maximal expected covering location problem," Computers and Operations Research, vol. 30-1, pp. 87-96. [Link to abstract]

Church, R. and ReVelle, C., 1974. The maximal covering locational problem. Papers of the Regional Science Association, vol. 32, pp. 101-108.

Church, R. and Weaver, J., 1985, "A median location model with non-closest facility service," Transportation Science, vol. 19, pp. 58-74. [Link to journal]

Cuninghame-Greene, R.A. and Harries, G., 1988, "Nearest-neighbour rules for emergency services," Zeitschrift fur Operations Research, vol 32-5, pp. 299-306.

Current, J., and Schilling, D., 1987, "Elimination of source of A and B errors in p-median location problems," Geographical Analysis, vol. 19, pp. 95-110. [Link to journal]

Current, J., Tatick, S., and ReVelle, C., 1997, "Dynamic facility location when the total number of facilities is uncertain: a decision analysis approach," European Journal of Operational Research, vol. 110, pp. 597-609. [Link to abstract]

Dai, J., Wang, S., and Yang, X., 1994, "Computerized support systems for emergency decision making," Annals of Operations Research, vol. 51-1, pp. 315-325. [Link to journal]

Daskin, M., 1983. A maximal expected covering location model: Formulation, properties, and heuristic solution. Transportation Science, vol. 17, pp. 48-69. [Link to journal]

Daskin, M., and Stern, E., 1981,"A hierarchical objective set covering model for emergency medical service vehicle deployment.," Transportation Science, vol. 15, pp. 137-152. [Link to journal]

Eaton, D., Church, R., Bennett, V., Hamon, B., and Lopez, L., 1981, "On deployment of health resources in rural Valle Del Cauca, Colombia," TIMS Studies in the Management Sciences vol. 17, pp. 331-359.

Eaton, D., Daskin, M., Simmons, D., Bulloch, B., and Jansma, G., 1985, "Determining emergency medical service vehicle deployment in Austin, Texas," Interfaces vol. 15-1, pp. 96-108. [Link to journal]

Eaton, D., Sanchez, H., Lantigua, R., and Morgan, J., 1986, "Determining ambulance deployment in Santo Domingo, Dominican Republic," Journal of the Operational Research Society, vol. 37, pp. 113-126. [Link to journal]

Erkut, E., and Bozkaya, B., 1999, "Analysis of aggregation errors for the p-median problem," Computers and Ooperations Research, vol. 26, pp.1075-1096. [Link to abstract]

Erkut, H. and Polat, S., 1992, "A simulation model for an urban fire fighting system," OMEGA, vol. 20-4, pp 535-542. [Link to abstract]

Fitzsimmons, J., 1971, "An emergency medical systems simulation model," Proceedings of the 1971 Winter Simulation Conference, pp. 18-25 New York. [Link to abstract]

Fitsimmons, J., and Stikar, B., 1982, "Emergency ambulance location using the contiguous zone search routine," Journal of Operations Management, vol. 21-4, pp. 225-237. [Link to abstract]

Francis, R., and Lowe, T., 1992, "On worst-case aggregation analysis for network location problems," Annals of Operations Research, vol. 40, 229-246. [Link to journal]

Fujiwara, M., Makjamroen, T., and Gupta, K., 1987, "Ambulance deployment analysis: a case study of Bangkok," European Journal of Operational Research, vol. 31, pp. 9-18. [Link to abstract]

Galvao, R., and ReVelle, C., 1996, "A Lagrangean heuristic for the maximal covering location problem," European Journal of Operational Research, vol. 88, pp. 114-123. [Link to abstract]

Gendreau, M., Laporte, G., and Semet, F., 1997, "Solving an ambulance location model by tabu search," Location Science, vol.5-2, pp.75-88. [Link to abstract]

Gendreau, M., Laporte, G., and Semet, F., 2001, "A dynamic model and parallel tabu search heuristic for real-time ambulance relocation," Parallel Computing, vol. 27, pp. 1641-1653. [Link to abstract]

Goldberg, J., Dietrich, R., Chen, J., Mitwasi, G., Valenzuela, T., and Criss, L., 1990, "Validating and applying a model for locating emergency medical vehicles in Tucson, Arizona," European Journal of Operational Research, vol. 49-3, 308-324. [Link to abstract]

Goldberg, J., Dietrich, R., Chen, J., Mitwasi, G., Valenzuela, T., and Criss, L., 1990a, "A Simulation Model for Evaluating a Set of Emergency Vehicle Locations: Development, Validation, and Usage," Socie-Economic Planning Sciences, vol. 24, 125-141. [Link to abstract]

Goldberg, J., and Paz, L., 1991, "Locating Emergency Vehicle Bases When Service Time Depends on Call Location," Transportation Science, vol. 25, pp. 264-280. [Link to journal]

Goldberg, J., and Szidarovsky, F., 1991, "Methods for solving nonlinear equations used in evaluating emergency vehicle busy probabilities," Operations Research, vol. 39-6, pp. 903-916. [Link to journal]

Halpern, J., 1977, "The Accuracy of estimates for the performance criteria of certain emergency service queueing systems," Transportation Science, vol. 11, 227-242 (1977). [Link to journal]

Hesse-Owen, S., and Daskin, M., 1998, "Strategic facility location: A review," European Journal of Operational Research, vol. 111-3, pp. 423-447. [Link to abstract]

Hillsman E., and Rhoda, R., 1978, "Errors in measuring distances from population to service centers," Annals of Regional Science, vol. 12, pp. 74-88. [Link to journal]

Hodgson, M., and Neuman, S., 1993, A GIS approach to eliminating source C aggregation error in p-median models," Location Science, vol. 1, pp. 55-70. [Link to journal]

Hogg, J., 1968, "The siting of fire stations," Operational Research Quarterly, vol. 19-3, pp. 275-287. [Link to journal]

Hogan, K., and ReVelle, C., 1986, "Concepts and applications of backup coverage", Management Science, vol. 32, pp. 1434-1444. [Link to journal]

Ignall, E., Carter, G. and Rider, K., 1982, "An algorithm for the initial dispatch of fire companies," Management Science, vol. 28-4, 366-378. [Link to journal]

Jarvis, J., 1975, "Optimization in stochastic systems with distinguishable servers," TR-19-75, Operations Research Center, M.I.T. [Link to insitiution]

Jarvis, J., 1985, "Approximating the equilibrium behavior of multi-server loss systems," Management Science, vol. 31-2, pp. 235-239. [Link to journal]

Jayaraman, V., and Stinastava, R., 1995, "A service logistics model for simultaneous siting of facilities and multiple levels of equipment," Computers and Operations Research, vol. 22-2, pp. 191-204. [Link to abstract]

Kamenetzky, R., Shuman, L., and Wolfe, H.,, 1982, "Estimating need and demand for prehospital care," Operations Research, vol. 30-6, 1148-1167. [Link to journal]

Kolesar, P., 1973, "Square root laws for fire engine response distances," Management Science, vol. 19, pp. 1368-1378. [Link to journal]

Kolesar, P., 1975, "A model for predicting average fire company travel times," New York City Rand Institute, R-1624-NYC, June. [Link to journal]

Lane, D.C., Monefeldt, C., and Rosenhead, J.V., 2000, "Looking in the wrong place for healthcare improvements: A system dynamics study of an accident and emergency department." Journal of the Operational Research Society, vol. 51-5, pp. 518-531. [Link to journal]

Larson, R., 1974, "A hypercube queuing model for facility location and re-sub-areaing in urban emergency services," Computers and Operations Research, vol. 1-1, pp. 67-95. [Link to journal]

Larson, R., 1975, Approximating the performance of urban emergency service systems," Operations Research, vol. 23, pp. 845-868. [Link to journal]

Larson, R., and Stevenson, K., 1972, "On insensitivities in urban redistricting and facility location," Operations Research, vol. 20, 595-612. [Link to journal]

Mabert, V., 1985, "Short interval forecasting of emergency phone call (911) work loads," Journal of Operations Management, vol. 5-3, pp. 259-271. [Link to abstract]

Mandell, M., 1998, "Covering models for two-tiered emergency medical services systems," Location Science, vol. 6-1, pp.355-368. [Link to abstract]

Marianov, V., and ReVelle, C., 1992, "The capacitated standard response fire protection siting problem: deterministic and probabilistic models," Annals of Operations Research, vol. 40-1, pp.303-322. [Link to journal]

Marianov, V. and ReVelle, C., 1995, "Emergency services," In: Drezner, Zvi, Editor, 1995. Facility Location: A Survey of Applications and Methods, Springer, New York, pp. 199-223. [Link to bookseller]

Marianov, V., and ReVelle, C., 1994, "The Queuing probabilistic location set covering problem and some extensions," Socio-Economic Planning Sciences, pp. 167-178. [Link to abstract]

Marianov, V., and ReVelle, C., 1996, "The Queueing maximal availability location problem: a model for the siting of emergency vehicles," European Journal of Operational Research, vol. 93-1, pp. 110-120. [Link to abstract]

Marianov, V., and Serra, D., 1998, "Probabilistic, maximal covering location-allocation models for congested systems," Journal of Regional Science, vol. 38-3, pp. 401-424. [Link to journal]

Marianov, V., and Serra, D., 2001, "Hierarchical location-allocation models for congested systems," European Journal of Operational Research, vol. 135-1, pp. 195-208. [Link to abstract]

Marianov, V., and Serra, D., 2002, "Location-allocation of multiple-server service centers with constrained queues or waiting times," Annals of Operations Research, vol. 111, pp. 35-50. [Link to journal]

Marsh, M., and Schilling, D., 1994, "Equity Measurement in Facility Location Analysis: A Review and Framework," European Journal of Operational Research, vol. 74-1, pp. 1-17. [Link to abstract]

McAleer, W., and Naqvi, I., 1994, "The relocation of ambulance stations: A successful case study," European Journal of Operational Research, vol. 75-3, pp. 582-588. [Link to abstract]

Monarchi, D., Hendrick, T. and Plane, D., 1977. Simulation for fire department deployment policy analysis. Decision Sciences, vol. 8, pp. 211-227. [Link to journal]

Morabito, R., and Mendonca, F., 2001, "Analyzing emergency medical service ambulance deployment on a Brazilian highway using the hypercube model," Journal of the Operational Research Society, vol. 52-3, pp. 261-270. [Link to journal]

Moore, G., and ReVelle, C., 1982, "The hierarchical service location problem," Management Science, vol. 28-7, pp. 775-780. [Link to journal]

Narasimhan, S., Pirkul, H., and Schilling, D., 1992, "Capacitated emergency facility siting with multiple levels of backup," Annals of Operations Research, vol. 40-1, 323-337. [Link to journal]

Neebe, A. W.,1988, “A procedure for locating emergency-service facilities for all possible response distances,” Journal of the Operational Research Society, vol. 39-8, 743-748. [Link to journal]

NFPA Fire Station Location Bibliography, July 1996. [Link to web page]

Pirkul, H., and Schilling, D., 1988, “The siting of emergency service facilities with workload capacities and backup service,” Management Science, vol. 37-7, pp. 896-908. [Link to journal]

Plane, D., and Hendrick, T., 1977, “Mathematical programming and the location of fire companies for the Denver Fire Department,” Operations Research, vol. 24-4, pp.563-578. [Link to journal]

Puig-Junoy, J., Saez, M.,, and Martinez-Garcia, E., 1998, “Why do patients prefer hospital emergency visits? A nested multinomial logit analysis for patient-initiated contacts,” Health Care Management Science, vol 1-1, pp. 39-52. [Link to journal]

Repede, J., and Bernardo, J., 1994, “Developing and validating a decision support system for locating emergency medical vehicles in Louisville, Kentucky,” European Journal of Operational Research, vol. 75-3, pp. 567-581. [Link to abstract]

Repede, J., Jeffries, C. and Hubbard, E., 1993, “ALIAS: A graphical user interface for an ambulance location model,” International Journal of Operations & Production Management, vol. 13-12, pp.36-46. [Link to journal]

ReVelle, C., 1989, “Review, extension and prediction in emergency service siting models,” European Journal of Operational Research, vol. 40-1, pp.58-69. [Link to abstract]

ReVelle, C., 1993, “Facility siting and integer-friendly programming,” European Journal of Operational Research, vol. 65, pp. 147-158. [Link to abstract]

Revelle, C., and Snyder, S., 1995, “Integrated fire and ambulance siting: A Deterministic Model,” Socio-Economic Planning Sciences, vol. 29-4, pp. 261-271. [Link to abstract]

ReVelle, C., Bigman, D., Schilling, D., Cohon, J., and Church, R., 1979, “Facility location: A review of context-free and EMS models,” Health Services Research, Summer 1979, pp. 129-146. [Link to journal]

ReVelle, C., and Hogan, K., 1989, “The maximum availability location problem,” Transportation Science, vol. 23-3, pp.192 - 200. [Link to journal]

ReVelle, C., and Hogan K., 1989a, “The maximum reliability location problem and ?-reliable p-center problem: Derivatives of the probabilistic location set covering problems,” Annals of Operations Research, vol. 18, pp. 155-174. [Link to journal]

ReVelle, C., Marks, D., and Liebman, J., 1970, “An analysis of private and public sector location models. Management Science, vol.. 16, pp. 692. [Link to journal]

ReVelle, C., Schweitzer, J., and Snyder, S., 1996, “The maximal conditional covering problem,” INFOR, vol. 34-2, pp.77-91. [Link to journal]

Richard, D., Beguin, H. and Peeters, D., 1990, "The location of fire stations in a rural environment: A case study," Environment and Planning A, vol.22-1, pp.39-52. [Link to abstract]
Savas, E., 1969, "Simulation and cost-effectiveness analysis of New York's emergency ambulance service," Management Science, vol. 15, pp. 608-627. [Link to journal]
Saydam, C., and Aytug, H., 2003, "Accurate estimation of expected coverage: revisited," Socio-Economic Planning Sciences, vol. 37-1, pp. 69-80. [Link to abstract]
Schilling, D., 1980, "Dynamic location modeling for public-sector facilities: A multi-criteria approach," Decision Sciences vol. 11, pp. 714-724. [Link to journal]
Schilling, D., Elzinga, Cohen, J., Church, R., and ReVelle, C., 1979, "The Team/Fleet models for simultaneous facility and equipment siting," Transportation Science, vol. 13-2, pp. 163-175. [Link to journal]
Schilling, D., ReVelle, C., Cohen, J. and Elzinga, D., 1980, "Some models for fire protection locational decisions," European Journal of Operational Research, vol. 5-1, pp. 1-7. [Link to abstract]
Schilling, D.A., Jayaraman, V. and Barkhi, R., 1993, "A review of covering problems in facility location," Location Science, vol. 1-1, pp.25-55. [Link to journal]
Schreuder, J., 1981, "Application of a location model to fire stations in Rotterdam," European Journal of Operational Research vol. 6-2, pp. 212-219. [Link to abstract]
Serra, D., 1996, "The coherent covering location problem," Papers in Regional Science, The Journal of the RSAI, vol. 75-1, pp. 79-101. [Link to journal]
Silva, F., and Serra, D., 2003, " Locating emergency services with priority rules: The priority queueing covering location problem," Economics Working Papers, Department of Economics and Business, Universitat Pompeu Fabra. [Link to full paper]
Storbeck, J., 1982, "Slack, natural slack, and location covering," Socio-Economic Planning Sciences, vol 16-3, pp. 99-105. [Link to abstract]
Swersey, A., 1982, "A Markovian decision model for deciding how many fire companies to dispatch," Management Science, vol. 28-4, pp. 352-365. [Link to journal]
Swersey, A.J., Goldring, L. and Geyer, E.D., 1993, "Improving fire department productivity-merging fire and emergency medical units in New Haven," Interfaces, vol. 23-1, pp.109-129. [Link to journal]
Swersey, A., 1994, ``The Deployment of Police, Fire, and Emergency Medical Units", in: S. M. Pollack, Handbooks in Operations Research and Management Science, vol. 6, Elsevier Science B. V., pp. 151-200. [Link to chapter outline]
Taylor, B.W., Baker, J.R. and Clayton, E.R., 1989, "A non-linear multi-criteria programming approach for determining county emergency medical service ambulance allocations," Journal of the Operational Research Society, vol. 40-5, pp. 423-432. [Link to journal]
Toregas, C., Swain, R., ReVelle, C. and Berman, L., 1971. The location of emergency service facilities. Operations Research, vol. 19-2, pp. 1363-1373. [Link to journal]
Trudeau, P., Rousseau, J., Ferland, J., and Choquette, J., 1989, "An operations research approach for the planning and operation of an ambulance service," INFOR, vol. 27-1, pp.95-113. [Link to journal]
Valenzuela, T., Goldberg, J., Keeley, K., and Criss, E., 1990, "Computer Modeling of Emergency Medical System Performance," The Annals of Emergency Medicine, vol. 19, pp. 898-901. [Link to abstract]
Van Buer, M., Venta, E., Hurter, A., and Lurigio, A., 1996, "The effect of vehicular flow patterns on crime and emergency services: The location of cul-de-sacs and one-way streets," Journal of the Operational Research Society, vol. 47-9, pp. 1110-1119. [Link to journal]
Volz, R., 1971, "Optimum ambulance location in semi-rural areas," Transportation Science, vol. 5-3, pp. 193-203. [Link to journal]
Weintraub, A., Aboud, J., Fernandez, C., Laporte, Gilbert, and Ramirez, E., 1999, "An emergency vehicle dispatching system for an electric utility in Chile," Journal of the Operational Research Society, vol. 50-7, 690-696. [Link to journal]
Zaki, A., Cheng, H., and Parker, B., 1997, "A simulation model for the analysis and management of an emergency service system," Socio-Economic Planning Sciences, vol. 31-3, pp.173-189. [Link to abstract]