Two-stage Stochastic Optimization for the Allocation of Medical Assets in Steady-state Combat Operations

Lawrence V. Fulton, Leon S. Lasdon, Reuben R. McDaniel, Jr and M. Nicholas Coppola

The Journal of Defense Modeling and Simulation: Applications, Methodology, Technology 2010 7: 89 originally published online 10 March 2010
DOI: 10.1177/1548512910364390

Original article

Journal of Defense Modeling and Simulation: Applications, Methodology, Technology
7(2) 89–102

DOI: 10.1177/1548512910364390
http://dms.sagepub.com

Two-stage Stochastic Optimization for the Allocation of Medical Assets in Steady-state Combat Operations

Lawrence V. Fulton, Leon S. Lasdon[1], Reuben R. McDaniel, Jr.[1] and M. Nicholas Coppola[2]

Abstract

In this study we describe a stochastic optimization model for the relocation of deployable military hospitals, the reallocation of hospital beds, and the emplacement of tactical medical evacuation assets (medical evacuation helicopters and ground ambulances) during steady-state military combat operations (stability operations). The network model is built around an intuitive objective function, one that is derived from military doctrine. The objective to be minimized is the time traveled, weighted by patient severity, from the evacuation site to the point of injury and onward to the hospital location. The optimal solution also determines the number of air and ground ambulances and the hospital beds of each type required at each selected site. Since future casualty locations, numbers, and severities are uncertain, this information is treated as a number of casualty scenarios with assigned scenario probabilities. The number, location, and severities of casualties can be randomly generated, or provided as part of a planning process. The model then seeks a single set of hospital and vehicle locations, plus the paths the evacuation assets should take in each scenario, which minimize expected travel time over all scenarios. The scenario generator is based on realistic historical data from Operation Iraqi Freedom. Since mobile hospitals provide the primary surgical treatment intervention while dedicated ground and air evacuation assets provide the transportation along evacuation paths, the study objective is important for military medical planners, especially those involved in tactical medical evacuation and treatment planning.

Keywords

emergency services, hospitalization, location problem, military, optimization

1. Background

The United States military relies on its mobile hospitals and tactical medical evacuation assets (medical evacuation helicopters and ground ambulances) to treat and evacuate its wounded from the battlefield. These assets provide the primary mechanisms for saving lives in combat. Mobile hospitals may be positioned on relatively short notice, while evacuation helicopters traverse vast expanses relatively quickly, providing the primary and preferred means of tactical medical evacuation support. For these reasons, relocation of hospitals and evacuation assets is an important decision that deserves careful analysis. In stability operations, military resources in a country are often fixed at relatively stable levels for several months. (For example, the coalition troop strength in Iraq from April 2008 through January 2009 averaged 147,000 with a standard deviation of 2300.)[1] Adjusting medical resources to support the uncertain future patient population becomes a focus of medical planners.

This study describes a stochastic optimization model for decision support in the relocation of medical assets during military stability operations. The model's objective is to minimize the sum of the penalty-weighted time traveled from potential evacuation sites to point of injury locations

[1]McCombs Business School, Austin TX 78712, USA
[2]Texas Tech University Health Sciences Center, 3601 4th Street, Lubbock, TX 79430, USA

Corresponding author:
Lawrence Fulton
Waco Baylor University, Waco Texas
larry_fulton@att.net

to possible hospital sites. These network paths are weighted by patient injury severity scores (ISSs) generated from Operation Iraqi Freedom (OIF) distributions. The casualty scenario generator distributes casualties around areas that are likely to see significant combat (casualty clusters), and the subsequent optimization determines the optimal location of assets given a series of appropriate constraints. The scenarios derive from a 3^{4-2} factorial design. The importance of this research is that military medical planners might optimize hospitalization and evacuation assets given a set of planning considerations, either simulated or real.

2. Motivation

One of the fundamental principles of military medical doctrine is the provision of far forward (proximal) medical care.[2] With this principle in mind, optimizing the locations of medical assets is both intuitively supportable and dictated by military doctrine. Solutions that achieve this objective should increase patient survivability and match capability with requirements. The authors were associated with a relocation of a military hospital in Iraq using methods similar but not identical to that discussed here,[3] which provided the initial motivation for this work. Since the development of this model, it has been adopted for use in post-hoc excursions of army experimentations by the Center for Army Medical Department Strategic Studies.

3. Relevant Previous Work

Many authors have used optimization models to determine emergency service facility locations, and these works provided the basis for development of this model. As early as 1971, Toregas et al.[4] minimized the number of emergency responders required based on response time or distance. They used integer linear programming and a set covering set formulation. Similarly, ReVelle and Hogan[5] employed probabilistic location set covering to demonstrate savings in terms of evacuation assets. More recently, Marianov and Taborga[6] determined the most effective locations for public health centers in Chile using a covering set formulation and heuristic. Batta and Mannur[7] also used a maximal covering location formulation to determine optimal allocation of emergency service units requiring multiple response units. This problem differs from covering location problems in that the objective function seeks to minimize patient travel time rather than provide maximal coverage. While the maximal coverage problem is useful for small area coverage, the problem in stability operations is often more complex. Emergency response may take hours, certain areas in which no troop densities exist require no coverage, and time to treatment has primacy in both doctrine and on the battlefield.

Aly and White[8] proffered a stochastic formulation for covering potential injury sites given that injuries occurred uniformly in given geographic areas. Aly and White's[8] work provided two underlying ideas for this study: stochastic components and uniform distribution of patients around geographically centered combat units, especially since military operations often involve considerable uncertainty requiring probabilistic excursions.

Neebe[9] provides a formulation for locating emergency services based upon all possible response distances. His formulation minimizes the number of facilities required based upon varying distance constraints. Neebe's concept of varying the distance provided some useful insight. In the case of this study, distances are randomly determined based upon known casualty clusters (e.g., areas historically using a phase-one simulation). The magnitude of the deviation may be manipulated but would normally be determined by distributions built from previously gathered empirics.

Eaton et al.[10] applied multi-objective programming procedures to determine ambulance locations in the Dominican Republic. While these authors provide an interesting model for consideration, the primacy of time provides a parsimonious objective function and ultimately the model for the work presented here.

Silva and Serra[11] discuss the importance of establishing different priority levels of patients in covering models. While not a covering model, this study uses a function of the ISS survival probabilities to weight the network paths and assigns bed requirement distributions to the population of injured patients in order to facilitate guidance on that path. This method is similar to that of Geoffrion and Graves[12] who assigned attributes in facility location models.

Bouma[13] discusses a multi-period model for evaluating flow through the military medical evacuation network, tactical through strategic. The author's study provides a theater analysis based upon known locations with the intent to optimize system flow rather than minimize first-responder time by generating the optimal locations for assets. The study presented here addresses a different yet complementary problem, that of proper asset location.

The addition of other scenarios has to be approached with techniques such as the Sample Average Approximation scheme discussed by Santoso et al.[14] In doing so, optimization over a set of scenarios is analyzed. In fact, this type of approach provides an excellent method for consideration when many scenarios are available for consideration, and this study also adopts this methodology.

The approach to this problem is to generate locations of patients based upon simulation and empirical distributions of previous casualties. These patients are generated for several scenarios. Since this study seeks to optimize the expected value based on the simulation over a set of scenarios, it logically falls into the category of stochastic optimization (rather than robust optimization) as defined by Snyder.[15] The objective function (minimization of the expected time traveled) is similar to the typical mini-sum location problem.

4. Context

Before specifying the model, some context is necessary. The Secretary of Defense has directed a one-hour evacuation standard for casualties. This one-hour standard requires coverage of those areas which have troop populations only, which provides the focus of the math programming objective. Serious casualties injured in combat are normally evacuated by dedicated air ambulance helicopters in order to provide fast evacuation to definitive care; however, on occasion ground ambulances provide that service. Patients are categorized by clinicians in both the military and civilian communities based on criticality using ISSs. ISSs range from 0 to 75 with high values resulting in near certain death.[16] Higher ISSs are generally associated with reduced patient survivability as illustrated in the National Trauma Databank,[17] so the speed in which a patient is transported becomes from point of injury to surgical care is critical to survival.

In military steady-state (stability) operations, transporting evacuation assets and crews reside at a set of finite evacuation sites. These sites are necessarily restricted to locations that have sufficient security and sustainment capabilities, as (for example) air evacuation assets require additional fuel, maintenance, and security. Further, the number of helicopters and ground ambulances available in steady-state operations is likely to be fixed over relatively long periods of time (just as in the troop populations discussed earlier), so distributing these assets across possible sites becomes largely a function of demand. Mobile hospitals require significant amounts of sustainment and security, so the number of feasible sites available for emplacement of these assets is small. If a site can support a mobile hospital, it is likely to be able to support ground and aeromedical evacuation assets.

One should note that some of the hospital and evacuation sites may be selected by planners, while others are fixed (i.e. one must maintain these sites regardless of contribution towards mission accomplishment). The sites to be selected may include both locations currently in use but subject to possible closing and new sites that may host either a new or a relocated facility. Sites are limited in that both security and logistics must be available for ground crews and helicopters. Hospital sites are generally feasible locations for evacuation sites as well. Military logistics and operational planning factors provide the feasible locations.

In addition to the previous considerations, lethality characteristics have an effect on the number of casualties generated during a relevant timeframe.[18] To account for such lethality, scenarios might be employed which allow for variation of the effectiveness in enemy attacks. This approach is employed in this study as one of the design factors.

This study seeks to minimize the penalty-weighted time traveled from evacuation sites to points of injury to hospital locations over all considered scenarios given constraints on the number of evacuation and hospitalization sites. The return path from the hospital to the evacuation site is deliberately not included in the objective function because this potential leg, unlike the other path components, is not associated with time to patient treatment unless demand exceeds the queue, e.g., a mass casualty situation in steady-state operations, in which case military commanders would logically assign other non-dedicated aircraft to support the crisis. The focus of this study is steady-state operations in which resources in-theater are relatively stable over the relevant range and where casualties are sparse in comparison to major combat operations but often severe.

5. Phase I Simulation

5.1. Hospital Components and Feasible Locations

To establish the appropriate data stream to be optimized, a simulation motivated from OIF was conducted. In this simulation, there were 10 possible hospital locations of which 5 were to be selected. Further, 20 possible evacuation locations exist with 10 to be selected. While one would normally know exactly which locations were feasible based on security, logistics, and other operational planning factors, the coordinates were randomly generated for all hospital and evacuation locations by sampling uniformly over a grid. These points were then fixed as part of all future simulation runs. The hospitalization sites were assumed to be a subset of the evacuation sites, since any site capable of sustaining a hospital would be expected under all conceivable circumstances to sustain evacuation capability as well, so the first 10 randomly generated coordinates applied to both hospital and evacuation sites. The coordinate system for the simulation was selected to represent the size (in nautical miles [NM]) of a large country (400 NM by 400 NM).

Total beds for all hospital designs by type were set to 1000 minimal care ward (MCW) beds (200 per hospital), 1000 intermediate care ward (ICW) beds (200 per hospital), and 240 intensive care unit (ICU) beds (48 per hospital); bed types need not be equally distributed among hospitals. The number of beds selected represents those beds that would be available in 5 separate 248-bed combat support hospitals with attached minimal care capability.[19] The assumption is made that, over the relevant range, hospital bed availability (e.g. those not being occupied) is 25% per type (an assumption that would be typically based on empirical observations during stability operations), leaving a daily average of 250 MCW beds, 250 ICW beds, and 60 ICU beds available.

5.2. Evacuation Components and Feasible Locations

Evacuation locations were generated in an initial simulation run but then fixed as part of all scenarios. The number

Table 1. Locations of casualty centers for simulation

Grid	Percent of Casualties
{100,100}	20%
{400,100}	35%
{100,400}	25%
{400,400}	20%

Table 2. Casualties at same location, data derived from Operation Iraqi Freedom Medical Evacuation Logs (source: Medical Evacuation Proponency Directorate)

Number of patients	$P(X{=}x)$
1 patient	0.574
2 patients	0.360
3 patients	0.050
4 patients	0.016

of missions per day that might be conducted by evacuation assets was based upon the presence of 60 air ambulance helicopters (about 5 × 12-aircraft companies) and 240 ground ambulance vehicles (about 10 × 24-vehicle companies).[21] Based on the operating characteristics of the platforms and limitations of human performance, a ground ambulance can conduct at most 10 lifts/day and air ambulances 5 lifts/day. Ground ambulances are assumed to operate uniformly on the interval 45 ± 15 NM/hour (about 52 ± 17 statute miles/hour) which accounts for patient loading, unloading, circuitous routing, urban environment, etc., while air ambulances are assumed to operate at 95 ± 20 nautical miles per hour (again, accounting for patient loading, unloading, threat, terrain, etc.). In stability operations, these parameters are less important.

5.3. Casualty Generation

Future locations and numbers of casualties might be approximated based on empirical locations of previous casualties. In this example, the assumption is that medical planners have estimated four areas (military brigade areas) that are likely to see high casualties based on either previous casualty patterns or known upcoming military operations. These areas would likely be associated with military brigades conducting operations in and around "hot-beds", for example Ramadi, Baghdad, Tikrit, or Mosul, Iraq during the years 2005 through 2007. For the purposes of this example, the following coordinate pairs are used for centers of operation: {100,100}, {400,100}, {100,400}, {400,400}. A frequency distribution then assigns the casualties occurring within each area based on unit mission and priority (e.g. main effort or supporting effort). In this example, the distribution in Table 1 provided the baseline distribution, although the actual distribution would be based on empirics.

A second stochastic element generates a casualty count for each event based on ongoing military operations. Table 2, which is based on OIF flight logs provided by the Army Medical Evacuation Proponency Directorate, provides the probability mass function for determining the number of casualties at a given location.

To model uncertainty regarding enemy capabilities in this area, a lethality multiplier (Factor A in the experimental design) was applied to the casualties generated at each location. The multipliers evaluated were {1.0, 1.5, 2.0} and served to evaluate the sensitivity of the location selection based on the number of injuries experienced on a site. The application of this multiplier also reflects uncertainty in enemy capability.

With the casualty centers identified, uniform randomness is applied to a casualty estimate. The distribution of casualties uniformly on a circle reflects a lack of knowledge but is reasonable in that brigades are assigned areas of operation in which to work. Empirical distributions of previous casualties and knowledge of upcoming security operations would generally provide the basis for determining future locations; however, a simple method provides a mechanism for assigning casualties to brigades when knowing only the previous frequency distribution of casualties. First, a casualty location center (n, o) is determined by sampling from Table 1. Then, a random angle, $p \sim U(0,360)$, is generated. A random magnitude, $q \sim U(-h,+h)$ reflecting the operating radius of a brigade is selected, and the casualty location then becomes $(n' = n + q\cos(p),\ o' = o + q\sin(p))$. The parameter, h, is Factor B in the experiment with values of {50 NM, 100 NM, 150 NM}. These magnitudes reflect the possible operating space of units in the four casualty cluster areas and hence serve as feasible values for investigation.

Another element used in the simulation involves casualty severity weighting, i.e. the determination of penalty weights which account for the most critically injured patient along each traversed path. The military maintains a database that provides the distribution of injury severity scores in current operations (Table 3),[18] while the National Trauma Databank provides survival probabilities based on injury severity scores (Table 4). This study takes samples from military distributions to estimate the severity of the most critically injured casualty in the group, and then uses the associated survival probability from the National Trauma Databank. After estimating the survival probability for the most critical patient, the inverse of that probability provided the weight for the network path. If one considers survival to be a geometric distribution, then the weights are the geometric expectation of the most critically injured person in the group. Further, patients from this event are

Table 3. Injury severity score percentages for Operation Iraqi Freedom (source: Devore[21])

Injury Severity Score	$P(X=x)$
1–8	0.45
9–15	0.32
16–24	0.12
>24	0.11

Table 4. National trauma databank survival by injury severity score (ISS), 2007

Injury severity score	P(Fatal)	P(Non-fatal)	Penalty weight
1–8	0.01	0.99	1.01
9–15	0.02	0.98	1.02
16–24	0.05	0.95	1.05
>24	0.29	0.71	1.41

assigned a bed category based on injury assignment. The most critically injured patients (based on the top groups of ISS scores in Table 4) are anticipated to occupy an ICU. Lower ISS scores associated with the bottom ranges in Table 4 are assigned to ICW and MCW beds. These distributions would normally be based on empirical observations of casualty types; however, the assignment based on severity is a logical simplification.

Finally, the percentage of ground ambulance use is Factor C in the fractional factorial design. This factor with values {0.2, 0.3, 0.4} appears directly in a constraint rather than the simulation and will be discussed as part of the optimization.

5.4. Simulation Run Estimates

The simulation allows the number of casualty locations and asset / location capabilities to be set for any given timeframe. In the example presented here, the number of casualty locations is motivated by the Brookings Institute Iraq Index which shows the monthly number of wounded-in-action from March 2005 through March 2007 to be approximately 500 per month[1] or about 117 in an average week. Based on the expected value of casualties per location, 80 locations were selected for runs. These locations would be expected to produce 121 casualties. To account for uncertainty in casualty generation, another experimental factor (Factor D, days of operation) was included, taking on values of {4, 7, 10}. This factor was manipulated to evaluate the effect of increasing and decreasing operational tempo on location and ambulance use by type *m*, if any.

We generated the scenario data and created and solved the optimization model using the General Algebraic Modeling System (GAMS, see http://www.gams.com). It is well suited to problems with multidimensional parameters, equations, and variables such as this. In estimating the number of runs required, the variable of interest was the minimum time required to traverse the network from evacuation site to patient pick-up to hospital. Based on an initial simulation run of size $n = 100$, the standard deviation for the time to travel along the network was 0.629 hours. Given a 95% confidence interval, a sample size of approximately $n = 98$ casualties was expected to bracket the mean time within ±7.5 minutes (15 minutes). The model provided here evaluated a 7-day period that included 80 independent casualty locations for 9 separate scenarios; applying the distribution of casualties per location results in expected values that would meet the 15 minute margin of error for all independent run, $E(\#\ \text{Casualties} \mid \text{Lethality}) = \{121 \mid 1.0, 181 \mid 1.5, 271 \mid 2.0\}$. All 9 runs with their estimated 1719 casualties should provide an interval around mean time of transport within ±2 minutes.

6. The Optimization Model

The optimization of all five scenarios provided a mechanism for assessing the optimal location for emplacement of assets. To formulate the problem, the following sets are defined.

6.1. Sets

The following sets provided both the path and the necessary attributes for modeling the scenario.

T = scenarios for evaluation with index $t \in T$.
I = set of locations where injuries may occur with index $i \in I$.
J = set of candidate helicopter evacuation sites with index $j \in J$.
K = set of candidate hospital sites with index $k \in K$.
L = type of bed requirement for a patient (minimal, intermediate, and intensive care) with $l \in L$.
M = type of evacuation assets, ground or air with $m \in M$.

6.2. Problem Data

The following problem data are assumed available.

p_t = probability that scenario t occurs (non-negative and summing to 1).
a_{it} = injury severity weight of most critically injured patient departing from site i for scenario t, assigned stochastically.
b_{ijt} = distance between injury location i and helicopter evacuation site j for scenario t.
c_{ikt} = distance between injury location i and hospital location k for scenario t.
d_{jkmt} = speed of transport from j to k by vehicle type m for scenario t, assigned stochastically.

e_{it} = number of total patients injured at location i for scenario t, assigned stochastically.
f_{lt} = the number of patients with type l bed requirements.
$ecap_{mt}$ = capacity during simulation time period for patient evacuation by m-type evacuation assets.
$hcap_{lt}$ = capacity during simulation time period for hospital acceptance of l-type patients.
$enod_{jt}$ = evacuation node capacity for each node j.
$hnod_{kt}$ = hospital node capacity for each node k.
u = maximum number of hospitalization sites to be occupied.
v = maximum number of evacuation sites to be occupied.
wx_{jt} = percentage of ground ambulances required for each evacuation site based on optimal mix factors.

The scenario probabilities p_t are generally subjective. In our computational experiments they are all equal. The weights (a_{it}) for the simulation are derived from inverse cumulative distribution function (CDF) sampling of g during phase I. The simple distance formula provides the distances between the randomly generated casualties and helicopter and evacuation sites (b_{ijt} and c_{ikt}). Two distributions represent fluctuations in vehicle operation based on inherent characteristics and the environment (d_{jkmt}). For ground ambulances, a uniform distribution between 30 and 60 NM/hour provides a base planning factor, while air ambulance speed is assumed to fluctuate between 75 and 105 NM/hour. While these planning factors are based on operational experience and include patient loading and unloading time, medical planners would manipulate them as necessary and for excursions. For this simulation, the number of patients injured at each location (e_{it}) is derived from OIF distributions. The number of patients assumed to require type l beds (f_{lt}) is generated from the sampling of injury severity as discussed. The $ecap_{mt}$ and $hcap_{lt}$ data were established as discussed previously. The *enod* and *hnod* values provide additional node capacity levels. Hospital and evacuation locations (u and v) are restricted to 5 and 10 sites, respectively. Finally, *wx* forces a mix of air and ground at each site in order to maintain evacuation capability in inclement weather (e.g. sandstorms, thunderstorms, etc.). This percentage is Factor C and takes the values {0.2, 0.3, 0.4} which reflects that air ambulance support is considered the primary and preferred method of evacuation.

6.3. Decision variables

To seek a set of evacuation and hospital sites that minimizes the total time traveled by all patients, the following decision variables are required.

x_{ijklmt} = number of patients traveling from injury location i with bed-type requirement l on vehicle m to hospital site k by a vehicle and crew located at evacuation site j for scenario t.
y_j = 1 if air evacuation site j is chosen and 0 otherwise.
z_k = 1 if hospitalization site l is chosen and 0 otherwise.

The decision variable counts the number of traveling patients along the network while maintaining visibility of their attributes. The integer variables y_j and z_j turn on or off the respective evacuation and hospitalization sites.

6.4. Constraints

The first constraint ensures that all patients are evacuated from each location i and scenario t:

$$\sum_j \sum_k \sum_l \sum_m x_{ijklmt} = e_{it}, \forall i, \forall t$$

The left-hand side of this constraint is the total number of casualties who are transported by all tactical medical evacuation aircraft to all hospitals. This sum must equal the number of casualties specified as originating at location i in scenario t.

The second constraint provides adequate evacuation capacity without exceeding system evacuation capability for any vehicle type in any scenario. The sum of the number of patients traveling the network on vehicle type m is forced to be less than or equal to the system capacity for providing evacuation assets of that type at all available locations. To determine how many to place at this site in an unrestricted planning environment, $ecap_{mt}$ might be set large:

$$\sum_i \sum_j \sum_k \sum_l x_{ijklmt} \leq ecap_{mt}, \forall m, \forall t$$

The third constraint provides adequate hospitalization capacity without exceeding system hospitalization capability by bed type and for every scenario. The sum of the number of patients traveling the network with bed-requirement m is forced to be less than or equal to the system capacity for providing beds of that type to all available locations for scenario t. (Hospitals may be manually configured to have varying bed capacities at different sites.) To determine how many to place at this site in an unrestricted planning environment, $hcap_{lt}$ might also be set large:

$$\sum_i \sum_j \sum_k \sum_m x_{ijklmt} \leq hcap_{it}, \forall l, \forall t$$

Two constraints allow for node throughput limits by restricting capacity at either evacuation or hospitalization nodes. To allow unconstrained nodes and post-hoc determination of what is needed, both *enod* and *hnod* might be set large:

$$\sum_i \sum_k \sum_l \sum_m x_{ijklmt} \leq enod_{jt} y_j, \forall j, \forall t$$
$$\sum_i \sum_j \sum_l \sum_m x_{ijklmt} \leq hnod_{kt} y_k, \forall k, \forall t$$

Table 5. Values for the 3^{4-2}_{III} fractional factorial design, $I = AB^2C$, $I = BCD$

Run	Lethality multiplier	Radius magnitude	Min % Evacuated by ground	Number of days
1	1.00	50.00	20%	7.00
2	1.50	50.00	40%	14.00
3	2.00	50.00	30%	21.00
4	1.00	100.00	30%	14.00
5	1.50	100.00	20%	21.00
6	2.00	100.00	40%	7.00
7	1.00	150.00	40%	21.00
8	1.50	150.00	30%	7.00
9	2.00	150.00	20%	14.00

When evacuation capacity is large such as in stability operations, solutions will generally select the military's preferred and primary source of evacuation assets, the air ambulance. Weather sometimes limits the ability of aircraft to fly, however. To account for limitations in use of air ambulances, an additional constraint was used. The left-hand side ensures that the evacuations conducted by ground ambulance are greater than a percentage of total evacuations at each evacuation site j:

$$\sum_i \sum_j \sum_k \sum_l \sum_m x_{ijkl,m='G',t} \geq wx_{it} \sum_i \sum_j \sum_k \sum_l \sum_m x_{ijklmt}, \forall t$$

The total number of all open evacuation sites and hospitals are limited using the following constraints. For this analysis, v represents the available evacuation sites, while u represents the available hospitalization sites. One can see that the decision variables, y and z, are not indexed by scenario. In this way, only a single set of open nodes are chosen which is necessary to find the optimal over all scenarios:

$$\sum_j y_j = v, \qquad \sum_k z_k = u$$

Finally, the flow variables x must be non-negative while the others must be binary.

$$x_{jikt} \geq 0, \qquad y_j \in \{0,1\}, \qquad z_k \in \{0,1\}$$

If certain nodes must remain open, one can add constraints forcing their associated y or z value to 1. Further, there is no need to restrict the x and y to be integer, because the problem has an integer optimal solution for any fixed set of binary y and z variables for which it is feasible due to its network structure.

6.5. Objective Function

The objective is to minimize the expectation over all scenarios of the total, penalty-weighted time traveled by all patients:

Minimize

$$\sum_i \sum_k \sum_l \sum_m X_{ijkl,m='G',t} \geq WX_{jt} * \sum_i \sum_k \sum_l \sum_m X_{ijklmt}, \forall t, j$$

The objective function sums the penalty-weighted time for evacuating patients via helicopter or ground ambulance along all paths for all scenarios. If some candidate locations are fixed, one simply does not define variables for these indices. The entire GAMS stochastic optimization less reporting functions is provided in Appendix A.

With both the simulation and optimization models in hand, the last remaining component prior to discussing results is the experimental selection. With 3 levels and 4 factors, the full 3^4 design would require optimization over 81 scenarios, a computationally expensive task considering the size of the problem. In order to evaluate the factors reasonably, a fractional factorial 3^{4-2}_{III} design was selected. While this design aliases second-order interactions, the main effects are clear and the design is often recommended when curvature is present.[22] The defining equations were $I = AB^2C$, $I = BCD$. Factor A was defined as the lethality multiplier, and Factor B was the magnitude in which the brigade was operating. Factor C was the ground evacuation percentage, while Factor D was the number of days in which the casualties were experienced. Table 5 summarizes the factors and levels for each simulation run. The results of the simulation, the optimization, and the post-experimental analysis follow.

Results

GAMS provided the platform for both the simulation and the optimization. Appendix A provides the entirety of the GAMS code for replication purposes. The simulation leveraged inverse CDF sampling and built-in functions while the optimization model was solved via the IBM-based OSL solver. GAMS provides reasonable capability for simulation via pseudo-random number generation, seed assignment, probability functions, and programming flow control; it also provides an excellent optimization modeling platform.

Prior to optimization, the generation of the model (the stage 1 simulation runs) took 256 MB of memory and 37.7 seconds for initialization and initial simulation runs. The model contained 1234 single equations with 1090 rows and 864,031 columns, 2.6 million single variables (30 discrete), and 6.0 million non-zero elements. On a Dell XPS 1730 laptop with 3.8 GHz Intel Core 2 Extreme processor with 4 GB of RAM and twin Raid 0 solid-state hard drives, the OLS solver completed 3400 simplex iterations before proceeding to branch and bound. After 43,225 iterations and 19.3 minutes, the OSL solver produced an integer optimal solution meeting the 0.01 optimality criterion. The solution included the following sets of evacuation and hospitalization locations: {E1, E2, E3, E4, E5, E7, E12, E13, E15, E20}

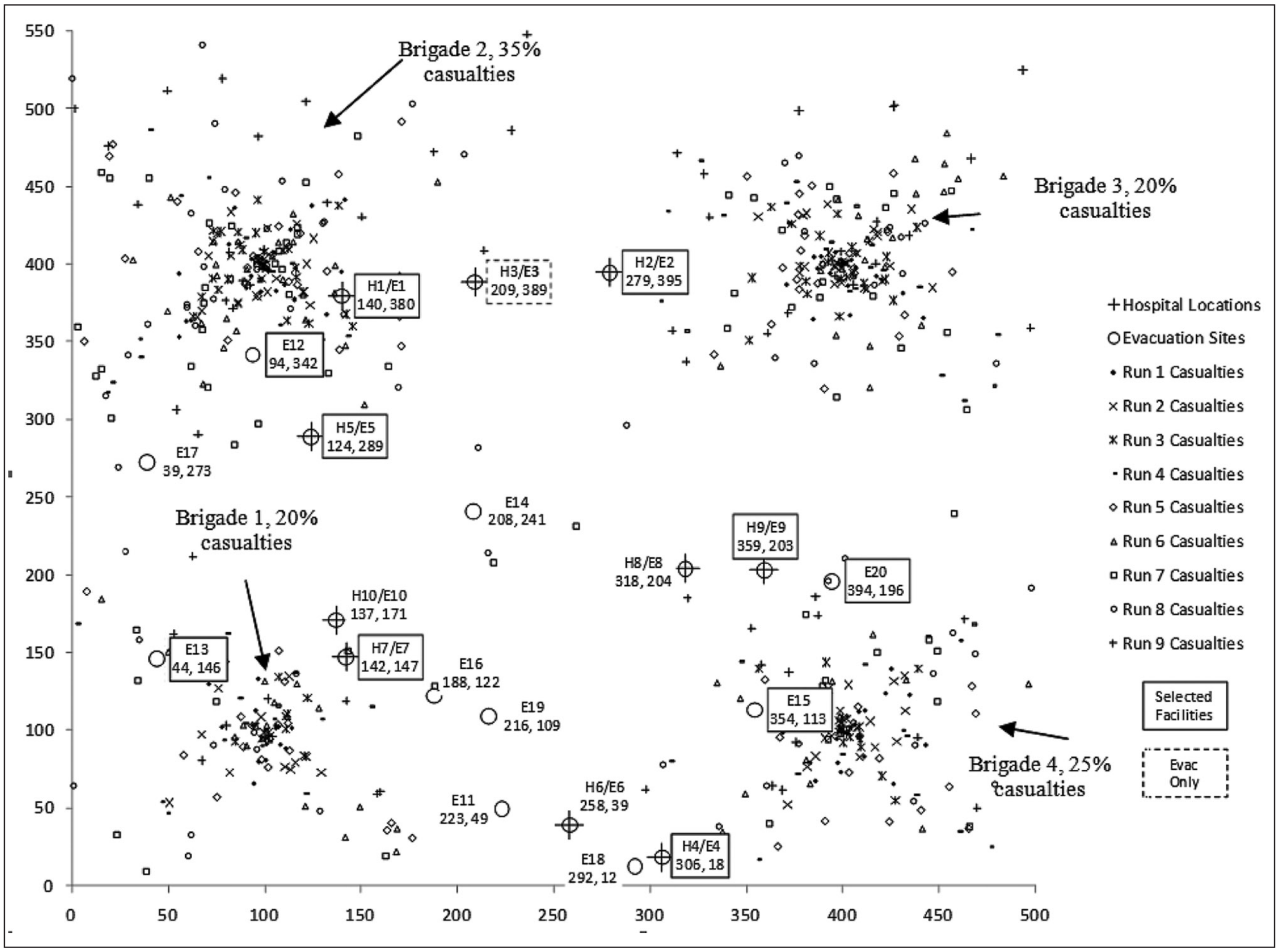

Figure 1. Graphical representation of the results of the optimization model.

and {H1, H2, H4, H7, H9}. Figure 1 depicts the optimization graphically, while Figure 2 graphically illustrates the computational complexity of adding additional casualties and scenarios.

For the multiple scenario locations randomly generated in the simulation, the average number of casualties was 203.60 per run. Simulation results provided an average time from evacuation site to point of injury of 1.93 hours with a standard error of 1.92 minutes. Average distance traveled was about 179.80 NM, and the average speed was 92.96 NM/hour.

Table 6 details the descriptive statistics for the combined scenarios. The table reveals interesting information in the form of recommended allocations of air ambulances, ground ambulances, MCWs, ICWs, and ICUs. Looking at the evacuation distribution, one can readily see that ground ambulance usage is largely centered around evacuation site 1 (22% of entire allocation of ground vehicles). Since large numbers of casualties are clustered near that location (i.e. the brigade located in that vicinity takes 35% of the casualties as shown in Figure 1) and since time in transit is the focus of the objective function, the result is understandable. The distribution of the largest percentages of air ambulance assets is {20%, 17%, 15%} for evacuation sites {E1, E2, E13} respectively. Again, E1 would likely receive more assets based on this distribution, which is intuitive. Interestingly, sites {E3, E4, E15} would have fewer than 5% of the air ambulance assigned based on demand, a finding of relevance to decision makers considering consolidation/downsizing efforts. In terms of hospital sites, H1 would be the largest as it requires 32% of the MCWs, 36% of the ICWs, and 32% of the ICUs. Since H1 is co-located with E1, the reason is clear. Conversely, H4 requires on 10% of the MCWs, 9% of the ICWs, and 10% of the ICUs. This type of allocation analysis provides significant utility for this model.

Post hoc analysis of the transport times along the network for all scenarios (9 scenarios*80 observations, $n = 720$)

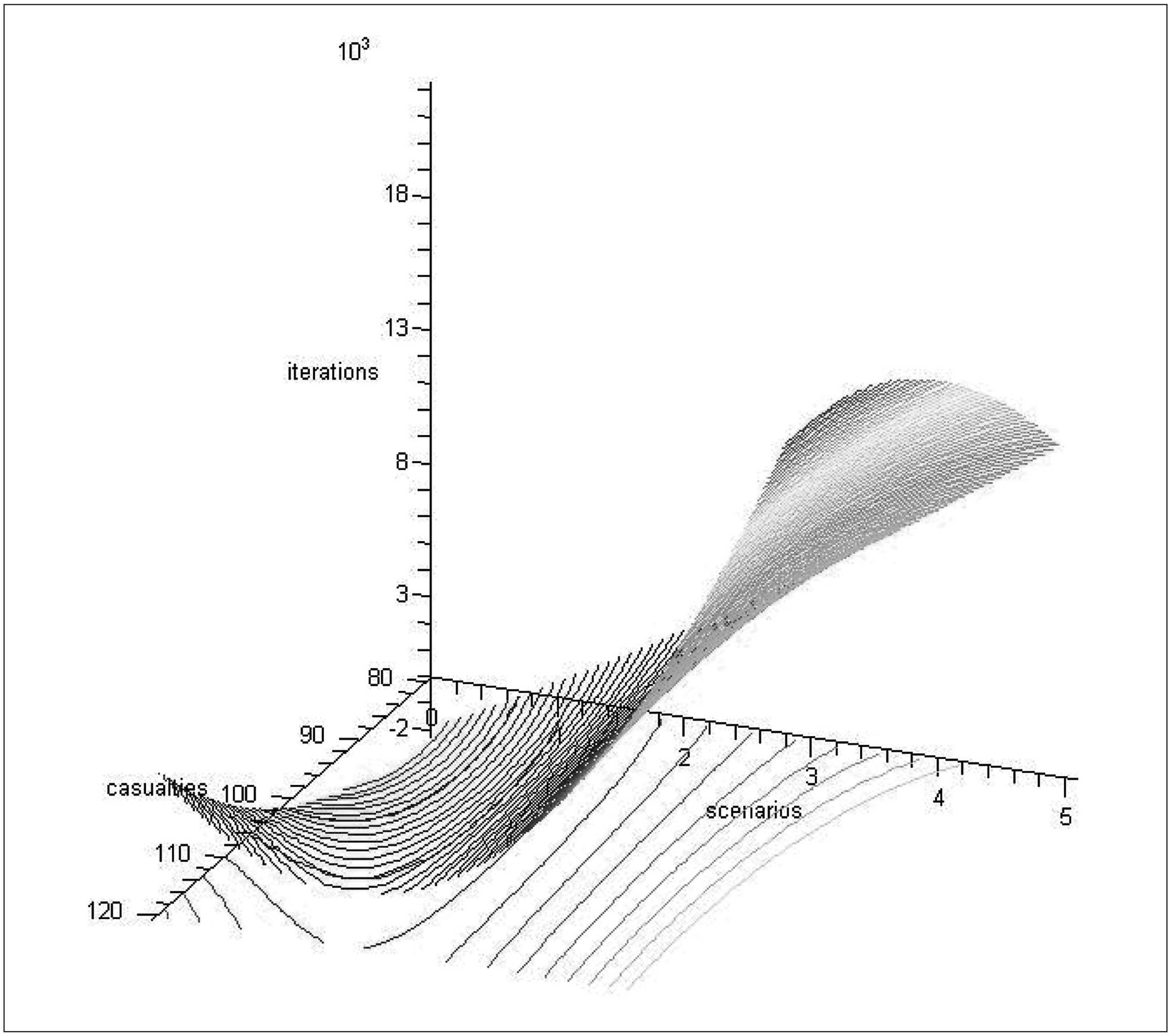

Figure 2. Contour plot of the stochastic optimization complexity by scenarios and casualty location numbers.

provides some interesting insight. Analysis of variance indicated that lethality and radial distance variations associated with different brigade operations were statistically significant variables in a full main effects model (see Table 7). The lethality effect of the lethality variables might be explained in that the network becomes saturated at points.

Limitations, Future Work, and Conclusions

This study uses only small subsets of potential patient, vehicle, and hospital attributes and considers grouped time rather than multi-period analysis. Adding attributes increases model complexity, so proper and parsimonious selection is important for future work. Multi-period analysis is less useful when determining the optimal geographic location of assets in stability operations given that asset flexibility is limited to that already on-hand within the theater of operations. Still, a multi-period extension with a different objective function is planned as an extension designed to model medical support for major combat operations rather than stability operations.

Future work will also include force structure analysis based on military stability operations experiments. By inserting parameters that represent future capabilities of evacuation and hospitalization assets, one can then analyze

Table 6. Descriptive statistics for all runs of the scenario. Averages are for the scenario (one month).

j=*evacuation site* *k*=*hospitalization site* *m*=*vehicle type* *l*=*patient type*	*m*=*G*, Avg. Patients Moved by Ground	*m*=*A*, Avg. Patients Moved by Air	*k*=H1, Avg. Patients Moved to Hosp. 1	*k*=H2, Avg. Patients Moved to Hosp. 2	*k*=H4, Avg. Patients Moved to Hosp. 4	*k*=H7, Avg. Patients Moved to Hosp. 7	*k*=H9, Avg. Patients Moved to Hosp. 9	$P(j{=}E^* \mid m{=}G)$	Ground Ambulance Asset Allocation at Site j	$P(j{=}E^* \mid m{=}A)$	Air Ambulance Asset Allocation at Site *j*
j=E1, Evac. Site 1	6.75	34.57	40.66	0.67				22%	53	20%	12
j=E2, Evac. Site 2	5.17	30.12	2.08	32.55			0.67	17%	41	17%	10
j=E3, Evac. Site 3	0.97	7.28	3.54	4.72				3%	8	4%	3
j=E4, Evac. Site 4	0.46	3.54			4.00			1%	4	2%	1
j=E5, Evac. Site 5	2.98	15.53			1.22	17.29		10%	23	9%	5
j=E7, Evac. Site 7	2.90	18.19	20.43			0.67		9%	23	11%	6
j=E12, Evac. Site 12	2.15	13.64				15.79		7%	17	8%	5
j=E13, Evac. Site 13	4.71	26.50			15.44		15.77	15%	37	15%	9
j=E15, Evac. Site 15	0.44	2.67				3.00	0.11	1%	3	2%	1
j=E20, Evac. Site 20	4.10	20.92		12.44	0.24		12.33	13%	32	12%	7
Totals	30.63	172.97	66.70	50.38	20.90	36.74	28.88			240	60
$E(l{=}MCW \mid column)$	26.89	142.55	55.06	45.27	17.68	26.59	24.84	169.4	Avg. Patients Moved to MCW		
$E(l{=}ICW \mid column)$	1.62	17.41	6.78	3.56	1.78	5.02	1.89	19.0	Avg. Patients Moved to ICW		
$E(l{=}ICU \mid column)$	2.12	13.01	4.87	1.56	1.44	5.13	2.15	15.1	Avg. Patients Moved to ICU		
$P(column \mid l{=}MCW)$	16%	84%	32%	27%	10%	16%	15%	100%	% Patients Moved to MCW		
$P(column \mid l{=}ICW)$	9%	21%	36%	19%	9%	26%	10%	100%	% Patients Moved to ICW		
$P(column \mid l{=}ICU)$	14%	51%	32%	10%	10%	34%	14%	100%	% Patients Moved to ICU		
MCW Asset Allocation for Hospital *k*			325	267	104	157	147	1000	Total MCW beds in scenario		
ICW Asset Allocation for Hospital *k*			356	187	93	264	99	1000	Total ICW beds in scenario		
ICU Asset Allocation for Hospital *k*			80	26	24	85	35	250	Total ICU beds in scenario		

Table 7. Analysis of variance results

Source	Type III sum of squares	df	Mean square	F	p
Model	2921.388[a]	9	324.599	366.793	<0.001
Lethality	6.160	2	3.080	3.481	0.031
Radius	7.348	2	3.674	4.152	0.016
% ground	1.908	2	.954	1.078	0.341
Days	3.506	2	1.753	1.981	0.139
Error	629.209	711	.885		
Total	3550.596	720			

[a] $R^2 = 0.823$ (adjusted $R^2 = 0.821$)

whether and how much of that capability is selected given the experimental data derived from military planning guidance. Although in steady-state stability operations, the rate-limiting factor is not availability of aircraft but rather distance from forecast casualty locations, a valuable extension might also be the use of a queuing hypercube similar to that proffered by Gerolominis et al.[23] The potential exists for widespread use of this approach in steady-state operations that involve logistical consolidation or relocation (e.g. military consolidation in Europe and in combat theaters) and as a complement to typical utility matrices employed by the military.

The model presented here provides important decision support information for those individuals deciding upon the placement of tactical medical evacuation and treatment assets. The example presented here illustrates that the use of stochastic optimization to support medical planning in and for stability operations has merit. The military medical community continues to adopt optimization methods for planning in conjunction with simulation, so the use of such methods are increasingly likely to become commonplace. While this model is certainly not a panacea for evaluating placement of assets in a stability-operations environment, it should be seen as complementary to the use of traditional simulations and utility matrices. The potential for use of increasingly advanced optimization techniques is excellent.

Appendix A: GAMS Code (Abbreviated)

```
$Title Evac Optimization
*$offlisting
option limrow=0;
option limcol=0;

*NOTE: all speeds and distances in nautical miles
options seed=100;
options reslim=100000;
options iterlim=100000;

Sets
t   scenarios                          /1*9/
i   casualty locations                 /1*80/
j   feasible evacuation locations      /E1*E20/
k   feasible hospital locations        /H1*H10/
l   type of bed required               /MCW, ICW, ICU/
m   type of evacuation asset           /G, A/
n number of brigades                   /1*4/
xy xy pairs for coordinates            /X,Y/
fact levels for each scenario          /lethality,casrad,
wx,days/

Parameter
*fixed parameters
b(i,j,t)       distance between injury location i and
               helicopter evacuation site j
c(i,k,t)       distance between injury location i and
               hospital location k
f(l,t)         type of patient
ecap(m,t)      capacity of ambulances to haul patients for a
               single time period
hcap(l,t)      bed limitation by type
mag(i,t)       radius around brigade for which casualties
               are likely to occur
lethality(t)   lethality multiplier for casualties
days           days included in the capacity estimates
aa             number of air ambulances /60/
ga             number of ground ambulances /240/
aatrips        average number of possible air ambulance
               trips in day /5/
gatrips        number of possible ground ambulance trips
               in day /10/
*mandatory use of ground ambulances may be necessary
if capacities are large but weather prohibits
*continuous flying of air ambulances
factors(t,fact) factors for each scenario

*stochastic components
a(i,t)         penalty weights for patient h of location i
d(j,k,m,t)     transport speed along from j to k for vehicle
               m for iteration t
e(i,t)         number of casualties by type at location i
               for iteration t
bde(xy,n)      brigade coordinates
cas(xy,i,t)    casualty location from center mass of
               brigade for iteration t
evac(j,xy)     coordinates for evacuation sites
hosp(k,xy)     coordinates for hospitalization sites
rand1(i,t)     random number
rand2(i,t)     random number
rand3(i,t)     random number
rand4(i,t)     random number
echo           number of casualties echoed during loop
casx(i,t)      random angle from brigade center for which
               casualties are likely to occur
casy(i,t)      random angle from brigade center for which
               casualties are likely to occur
u              hospitalization sites
```

v evacuation sites
;

*3^(4-2) fractional factorial design table

Table factors(t,fact)

	lethality	casrad	wx	days
1	1.00	50.00	0.20	4.00
2	1.50	50.00	0.40	7.00
3	2.00	50.00	0.30	10.00
4	1.00	100.00	0.30	7.00
5	1.50	100.00	0.20	10.00
6	2.00	100.00	0.40	4.00
7	1.00	150.00	0.40	10.00
8	1.50	150.00	0.30	4.00
9	2.00	150.00	0.20	7.00

;

*hospital and evac sites to be selected. Note: t-index not required but used for other processing
u=5; v=10;
*assign spread of casualties around brigade
*Total number of movements available to aircraft and to ground vehicles in allotted time
ecap('G',t)=ga*factors(t,'days')*gatrips;
ecap('A',t)=aa*factors(t,'days')*aatrips;

*radius around brigade for casualties
mag(i,t)=uniform(-factors(t,'casrad'),factors(t,'casrad'));

*Set up brigade locations on grid
bde('X','1')=100; bde('Y','1')=100;
bde('X','2')=100; bde('Y','2')=400;
bde('X','3')=400; bde('Y','3')=100;
bde('X','4')=400; bde('Y','4')=400;

*Set up evacuation and hospitalization locations on grid
*Note: Locations 1-10 of evac match 1-10 of hospitals.
*If a site can support a hospital, it can support evacuation.
Table evac(j,xy)

	X	Y
E1	140	380
E2	279	395
E3	209	389
E4	306	18
E5	124	289
E6	258	39
E7	142	147
E8	318	204
E9	359	203
E10	137	171
E11	223	49
E12	94	342
E13	44	146
E14	208	241
E15	354	113
E16	188	122
E17	39	273
E18	292	12
E19	216	109
E20	394	196

;
Table hosp(k,xy)

	X	Y
H1	140	380
H2	279	395
H3	209	389
H4	306	18
H5	124	289
H6	258	39
H7	142	147
H8	318	204
H9	359	203
H10	137	171

;

$ontext
Set up total theater hospital capacities by type
Based on mean ICU availability of 25 per day
Mean ICW availability of 50 per day
Mean MCW availability of 100 per day
30 day run
$offtext

hcap('MCW',t)=250*factors(t,'days');
hcap('ICW',t)=250*factors(t,'days');
hcap('ICU',t)=60*factors(t,'days');

*generate some random uniform seeds for inverse CDF sampling
rand1(i,t)=uniform(0,1);
rand2(i,t)=uniform(0,1);
rand3(i,t)=uniform(0,1);
casx(i,t)=uniform(0,6.28);
casy(i,t)=uniform(0,6.28);
*rand4(i,t)=uniform(0,1);

*..set distribution for vehicle speeds
d(j,k,'G',t)=uniform(30,60);
d(j,k,'A',t)=uniform(75,115);

loop(t,
*set casualty locations
loop(i,
if(rand1(i,t)<=.2,
cas('X',i,t)=bde('X','1')+mag(i,t)*cos(casx(i,t));
cas('Y',i,t)=bde('Y','1')+mag(i,t)*sin(casy(i,t));

```
elseif (rand1(i,t)<=.55),
cas('X',i,t)=bde('X','2')+mag(i,t)*cos(casx(i,t));
cas('Y',i,t)=bde('Y','2')+mag(i,t)*sin(casy(i,t));
elseif (rand1(i,t)<=.75),
cas('X',i,t)=bde('X','3')+mag(i,t)*cos(casx(i,t));
cas('Y',i,t)=bde('Y','3')+mag(i,t)*sin(casy(i,t));
else
cas('X',i,t)=bde('X','4')+mag(i,t)*cos(casx(i,t));
cas('Y',i,t)=bde('Y','4')+mag(i,t)*sin(casy(i,t));
); );

*determine distances using distance formula
loop(j,
b(i,j,t)=sqrt((power(cas('X',i,t)-
evac(j,'X'),2)+power(cas('Y',i,t)-evac(j,'Y'),2)));
);
loop(k,
c(i,k,t)=sqrt((power(cas('X',i,t)-
hosp(k,'X'),2)+power(cas('Y',i,t)-hosp(k,'Y'),2)));
);

loop(i,
*assign distributions for number of casualties
    if(rand2(i,t)<=.574,
    e(i,t)=round(1*factors(t,'lethality'));
    echo=e(i,t);
    elseif (rand2(i,t)<=.914),
    e(i,t)=round(2*factors(t,'lethality'));
    echo=e(i,t);
    elseif (rand2(i,t)<=.964),
    e(i,t)=round(3*factors(t,'lethality'));
    echo=e(i,t);
    else
    e(i,t)=round(4*factors(t,'lethality'));
    echo=e(i,t);
    );

*assign penalty weights based on worst casualty at i
*11% of the casualties are high ISS. Since there are echo
 casualties, the
*probability that the worst one is in this group is
1-(1-.11)**echo.
    if(rand3(i,t)>=1-(1-.11)**echo,
    a(i,t)=1.41; f('ICU',t)=echo;
    elseif (rand3(i,t)>=1-(1-.24)**echo),
    a(i,t)=1.05; f('ICU',t)=echo;
    elseif (rand3(i,t)>=1-(1-.55)**echo),
    a(i,t)=1.02; f('ICW',t)=echo;
    else
    a(i,t)=1.01;f('MCW',t)=echo);
*close loop on i
 );
*patient type
*f('MCW',t)=round(.4*sum(i,e(i,t)));
*f('ICW',t)=round(.35*sum(i,e(i,t)));
*f('ICU',t)=sum(i,e(i,t))-f('MCW',t)-f('ICU',t);
*Close the loop on t
);

display 'casualty coordinates ',cas;
display 'distances from evac locations to casualty loca-
tions by scenario',b;
display 'distances from casualty locations to hospitals by
scenario',c;

display 'number of casualties by scenario',e;
display 'penalty weights (based on most severe
casualty)',a;
display ' numbers of casualties of each type by scenario',f;
*The positive variable identifying the number of patients
transported
Positive variables
x(i,j,k,l,m,t) patients of type l traversing track ijk on
vehicle m
*The objective function variable

Free variables
objf obj function
*The binary variables that are used to turn on and off
evacuation / hospitalization assets

Binary Variables
*Sets whether the evacuation or hospitalization site is
open
y(j) air evacuation assets
z(k) hospital site open
;

*The set of equations.
Equations
obj,c1,c2,c3,c4,c5,c6,c7,c8;

*The objective function is the sum over all paths of the
patients of the estimated time for aircraft from site j to
traverse the path
obj..     objf=e=sum((i,j,k,l,m,t),x(i,j,k,l,m,t)*a(i,t)*
          (b(i,j,t)+c(i,k,t))/d(j,k,m,t));
*The first constraint requires that all patients be
evacuated.
c1(i,t)..     sum((j,k,l,m),x(i,j,k,l,m,t))=g=e(i,t) ;
*These constraints provide limitations on open nodes
(evacuation)
c2(j,t)..     sum((i,k,l,m),x(i,j,k,l,m,t))=l=1000000*y(j);
*These constraints provide limitations on open nodes
(hospitalization)
c3(k,t)..     sum((i,j,l,m),x(i,j,k,l,m,t))=l=1000000*z(k);
*The second constraint provides adequate evacuation
capacity without exceeding system evacuation capability.
c4(m,t)..     sum((i,j,k,l),x(i,j,k,l,m,t))=l=ecap(m,t);
```

```
*This constraint ensures adequate hospital assets by type
on each site k for each patient type l.
c5(l,t)..      sum((i,j,k,m),x(i,j,k,l,m,t))=l=hcap(l,t) ;
*Forces the proper distribution of patient types
*c6(l,t)..      sum((i,j,k,m),x(i,j,k,l,m,t))=e=f(l,t);
*This constraint restricts the number of hospital sites to 5
of the 10
c6(t)..      sum(k,z(k))=l=u;
*This constraint restricts the number of evacuation sites to
10 of the 20
c7(t)..      sum(j,y(j))=l=v;
*Set mixture for air to ground ambulances
c8(j,t)..      sum((i,k,l,m),x(i,j,k,l,'G',t))=g=factors
(t,'wx')*(sum((i,k,l,m),x(i,j,k,l,m,t)));

model minit /all/ ;
*objective function variable.
Minit.optcr = .01

*The solve statement
SOLVE minit using mip min objf;
```

References

1. Brookings Institute, 'Iraq Index' http://www.brookings.edu/saban/~/media/Files/Centers/Saban/Iraq%20Index/index.pdf (2009, accessed 14 February 2009).
2. US Army, Force Health Protection in a Global Environment, Field Manual 4-02, October 2003.
3. Heil J. Al Asad hospital a reality. Operation Iraqi Freedom (official website of Multi-National Force-Iraq), http://www.mnf-iraq.com/index.php?option=com_content&task=view&id=11237&Itemid=128 (accessed 15 April 2007).
4. Toregas C, Swain R, ReVelle C, Bergman L. The location of emergency service facilities. *Operations Research* 1971; 19: 1363-1373.
5. ReVelle C, Hogan K. The maximum reliability location problem and alpha - reliability *p*-center problem: derivatives of the probabilistic location set covering problem. *Annals of Operations Research* 1989; 18: 58-69.
6. Marianov V, Taborga P. Optimal location of public health centres which provide free and paid services. *The Journal of the Operational Research Society* 2001; 52: 391-400.
7. Batta R, Mannur N. Covering location models for emergency situations that require multiple response units. *Management Science* 1990; 36: 16-28.
8. Aly A, White J. Probabilistic formulation of the emergency services problem. *Journal of Operational Research Society* 1978; 29: 1167-1179.
9. Neebe A. A procedure for location emergency service facilities for all possible response distances. *Journal of Operational Research Society* 1988; 39: 743-748.
10. Eaton D, Sanchez H, Lantigua R, Morgan J. Determining ambulance deployment in Santo Domingo, Dominican Republic. *Journal of Operational Research Society* 1986; 37: 113-126.
11. Silva F, Serra D. Locating Emergency Services with Different Priorities: The Priority Queuing Covering Location Problem, May 2008, available at: http://ssrn.com/abstract=1143278
12. Geoffrion A, Graves G. Multicommodity distribution system design by Bender's decomposition. *Management Science* 1974; 20: 822-844.
13. Bouma M. Medical Evacuation And Treatment Capabilities Optimization Model (Metcom). Naval Postgraduate Thesis, 2005, available at http://www.dtic.mil.
14. Santoso T, Ahmed S, Goetschalckx M, Shapiro A. A stochastic programming approach for supply chain network design under uncertainty. *European Journal of Operational Research* 2005; 167: 96-115.
15. Snynder L. Facility location under uncertainty: a review. *IIE Transactions* 2006; 38: 537-554.
16. Baker SP, et al. The Injury Severity Score: a method for describing patients with multiple injuries and evaluating emergency care. *Journal of Trauma* 1974; 14: 187-196.
17. Clark E, Fantus R (eds). *National Trauma Databank Annual Report 2007*, American College of Surgeons, 2007.
18. Tyson S, Anderson J. Attacks on US troops in Iraq grow in lethality, complexity. Bigger Bombs a key cause of May's high death toll. *Washington Post*, 3 June 2007.
19. Burris D et al. (eds). *Emergency War Surgery*, Borden Institute, US Army, 2003.
20. US Army. *Medical Evacuation*. Army Field Manual 4.02.2, May 2007.
21. Devore R. Data provided from Center for Army Medical Department Strategic Studies, 2007.
22. Montgomery D. *Design and Analysis of Experiments*. New York: John Wiley & Sons, 2001.
23. Geroliminis N, Karlaftis M, Skabardonis A. Generalized hypercube queueing model for locating emergency response vehicles in urban transportation networks. *Transportation Research Board 85th Meeting*, 2006.