Optim Lett
DOI 10.1007/s11590-014-0819-6

ORIGINAL PAPER

A maximum expected covering problem for locating and dispatching two classes of military medical evacuation air assets

Benjamin C. Grannan · Nathaniel D. Bastian · Laura A. McLay

Received: 19 March 2014 / Accepted: 10 October 2014

Abstract Military medical evacuation (MEDEVAC) systems respond to casualty incidents and transport the most urgent casualties to a medical treatment facility via multiple types of air ambulance assets. Military MEDEVAC systems are subject to an uncertain number of service calls and each service call demands different system operations depending on type and the priority level. Therefore, military medical planners need an air MEDEVAC asset management system that determines how to dispatch multiple types of air assets to prioritized service calls to maintain a high likelihood of survival of the most urgent casualties. To reach this goal, we propose a novel binary linear programming (BLP) model to optimally locate two types of air assets and construct response districts using a dispatch preference list. Additionally, the BLP model balances the workload among assets and enforces contiguity in the first assigned locations for each air asset. The objective of the BLP model is to maximize the proportion of high-priority casualties responded to within a pre-specified time threshold while meeting performance benchmarks to other types of casualties. A spatial queuing approximation model is derived to provide inputs to the BLP model, which thus reflects the underlying queuing dynamics of the system. We illustrate the

B. C. Grannan (✉)
Department of Economics and Business, Virginia Military Institute, 341 Scott Shipp Hall, Lexington, VA 24450, USA
e-mail: grannanbc@vmi.edu

N. D. Bastian
Department of Industrial and Manufacturing Engineering, The Pennsylvania State University, 355 Leonhard Building University Park, State College, PA 16802, USA
e-mail: ndbastian@psu.edu

L. A. McLay
Department of Industrial and Systems Engineering, University of Wisconsin-Madison, 3218 Mechanical Engineering Building, 1513 University Avenue, Madison, WI 53706, USA
e-mail: lmclay@wisc.edu

Published online: 24 October 2014

model and algorithms with a computational example that reflects realistic military data.

Keywords Air MEDEVAC asset management · Casualty triage · Spatial queuing · Binary linear programming · Military medicine

1 Introduction

Military medical evacuation (MEDEVAC) systems save lives by responding to casualty incidents and transporting the most urgent casualties to a medical treatment facility (MTF) via a fleet of distinguishable, geographically-dispersed air ambulance assets (servers) dedicated to the response of casualties. It is important to know where to locate distinct classes of air assets and how to dispatch them in order to maximize the likelihood of survival of the most urgent casualties. To do so, it becomes necessary to triage casualties so that soldiers with the most severe, life-threatening injuries—where death can occur if response is not immediate—receive the most timely treatment. This paper examines the following hard problems found in military MEDEVAC systems: (1) how to geographically locate two classes of air assets and (2) how to construct response districts for all air assets in the system. The problems mentioned above are hard because it is not obvious where to locate scarce air assets or when to dispatch further air assets (and ration closer ones) to casualty incidents in response time-dependent military medical systems.

Optimal decision-making within MEDEVAC systems is complicated due to multiple types of casualties and distinguishable air assets. The ranking of evacuation precedence is handled through a straightforward casualty triage scheme:

- *CAT A* alpha category includes urgent and urgent-surgical casualties that need to be treated within one hour.
- *CAT B* bravo category includes priority casualties that need to be treated within four hours.
- *CAT C* charlie category includes routine casualties that need to be treated within 24 h.

The fleet includes two classes of air assets to manage:

- US Army helicopters such as the UH-60A/L Black Hawk or HH-60 MEDEVAC platforms are normally leveraged to respond to casualties based on their speed, range, and en route medical capability.
- US Air Force helicopters such as the HH-60G Pave Hawk are also available to respond to casualties but are slower because they carry heavy weapon systems and have a larger flight crew.

When effectively employed together, both Army and Air Force air MEDEVAC assets can provide a more efficient and responsive military MEDEVAC system for overseas combat and stability operations.

We propose a novel binary linear programming (BLP) model that optimally locates two classes of air assets within a military medical system deployed overseas and assigns these assets to casualty locations using a dispatch preference list. A dispatch preference list is an ordered ranking of air assets for each casualty location. Additionally, the BLP model balances the workload among assets and enforces contiguity amongst the first assigned location for each air asset. The objective of the BLP model is to maximize the proportion of high-priority casualties (CAT A) responded to within a given response time threshold. The BLP model incorporates coverage thresholds for low-priority casualties to encourage prompt service for all casualties. Accounting for the queuing dynamics in military medical systems results in a more realistic optimization model in which air assets are not always available for service. Therefore, we also propose a queuing model to derive performance statistics, including busy probabilities and independence correction factors, for two classes of air assets. Realistic busy probabilities and removing the assumption of asset independence leads to more effective resource allocation decision-making for military medical planners. The main contribution of this paper is a BLP model that constructs a complete set of MEDEVAC response districts (modeled as dispatch preference lists) of air assets to casualties that explicitly account for two types of air assets and multiple casualty types.

This paper is organized as follows. Section 2 provides a literature review of service system location models and the probabilistic nature of military MEDEVAC systems. Section 3 introduces the approximate Hypercube algorithm [14,16] used to compute queuing system performance statistics. Section 4 introduces the BLP model to locate air assets and construct dispatch preference lists. We present computational results for a military medical system in Sect. 5. Conclusions and future work are presented in Sect. 6.

2 Literature review

A number of existing research papers focus on air MEDEVAC asset optimization models for military medical systems. Bastian [2] presents a multi-criteria decision analysis (MCDA) model using stochastic, mixed-integer goal programming to determine the minimum number of air assets needed at each MTF. This in turn maximizes the coverage of the theater-wide casualty demand and the probability of meeting that demand. Simultaneously, Bastian's model also minimizes the maximal MTF evacuation site total vulnerability to enemy attack. Fulton et al. [13] introduce a two-stage stochastic optimization modeling framework for the medical evacuation (ground and air) of casualties, which identifies optimal casualty evacuation sites and MTF sites in response to stochastic demands for service. While Bastian's and Fulton's papers also account for the stochastic nature of military medical systems, this paper is distinct because it leverages queuing theory to derive busy probabilities to more accurately model the stochastic nature and availability of air assets. Bastian and Fulton [5] present a geospatial-based decision-support tool that identifies which air asset to launch in response to a casualty event given knowledge of terrain, aircraft location, and aircraft capabilities. Zeto et al. [24] examine the pre-location of air assets, along with type and quantity, to maximize the theater-wide coverage while balancing air asset reliability. Bastian and Fulton [5] and Zeto et al. [24] examine location and dispatch problems, but differ from the model presented in this paper because they do not construct a dispatch preference list and complete ordering of all air assets for each casualty location.

Bastian et al. [3,4] examine medical evacuation platforms of the future of vertical lift (FVL) program in support of brigade operations. The specific operations analyzed include the zero-risk aircraft ground speed; different aircraft engine trade-off considerations; and the effects of weaponizing the current air asset fleet will have on the fleet's range, coverage radius, and response time. In contrast to the evaluation of future platforms, this paper presents a model that can be used at the tactical or operational level to decide where to locate air assets and to identify how to dispatch air assets to casualty incidents. Bouma [7] develops the medical evacuation and treatment capabilities optimization model (METCOM) that considers policy effects on key measures of effectiveness, and then optimizes treatment and facility capacities for given casualty flows. This paper does not consider the impact of policy on military medical systems or examine the impact of casualty flows on facility capacities. Fulton et al. [12] present a Monte Carlo simulation to evaluate rules of allocation and planning considerations for Army air ambulance companies during major combat operations (MCO). Rather than incorporating a simulation model, this paper uses queuing theory and binary linear programming to model and examine military medical systems.

There has been much research on civilian emergency medical service (EMS) systems which share similar characteristics as military medical systems. There are many ambulance location models that focus on "covering" patient demand based on response time thresholds that are (usually) nine minutes from dispatch. Many of the models are extensions and variations on the seminal paper by Church and ReVelle [10] that introduces the maximal covering location problem (MCLP) as a basic facility location model for EMS systems. An extension to MCLP is the maximal expected covering location problem (MEXCLP) developed by Daskin [11], which attempts to maximize the expected number of calls covered in a given amount of time, and makes the following assumptions: servers operate independently, each server has the same busy probability, and busy probabilities do not depend on server location. The model in this paper lifts Daskin's last two assumptions. "Assets of different types may be used in ways that result in large differences in the busy probabilities between the asset types. Likewise, there are often "hot-beds" of activity within military systems (e.g., near an enemy stronghold) that would naturally result in some assets to have larger busy probabilities due to their frequent selection as the responding assets. This paper addresses both of these issues." Batta et al. [6] further extend MEXCLP by introducing the adjusted MEXCLP (AMEXCLP) to lift the three previously mentioned assumptions of MEXCLP by use of Hypercube correction factors (see below) for one type of server and a single casualty type. Our model extends the work by Batta et al. [6] by examining more than one type of server, multiple casualty types, and probabilistic travel times. The papers thus far do not consider distinguishable types of ambulances as we do in this paper.

ReVelle and Marianov [23] and Marianov and ReVelle [19] examine the location of multiple types of servers—fire engines and fire trucks—to maximize the number of calls requiring both fire engines and fire trucks. Revelle and Marianov's work assumes that one server of both types is required for each call for service, where this model dispatches one server to each call for service. Mandell [18] also considers multiple types of medical units but does not consider multiple casualty types. Marianov and

Serra [20] introduce a model that maximizes the demand covered when the customer does not have to wait in line (due to congestion) with more than a prespecified number of other customers. In contrast to Marianov and Serra's work, there is a zero length queue in this model. McLay [21] introduces a model that locates two types of ambulances while responding to multiple types of customers but does not assign response zones to ambulances. This paper extends the work of McLay [21] to construct dispatch preference lists modeled as contiguous response zones while also balancing asset workloads.

Spatial queuing models have been widely used to describe the underlying dynamics of military medical and civilian EMS systems. The exact and approximate Hypercube models by Larson [15,16] are the most well-known spatial queuing models. The Hypercube model assumes a multi-server queuing system with indistinguishable servers. Jarvis [14] extends the Hypercube queuing model with an approximation algorithm to incorporate dependencies on casualty call type for service times with servers that are distinguishable by location. Chelst and Barlach [9], Larson and McKnew [17], and Budge et al. [8] consider other extensions to the Hypercube model. Several papers use Hypercube model outputs as BLP model inputs (see Batta et al. [6], McLay [21], and Ansari et al [1]).

This work adds to the literature by introducing a new BLP model that optimally locates two classes of air assets and assigns assets to casualty locations in a preference list. This work also includes an extension to the approximate Hypercube queuing model that accounts for multiple casualty types to derive system statistics that are used as input to the BLP. The new BLP model in this work better captures the military medical evacuation problem studied than previous models. For example, recently military medical planners and senior decision-makers must manage two distinct classes of air assets (both US Army and US Air Force) to respond to triaged casualty events. Further, contiguous response zones are a realistic detail of this model that reflect the practice of a determining dedicated, compact area of responsibility for helicopter pilots. Accounting for the queuing dynamics in military medical systems results in a more realistic optimization model, in which air assets are not always available for service.

3 Queuing model

In this section, we introduce a queuing model that computes system performance statistics to reflect the complex dynamics of military medical systems. Air assets are not always available for service and, therefore, computing busy probabilities allows for more realistic air asset management in military systems. Servers do not act independently of one another in service systems; thus, the queuing model in this section also develops independence correction factors to account for the assumption of server independence. A list of the symbols used in this paper is given next, followed by a discussion of the system dynamics and assumptions of the queuing model.

s_A (s_B) – Number of type A (type B) air assets in the system with $s_A + s_B = s$,
I_A (I_B) – Set of all potential type A (type B) MTF locations,
J – Set of all casualty locations,

C – Set of casualty types that can arrive partitioned into four classes $C_A,\ C_{AB},\ C_{BA},\ C_B$,
$C_A\ (C_B) \subseteq C$ – Set of casualties that require type A (type B) air asset evacuation,
$C_{AB}\ (C_{BA}) \subseteq C$ – Set of casualties that prefer type A air asset evacuation over type B air evacuation (prefer type B over type A),
λ_c – Call arrival rate of casualty type c, for $c \in C$, with $\lambda_c = \sum_{j\in J} \lambda_{jc}$, and λ_{jc} is the call arrival rate of a casualty at location j of type c, for $j \in J,\ c \in C$,
λ – System-wide total call arrival rate with $\lambda = \sum_{j\in J, c\in C} \lambda_{jc}$,
$\tau^A_{ijc}\ (\tau^B_{ijc})$ – Service time when type A (type B) air asset location i responds to casualty of type c at location j, for $c \in C, j \in J,\ i \in I_A(I_B)$,
$\tau^A\ (\tau^B)$ – System-wide average type A (type B) air asset service time,
θ_c – Minimum performance benchmark for the fraction of casualty type c calls that must be responded to in a prespecified time threshold, for $c \in C$,
$\rho^A\ (\rho^B)$ – Traffic intensity for type A (type B) air assets,
$r^A_i\ (r^B_i)$ – Type A (Type B) air asset location i busy probability, for $i \in I_A(I_B)$,
$r^A\ (r^B)$ – System-wide average type A (type B) air asset busy probability,
a_{jck} – Ordered dispatch preference list where a_{jck} is the k^{th} preferred type A air asset for casualty of type c at location j, for $j \in J,\ c \in C_A \cup C_{AB},\ k = 1, \ldots, s_A$ or for $c \in C_A \cup C_{BA},\ k = s_B + 1, \ldots, s_B + s_A$,
$f^A_{ijck}\ (f^B_{ijck})$ – The dispatch probability of type A air asset location $i in I_A$ (type B air asset $i \in I_B$) to casualty $j \in J$ of type $c \in C$ as the $k^{th} = 1, \ldots, s_A\ (s_B)$ priority,
$P^A_k\ (P^B_k)$ – Loss probability that k type A (type B) air assets are busy,
$R^A_{ijc}\ (R^B_{ijc}\ )$ – Fraction of time type A (type B) air asset location i can respond to casualties at location j of type c in a prespecified time threshold, for $i \in I_A(I_B)$, $j \in J,\ c \in C$,
$Q^A(s_A, \rho^A, k)$ or $Q^B(s_B, \rho^B, k)$– Hypercube correction factor as a function of the number of air assets s_A or s_B, traffic intensity ρ^A or ρ^B, and priority k, for $k = 0, ..., s_A - 1$ or $s_B - 1$,
N_j – A neighbor incident matrix used for contiguity, i.e., the set of casualty locations that are adjacent to casualty location j in the usual sense, for $j \in J$.

Our queuing model is an extension to the approximate Hypercube queuing model given in Jarvis [14] that preserves the location of assets. The extension considers two types of air assets—type A and type B—and multiple casualty types in a military MEDEVAC system. Consider a system with s_A type A air assets, s_B type B air assets, J casualty locations, and C casualty types. Classification by triage leads to four mutually exclusive casualty types: (1) casualties that require type A air asset response denoted C_A, (2) casualties that prefer type A air asset response over type B air asset response denoted C_{AB}, (3) casualties that prefer type B air asset response over type A air asset response denoted C_{BA}, and (4) casualties that require type B air asset response denoted C_B. Note that $C = C_A \cup C_{AB} \cup C_{BA} \cup C_B$. Casualties at location $j \in J$ of type $c \in C$ arrive according to a Poisson process with rate λ_{jc} independent of other casualty locations or types. The total arrival rate of casualties in the system is $\lambda = \sum_{j\in J} \sum_{c\in C} \lambda_{jc}$. Exactly one air asset is assigned to each casualty for evacuation unless all air assets are busy, in which case a non-traditional evacuation air asset will be called upon for assistance. In military medical systems there is a zero length queue for service. Note that in this section we develop formulas for the type A air assets. The equations for the type B air assets are omitted but computed analogously.

In military medical systems, air assets do not act independently from one another and thus an optimization model to locate and dispatch air assets should account for asset dependencies. As in Jarvis [14], the type A air asset independence correction factors $Q_A(s_A, \rho^A, j)$ quantify the correction to the probability of obtaining j busy type A air assets followed by an available type A air asset when assuming that air assets operate independently. Note that $P^A_{s_A}$ is the probability that all s_A type A air assets are busy.

$$Q^A(s_A, \rho^A, k) = \sum_{j=k}^{s_A-1} \frac{(s_A - k - 1)!(s_A - j)(s_A)^j (\rho^A)^{j-k} P^A_0}{(j - k)!s_A!(1 - P^A_{s_A})^k (1 - \rho^A(1 - P^A_{s_A}))},$$
$$\text{for} \quad k = 0, 1, \ldots, s_A - 1. \tag{1}$$

It is desirable to know the probability of assigning specific air assets to specific customers. Denote the dispatch probability of type A air asset location $iinI_A$ to casualty $j \in J$ of type $c \in C$ as the $k^{th} = 1, \ldots, s_A$ priority as f^A_{ijck}. For type A air assets, f^A_{ijck} is computed differently for casualty types C_A and C_{AB} than casualty type C_{BA} since type A air assets can only be dispatched to casualties of type C_{BA} if all type B air assets are busy in service. The remaining system statistics depend on f^A_{ijck}. Assume there is a known dispatch preference list, and let a_{jck} equal the air asset that is k^{th} preferred to respond to a casualty of type c at location j. For example if $a_{1A2} = 3$, then air asset 3 is the second preferred air asset to respond to a casualty of type A at location 1. In the computation of f^A_{ijck}, r^A_i is the busy probability for the type A air asset location $iinI_A$. Note that $f^A_{ijck} = 0$ for $c \in C_B$ since type B air assets are required to respond to casualties of type C_B.

$$f^A_{ijck} \approx Q^A(s_A, \rho^A, k-1)(1 - r^A_i) \prod_{l=1}^{k-1} r^A_{a_{jcl}}, \text{ for } c \in C_A \cup C_{AB},$$
$$j \in J,\ k = 1, \ldots, s_A,\ i \in I_A \tag{2}$$

$$f^A_{ijck} \approx P^B_{s_B} Q^A(s_A, \rho^A, k-1)(1 - r^A_i) \prod_{l=s_B+1}^{k-1} r^A_{a_{jcl}}, \text{ for } c \in C_{BA},$$
$$j \in J,\ k = s_B + 1, \ldots, s_B + s_A,\ i \in I_A \tag{3}$$

The dispatch probabilities computed in Eqs. (2) and (3) depend on three values: (1) correction of the assumption of independence among air assets $Q^A(s_A, \rho^A, k-1)$, (2) the probability that the specific type A air asset is available $(1 - r^A_i)$, and (3) the probability that the $k-1$ more preferred air assets are busy $\prod_{l=1}^{k-1} r^A_{a_{jcl}}$. Note that $a_{jcl} = i$ is a dispatch preference list, so the formula for f^A_{ijck} is multiplying together the busy probabilities of the $k-1$ more preferred air assets. Lastly, Eq. (3) includes the $P^B_{s_B}$ term because a type A air asset is only assigned to a C_{BA} type casualty once all s_B type B air assets are busy.

The computation of the type A air asset busy probabilities r^A_i is given by Eq. (4).

$$r_i^A = \frac{V_i^A}{(1 + V_i^A)}, \ i \in I_A \tag{4}$$

In Eq. (4), V^A_i is the rate at which type A air asset i is assigned to a call for service and is computed by Eq. (5):

$$\begin{aligned} V_i^A = &\sum_{k=1}^{s_A} \left(\sum_{c \in C_A \cup C_{AB}} \sum_{j: a_{jck}=i} \lambda_{jc} \tau_{ijc}^A Q^A(s_A, \rho^A, k-1) \prod_{l=1}^{k-1} r_{a_{jcl}}^A \right) \\ &+ P_{s_B}^B \sum_{k=s_B+1}^{s_B+s_A} \left(\sum_{c \in C_{BA}} \sum_{j: a_{jck}=i} \lambda_{jc} \tau_{ijc}^A Q^A(s_A, \rho^A, k-1) \prod_{l=s_B+1}^{k-1} r_{a_{jcl}}^A \right), \end{aligned} \tag{5}$$

for $i \in I_A$. Equation (5) is comprised of two pieces separated by addition. The first piece captures the rate at which type A air asset i is assigned to casualty types C_A and C_{AB}. Type A air asset i is assigned to casualties only when the more preferred air assets are busy, captured by $\prod_{l=1}^{k-1} r^A_{a_{jcl}}$. The first summation shows that type A air asset i can be any of the first $1...s_A$ preferred air assets. The third summation is over the casualty location j only if the dispatch preference list maps type A air asset i to casualty location j as the kth preferred asset. The second piece of Eq. (5) captures the rate at which type A air asset i is assigned to casualty types C_{BA}. Note that the preference index k start at $s_B + 1$ because type A air assets cannot be assiged to the first s_B spots of the dispatch list for casualty type C_{BA}. Also, $P^B_{s_B}$ is included in this second piece because a type A air asset is only assigned to a C_{BA} type casualty once all s_B type B air assets are busy.

The average service time for type A air assets τ^A is given by Eq. 6.

$$\tau^A = \sum_{j \in J} \sum_{c \in C} (\lambda_{jc}/\lambda) \sum_{i=1}^{s_A} \sum_{k=1}^{s} \tau_{ijc}^A f_{ijck}^A / (1 - P_{s_A}^A) \tag{6}$$

In Eq. (6), The average service time is a product of the service times τ^A_{ijc} and the normalized dispatch probabilities $f^A_{ijck}/(1-P^A_{s_A})$. The average service time is weighted by the call arrival rate to casualties (λ_{jc}/λ).

The approximate algorithm is run until the maximum change in air asset busy probabilities falls below a pre-specified ϵ. The approximate algorithm contains an initialization step where the closest air asset is assumed always available.

Given:

1. Call arrival rates λ_{jc}, for $j \in J, c \in C$
2. Type A air asset service times τ^A_{ijc}, for $i \in I_A$, $j \in J, c \in C_A \cup C_{AB} \cup C_{BA}$

Initialize:

$$r_i^A = \sum_{c \in C} \sum_{j \text{ such that } i \text{ is the closest}} \lambda_{jc} \tau_{ijc}^A, \ i \in i_A \tag{7}$$

$$\tau^A = \sum_{j \in J} \sum_{c \in C} \sum_{k=1}^{s} (\lambda_{jc}/\lambda) \tau_{ijc}^A \tag{8}$$

Iterate:

Step 1. Compute $Q^A(s_A, \rho^A, k)$ for $k = 0, 1, \ldots, s_A - 1$ using Eq. (1). Use $\rho^A = \lambda \tau^A / s_A$.
Step 2. For $i = 1, \ldots, s_A$, compute r_i^A with Eq. (4), where V_i^A is given by Eq. (5).
Step 3. Stop if maximum change in r_i^A is less than a tolerance.
Step 4. Else, compute τ^A by Eq. (6), where f_{ijck}^A is given by Eqs. (2) and (3).
Step 5. Return to Step 1.

To compute the type B correction factors and average air asset busy probabilities also used as inputs to the BLP model, the analogous type B approximate algorithm is run with the corresponding type B air asset equations. The following differences are necessary to derive the type B equations. Type B air assets are preferred first when responding to casualties of type $c \in C_B \cup C_{BA}$ as priority $k = 1, \ldots, s_B$. The type B air assets responding to casualties of type $c \in C_{AB}$ are preferred after type A air assets in the dispatch list as priority $k = s_A + 1, \ldots, s_A + s_B$. The type A air asset independence correction factors $Q^A(s_A, \rho^A, k)$ and average air asset busy probabilities $r^A = \sum_{i \in I_A} r_i^A / s_A$ are used as input to the BLP model formulated in the next section. Likewise, the type B air asset independence correction factors $Q^B(s_B, \rho^B, k)$ and average air asset busy probabilities $r^B = \sum_{i \in I_B} r_i^B / s_B$ are also used as input to the BLP model. The next section introduces an optimization model that utilizes the queuing correction factors and dispatch probabilities.

The outputs of the spatial queuing approximation can be used to compute several inputs to the BLP model derived in the next section. The coefficients h_{ijck}^A and h_{ijck}^B found in the objective function and constraints of the BLP in the next section represent the fraction of casualties of type c at location j that air asset i can reach in a prespecified time threshold as the kth preferred asset to send, weighed by the demand at location j (captured by the λ_{jc}/λ_c term). These coefficients utilize dispatch probabilities (see (2 and 3) when assuming that busy probabilities of assets of the same type are equal. The coefficients for type A assets h_{ijck}^A are as follows:

$$h_{ijck}^A = (\lambda_{jc}/\lambda_c) R_{ijc}^A \left[Q^A(s_A, \rho^A, k)(1 - r^A) \left(r^A\right)^{k-1} \right],$$
$$i \in I_A, j \in J, c \in C_A \cup C_{AB}, k = 1, \ldots, s_A \tag{9}$$

$$h_{ijck}^A = (\lambda_{jc}/\lambda_c) R_{ijc}^A P_{s_B}^B \left[Q^A(s_A, \rho^A, k)(1 - r^A) \left(r^A\right)^{k-1} \right],$$
$$i \in I_A, j \in J, c \in C_{BA}, k = s_B + 1, \ldots, s_B + s_A. \tag{10}$$

These coefficients take two sources of uncertainty into account: (1) uncertain asset availability reflected in the dispatch probabilities and correction factors and (2) uncertain travel times that lead to imperfect coverage. (reflected by the terms R^A_{ijc} that represent the fraction of type c calls that can be reached in the fixed response time threshold).

In a similar manner, the coefficients g^A_{ijck} and g^B_{ijck} represent the asset workloads for type A and type B air assets, respectively, and the coefficients for type A assets g^A_{ijck} are defined in the following equations:

$$g^A_{ijck} = \left(\lambda_{jc}/\lambda_c\right) \tau^A_{ijc} \left[Q^A(s_A, \rho^A, k)(1 - r^A) \left(r^A\right)^{k-1} \right],$$
$$i \in I_A, j \in J, c \in C_A \cup C_{AB}, k = 1, \ldots, s_A \tag{11}$$
$$g^A_{ijck} = \left(\lambda_{jc}/\lambda_c\right) \tau^A_{ijc} P^B_{s_B} \left[Q^A(s_A, \rho^A, k)(1 - r^A) \left(r^A\right)^{k-1} \right],$$
$$i \in I_A, j \in J, c \in C_{BA}, k = s_B + 1, \ldots, s_A + s_B \tag{12}$$

where τ^A_{ijc} and τ^B_{ijc} are the service times associated with type A and type B assets, respectively. These coefficients are used to balance the workload among the servers of the same type, and therefore, they also reflect the queuing dynamics.

4 Binary linear programming model

We next present the BLP model used to locate two classes of air MEDEVAC assets and construct dispatch preference lists while balancing the workload among assets and enforcing contiguity amongst the first assigned location for each air asset. The objective of the BLP model is to maximize the proportion of high-priority casualties (C_A calls) responded to within a pre-specified time threshold. The BLP model also incorporates coverage thresholds for low-priority casualties to provide prompt service for all casualties.

To describe the BLP model we first introduce the decision variables. The first two sets of binary variables y^A_i and y^B_i indicate the set of MTF locations utilized: $y^A_i = 1$ if type A MTF location $i \in I_A$ is utilized and 0 otherwise. Likewise, $y^B_i = 1$ if type B MTF location $i \in I_B$ is utilized and 0 otherwise.

The next sets of binary variables x^A_{ijck} and x^B_{ijck} capture the dispatch preference list assignment of type A and type B air assets to casualties. These decision variables are defined to reflect the response protocols to casualties of different types. For example, only type A air assets can be dispatched to casualties of type C_A, and type A air assets must be dispatched to casualties of type C_{AB} before type B air assets are dispatched as backup coverage to casualties of type C_{AB}. Likewise, only type B air assets can be dispatched to casualties of type C_B, and type B air assets must be dispatched to casualties of type C_{BA} before type A air assets are dispatched in secondary support to casualties of type C_{AB}. Therefore, $x^A_{ijck} = 1$ if type A air asset $i \in I_A$ is the kth, $k = 1, \ldots, s_A$, preferred air asset to respond to a casualty at location $j \in J$ of type $c \in C_A \cup C_{AB}$, or if type A air asset location $i \in I_A$ is the kth, $k = s_B + 1, \ldots, s_B + s_A$, preferred air asset to respond to a casualty of type $c \in C_{BA}$, and 0 otherwise. Likewise,

$x^B_{ijck} = 1$ if type B air asset location $i \in I_B$ is the kth, $k = 1, \ldots, s_B$, preferred air asset to respond to a casualty at location $j \in J$ of type $c \in C_B \cup C_{BA}$, or if type B air asset location $i \in I_B$ is the kth, $k = s_A + 1, \ldots, s_A + s_B$, preferred air asset to respond to a casualty of type $c \in C_{AB}$, and 0 otherwise.

Now, we formally state the binary linear program.

$$\max_{c \in C_A} \quad \sum_{j \in J} \sum_{i \in I_A} \sum_{k=1}^{s_A} h^A_{ijck} x^A_{ijck} \tag{13}$$

$$\text{subject to} \quad \sum_{i \in I_A} x^A_{ijck} = 1 \quad \forall j \in J,\ c \in C_A \cup C_{AB},\ k = 1, \ldots, s_A \tag{14}$$

$$\sum_{i \in I_A} x^A_{ijck} = 1 \quad \forall j \in J,\ c \in C_{BA},\ k = s_B + 1, \ldots, s_B + s_A \tag{15}$$

$$\sum_{i \in I_B} x^B_{ijck} = 1 \quad \forall j \in J,\ c \in C_B \cup C_{BA},\ k = 1, \ldots, s_B \tag{16}$$

$$\sum_{i \in I_B} x^B_{ijck} = 1 \quad \forall j \in J,\ c \in C_{AB},\ k = s_A + 1, \ldots, s_A + s_B \tag{17}$$

$$\sum_{k=1}^{s_A} x^A_{ijck} = y^A_i \quad \forall j \in J,\ c \in C_A \cup C_{AB},\ i \in I_A \tag{18}$$

$$\sum_{k=s_B+1}^{s_B+s_A} x^A_{ijck} = y^A_i \quad \forall j \in J,\ c \in C_{BA},\ i \in I_A \tag{19}$$

$$\sum_{k=1}^{s_B} x^B_{ijck} = y^B_i \quad \forall j \in J,\ c \in C_B \cup C_{BA}\ i \in I_B \tag{20}$$

$$\sum_{k=s_A+1}^{s_A+s_B} x^B_{ijck} = y^B_i \quad \forall j \in J,\ c \in C_{AB},\ i \in I_B \tag{21}$$

$$\sum_{j \in J} \left(\left(\sum_{i \in I_A} \sum_{k=1}^{s_A} h^A_{ijck} x^A_{ijck} \right) + \left(\sum_{i \in I_B} \sum_{k=s_A+1}^{s_A+s_B} P^A_{s_A} h^B_{ijck} x^B_{ijck} \right) \right) \geq \theta_c \quad \forall c \in C_{AB} \tag{22}$$

$$\sum_{j \in J} \left(\left(\sum_{i \in I_A} \sum_{k=s_B+1}^{s_B+s_A} P^B_{s_B} h^A_{ijck} x^A_{ijck} \right) + \left(\sum_{i \in I_B} \sum_{k=1}^{s_B} h^B_{ijck} x^B_{ijck} \right) \right) \geq \theta_c \quad \forall c \in C_{BA} \tag{23}$$

$$\sum_{j \in J} \sum_{i \in I_B} \sum_{k=1}^{s_B} h^B_{ijck} x^B_{ijck} \geq \theta_c \quad \forall c \in C_B \tag{24}$$

$$\sum_{i \in I_A} y^A_i = s_A \tag{25}$$

$$\sum_{i \in I_B} y^B_i = s_B \tag{26}$$

$$x_{ijck}^A \leq y_i^A \quad \forall i \in I_A, j \in J, c \in C_A \cup C_{AB}, k = 1, \ldots, s_A \tag{27}$$

$$x_{ijck}^A \leq y_i^A \quad \forall i \in I_A, j \in J, c \in C_{BA}, k = s_B + 1, \ldots, s_B + s_A \tag{28}$$

$$x_{ijck}^B \leq y_i^B \quad \forall i \in I_B, j \in J, c \in C_B \cup C_{BA}, k = 1, \ldots, s_B \tag{29}$$

$$x_{ijck}^B \leq y_i^B \quad \forall i \in I_B, j \in J, c \in C_{AB}, k = s_A + 1, \ldots, s_A + s_B \tag{30}$$

$$\sum_{j \in J} \sum_{c \in C_A \cup C_{AB}} \sum_{k=1}^{s_A} g_{ijck}^A x_{ijck}^A + \sum_{j \in J} \sum_{c \in C_{BA}} \sum_{k=s_B+1}^{s_B+s_A} P_{s_B}^B g_{ijck}^A x_{ijck}^A \leq (r^A + \delta^A) y_i^A \quad \forall i \in I_A \tag{31}$$

$$\sum_{j \in J} \sum_{c \in C_B \cup C_{BA}} \sum_{k=1}^{s_B} g_{ijck}^B x_{ijck}^B + \sum_{j \in J} \sum_{c \in C_{AB}} \sum_{k=s_A+1}^{s_A+s_B} P_{s_A}^A g_{ijck}^B x_{ijck}^B \leq (r^B + \delta^B) y_i^B \quad \forall i \in I_B \tag{32}$$

$$\sum_{j' \in N_j} x_{ij'c1}^A \geq x_{ijc1}^A \quad \forall i \in I_A, c \in C_A \cup C_{AB}, j \in J \tag{33}$$

$$\sum_{j' \in N_j} x_{ij'c1}^B \geq x_{ijc1}^B \quad \forall i \in I_B, c \in C_B \cup C_{BA}, j \in J \tag{34}$$

$$x_{ijck}^A \in \{0, 1\} \quad \forall i \in I_A, j \in J, c \in C, k = 1, \ldots, s_A + s_B \tag{35}$$

$$x_{ijck}^B \in \{0, 1\} \quad \forall i \in I_B, j \in J, c \in C, k = 1, \ldots, s_A + s_B \tag{36}$$

$$y_i^A \in \{0, 1\} \quad \forall i \in I_A \tag{37}$$

$$y_i^B \in \{0, 1\} \quad \forall i \in I_B \tag{38}$$

The objective function in (13) maximizes the expected coverage of C_A casualties. In the BLP model, only type A air assets respond to C_A casualties. C_{AB} casualties prefer type A air assets due to operational characteristics such as in-flight medical treatment ability and flight speed. Constraints (14) assign one type A air asset in each of the first $1, \ldots, s_A$ positions in the preference list to casualties of type C_A or C_{AB}. Casualties of type C_{BA} are assigned all of the type B air assets first in the preference list before the type A air assets are assigned in the preference list as seen in constraints (15). Likewise, only type B air assets respond to C_B casualties and C_{BA} casualties prefer type B air assets. Constraints (16) assign one type B air asset in each of the first $1, \ldots, s_B$ positions of the preference list to casualties of type C_B or C_{BA}. Casualties of type C_{AB} are assigned all of the type A air assets in the preference list first before the type B air assets are assigned in the preference list as seen in constraints (17). Constraints (18)–(21) assign each air asset in the system to only one position in the preference list. In other words, a specific air asset cannot be both the first and third preferred air asset for the same casualty location. Constraints (22), (23), and (24) enforce minimum prespecified thresholds θ_c of the expected coverage for the non-objective casualty types

$c \in C_{AB} \cup C_{BA} \cup C_B$. The minimum levels of service are typically defined by a division surgeon cell or senior medical planners. Constraints (25) and (26) locate s_A type A and s_B type B air assets at MTFs in the system, respectively. Constraints (27)–(30) make sure that only air assets located at open MTF facility locations are utilized in construction of the preference list. Constraints (31) compute and balance the type A workloads within a pre-specified δ^A of the average workload. In a similar manner, constraints (32) compute and balance the type B workloads. Constraints (33) and (34) enforce contiguity amongst each type A and type B air asset's set of first priority casualty locations. Two locations are contiguous to each other if the two locations differ by one unit in either the x or y direction. Lastly, constraints (35)–(36) and (37), (38) require the preference list and the locating variables to be binary, respectively. In the next section, we present an example to apply the BLP model to a military medical system.

5 Computational example

We now present a computational example to illustrate how to locate two classes of air assets and assign them to multiple types of casualties in a dispatch preference list. In the ongoing stability operations in support of operation enduring freedom (OEF) in Afghanistan, coalition forces from the North Atlantic Treaty Organization (NATO) leverage both US Army and US Air Force helicopters in support of combat troops conducting operations on the ground. Specifically, the US Army air assets (UH-60A/L and HH-60 MEDEVAC) and the US Air Force air asset (HH-60G Pave Hawk) are primarily responsible for the aeromedical evacuation of casualties [2]. The HH-60 MEDEVAC platform—referred to as air asset type A—has been specifically modified with a state-of-the-art medical interior to accommodate acute care casualties and has a cruising speed of 278 km/h. The HH-60G Pave Hawk air asset—referred to as air asset type B—is a slower air asset with a cruising speed of 240 km/h and does not have the in-flight medical capabilities to accommodate acute care casualties [22]. In this paper, we assume that both type A and type B air assets fly single-ship (i.e., solo) during MEDEVAC operations, in contrast to tandem operations in which type A air assets must be accompanied by a security/escort asset due to risk of enemy action. Knowing where to locate both type A and type B air assets and how to dispatch them to battle triaged casualties will help increase survivability of the most urgent casualties.

We apply our air MEDEVAC asset optimization model to NATO Regional Command-South (RC-South), one of the four regional commands which span several Afghanistan provinces, to illustrate how air asset management can lead to more effective military medical systems. Figure 1 depicts the area of RC-South divided into seventy-five distinct casualty locations. Twenty-five potential sites for medical treatment facilities (MTF) are distributed throughout RC-South, represented by the red crosses in Fig. 1. The base case example solved in this section will optimally locate ten type A air assets and four type B air assets at MTFs within RC-South. Note that the casualty data (from 2001 through 2012) was one-time extracted from the defense casualty analysis system (DCAS).

The following organization of casualty types is used to be consistent with the notation of Sects. 3 and 4. Casualty type C_A contains the CAT A casualties, C_{AB} contains

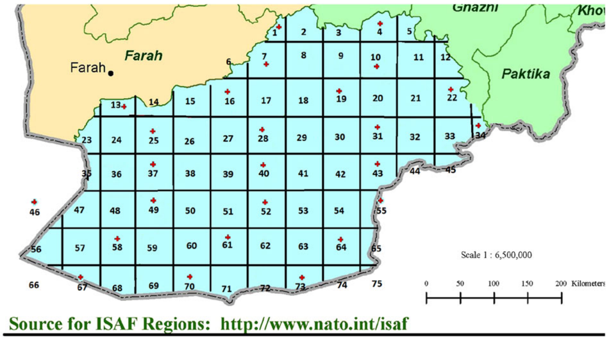

Fig. 1 25 Potential MTF sites in NATO Regional Command-South (represented by *red crosses*) (color figure online)

the CAT B casualties, and C_B contains the CAT C casualties. Note that there are no C_{BA} call types in this example.

We next derive the BLP model inputs. The type A service times τ^A_{ijc} (hours) are computed by dividing the Euclidean distance (kilometers) between a MTF location $i \in I_A$ and a casualty location $j \in J$ by the type A air asset speed (km/h). The type B service times τ^B_{ijc} are computed similarly except that the service time calculation is multiplied by two since medical treatment does not begin until a casualty returns back to the MTF (in contrast to the medical treatment provided at the casualty location with a type A air asset). The distribution of casualty type is as follows: CAT A 75 %, CAT B 20 %, and CAT C 5 %. Further, the total casualty arrival rate $\lambda = 3.0$ and each casualty location $j \in J$ has equal chance of a casualty event. Evacuation of a CAT A casualty by a type A air asset results in an expected coverage between 0.8 and 1.0, depending on how many minutes have elapsed up to the threshold of one hour. Therefore, we use an exponential decreasing curve to model the rapid deteriorating health of CAT A casualties after 1 h. Similar translated exponential decreasing curves are used for the evacuation of CAT B and CAT C casualties to reflect time-standards of 4 and 24 h, respectively. We use a time-standard of 6 h for CAT C casualties to encourage the rapid evacuation of all casualty types. R^A_{ijc} and R^B_{ijc} depend on the expected type A and type B air asset service time between casualty location j and air asset location i (denoted t in the equations below) as well as the casualty type $c \in C$. R^A_{ijc} and R^B_{ijc} and are given by the following equations. For $c \in$ CAT A:$R_{ijc} = -0.2t + 1$ if $t \leq 1$ and $e^{-1.6(t-1)}$ otherwise , For $c \in$ CAT B:$R_{ijc} = -0.05t + 1$ if $t \leq 4$ and $e^{-1.6(t-4)}$ otherwise , and For $c \in$ CAT C:$R_{ijc} = -0.05t + 1$ if $t \leq 6$ and $e^{-1.6(t-6)}$ otherwise.

We perform all computations on laptop with a AMD A6 2.10 GHz processor and 8GB RAM, leveraging Python 2.6 for parameter calculations and Gurobi 5.6.2 for optimization of the BLP model. The following algorithm is executed to solve the optimization model. Note that each optimization iteration contains an instance of the approximate Hypercube algorithm. Optimize the BLP using the most recent values of the following system statistics: (1) the average busy probabilities r^A and r^B, (2) the independence correction factors Q_A and Q_B, and (3) the offered load imbalance thresholds δ_A and δ_B. Optimize a new instance of the problem until the workload of each type A and type B air assets is less than 5 % more than the average air asset workload. After each optimization iteration, the current solution contains the location of the air assets and the dispatch preference list. Using the dispatch preference list, execute the approximate Hypercube algorithm until the max change in type A and type B busy probabilities is less than 0.001. The system performance statistics $(r^A,\ r^B,\ Q_A(s_A, \rho^A, k),\ Q_B(s_B, \rho^B, k))$ are updated and passed as input to the next instance of the BLP model. The offered load imbalance thresholds δ_A and δ_B are tightened to the new standard deviation of the type A and type B busy probabilities. For the base case example, the optimization algorithm converged in four iterations. The approximate Hypercube algorithm converged in six or fewer iterations within each iteration of the optimization algorithm. The base case example corresponds to a BLP with 67,757 constraints and 52,602 variables. Solving the optimization algorithm on the base case ten times resulted in the following run time metrics: a minimum CPU time of 908 s (15.1 min), a maximum CPU time of 1,132 s (18.9 min), and an average CPU time of 1,001 s (16.7 min).

Table 1 presents the ten type A and four type B air asset locations selected by the BLP model as well as the corresponding busy probabilities. MTF43 is the only location with both a type A and a type B air asset. The type A busy probabilities range from the low of 0.051 (MTF07) to the high of 0.103 (MTF25). The type B busy probabilities range from the low of 0.224 (MTF37) to the high of 0.358 (MTF43). All air asset busy probabilities are balanced to be no greater than 5 % more than the respective type A or type B average busy probability.

Table 2 presents the independence correction factors for type A and type B air assets. The type A factors quantify the correction to the probability of obtaining k busy type

Table 1 Location and busy probabilities for Type A and Type B air assets in the base case example

The location of the $s_A = 10$ type A air assets	MTF25	MTF22	MTF28	MTF07	MTF46	MTF61	MTF64	MTF43	MTF10	MTF58
Type A air asset busy probabilities r_i^A	0.103	0.058	0.098	0.051	0.068	0.086	0.076	0.061	0.081	0.070
Average type A air asset busy probability r^A						0.075				
The location of the $s_B = 4$ type B air assets				MTF43	MTF73	MTF37	MTF19			
Type B air asset busy probabilities r_i^B				0.358	0.339	0.224	0.320			
Average type B air asset busy probability r^B						0.310				

Table 2 Independence correction factors for Type A and Type B air assets in the base case example

$Q_A(s_A, \rho^A, k)$	$k=0$	$k=1$	$k=2$	$k=3$	$k=4$	$k=5$	$k=6$	$k=7$	$k=8$	$k=9$
	1.0	0.907	0.878	0.907	0.995	1.155	1.411	1.805	2.405	3.318
$Q_B(s_B, \rho^B, k)$				$k=0$	$k=1$	$k=2$	$k=3$			
				1.0	0.856	0.807	0.830			

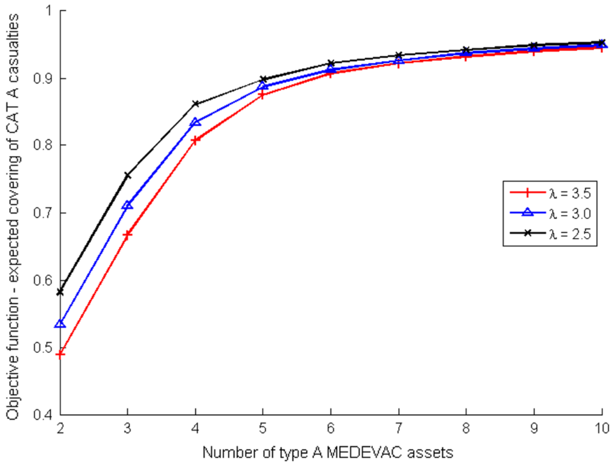

Fig. 2 Sensitivity analysis of the objective function for the number of Type A air assets and call arrival rate λ (color figure online)

A air assets followed by an available type A air asset when assuming that air assets operate independently. Note that $Q_A(s_A, \rho^A, 0) = Q_B(s_B, \rho^B, 0) = 1.0$ because the probability of obtaining an available air asset when no air assets are busy is equal to one.

The objective function of the BLP is the expected fraction of calls covered when dispatching type A air assets to CAT A casualties. It is desirable to know the performances changes in the system as additional type A air assets are introduced or removed from the system. Figure 2 presents a sensitivity analysis on the expected coverage of CAT A casualties with respect to the number of type A air assets in the system as well as the call arrival rate λ. Figure 2 can be used by decision-makers to identify how many type A air assets are needed for a given benchmark. For example, if the desired system performance is 90 % expected coverage, then six type A air assets will be enough when the call volume λ is between 2.5 and 3.5 calls per hour. However, to achieve 95 % expected coverage in the same call volume range, nine type A air assets are required. In Fig. 2, there is a diminishing return on the objective function value with respect to the number of type A air assets. For example, when $\lambda = 3.0$, consider the large increase in expected coverage when $s_A = 2$ (53 %) and when $s_A = 5$ (89 %) to the small increase when $s_A = 8$ (94 %) and when $s_A = 10$ (95 %).

Figure 3a–c illustrate the contiguity of the 1st priority response districts created for CAT A, CAT B, and CAT C casualties, respectively. Figure 3a, b show type A

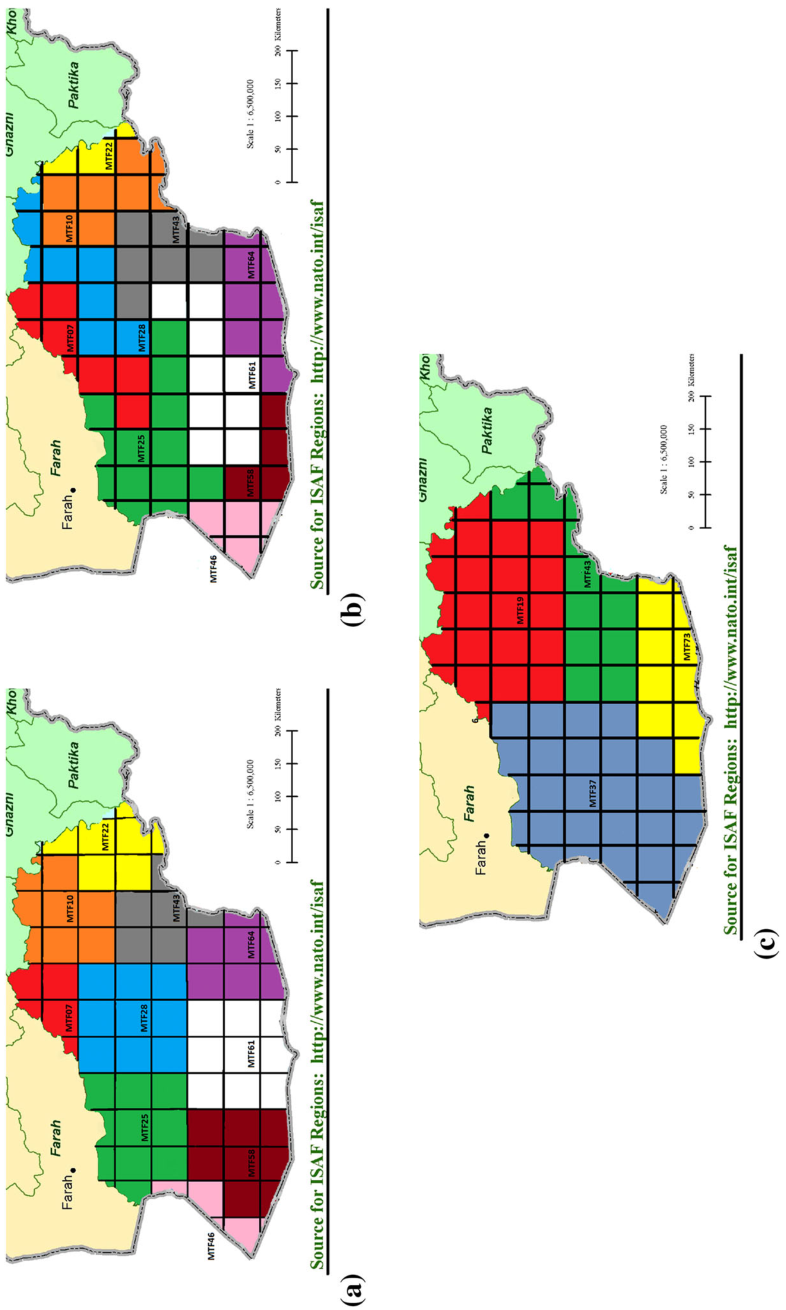

Fig. 3 Contiguous 1st priority dispatching assignments for type A and type B air assets in the base case example. Casualty locations that comprise an air asset's 1st priority district are the *same color*. **a** 1st priority response districts of Type A air assets assigned to CAT A casualties, **b** 1st priority response districts of Type A air assets assigned to CAT B casualties, **c** 1st priority response districts of Type B air assets assigned to CAT C casualties (color figure online)

Table 3 Optimal solution versus closest server heuristic solution with additional Type A air assets

s_A	s_B	Heuristic obj value	Optimal obj value	Obj improvement (%)
2	4	0.533	0.533	0
3	4	0.710	0.710	0
4	4	0.805	0.834	3.60
5	4	0.853	0.887	3.99
6	4	0.884	0.912	3.17
7	4	0.906	0.926	2.21
8	4	0.916	0.937	2.29
9	4	0.922	0.944	2.39
10	4	0.933	0.949	1.71

air assets because the first priority for CAT A and CAT B casualties must be a type A air asset. The BLP model identifies the set of type A air asset locations that maximize the fraction of CAT A casualties covered within 60 min. The set of locations selected must also balance air asset workloads while considering busy probabilities and accounting for air assets that are dependent. Likewise, Fig. 3c shows type B air assets because the first priority for each CAT C casualty must be a type B air asset.

The objective function used in this computational example encourages dispatch of the closest air asset to a casualty incident. Dispatching the closest air asset leads to the shortest service time and subsequently the largest expected coverage. However, due to the unavailability of air assets in a crowded system, it is sometimes optimal to ration the closest available air asset. Table 3 presents the difference in objective function value between optimality and the heuristic policy of dispatch the closest air asset to a casualty. To implement the closest air asset heuristic, s_A type A and s_B type B air assets must be selected for use in the priority list. Therefore, a rank ordering of all potential MTF sites is created based upon the average distance to all seventy five casualty locations. Once the air asset locations are known, the priority list is created based on proximity between each casualty location and each asset location. To evaluate the objective function value of the heuristic, to locating and dispatching variables are set to 1.0, the system input parameters are calculated, and the instance is solved in Gurobi. Table 3 shows that the heuristic of dispatching the closest air asset performs very well. The largest difference in objective function value between optimality and the heuristic over the range of s_A and s_B values considered is 3.99 %. In some scenarios, such as $s_A = 2$ and $s_B = 4$, the heuristic policy is optimal. Note that in Table 3 the base case value of $\lambda = 3$ is used.

The comparison of optimality to the dispatch closest heuristic is examined further in Table 4. The call arrival rate is varied from $\lambda = 1$ to 10. Under this range of values, the heuristic of dispatching the closest air asset again performs very well. The largest difference in heuristic objective function value from optimality is 1.71 %. Note that in Table 4 the base case values of $s_A = 10$ and $s_B = 4$ are used. These improvements,

Table 4 Optimal solution versus closest server heuristic solution with varying call arrival rate λ

Call arrival rate λ	Heuristic obj value	Optimal obj value	Obj improvement (%)
1	0.941	0.957	1.70
3	0.933	0.949	1.71
6	0.918	0.932	1.53
10	0.908	0.911	0.33

while small in absolute value, are large relative to other methodological improvements in this domain.

The strength of the model in this paper lies in the construction of an optimal dispatch preference list that considers the negative effect of busy assets on the overall system. For example, the model identifies the non-obvious opportunities to dispatch a further air asset (instead of the nearest air asset) to a casualty location to help lower air asset busy probabilities. While the afore mentioned closest air asset dispatching heuristic does perform well, strategically spreading the work load among all of the air assets leads to shorter wait times for an available air asset system wide and ultimately greater expected coverage of casualties.

This computational example in this section illustrates how to locate two classes of air assets and construct an air MEDEVAC asset dispatch preference list for all casualty locations. The base case example developed in this section illustrates how to locate and dispatch 10 type A and 4 type B air assets to maximize the expected coverage of CAT A casualties within sixty minutes, while balancing the air asset workload. A dispatch preference list is valuable to senior military decision-makers and medical planners to identify preferred back-up air assets in a crowded military MEDEVAC system.

6 Concluding remarks

In this paper, we introduced a binary linear programming model that optimally locates two classes of military air MEDEVAC assets and assigns air assets to casualties in a dispatch preference list. Queuing system performance statistics, including busy probabilities and independence correction factors, are derived for air assets and are leveraged as inputs to the BLP model. Accounting for the queuing dynamics in military medical systems results in a more realistic optimization model in which air assets are not always available for service. The problem of military air MEDEVAC asset management is important because military medical systems save lives by responding to multiple types of casualty incidents. By providing prompt en route medical care to the most urgent casualties, the likelihood of survival is increased. The effective and efficient medical evacuation of casualties serves to keep troop morale high while maintaining a healthy and sustained fighting force.

The model in this paper can be extended to consider more than two types of air assets. The resulting spatial queuing model and BLP can be derived using methods that are analogous to those in this paper.

This paper also can be extended to account for batch arrivals of casualties to the system where the volume of casualties requires the dispatch of more than one air asset. Also, the initial triage of casualties is not perfect and can lead to misclassification of the severity of a casualty incident. The logistics of military medical systems are even more complicated when there are multiple air assets and then the accuracy of triage affects dispatch decisions. Also, in some cases, air assets must be assigned to and wait for security/escort assets. An optimization model to identify which air asset and security/escort asset pair to dispatch to casualty incidents to minimize MEDEVAC response time of urgent casualties would be of value to military decision-makers. Work is currently in progress to address these issues.

Acknowledgments The first and third authors were supported by the U.S. Department of the Army under Grant Award Number W911NF-10-1-0176. The second author was supported by the National Science Foundation under Grant No. DGE1255832. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the United States Army, National Science Foundation, Virginia Military Institute, Pennsylvania State University, or University of Wisconsin-Madison.

References

1. Ansari, S., McLay, L.A., Mayorga, M.E.: A Maximum Expected Covering Problem for Locating and Dispatching Servers. Under revision at Transportation Science, Virginia Commonwealth University (2014)
2. Bastian, N.: A robust, multi-criteria modeling approach for optimizing aeromedical evacuation asset emplacement. J. Def. Model. Simul. **7**(1), 5–23 (2010)
3. Bastian, N., Brown, D., Fulton, L., Mitchell, R., Pollard, W., Robinson, M., Wilson, R.: Analyzing the future of army aeromedical evacuation units and equipment: a mixed methods, requirements-based approach. Mil. Med. **178**(3), 321–329 (2013)
4. Bastian, N., Fulton, L.V., Mitchell, R., Pollard, W., Wierschem, D., Wilson, R.: The future of vertical lift: initial insights for aircraft capability and medical planning. Mil. Med. **177**(7), 863–869 (2012)
5. Bastian, N., Fulton, L.V.: Aeromedical evacuation planning using geospatial decision-support. Mil. Med. **179**(2), 174–182 (2014)
6. Batta, R., Dolan, J.M., Krishnamurthy, N.N.: The maximal expected covering location problem: revisited. Transp. Sci. **23**(4), 277–287 (1989)
7. Bouma, M.F.: Medical evacuation and treatment capabilities optimization model metcom. In: Proceedings of Technical Report, Naval Postgraduate School, California (2005)
8. Budge, S., Ingolfsson, A., Erkut, E.: Approximating vehicle dispatch probabilities for emergency service systems with location-specific service times and multiple units per location. Op. Res. **57**(1), 251–255 (2009)
9. Chelst, K.R., Barlach, Z.: Multiple unit dispatches in emergency services: models to estimate system performance. Manag. Sci. **27**(12), 1390–1409 (1981)
10. Church, R., ReVelle, C.: The maximal covering location problem. Pap. Reg. Sci. Assoc. **32**, 101–108 (1974)
11. Daskin, M.S.: A maximum expected covering location model: formulation, properties and heuristic solution. Transp. Sci. **17**(1), 48–70 (1983)
12. Fulton, L., Lasdon, L., McDaniel, R., Coppola, M.: Two-state stochastic optimization for the allocation of medical assets in steady-state combat operations. J. Def. Model. Simul. **7**(2), 89–102 (2010)
13. Fulton, L., McMurry, P., Kerr, B.: A monte carlo simulation of air ambulance requirements during major combat operations. Mil. Med. **174**(6), 610–614 (2009)

14. Jarvis, J.P.: Approximating the equilibrium behavior of multi-server loss systems. Manag. Sci. **31**(2), 235–239 (1985)
15. Larson, R.C.: A hypercube queuing model for facility location and redistricting in urban emergency services. Comput. Op. Res. **1**(1), 67–95 (1974)
16. Larson, R.C.: Approximating the performance of urban emergency service systems. Op. Res. **23**(5), 845–868 (1975)
17. Larson, R.C., McKnew, M.A.: Police patrol-initiated activities within a systems queuing model. Manag. Sci. **28**(7), 759–774 (1982)
18. Mandell, M.B.: Covering models for two-tiered emergency medical service systems. Locat. Sci. **6**(1–4), 355–368 (1998)
19. Marianov, V., ReVelle, C.: A probabilistic fire-protection siting model with joint vehicle reliability requirements. Pap. Reg. Sci. **71**(3), 217–241 (1992)
20. Marianov, V., Serra, D.: Hierarchical location-allocation models for congested systems. Eur. J. Op. Res. **135**(1), 195–208 (2001)
21. McLay, L.A.: A maximum expected covering location model with two types of servers. IIE Trans. **41**(8), 730–741 (2009)
22. Federation of American Scientists. Uh-60 black hawk, uh-60l black hawk, uh-60q medevac, mh-60g pave hawk, hh-60g pave hawk, ch-60 sea hawk. Accessed 13 Nov 2013
23. ReVelle, C., Marianov, V.: A probabilistic FLEET model with individual vehicle reliability requirements. Eur. J. Op. Res. **53**(1), 93–105 (1991)
24. Zeto, J.F., Yamada, W., Collins, G.: Optimizing the emplacement of scarce resources to maximize the expected coverage of a geographically variant demand function. In: Proceedings of Technical Report, US Center for Army Analysis, Ft. Belvoir (2006)