



**Resupply Operations of a Dispersed Infantry
Brigade Combat Team using Approximate
Dynamic Programming**

THESIS

Camero K. Song, MAJ, U.S. Army
AFIT-ENS-MS-21-M-184

**DEPARTMENT OF THE AIR FORCE
AIR UNIVERSITY**

AIR FORCE INSTITUTE OF TECHNOLOGY

Wright-Patterson Air Force Base, Ohio

DISTRIBUTION STATEMENT A
Approved for public release: distribution unlimited.

The views expressed in this document are those of the author and do not reflect the official policy or position of the United States Department of Defense or the United States Government. This material is declared a work of the U.S. Government and is not subject to copyright protection in the United States.

AFIT-ENS-MS-21-M-184

RESUPPLY OPERATIONS OF A DISPERSED INFANTRY BRIGADE COMBAT
TEAM USING APPROXIMATE DYNAMIC PROGRAMMING

THESIS

Presented to the Faculty
Department of Operational Sciences
Graduate School of Engineering and Management
Air Force Institute of Technology
Air University
Air Education and Training Command
in Partial Fulfillment of the Requirements for the
Degree of Master of Science in Operations Research

Camero K. Song, BS
MAJ, U.S. Army

March 2021

DISTRIBUTION STATEMENT A
Approved for public release: distribution unlimited.

AFIT-ENS-MS-21-M-184

RESUPPLY OPERATIONS OF A DISPERSED INFANTRY BRIGADE COMBAT
TEAM USING APPROXIMATE DYNAMIC PROGRAMMING

THESIS

Camero K. Song, BS
MAJ, U.S. Army

Committee Membership:

Matthew J. Robbins, PhD
Chair

Brian J. Lunday, PhD
Member

Abstract

Military sustainment planners must consider the effective employment of Cargo Unmanned Aerial Vehicles (CUAVs) while performing resupply operations under austere threat conditions. Viewed as an inventory routing problem, the complexities of resupplying dislocated Forward Operating Bases (FOBs) under dangerous and dynamic conditions require joint consideration of several challenging inventory management and logistical features. This research formulates a military stochastic inventory routing problem with multiple routing as a Markov decision process (MDP) model to understand the effects of various problem features on sustainment performance. The completion of CUAV resupply itineraries, the servicing of FOB supply demands, and the evolution of FOB threat levels all occur in a stochastic manner. Multiple routing options for CUAVs allow sustainment planners to experiment with different route configurations to maximize the safety of the CUAV fleet while maintaining adequate supply levels.

The MDP model appropriately frames the problem of interest. However, the high dimensionality and large size of both the state and decision spaces render classical dynamic programming solution methods intractable. Instead, approximate dynamic programming (ADP) techniques, supplemented with ordinal optimization, are developed to produce high-quality resupply dispatching policies for the CUAV fleet. An approximate policy iteration (API) algorithmic framework is adopted, within which kernelized least squares temporal difference (KLSTD) provides the learning mechanism. Ordinal optimization reduces computational effort by effectively reducing the size of the decision space, providing another layer of value function approximation. Varying problem features such as CUAV fleet size, number of threat maps, and num-

ber of FOBs enables generation of 15 problem instances, each of which undergoes a sequence of designed experiments to understand the effects of both problem and algorithm features on solution quality. A solution procedure consisting of repeatedly solving a deterministic set partitioning vehicle routing problem (SPVRP) —which seeks to minimize the risk of CUAV loss and the supply deficit across all FOBs —serves as a benchmark policy to compare with ADP policies generated by the proposed API-KLSTD algorithm. ADP policies outperformed the benchmark policies in 13 of the 15 problem instances investigated.

The current method of generating UAV missions primarily depends on demand requests from subordinate units. The proffered researched methods incorporate the operational environment to offer timely, intelligent, and high-quality itinerary recommendations to military sustainment and UAV planners. By leveraging ADP techniques to aid sequential decision making under uncertainty, units can increase fleet efficiency and readiness.

Key words: inventory routing problem, Markov decision process, approximate dynamic programming, kernelized least squares temporal difference

*I dedicate this work to the men and women who served before me, and those who are
currently serving this great Nation.*

Acknowledgements

Foremost, I would like to thank the Lord God for providing this opportunity to learn and grow as a military professional. I would also like to thank my family for providing balance throughout this endeavor. My sincere gratitude to Dr. Matthew J. Robbins and Dr. Brian J. Lunday for their mentoring guidance.

Camero K. Song

Table of Contents

	Page
Abstract	iv
Dedication	vi
Acknowledgements	vii
List of Figures	ix
List of Tables	x
I. Introduction	1
II. Literature Review	6
2.1 Vehicle Routing Problem	6
2.2 Inventory Routing Problem	8
2.3 Military Inventory Routing Problem	10
III. Solution Methodology	12
3.1 Problem Overview	12
3.2 MDP Formulation	13
3.3 ADP Formulation	19
3.3.1 Reduced Decision Space	22
3.3.2 ADP Algorithm	23
IV. Testing, Results, and Analysis	27
4.1 Problem Instances	27
4.2 Benchmark Policy	30
4.3 Algorithm Features	31
4.4 Results and Analysis	32
4.4.1 Summary of Testing Results	37
V. Conclusions and Recommendations	39
Bibliography	41

List of Figures

Figure		Page
1	Contribution Function Graphs	18
2	Combat Outpost Keating, Afghanistan	28
3	CUAV Fleet OR% Comparison across Problem Instances	33
4	Daily Class I above Threshold across FOBs	34
5	FOB Locations	35
6	Threat Environment M_t^{Env} Scenario	36

List of Tables

Table		Page
1	Ordinal Optimization Sample Sizes	23
2	CUAV Characteristics	28
3	Experimental Design Factor Levels	31
4	Experimental Design Results	32
5	Experimental Design for 6 FOBs	36
6	ADP Policy for 6 FOBs	37

RESUPPLY OPERATIONS OF A DISPERSED INFANTRY BRIGADE COMBAT TEAM USING APPROXIMATE DYNAMIC PROGRAMMING

I. Introduction

A Brigade Combat Team (BCT) is the Army's primary decisive action force, modular enough to operate in many conditions and execute diverse missions [14]. The Army fields three types of BCTs: Armored, Stryker, and Infantry (IBCT), with sizes ranging from 3900 to 4100 soldiers [39]. Of particular interest is the IBCT, a combined arms force structured around dismounted infantry and optimized for operating in complex terrain, which is comprised of two or more forms of restrictive terrain [16]. An IBCT's ability to operate dismounted in complex terrain is what distinguishes itself from the other BCTs. An IBCT's primary purpose is to "close with the enemy using fire and movement to destroy or capture enemy forces, or to repel enemy attacks by fire, close combat, and counterattack" [14]. During recent deployments, IBCTs spread geographic influence by managing dislocated bases known as Forward Operating Bases (FOBs). An IBCT is known to operate as many as 20 FOBs during deployments, each with approximately 100-250 personnel [24]. Subordinate to the IBCT, the brigade support battalion (BSB) directly supports IBCT operations by providing self-sustainment for the IBCT through the BSB's robust support network. The BSB plans, prepares, and executes all sustainment operations to include providing classes of supply, maintenance, and general field services.

Sustaining an IBCT in complex environments often requires supporting dislocated units that operate at the edges of extended lines of communication. To ensure the IBCT can effectively complete its mission, enable freedom of action, and extend

the endurance of subordinate units, supply and distribution systems must be able to deliver supplies throughout the IBCT's Area of Operation (AO) at the proper location and at the right time. The BSB remains informed of current logistics statuses and forecasted requests through unit logistics reports (LOGSTAT). Reporting through a hierarchical system, all subordinate elements report LOGSTATs through the unit supply echelons, which are subsequently managed by the IBCT's primary logistic staff organization, the S4.

The primary element charged with executing the IBCT's sustainment support plan is the Distribution Company, a subordinate organization within the BSB [17]. Located at the Brigade Support Area (BSA) and equipped with medium to large trucks and fuel tankers, this company manages the physical receipt and distribution of all classes of supply, except for Classes VI (personal items), VII (major end items), and VIII (medical supplies). Due to the large amounts—in terms of both weight and volume—of requested supplies necessary to sustain personnel, support operations, and maintain equipment, and given the organic capabilities of the BSB, ground resupply is the primary means for resupply missions. Successfully delivering supplies via ground lines of communication (GLOCs) depends on relatively favorable terrain, weather, and threat activities. GLOCs rely on load bearing bridges, routes through city centers, and roads that are susceptible to road freeze, washouts, and blockades. Moreover, IBCTs establish bases on or near key terrain; such terrain is often remote from urban areas and well-developed road networks, which means access to these locations is difficult. In Iraq and Afghanistan, GLOCs were known to be insurgent engagement areas, riddled with complex IEDs and ambushes. For example, a unit in Afghanistan conducting a 60 kilometer resupply mission took 54 hours and lost two combat vehicles and one logistics vehicle due to striking three different IEDs [33].

Although ground resupply is the primary means for sustaining IBCTs, aerial re-

supply becomes very important when challenging environmental factors are present. External aerial assets such as CH-47 Chinooks and UH-60 Blackhawks provide general resupply support. For example, a sustainment unit in Afghanistan reported several combat outposts (COPs) being resupplied by air during the summer. The number of COPs requiring aerial resupply more than doubled in the winter due to heavy snows closing mountain passes [12]. From 2002 through 2012, Forward Operating Base (FOB) Tillman in eastern Afghanistan relied almost solely on aerial resupply due to the threat conditions and unreliable road infrastructure [4]. Like ground resupply, this mode of transportation is susceptible to attacks. The local air threat always poses a risk to aerial resupply missions with the most prevalent threats being man portable air defense systems (MANPADS), rocket propelled grenades, and rockets. Sponsored by the Office of the Under Secretary of Defense for Acquisition, Technology and Logistics, a 2009 study reported 496 fatalities from 375 rotorcraft losses supporting Operation Iraqi Freedom and Operation Enduring Freedom within the timeframe of October 2001 to October 2009 [9]. With the loss of lives and supplies, the military has looked to alternate means of delivery, namely cargo unmanned aerial vehicles (CUAV).

The Army acknowledges this requirement as it highlights the growing importance of UAVs in its current doctrine for sustainment operations [15]. The Department of Defense envisions that, within 25 years, military services will rely on enhanced UAV capabilities to sustain the force [13]. This concept is already a reality for some service components. The Marine Corps tested the use of CUAVs to resupply its bases in 2009-2012. The K-Max CUAV successfully delivered 4.5 million pounds of supplies during its tenure in Afghanistan [11].

The problem of effectively operating a fleet of CUAVs for sustainment can be modeled as a vehicle routing problem (VRP). A large volume of research exists sup-

porting the civilian sector because effectively modeling and solving the VRP benefits commodity producing organizations. In the VRP, companies have at least one central depot and must route vehicles carrying supplies to either distribution centers or customers to meet demand requirements while minimizing transportation costs. Similarly, for the problem of interest in this research, the BSB acts as the central depot, and FOBs serve as demand nodes. However, very few efforts in the technical literature address the dynamic and stochastic nature of successful supply deliveries in a contested environment. Research exists regarding CUAVs delivering supplies along contested routes with the constraint of direct delivery [27]. This situation is represented as an inventory routing problem (IRP) variant, defined as a military inventory routing problem (MILIRP). However, the relaxation of direct delivery under the same environmental conditions has yet to be considered.

This research addresses the following problem. Sustainment planners within an IBCT must develop daily air tasking order (ATO) policies to resupply a set of geographically dispersed FOBs using a fleet of homogeneous CUAVs. An ATO policy indicates which CUAVs visit which FOBs, the FOB-to-FOB route(s) each CUAV should traverse, and the load (i.e., the class and amount of supplies) each CUAV should carry and distribute. A single CUAV may deliver supplies to multiple FOBs during a single mission (i.e., multiple delivery). A challenging feature of this problem results from the uncertain outcomes of the resupply sorties. The IBCT operates in a hostile, combat environment, so vulnerable CUAVs may be destroyed while flying resupply missions.

This research provides initial insight into incorporating a fleet of CUAVs into sustainment operations for a military unit. The relaxation of the direct delivery constraint offers sustainment planners more options to effectively resupply their units using vehicles developed via the Army’s cargo UAV program. Under the Department

of Defense Unmanned Logistics System-Air (ULS-A) program, the Army has completed the first year of demonstrating the current prototypes; a small variant has a capacity requirement of 50-300 pounds, a medium variant has a capacity requirement of 300-800 pounds, and a large variant's capacity requirement is yet to be defined [31]. Given the Army's intent to experiment with larger capacity CUAVs, multiple delivery will be a requirement for the vehicles. This research offers a deliberate process to inform implementation of this future capability.

The remainder of this thesis is organized as follows. Chapter 2 provides a literature review of relevant research on the VRP and IRP, with an emphasis on the MILIRP variant. Chapter 3 narrows the problem description with realistic assumptions gained through research and discussions with Army acquisition experts. These insights help formulate the Military Stochastic Inventory Routing Problem- Multiple Routing (MIL SIRP-MR) as a Markov decision process (MDP) model. This chapter also introduces and discusses the approximate dynamic programming solution technique, coupled with ordinal optimization, developed to obtain high-quality dispatching policies. Chapter 4 provides a series of designed experiments that produce meaningful and insightful computational results. Chapter 5 concludes the thesis, indicating salient points and possible areas for future research.

II. Literature Review

Scholars have contributed notably and extensively to the VRP and IRP threads of research in the open, technical literature. This attention is due in part to the models' inherent applicability to real problems. Researchers have applied variants of these problems to a myriad of industrial activities, including shipping, e-Commerce, urbanization, and telecommunication infrastructures [43]. The diverse applications of the VRP and IRP have led to numerous surveys to categorize and classify the myriad works.

This chapter provides a review of the literature relevant to this research, focusing on the stochastic and dynamic VRP (SDVRP) and the stochastic IRP (SIRP). Researchers examine SDVRPs and SIRPs through various modeling methods, but this chapter focuses on MDP models with an approximate dynamic programming solution methodology. This chapter concludes by discussing recent research concerning the MILIRP, a novel IRP variant.

2.1 Vehicle Routing Problem

Introduced in 1959 by Dantzig and Ramser as the truck dispatching problem, the VRP formulation generalizes the Traveling Salesman Problem (TSP) [10]. The objective of the VRP is to optimize the delivery schedule of a vehicle fleet delivering products from a single or multiple depots to a set of customers [34]. Many VRP variants have developed over time to capture real world problem features, including time windows, capacitated vehicles, and stochastic demand. Researchers have surveyed the supporting research efforts using such variants [5]. Laporte [23] provides a general progression of a TSP model formulation to a capacitated VRP and offers three categories of solving VRPs with exact algorithms: direct tree search methods,

dynamic programming, and integer linear programming.

Laporte [23] also recognizes the value of solving various VRP variants with heuristics. To solve a capacitated VRP (CVRP), Laporte introduces the Clarke and Wright algorithm, wherein vehicle routes are generated by relaxing the fleet size and iteratively merging two routes to maximize savings. The sweep algorithm identifies vehicle routes for CVRPs with one or several depots by arraying vertices —preferably represented in polar coordinates—in a Euclidean plane, assigning vertices to vehicles based on polar angles, and continually optimizing the corresponding Traveling Salesman Problem to save distance. The Christofides-Mingozi-Toth two-phase algorithm is applied to both the CVRP and the distance constrained VRP (DVRP). This algorithm produces two solutions (sequential and parallel route construction), which users can then choose the best solution based off given parameters. A tabu search algorithm solves classical VRPs producing a sequence of solutions and improvements until termination. Laporte reports the tabu search algorithm producing the best computational results within his research. Cordeau *et al.* [8] solve the classical VRP with the savings, sweep, and Fisher and Jaikumar algorithms as well as the metaheuristic algorithms simulated annealing and genetic search. They agree with Laporte’s [23] findings that tabu search, specifically the Taillard tabu search algorithm, provides the most accurate results.

Researchers also categorize VRPs based on the timeliness and level of uncertainty of information. Pillac *et al.* [34] offer four distinct categories: static and deterministic, static and stochastic, dynamic and deterministic, and dynamic and stochastic. Because this research investigates a stochastic environment, stochastic VRPs are emphasized. Vehicle routes must continually be determined as time progresses and information captured by random variables, is revealed. Researchers refer to these VRPs as stochastic VRPs (SVRP) or stochastic dynamic VRPs (SDVRP). Berharn *et al.*

[3] provide a brief summary of research conducted on stochastic customer demand, stochastic customers, stochastic customer and demand, and stochastic service times. The authors note that SVRPs also characterize other areas of uncertainty such as road conditions, work hours, split customer demands, and other considerations. Ulmer [43] provides a comprehensive review of SDVRP applications, SDVRP formulations as an MDP, and recommended approximate dynamic programming algorithms based on problem constraints.

2.2 Inventory Routing Problem

IRPs seek to optimize the decision of combining customers in routes, the amount to deliver per customer, and when to serve the customer within a time horizon. IRPs incorporate the dimension of time into VRPs, which makes IRPs difficult to solve. Researchers have conducted extensive literature reviews on this general topic, which is classified by two characteristics: the structure of the IRP and the availability of information when making decisions. Coelho *et al.* [7] provides a comprehensive list of research efforts on the varying IRP models. They also provide an IRP taxonomy this research adopts.

Coelho *et al.* [7] categorize the IRP variants using seven criteria. First, the time horizon can be either finite or infinite. Second, structures range from single depot to single demand node, multiple demand nodes, or may also have multiple depots to multiple demand nodes. Third, routing can be limited to direct delivery, relaxed to multiple, or fully relaxed with continuous routes. Fourth, inventory policies are predetermined resupply policies, which are most commonly set as maximum level and order-up policy. Under the maximum level policy, resupply levels become flexible but bounded by the capacity of nodes, whereas an *order-up policy* delivers an amount of supplies to meet node capacity. Inventory decisions refer to the inventory level

at demand nodes. If back-ordering is allowed, node inventories are relaxed to allow negative levels, whereas lost sales refer to overages of inventory. Fleet composition and size refer to the transport vehicles. A homogeneous fleet maintains the same vehicle capacity, whereas non-homogeneous relaxes this constraint. IRP fleet sizes may be categorized as single, multiple, and unconstrained.

The second classification characteristic for IRPs pertains to when demand requirements are revealed during the decision-making process. If all information is known at the beginning of a time epoch, the IRP falls under a deterministic category. Stochastic IRPs (SIRP) are appropriate models when demand is unknown and generally follows a probability distribution. A dynamic IRP (DIRP) occurs when demand is revealed throughout the time epoch. A dynamic and stochastic inventory routing problem (DSIRP) arises when both categories apply.

If the problem set is small enough, IRP exact solutions provide optimal decision policies. Archetti *et al.* [2] solve a single vehicle IRP using a form of the branch and cut algorithm. Coelho *et al.* [6] expand on the work of Archetti *et al.* [2] by solving a classic multiple vehicle IRP (MIRP) using branch and cut, and the authors offer an improvement algorithm, noted as the solution improvement algorithm. This algorithm approximates the cost of a new solution that results in removing and reinserting nodes with the base branch and cut algorithm.

Focusing on SIRPs, many contributions in the literature have explored this problem with alternative modeling variants and solution techniques. Trudeau and Dror [41] introduce the possibility of route failures when route demands exceeds vehicle capacities. Solyali and Süral [38] formulate an SIRP with two mixed integer program models and solve using a branch and cut algorithm. Minkoff [28] models an unconstrained fleet size SIRP as an MDP and solves using a decomposition heuristic algorithm. Kleyweight *et al.* [21] formulate an SIRP as an MDP and propose

an approximate dynamic programming algorithm for a direct delivery environment, and Kleyweight *et al.* [22] extend on this work by considering multiple delivery relaxation. Adelman [1] constructs an IRP MDP model and proposes an algorithm that both creates and eliminates routes while solving the linear programs and control policy.

2.3 Military Inventory Routing Problem

A niche research avenue within the family of IRPs is the military inventory routing problem (MILIRP). Very few research efforts review the probability of successfully traversing arcs and returning to depots for IRPs, an important characteristic of the MILIRP. Mu *et al.* [30] introduce a new class of VRP noted as the disrupted VRP that deals with uncertainty during the execution stage. Such disruptions occur as vehicle breakdowns, traffic accidents, and canceled orders. The authors propose two tabu search algorithms that provide promising results.

McCormack [24] introduces the MILIRP concept and captures the inherent threat of a deployed operational environment using a hexagonal grid of the IBCT's AO. The probability of traversing each arc is derived from historical significant action reports. A threat map is generated for each fighting season, and the transition of maps follows a Markovian transition function, wherein the author solves a shortest path problem *a priori* using Dijkstra's Algorithm [20]. The work uses a two-step method of solving the turnpike sub-problem for the optimal infinite horizon decision rule to leverage within a finite horizon model that mimics a 12-month deployment. McKenna [26] expands this research by formulating an infinite time horizon model and solves instances of it using approximate dynamic programming. Using the same direct delivery construct, the author identifies the number of UAVs per decision epoch as part of the decision maker's policy. Salgado [37] continues this research thread by incorporating stochastic

demand to model a direct delivery SIRP and uses approximate dynamic programming to offer solutions. Currently, a void of MILIRP with *multiple* delivery conditions exists; this research addresses this gap in the literature.

III. Solution Methodology

3.1 Problem Overview

Within the family of IRP classes, structural variants allow researchers to compactly define specific problem instances. The MIL SIRP-MR follows an infinite time horizon formulation. A one-to-many structure represents a single supplier (i.e., the BSB) delivering to multiple customers (i.e., IBCT FOBs). Assume the BSA has adequate supplies to fully restock all FOBs at any given time.

Multiple routing relaxation allows CUAVs to visit multiple customers along a single route. A maximum level inventory policy offers replenishment level flexibility, and it is informed by the FOB capacity. Problem features such as lost sales, back-ordering, and non-negative inventory levels impact possible inventory decisions and enable further problem instance refinement. Lost sales penalize surplus supplies, whereas back-ordering allows supply levels to become negative, effectively allowing demand to be served at later stages. For MIL SIRP-MR problem instances, non-negative decisions with stockout penalties apply. If stock levels at FOBs reach negative quantities, the FOB is immediately resupplied to its maximum capacity via ground convoy, but doing so incurs a stockout penalty. Given the application of a military unit in combat operations, decision makers will never allow FOBs to reach negative supply levels because lives are at stake. Within this context, stockout penalties portray the danger of sending ground convoys on dangerous routes.

A homogeneous, unconstrained CUAV fleet with equal lift capacities, $Q \in \mathbb{N}$ service the FOBs. Fully loaded to Q when dispatched, each CUAV delivers only one supply class (i.e., Class I (food and water)) in this initial research endeavor into the MIL SIRP-MR problem. The BSB dispatches a limited number of CUAVs per decision epoch. The unconstrained CUAV fleet is limited to a maximum number of

available CUAVs, $K \in \mathbb{N}$ for each decision epoch based on the limitations imposed by maintenance, support, and flight crews. The assumption of an unconstrained fleet depicts the instance of having a large number of spare CUAVs with which to replenish the fleet to size K at the beginning of each epoch. An unconstrained fleet allows further investigation into decision policies of infinite time horizon inventory routing options. Depending on the threat situation, each FOB has an associated threat level that accounts for the probability of successfully arriving. The BSB is penalized for losing CUAVs, but it is replenished to fleet size K at the start of each epoch. The stochastic system represents the demand and probability of successfully arriving at each FOB, characterized by a probability distribution that changes over time as a function of the system state.

3.2 MDP Formulation

Let $t \in \mathcal{T} = \{1, 2, \dots\}$ represent a decision epoch depicting the start of a 24-hour ATO cycle, wherein $\mathcal{T}$ represents the set of decision epochs. CUAVs prepare, execute itineraries, and conduct post operation maintenance all within the same period. This allows K CUAVs to be available for resupply operations at each decision epoch.

Comprised of two components, the state variable represents the inventory status at each FOB, R_t , and an index corresponding to the applied threat map, M_t . Let $R_t = (R_{tb})_{b \in \mathcal{B}'} \equiv (R_{t1}, R_{t2}, \dots, R_{tB})$, wherein $B \in \mathbb{N}$ is the total number of FOBs. Let $\mathcal{B} = \{0, 1, 2, \dots, B\}$ be the set of all nodes to and from which the CUAVs must travel, wherein the depot (BSB) is represented as $b = 0$. Let $\mathcal{B}' = \mathcal{B} \setminus \{0\}$ represent all FOBs except the depot. Let $R_{tb} \in \{0, 1, 2, \dots, R_b^{cap}\}$ represent the Class I inventory level in 1000 lb increments at FOB $b \in \mathcal{B}'$ at time t , and let $R_b^{cap} \in \mathbb{N}$ represent the maximum capacity of Class I supplies at FOB $b \in \mathcal{B}$. Each FOB operates in a local threat level $(M_{tb})_{b \in \mathcal{B}} \in \mathcal{L} = \{L, M, H\}$. The threat level index impacts the demand for each

FOB, the FOB's respective threshold supply level, the likelihood of transitioning to another threat level, and the probabilities of successful travel to each node. The threat situation within the BCT's AO at time t is represented by threat map $M_t \in \{1, 2, \dots, M\}$, wherein $M \in \mathbb{N}$ represents the number of threat maps modeled by the military intelligence staff. The threat map index indicates the likelihood of threat levels M_{tb} transitioning to other levels in set $\mathcal{L}$. Let $S_t = (R_t, M_t) \in \mathcal{S}$, wherein $\mathcal{S}$ is the set of all possible states.

The decision space is the set of all CUAV fleet itineraries each 24-hour ATO cycle. The BSB decides which FOBs to resupply, the amount of supplies to deliver at each FOB, and the sequence of FOBs resupplied on each CUAV route. The decision space observes the following assumptions: CUAVs are limited in fuel capacity and can only execute one mission comprised of resupplying at most two FOBs per ATO cycle; CUAVs executing resupply missions travel with their maximum supply capacity, $Q \in \mathbb{N}$; and each CUAV evenly distributes its load to visited FOBs within itineraries.

A 2-tuple of vectors represents the decision space $\mathcal{X}$. Let $\mathcal{E} = \mathcal{B} \times \mathcal{B}$ be the set of edges, i.e., base-to-base, depot-to-base, or base-to-depot movements. Let $\mathcal{K} = \{0, 1, 2, \dots, K\}$ be the set of CUAVs used for itineraries at time t . The BSB has the option of not employing any CUAVs. The decision variable $y_t = (y_{tij}^k)_{i,j \in \mathcal{E}, k \in \mathcal{K}, t \in \mathcal{T}}$, wherein $y_{tij}^k \in \{0, 1\}$, indicates whether CUAV $k \in \mathcal{K}$ traverses on edge $(i, j) \in \mathcal{E}$ or not at time t . The decision variable $z_t = (z_{tj}^k)_{j \in \mathcal{B}', k \in \mathcal{K}, t \in \mathcal{T}}$, wherein $z_{tj}^k \in \{0, 1, 2, \dots, Q\}$ represents supplies delivered to location $j \in \mathcal{B}'$ by CUAV $k \in \mathcal{K}$ at time t . The amount of supplies delivered is dictated by the number of FOBs visited for each CUAV within the itinerary. Loaded with max capacity Q , the decision variable z_{tj}^k is

defined as

$$z_{tj}^k = \max \left\{ \left\lfloor \frac{Q}{\sum_{i:(i,j) \in \mathcal{E}} y_{tij}^k - 1} \right\rfloor, 0 \right\}, \forall k \in \mathcal{K}, j \in \mathcal{B}', t \in \mathcal{T}.$$

Let $x_t = (y_t, z_t)$ compactly represent the decision at time t . Because each FOB has a capacity, R_b^{cap} , actions cannot result in FOBs exceeding Class I capacity. Therefore, the decision space $\mathcal{X}$ relies on the current system state S_t . Let

$$\mathcal{X}_{S_t} = \left\{ x_t : R_{tb} + \sum_{k \in \mathcal{K}} z_{tk}^k \leq R_b^{cap}, \forall b \in \mathcal{B}' \right\}, \forall S_t \in \mathcal{S}, t \in \mathcal{T} \quad (1)$$

represent the set of all feasible decisions.

The system evolves over time according to the following state transition function, $S_{t+1} = S^M(S_t, x_t, W_{t+1})$, wherein $W_{t+1} = (\hat{d}_{tb}, \hat{M}_{t+1}, \hat{K}_{t+1}^{lost})$ captures the exogenous information with three random variables, $\hat{d}_{tb}$, $\hat{M}_{t+1}$, and $\hat{K}_{t+1}^{lost}$. Let $\hat{d}_{tb} = \hat{d}_{tb}(M_{tb}) \equiv (\hat{d}_{tb}^L, \hat{d}_{tb}^M, \hat{d}_{tb}^H)$, wherein $\hat{d}_{tb}(M_{tb})$ represents the random demand of FOB b at time t , conditioned on the FOB threat level M_{tb} . Each demand threat level dictates a range of Class I consumption, $\hat{d}_{tb}(M_{tb}) \in \{1, 2, \dots, R_b^{cap}\}$, $\forall b \in \mathcal{B}'$, informed by historical data on respective threat levels specific to unit type. The random variable $(\hat{d}_{tb}^\ell)_{\ell \in \mathcal{L}}$ represents a uniform distribution of demand if FOB b is in threat category $\ell \in \mathcal{L}$. The threat map M_t transitions following $\hat{M}_{t+1} = P^M(M_t, M_{t+1})$, wherein P^M is a transition probability function that governs the evolution of threat maps based on information revealed at $t+1$. The military intelligence staff derives P^M from multiple information sources to include sensors, weather reports, and the aviation maintenance cell. Assume FOB threat levels cannot transition by more than one level of threat between time epochs. For example, $M_{tb} = L$ can transition to M but not to H in the next epoch, $t+1$. With the intelligence community providing periodic intelligence

updates to the BSB, this structure seems reasonable to implement.

The transition function also relies on the decision, x_t , along with the exogenous revealed information to determine S_{t+1} . The successful travel outcomes of CUAV $k \in \mathcal{K}$ follows a multinomial distribution with parameters $p_{ij}(M_{tb})$, which represents all edge probabilities of successful travel with respect to FOB threat level M_{tb} . Given an itinerary of CUAV k , denote the set of edges traversed as $e \in \mathcal{E}'_k = \{1, 2, \dots, E\}$, wherein the elements of the set correspond to the traversed arcs $(i, j) \in \mathcal{E}$, ordinaly ranked in order of traversal. Let p_e^k represent the respective probabilities of the CUAV successfully traversing edge $e \in \mathcal{E}'_k$. Let $E \in \mathbb{N}$ represent the last edge of the itinerary for CUAV k returning to the depot. Each CUAV k itinerary within decision x_t follows an independent multinomial trial with parameters $p_1^k, \dots, p_E^k$ that results in exactly one of E possible outcomes. Let

$$\rho_e^k = \begin{cases} (1 - p_e^k) & \text{for } e = 1 \\ \prod_{a=1}^{e-1} p_a^k (1 - p_e^k) & \text{for } 1 < e < E \\ \prod_{e \in \mathcal{E}'_k} p_e^k & \text{for } e = E \end{cases}, \forall e \in \mathcal{E}'_k, k \in K,$$

wherein ρ_e^k represents the probability that CUAV k will fail on edge e on a given trial. For each CUAV k , the E -tuple $(X_1^k, \dots, X_E^k)$ represents a multinomial random variable with a single independent multinomial trial with parameter ρ_e^k , wherein X_e^k denotes the edge $e \in \mathcal{E}'_k$ on which the CUAV fails. Let X_E^k indicate the CUAV has successfully completed its itinerary and has safely returned to the BSB. Let

$$\mathbb{E}(\hat{K}^{lost}, K | S_t, x_t) = \sum_{k \in K} (\mathbb{E}(X_e^k) : e \neq E)$$

represent the number of CUAVs lost given the system state S_t with decision x_t and

compactly denoted as $\mathbb{E}(\hat{K}^{lost})$. Let $K' = K - \mathbb{E}(\hat{K}^{lost})$, wherein K' represents the remaining fleet size based off the outcomes of choosing x_t .

The contribution function is comprised of two components: the cost of losing CUAVs from enemy threats (i.e., MANPADS, small arms fire (SAF), or rockets), inclement weather, or vehicle malfunction, denoted as C^{UAV} , and the contribution of delivering supplies to FOBs with respect to threshold levels, denoted as C^{Supply} . Both component functions rely on the current state S_t , decision x_t , resulting state S_{t+1} , and two parameters: a threshold value $T_b = T_b(M_{tb}) \in \mathbb{N}$ conditioned on the threat level and a stockout penalty $O_b \in \mathbb{N}^+$ at FOB b . The cost of losing CUAVs is represented as

$$C^{UAV}(S_t, x_t) = -u^{cost} + \frac{u^{cost} K'}{K},$$

and applies $\forall t \in \mathcal{T}$. This cost function consists of a penalty parameter, $u^{cost} \in \mathbb{N}^+$, for losing all CUAVs due to decision x_t . The supply contribution, C^{Supply} , offers a reward for delivering to FOBs if delivered supplies sustain FOB b above the threshold T_b and penalizes the BSB if supplies fall below this parameter. The size of reward relies on the severity of threat at each FOB b ; the higher threat offers greater rewards. Likewise, a higher threat leads to a higher penalty if supply levels fall below T_b . The threshold is conditioned on M_{tb} as the threat varies from FOB to FOB. The lower the threat, the closer to depletion FOBs are allowed to consume. The contribution for delivering supplies is represented as

$$C^{Supply}(S_t, x_t, S_{t+1}) = \begin{cases} \sum_{b \in \mathcal{B}'} \left(\frac{O_b}{T_b} (R_{t+1,b} - T_b) \right) & \text{for } R_{t+1,b} < T_b \\ \sum_{b \in \mathcal{B}'} \left(\frac{O_b R_b^{cap} - T_b^2 + O_b T_b}{R_b^{cap} - T_b} (R_{t+1,b} - R_b^{cap}) + O_b R_b^{cap} \right) & \text{for } R_{t+1,b} > T_b. \\ 0 & \text{otherwise} \end{cases}$$

The final contribution function is expressed as

$$C(S_t, x_t, S_{t+1}) = C^{Supply}(S_t, x_t, S_{t+1}) + C^{UAV}(S_t, x_t)$$

and is graphically depicted in Figure 1.

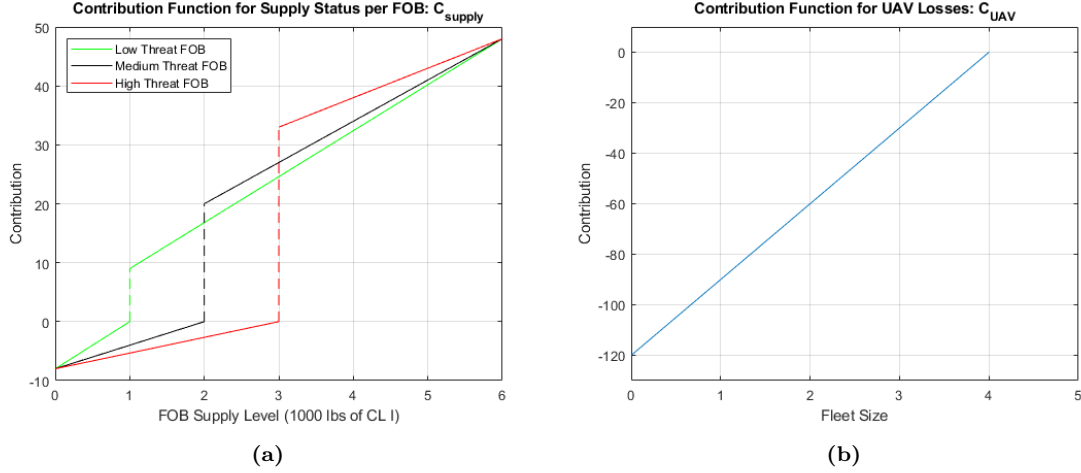


Figure 1. Contribution Function Graphs

Figure 1(a) provides an example of C^{supply} with low, medium, high threshold supply levels $T_b = [1000, 2000, 3000]$ respectively, a stockout penalty $O_b = 8$ and $R_b^{cap} = 6000$ lbs $\forall b \in \mathcal{B}'$. Figure 1(b) provides an example of C^{UAV} with a max fleet size $K = 4$ and penalty for losing entire fleet, $u^{cost} = 120$.

The objective of the MDP model is to determine a policy that maximizes the expected total discounted reward, expressed as

$$\max_{\pi \in \Pi} \mathbb{E}^{\pi} \left[\sum_{t=1}^{\infty} \gamma^{t-1} C(S_t, X^{\pi}(S_t), S_{t+1}) \right],$$

wherein $\gamma \in [0, 1)$ is a fixed discount rate and $X^{\pi}(S_t)$ represents the decision function that returns a decision x_t given state S_t based on policy π . The MDP model seeks to find the optimal policy π^* from a set of policies $(X^{\pi}(S_t))_{\pi \in \Pi}$. The optimal policy is found by solving the Bellman Equation,

$$V(S_t) = \max_{x_t \in \mathcal{X}_{S_t}} \mathbb{E} \left[C(S_t, x_t, S_{t+1}) + \gamma \left(V(S_{t+1}) | S_t, x_t \right) \right]. \quad (2)$$

Unfortunately, the high dimensionality and immense size of the state and decision spaces makes solving Equation (2) for the optimal policy computationally intractable. Therefore, approximate dynamic programming is implemented to attain improved inventory routing policies relative to a policy attained from a myopic, benchmark policy.

3.3 ADP Formulation

An approximate policy iteration (API) algorithm structure is used to approximate the value function, Equation (2). Post-decision state S_t^x refers to the state of the system after being in pre-decision state S_t and taking action x_t . Leveraging a post-decision state convention provides a formidable computational advantage by avoiding

explicitly computing expectations [36]. The optimality equation can be modified to incorporate the post-decision state variable convention. Let

$$V^x(S_t^x) = \mathbb{E}[V(S_{t+1})|S_t^x], \quad (3)$$

represent the value of being in post-decision state S_t^x . The optimality equation can now be represented in terms of the post-decision state variable by substituting Equation (3) into Equation (2),

$$V(S_t) = \max_{x_t \in \mathcal{X}_{S_t}} \left(C(S_t, x_t, S_{t+1}) + \gamma V^x(S_t^x) \right). \quad (4)$$

Identifying the value of being in post-decision state S_{t-1}^x denoted as

$$V^x(S_{t-1}^x) = \mathbb{E}[V(S_t)|S_{t-1}^x], \quad (5)$$

and substituting this recognition into Equation (4), the optimality equation in terms of post-decision state, S_{t-1}^x , is represented as

$$V^x(S_{t-1}^x) = \mathbb{E} \left[\max_{x_t \in \mathcal{X}_{S_t}} \left(C(S_t, x_t, S_{t+1}) + \gamma V^x(S_t^x) \right) \middle| S_{t-1}^x \right]. \quad (6)$$

Although utilizing the post-decision state convention does offer computational advantages, the modified optimality equation represented in Equation (6) remains computationally intractable due to size of the motivated problem instances. Therefore, another approach must be applied to approximate the value function. A notable method is value function approximation (VFA) through a set of fixed set of basis functions and implementing an API strategy to the basis function results to minimize the mean-squared projected Bellman error [18]. Least squares temporal difference (LSTD) remains a popular option because it exploits the linear architecture

assumption and displays fast convergence for policy evaluation [42]. However, such a technique’s success hinges on the manual selection of basis functions.

To mitigate the risk of affecting the convergence properties, kernel methods offer another solution methodology option. Compared to Temporal Difference (TD) and LSTD algorithms, kernelized least squares temporal difference (KLSTD) can offer improved approximations, especially in nonlinear VFA instances [45]. Let $\phi : \mathbb{R}^n \rightarrow \mathcal{D}$ represent the n dimension state space in a high-dimensional feature space or dictionary size $\mathcal{D}$, wherein ϕ serves as a basis function. Using the Mercer Theorem [45], the kernel function maps the state space to the basis function

$$K_h(s, s') = \phi(s)^\top \phi(s').$$

Replacing all inner products with the kernel function computations greatly reduces the computational costs incurred by high dimensional state spaces. This research effort implements the Gaussian kernel, given by

$$K_h(S_{t-1}^x, S_t^x) = e^{\left(-\frac{\|S_{t-1}^x - S_t^x\|^2}{h}\right)}, \quad (7)$$

wherein $\|\cdot\|$ represents the Euclidean norm and h is the bandwidth, a tunable parameter that adjusts S_t^x range of influence [35]. Capitalizing on the Kernel Representer Theorem [35], the value of transitioning from one state to the next based on decision x_t can be approximated using a VFA

$$\bar{V}^x(S_t^x|\theta) = K_h(S_t^x)^\top \theta, \quad (8)$$

wherein $\theta = (\theta_d)_{d \in |\mathcal{D}|}$ is a column vector of weights representing the set of all kernel features, depending on dictionary $|\mathcal{D}|$. The θ vector represents the policy π that

updates the decision function

$$X^\pi(S_t|\theta) = \arg \max_{x_t \in \mathcal{X}_{S_t}} \left(C(S_t, x_t, S_{t+1}) + \gamma \bar{V}^x(S_t^x|\theta) \right). \quad (9)$$

3.3.1 Reduced Decision Space

The MIL SIRP-MR relaxation of multiple routing generates a large number of state action pairs, which in large problem instances makes it difficult to attain quality solutions in a timely manner. Reducing the decision space is required to solve problem instances of interest. Implementing *ordinal optimization* effectively generates high-quality results by selecting a subset of the decision space to consider.

This research implements such methods to select a subset of decisions for each state such that ω percent of the decision space contains a *good enough* decision [32] i.e., within a predetermined confidence threshold ψ . Using Equation (1), let $\tilde{\mathcal{X}}_{S_t} \subset \mathcal{X}_{S_t}$ denote a reduced decision space of size $|\tilde{\mathcal{X}}_{S_t}| = Z$. Leveraging Equation (8), define

$$\tilde{V}^x(S_t^x|\theta) = \max_{x_t \in \tilde{\mathcal{X}}_{S_t}} (K_h(S_t^x)^\top \theta) \quad (10)$$

as an approximation for $\bar{V}^x(S_t^x|\theta)$.

Using this approach requires additional explanation regarding Z and the method for selecting Z decisions from $\mathcal{X}_{S_t}$. Previous work [19] implementing ordinal optimization proves

$$Z \geq \left\lceil \frac{\log(1 - \psi)}{\log(\omega)} \right\rceil. \quad (11)$$

For each state, a subset of decisions with size Z is investigated such that with ψ confidence, the best decision produces a value better than a pre-determined population percentage ω . Table 1 provides the combinations of ψ and ω used in this research.

The CUAV fleet size determines the level of ordinal optimization implementation. If

Table 1. Ordinal Optimization Sample Sizes

ω	ψ	Z
0.95	0.95	59
0.95	0.99	90
0.99	0.95	299
0.99	0.99	459

the size of original decision space is larger than Z (i.e., $|\mathcal{X}_{S_t}| > Z$), a higher level of confidence and population percentage may be used. However, if $|\mathcal{X}_{S_t}|$ is less than any Z values listed in Column 3 of Table 1, ordinal optimization is not implemented, and the full decision space is evaluated.

The next task requires a sampling scheme to appropriately select Z decisions. For the purpose of this research, a uniform distribution $\mathcal{U}(1, |\mathcal{X}_{S_t}|)$ and binomial distribution $\mathcal{BIN}(|\mathcal{X}_{S_t}|, \delta)$ with varying levels of δ binomial probabilities provide the sampling scheme index for $\tilde{\mathcal{X}}_{S_t}$. Therefore, Equation (9) can now be represented as

$$X^\pi(S_t|\theta) = \arg \max_{x_t \in \tilde{\mathcal{X}}_{S_t}} \left(C(S_t, x_t, S_{t+1}) + \gamma \tilde{V}^x(S_t^x|\theta) \right). \quad (12)$$

3.3.2 ADP Algorithm

The API algorithm generates a sequence of policies and associated approximate value functions by iteratively applying two hierarchically nested procedures: policy evaluation and policy improvement. Within the policy evaluation steps, KLSTD approximates the value function for a fixed policy π through simulation and KLSTD reinforcement learning. Within the policy improvement steps, the algorithm identifies a new policy based on the previous policy evaluation steps' associated data. The API algorithm is shown in Algorithm 1.

Algorithm 1 API-KLSTD with Ordinal Optimization

Initialize $\mathcal{D}$ of reference points in $|S_t^x|$

Initialize θ

Determine Ordinal Optimization sample size Z using Table 1 and sampling distribution

for $m = 1$ to M (*Policy Improvement Loop*) **do**

 Simulate N random post-decision states $S_{t-1,n}^x$

for $n = 1$ to N (*Policy Evaluation Loop*) **do**

 Record $\phi_{t-1} = K_h(S_{t-1,n}^x)$

 Simulate state transition to next pre-decision state $S_{t,n}$

 Determine decision at next pre-decision state

if $Z > 0$ **then**

 Draw Z samples from $\tilde{\mathcal{X}}_{S_t}$ with chosen sampling dist.

 Solve for x_t^n using Eq. (12)

else

 Solve for x_t^n using Eq. (9)

end

 Record $C(S_{t,n}, x_t, S_{t+1,n})$

 Record $\phi_t = K_h(S_{t,n}^x)$

end

 Compute $\hat{\theta}$ using Eq. (13)

 Update θ^m using Eq. (14)

end

Return decision function $X^{\pi^{KLSTD}}(\cdot | \theta^M)$ for policy π^{KLSTD}

The algorithm begins by initializing the dictionary $\mathcal{D}$ and θ vector, which represents the initial policy. Based on $|\mathcal{X}_{S_t}|$, an appropriate ordinal optimization sample size and sampling distribution are identified. The performance evaluation of the current policy begins wherein the following steps occur for each iteration $n = 1, 2, \dots, N$. A

set of random post-decision states $(S_{t-1,n}^x)_{n \in N}$ are generated, and the kernel function evaluations $K_h(S_{t,n}^x)$ are recorded. Next, a one step simulation forward generates pre-decision state, $S_{t,n}$. The best decision, x_t , is determined using ordinal optimization with sample size Z , subject to the aforementioned sample quality conditions, else a full decision space evaluation. Once x_t is determined, the contribution $C(S_{t,n}, x_t, S_{t,n+1})$ and kernel function evaluation $K_h(S_{t,n}^x)$ are recorded. A total of N temporal difference sample realizations is collected, where $C(S_t, x_t, S_{t+1}) + \gamma \bar{V}^x(S_t^x | \theta)$ is the n th temporal difference, given the kernel function weight vector θ .

Policy improvement is the next phase of the algorithm, where the following steps occur for $m = 1, 2, \dots, M$. The vector $\hat{\theta}$ is computed via least squares regression, seeking a $\hat{\theta}$ weight vector that minimizes the mean squared error. This computation is performed by solving the normal equation

$$\hat{\theta} = [(\Phi_{t-1} - \gamma \Phi_t)^\top (\Phi_{t-1} - \gamma \Phi_t) + R]^{-1} (\Phi_{t-1} - \gamma \Phi_t)^\top C_t, \quad (13)$$

wherein Φ_{t-1} and Φ_t represent the basis function matrices and C_t represents the contribution vector

$$\Phi_{t-1} = \begin{bmatrix} \phi(S_{t-1,1}^x)^\top \\ \vdots \\ \phi(S_{t-1,N}^x)^\top \end{bmatrix}, \quad \Phi_t = \begin{bmatrix} \phi(S_{t,1}^x)^\top \\ \vdots \\ \phi(S_{t,N}^x)^\top \end{bmatrix}, \quad C_t = \begin{bmatrix} C(S_{t,1}, x_t) \\ \vdots \\ C(S_{t,N}, x_t) \end{bmatrix}.$$

Let $R = \eta \mathbf{I}$ represent the regularization matrix of size $|\mathcal{D}| \times |\mathcal{D}|$ with diagonal penalty values given by the regularization parameter $\eta \geq 0$. Regularization is used to avoid non-singularity matrix inversion issues and to reduce overfitting. The weight vector θ is updated by smoothing in $\hat{\theta}$ using

$$\theta^m \leftarrow (1 - \alpha_m)\theta^{m-1} + \alpha_m\hat{\theta}. \quad (14)$$

The step-size rule α_m follows a polynomial step-size rule, wherein $\beta \in (\frac{1}{10}, 1]$ and

$$\alpha_m = \frac{1}{m^\beta}.$$

The rate of α_m declines as the policy improvement loop m increases. Moreover, the smaller the β -value, the slower α_m declines. As the policy improvement phase progresses, greater relative weight is applied to θ compared to $\hat{\theta}$. The algorithmic parameters $M, N, h, |\mathcal{D}|, \eta$, and β are tunable, where M is the number of policy improvement iterations, N is the number of policy evaluation iterations, h is the Gaussian kernel function bandwidth, $|\mathcal{D}|$ is the dictionary size, η is the regularization parameter, and β is the step-size parameter. The ordinal sampling distribution and sample size $Z(|\mathcal{X}_{S_t}|)$ conditioned on the decision space size are tunable parameters specific to the ordinal optimization procedure.

IV. Testing, Results, and Analysis

This chapter demonstrates the applicability of the proposed MDP model for analyzing a military logistical planning problem, represented as MIL SIRP-MR instances. The chapter also examines the efficacy of the proposed ADP solution technique. A series of computational experiments examines the performance of the ADP technique effects by altering algorithm parameters within different problem instances. The experiments were conducted on an Intel Xeon E5-2687W workstation with 64GB RAM, 10 cores, and MATLAB Parallel Computing Toolbox.

4.1 Problem Instances

Several MIL SIRP-MR problem instances are generated to depict a BSB implementing a CUAV fleet in their resupply operations for an IBCT. The IBCT conducts operations in eastern Afghanistan with varying levels of threat maps M_t . The BSB tasks CUAVs to resupply FOBs that are located in areas less conducive to ground resupply either due to ground convoy threats or inhospitable road conditions. An example of an eligible location is Combat Outpost Keating, depicted in Figure 2 [44]. A combat outpost is a term for a relatively small, austere FOB with typically higher threat levels.

Nestled in a "bowl" in Nuristan province near the intersection of two rivers, Combat Outpost Keating is surrounded by steep mountains that provide adversaries a bird's eye view of friendly activity [25]. A combination of narrow-winding roads, often impassable during the winter seasons and active insurgents make ground resupply a treacherous task. Within this context, the series of computational experiments focus on a subset of the IBCT FOBs. Hereafter, this subset is considered the set of FOBs $\mathcal{B}'$ that the BSB must plan to resupply with the CUAV fleet.



Figure 2. Combat Outpost Keating, Afghanistan

This research explores problem instances with three and four FOBs, each having the same storage capacity $(R_b^{cap})_{b \in \mathcal{B}'}$ equal to 6000 lbs of Class I. The threat map M_t is aggregated into three categories: Low, Medium, and High. For example, if $M_t = \text{Low}$, the FOBs have a higher probability of transitioning to lower threat levels. Each FOB has an initial starting threat level based on surrounding enemy activity, weather, and location.

The CUAV fleet size ranges from two to four vehicles, each with different capacities Q . Table 2 provides a summary of the CUAV characteristics investigated.

Table 2. CUAV Characteristics

Type	K	Q (1000 lbs)
UH-60	2	10
Variation	3	7
K-Max	4	5

All vehicles represent viable options within the ULS-A Large category. The types of CUAVs were selected based on either current CUAVs such as the K-Max or potential prototypes based on existing rotary wing assets such as the UH-60 Blackhawk. The CUAV quantities and capacities were based on a problem constraint of an entire fleet delivering approximately 20,000 lbs. Moreover, the larger capacity vehicles were

limited in quantity based on the intuition that larger, capable CUAVs may be more expensive and fewer in number, given greater maintenance and support requirements. A CUAV can only travel up to two FOBs, departing the BSB at maximum capacity and delivering the same amount to each FOB within its multi-route itinerary. Three problem features are varied to generate the problem instances: number and type of CUAVs, threat maps, and number of FOBs. The problem instances wherein the BSB maintains more CUAVs than FOBs were withdrawn from the set of problem instances as these scenarios are unlikely. This research investigates a total of 15 problem instances.

BSB planners must balance supply levels while maintaining operational readiness of the CUAV fleet. Therefore, two response variables are recorded during each experiment. The supply status of $\mathcal{B}'$ is best represented by the daily amount of supplies above the thresholds across all FOBs,

$$Supply = \sum_{b \in \mathcal{B}'} (R_b - T_b) . \quad (15)$$

The health of the CUAV fleet is represented using a daily operational readiness percentage (OR%). This rate is simply calculated by

$$OR\% = \frac{K'}{K} , \quad (16)$$

wherein K' represents the remaining fleet size based on the decision and subsequent outcomes and K represents the max fleet size. A simulation based on a 365-day deployment is implemented to obtain these performance measures. The quality of the ADP policies generated by the API-KLSTD algorithm is determined using this information.

4.2 Benchmark Policy

To assess the solution quality of the ADP algorithm, a benchmark policy is used to compare results. This benchmark is myopic in the sense that the decision does not account for future realizations of the system state. A set partitioning VRP (SPVRP) [40] captures the myopic view point of balancing both the OR% and maximizing supply at all FOBs. This benchmark policy $X^{\pi^{Benchmark}}(S_t)$ is constructed as a binary integer program,

$$\min \sum_{f \in \mathcal{F}} c_f a_f \quad (17)$$

$$\text{s.t. } \sum_{f \in \mathcal{F}} a_f = 1, \quad (18)$$

$$a_f \in \{0, 1\}, \quad \forall f \in \mathcal{F}. \quad (19)$$

Let $\mathcal{F} = \{1, 2, \dots, |\mathcal{X}_{S_t}|\}$ denote the set of all possible decisions or circuits possible for the CUAV fleet with associated deliveries. Let c_f represent the cost of each decision

$$c_f = -\ln \left(\sum_{k \in K} P_{k,f} \right) + \sum_{b \in \mathcal{B}'} \frac{R_b^{cap} - d_b + D_{b,f}}{R_b^{cap}}, \quad \forall f \in \mathcal{F},$$

wherein $P_{k,f} = (\prod p_e^k, \forall e \in \mathcal{E}')$ represents the probability of k CUAV successfully completing its itinerary. The variable $P_{k,f}$ is summed across $k \in K$ to provide an overall probability of success for choosing a_f . Let d_b represent the realized demand and $D_{b,f}$ represent the intended delivery amount to FOB b using decision a_f . Equation (18) imposes the constraint that only one action can be made and Equation (19) indicates the decision variable a_f can only be binary. The Benchmark Policy is based off the goal of minimizing the risk to the CUAV fleet while minimizing the percentage

of supply deficit from R^{cap} across all FOBs.

4.3 Algorithm Features

Two categories divide the algorithm features: the ADP algorithm features and the ordinal optimization features. The ordinal optimization sample size is conditioned on $|\mathcal{X}_{S_t}|$. For any problem instance that is less than the ordinal optimization sample size Z when $\psi = \omega = 0.95$, ordinal optimization is not implemented (i.e., it only applies to problem instances having 2 CUAVs). Table 3 provides a summary of the problem and algorithm features.

Table 3. Experimental Design Factor Levels

Category	Feature	Parameter Settings
Problem Instance	K	$\{2,3,4\}$
	M_t	$\{L,M,H\}$
API-KLSTD	$\mathcal{B}'$	$\{3,4\}$
	N	$\{0.1, 0.2, 0.3, 0.5\} \cdot \mathcal{S}_t $
	$ \mathcal{D} $	$\{2,5,6,20\}$
	h	$\{0.1, 0.5, 1, 10\}$
	β	$\{0.1, 0.8\}$
	η	$\{0.001, 0.1\}$
	Ordinal Optimization	$\{\mathcal{U}(1, \mathcal{X}_{S_t}),$
		$\mathcal{BIN}(\mathcal{X}_{S_t} , 0.3),$
		$\mathcal{BIN}(\mathcal{X}_{S_t} , 0.5),$
		$\mathcal{BIN}(\mathcal{X}_{S_t} , 0.7)\}$
	Z	$\{(0,0) : K = 2,$ $(59,90) : K = 3,4\}$

By using a replacement method [29] to translate a four-level feature to a two-level, seven two-level features translate into eleven two-level features by implementing the method to $N, |\mathcal{D}|, h$ and the ordinal optimization sampling distribution method. For each problem instance, a 2_{IV}^{11-6} design of experiment examines the solution quality of policies derived by the API-KLSTD algorithm. Each problem instance-specific 32-run experimental design was replicated 50 times to achieve a desired level of confidence.

These experiments provide insight into the ADP algorithm performance for the MIL SIRP-MR.

4.4 Results and Analysis

Table 4 summarizes the results of the problem instance experiments. Starting from the left, Columns 1-4 indicate the problem feature levels. Columns 5-12 tabulate the superlative parameter settings and associated computation times for the API-KLSTD algorithm and associated ordinal optimization for each problem instance. Columns 13-14 provide a daily average CUAV fleet OR% interval over the 50 simulations for both the Benchmark and ADP Policies, respectively. Column 15 provides a solution quality measurement of the ADP algorithm expressed in terms of a 95% confidence interval of percent improvement over the Benchmark Policy of supply levels across all FOBs. The supply levels are determined using Equation (15), the amount of supplies at each FOB at the end of time t in relation to the FOB's respective threshold.

Table 4. Experimental Design Results

Problem Instance Parameter Settings				KLSTD-API Parameter Settings						Ordinal Optimization Parameter Settings		Daily Fleet OR %		Daily Class I above Threshold for FOBs $\mathcal{B}'$
Instance	K	M_t	B	N	$ \mathcal{D} $	h	β	η	Z	Sampling Distribution	Computational Efficiency (min)	Benchmark Policy	ADP Policy	Improvement over Benchmark (%)
1	2	L	3	$0.4 \cdot \mathcal{S} $	2	0.1	0.1	0.001	-	-	0.36	94.84 ± 0.56	98.63 ± 0.21	2.66 ± 0.14
2	2	M	3	$0.4 \cdot \mathcal{S} $	2	0.1	0.1	0.001	-	-	0.36	86.89 ± 0.70	95.9 ± 0.33	13.64 ± 0.32
3	2	H	3	$0.4 \cdot \mathcal{S} $	2	0.1	0.1	0.001	-	-	0.37	85.64 ± 0.67	95.15 ± 0.30	11.79 ± 0.15
4	3	L	3	$0.4 \cdot \mathcal{S} $	5	1	0.8	0.001	90	$\mathcal{BIN}(Z, 0.5)$	0.64	94.09 ± 0.46	99.07 ± 0.14	-4.41 ± 0.15
5	3	M	3	$0.5 \cdot \mathcal{S} $	5	0.1	0.8	0.001	59	$\mathcal{BIN}(Z, 0.7)$	0.84	87.83 ± 0.54	96.97 ± 0.21	7.98 ± 0.56
6	3	H	3	$0.5 \cdot \mathcal{S} $	5	0.1	0.8	0.001	59	$\mathcal{BIN}(Z, 0.7)$	0.92	86.22 ± 0.52	96.63 ± 0.19	10.92 ± 0.69
7	2	L	4	$0.4 \cdot \mathcal{S} $	2	1	0.1	0.1	-	-	14.51	93.09 ± 0.72	98.44 ± 0.23	4.00 ± 0.37
8	2	M	4	$0.4 \cdot \mathcal{S} $	2	1	0.1	0.1	-	-	11.80	85.36 ± 0.55	96.96 ± 0.20	8.39 ± 0.48
9	2	H	4	$0.4 \cdot \mathcal{S} $	2	1	0.1	0.1	-	-	9.90	84.79 ± 0.62	96.47 ± 0.28	4.50 ± 0.52
10	3	L	4	$0.2 \cdot \mathcal{S} $	6	0.1	0.8	0.1	90	$\mathcal{U}(1, \mathcal{X}_{S_t})$	25.63	91.05 ± 0.83	99.59 ± 0.07	0.57 ± 0.84
11	3	M	4	$0.3 \cdot \mathcal{S} $	5	0.1	0.8	0.1	90	$\mathcal{U}(1, \mathcal{X}_{S_t})$	48.30	82.57 ± 0.51	99.00 ± 0.11	12.29 ± 0.77
12	3	H	4	$0.3 \cdot \mathcal{S} $	5	0.1	0.8	0.1	90	$\mathcal{U}(1, \mathcal{X}_{S_t})$	48.48	82.33 ± 0.49	98.77 ± 0.12	11.06 ± 0.63
13	4	L	4	$0.4 \cdot \mathcal{S} $	5	0.1	0.8	0.1	90	$\mathcal{U}(1, \mathcal{X}_{S_t})$	77.72	91.15 ± 0.83	99.87 ± 0.04	3.58 ± 0.25
14	4	M	4	$0.3 \cdot \mathcal{S} $	20	1	0.1	0.1	59	$\mathcal{BIN}(Z, 0.5)$	36.80	82.67 ± 0.54	99.75 ± 0.05	8.96 ± 0.37
15	4	H	4	$0.2 \cdot \mathcal{S} $	5	0.1	0.8	0.1	90	$\mathcal{U}(1, \mathcal{X}_{S_t})$	52.95	83.77 ± 0.47	96.82 ± 0.27	-0.77 ± 0.83

The ADP algorithm appropriately anticipates transition of the system state, specif-

ically the FOB threat levels associated with threat maps. Comparing the Benchmark Policy and ADP Policy Daily Fleet OR% provides evidence in support of this finding. The lowest ADP Policy OR% occurs in Problem Instance 3 (95.15%), but it still outperforms the highest Benchmark Policy OR% of 94.84%. Furthermore, a 12.82% difference exists when comparing the lowest OR% of both policies. Problem Instances 4-6 model 3 CUAVs with 3 FOBs to resupply during Low, Medium, and High threat maps, respectively. The difference between the Benchmark Policy and ADP Policy OR% increases as threat level increases. The increasing difference in relation to increasing threat is more pronounced with fewer CUAVs than FOBs to resupply (e.g., Problem Instances 7-9). Figure 3 visually depicts the marked advantage of the ADP Policy over the Benchmark Policy across all problem instances.

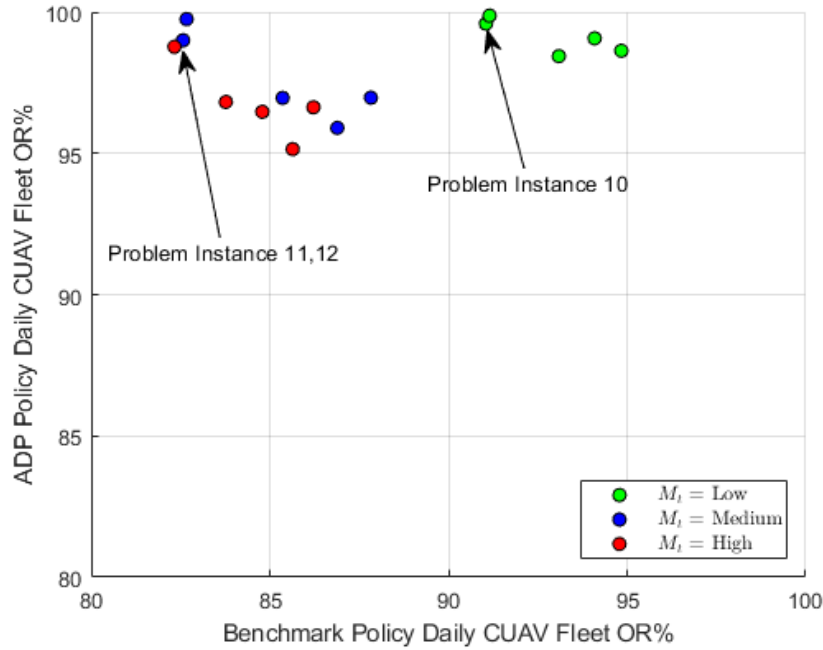


Figure 3. CUAV Fleet OR% Comparison across Problem Instances

In Problem Instances 10-12 wherein 3 CUAVs service 4 FOBs, both policies exhibit a decline in OR% as threat increases. However, the ADP Policy continues to maintain a high OR% compared to the Benchmark Policy.

The ADP Policy maintains Class I supplies above thresholds better than the Benchmark Policy in 13 out of the 15 problem instances investigated. Figure 4 provides a visual summary of the problem instances, segmented by threat maps.

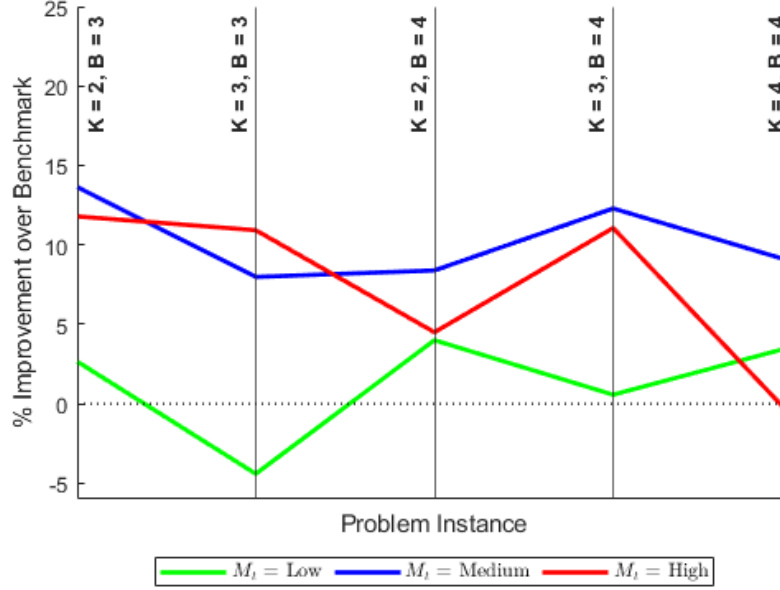


Figure 4. Daily Class I above Threshold across FOBs

From left to right, the number of FOBs (B) increases from 3 to 4, with each respective problem instance, and the number of CUAVs increases, for a given number of FOBs, from left to right as well. The ADP Policy provides the best improvement over the Benchmark Policy when the threat map M_t begins at Medium, whereas the worst improvement for all but one scenario is obtained when M_t begins at Low. Although the ADP Policy outperforms the Benchmark Policy in low threat maps (e.g., low intensity, steady state conflicts), implementing the SPVRP for planning becomes a viable option. Similar to the ADP Policy behavior for the OR%, the policy generally performs better when the number of FOBs surpasses the number of CUAVs. Of the two instances where the Benchmark Policy outperforms the ADP Policy, both the number of CUAVs and FOBs were equal. This observation suggests that the ADP

policy performs better than the Benchmark Policy in both maintaining fleet OR % and supply levels when the problem instances require CUAVs be routed to more than one FOB for resupply missions. Operating in high threat maps, the ADP Policy consistently outperforms the Benchmark Policy, but performance improvement follows a decreasing trend as the problem becomes more complicated. Operating in low threat maps, the percent improvement over the Benchmark Policy begins to increase.

The results of the experiments indicate the ADP algorithm provides high-quality and robust solutions for BSB planners. An additional series of experiments tested the ADP solution technique’s efficacy on larger problem instances. Furthermore, using the best API-KLSTD algorithms for each fleet size $K = \{2, 3, 4\}$ provides another layer of algorithm robustness evaluation. The number of FOBs was increased to $B = 6$, and an additional layer of threat stochasticity was introduced. Figure 5 shows the scenario for the follow on experiments.

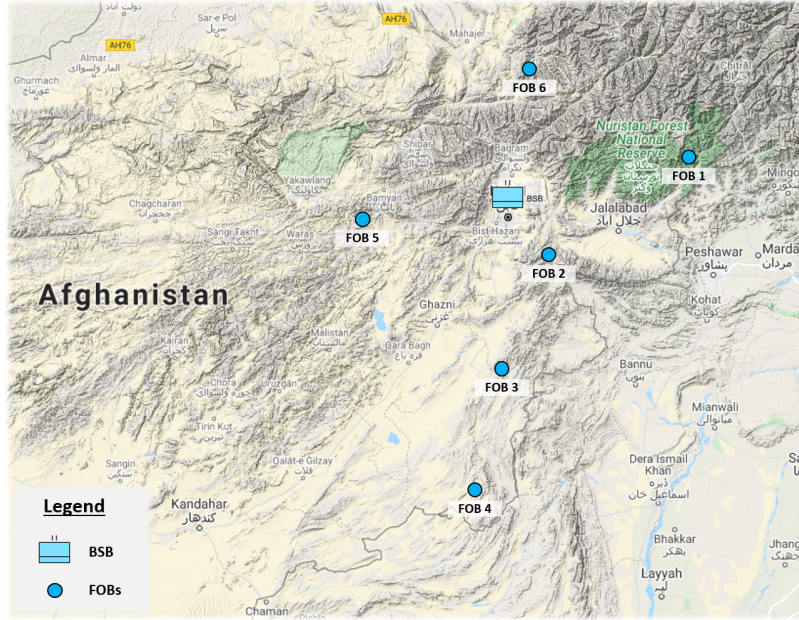


Figure 5. FOB Locations

The following series of simulations changes the threat maps based on a season-specific threat transition variable M_t^{Env} that follows a threat scenario depicted in

Figure 6. The variable M_t^{Env} represents a notional scenario wherein the threat maps and subsequent FOB threat levels again transition with probabilities informed by both the effects of weather and insurgent activity. This 365-day deployment scenario ran for 50 simulations.

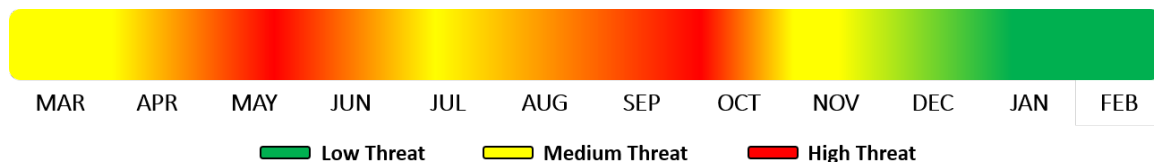


Figure 6. Threat Environment M_t^{Env} Scenario

With respect to percent improvement over the Benchmark Policy in maintaining supplies, the best performing ADP algorithm parameters within each set of K were tested in this experiment. The exponentially larger decision space warranted a supplemental experimental design to identify the appropriate ordinal optimization parameters. Holding all API-KLSTD algorithm parameters constant and maintaining the respective constant threat map levels of problem instances chosen, the sample size Z was considered for $\omega = 0.99, \psi = 0.95$ and for $\omega = \psi = 0.99$. Reviewing the same response variables to determine solution quality, Table 5 summarizes the results.

Table 5. Experimental Design for 6 FOBs

Problem Instance Parameter Settings				KLSTD-API Parameter Settings					Ordinal Optimization Parameter Settings		Daily Fleet OR%	Daily Fleet OR%	Daily Class I above Threshold for FOBs $\mathcal{B}'$	
Instance	K	M_t	B	N	$\mathcal{D}$	h	β	η	Z	Sampling Distribution	Computational Efficiency (min)	Benchmark Policy	ADP Policy	Improvement over Benchmark (%)
1	2	Mixed	6	$0.4 \cdot \mathcal{S} $	2	0.1	0.1	0.001	299	$\mathcal{U}(1, \mathcal{X}_{S_t})$	98.28	87.69 ± 0.81	99.24 ± 0.12	8.47 ± 0.88
2	3	Mixed	6	$0.3 \cdot \mathcal{S} $	5	0.1	0.8	0.1	459	$\mathcal{U}(1, \mathcal{X}_{S_t})$	166.69	85.95 ± 0.85	99.70 ± 0.03	5.95 ± 0.97
3	4	Mixed	6	$0.3 \cdot \mathcal{S} $	20	1	0.1	0.1	459	$\mathcal{U}(1, \mathcal{X}_{S_t})$	303.45	85.78 ± 0.83	99.99 ± 0.01	8.23 ± 0.79

The best percent improvement over the Benchmark Policy occurred in Problem Instance 1. However, in Problem Instance 3, a nearly flawless itinerary for maintaining the CUAV fleet was observed. This configuration comes at a cost of computational

effort, as more CUAVs generate an exponentially larger decision space for the algorithm to address. Acknowledging the ADP Policy outperforms the Benchmark Policy, reviewing the ADP Policy’s average supply values provide additional insight. Table 6 provides the Daily Fleet OR% and the average supply level across all FOBs with respect to the threshold levels over the 365-day horizon.

Table 6. ADP Policy for 6 FOBs

Problem Instance Parameter Settings				Daily Fleet OR%	Daily Class I above Threshold for FOBs $\mathcal{B}'$
Instance	K	M_t	B	ADP Policy	Average (1000 lbs)
1	2	M_t^{Env}	6	99.24 ± 0.12	13.09 ± 0.17
2	3	M_t^{Env}	6	99.70 ± 0.03	16.25 ± 0.27
3	4	M_t^{Env}	6	99.99 ± 0.01	17.43 ± 0.33

The set of experiments suggests shorter itineraries are more advantageous. A combination of direct deliveries and multi-route deliveries offers a healthier supply situation for the unit and better maintains the CUAV fleet. Furthermore, these experiments also provide insight into CUAV procurement options, given the problem settings. If a service component has the option of choosing between fewer CUAVs with greater capacity and more CUAVs with less capacity, the latter would prove more beneficial.

4.4.1 Summary of Testing Results

The results from the series of computational experiments show the API-KLSTD with ordinal optimization outperforms a set partitioning vehicle replacement problem benchmark policy in 13 out of the 15 explored problem instances. The advantage of such techniques guarantee safer itineraries for the CUAVs and maintain more supplies at the dislocated FOBs. Able to anticipate future states of the system, the ADP solution technique manages the stochastic threat levels better than the Benchmark

Policy by dispatching longer itineraries. The ADP technique performs better when the CUAV fleet is relatively smaller than the number of FOBs and threat levels at FOBs are higher.

The notional scenario suggests the ADP Policy outperforms the Benchmark Policy with a larger set of FOBs and a varying threat situation within the area of operation. Furthermore, this scenario provides insight into the potential procurement of CUAVs. Specifically, the size and type of CUAVs providing direct support to an IBCT was examined. A larger fleet of smaller CUAVs offers the best solution of maintaining fleet operational readiness and supply levels at the FOBs. However, this configuration comes with a large computational requirement. If time sensitive itineraries must be generated or cyclical, ad-hoc requirements are anticipated, a smaller fleet with more lift capacity will prove more benefit to the sustainment planners. Such comparisons can inform the design, analysis, and purchase of Department of Defense sponsored Unmanned Logistics System-Air program CUAVs.

V. Conclusions and Recommendations

This research examines a multi-route stochastic inventory routing problem (SIRP) within a military context. Typical IRPs focus on meeting customer demands while minimizing cost. However, such goals are not necessary for military planners and commanders. Routes possess inherent dangers, producing non-deterministic results for planned itineraries. Furthermore, demands are stochastic, leaving planners to construct route itineraries for fleets with information that is subject to change. Implementing CUAVs into resupply operations will inherently save lives and bolster the supply status of dispersed units at FOBs.

The intent of this research is to determine dispatching policies that improve the BSB's assigned CUAV fleet's operational readiness and the directed responsibility of managing demand across dislocated FOBs for an extended amount of time. A discounted, infinite-horizon Markov decision process model of the MIL SIRP-MR enables a thorough examination of problem features relating to different military scenarios. Although the MDP model provides an appropriate framework for the problem, the large size of the problem instances required the design and implementation of ADP techniques, joined with ordinal optimization to solve problems of interest within a reasonable amount of time. As such, the kernelized least squares temporal difference approach for policy evaluation leveraged the Mercer Theorem [45] to use the inner products and reduce computational requirements. Ordinal optimization allowed the ADP technique to increase computational efficiency by reducing the decision space. Applying these two methods within the ADP framework resulted in high-quality results for a set of motivated problems. A notional, representative scenario based on a 365-day deployment in eastern Afghanistan with seasonal threats demonstrated the applicability of the MDP model, the efficacy of the ADP solution techniques, and insight into CUAV procurement.

There are several areas for future research on the MIL SIRP-MR. A heterogeneous CUAV fleet may offer insights into other scenarios military planners may encounter. With the advancement and proven benefits of UAVs within the military, a sustainment planner may have to manage different types of CUAVs, constrained with varying levels of flight time, maintenance requirements, sensitivity to weather effects, and lift capacity. Another recommended change to the problem feature is the ability for CUAVs to deliver more than one type of supply class. This change would offer an additional layer of realism because resupply vehicles rarely deliver only a single class of supply. The proposed problem feature changes would undoubtedly increase the decision spaces to an intractable size of thousands to millions of possible decisions. Converting this space to the continuous domain may prove beneficial. Leveraging API-KLSTD produced promising results. Implementing kernel sparsification may yield faster convergence rates, which may allow for larger problem instances to be investigated.

Bibliography

1. Adelman, Daniel. A Price-Directed Approach to Stochastic Inventory/Routing. *Operations Research*, 52(4):499–514, 2004.
2. Archetti, Claudia, Luca Bertazzi, Gilbert Laporte, and Maria Grazia Speranza. A Branch-And-Cut Algorithm for a Vendor-Managed Inventory Routing Problem. *Transportation Science*, 41(3):382–391, 2007.
3. Berhan, Eshetie, Birhanu Beshah, Daniel Kitaw, and Ajith Abraham. Stochastic Vehicle Routing Problem: A Literature Survey. *Journal of Information & Knowledge Management*, 13(03):1450022–1–12, 2014.
4. Betson, Andrew P. The Case Against a Cargo Unmanned Aircraft System. *Army Sustainment*, 44(5), 2012.
5. Cherif-Khettaf, Wahiba Ramdane, Mais Haj Rachid, Christelle Bloch, and Pascal Chatonnay. New Notation and Classification Scheme for Vehicle Routing Problems. *RAIRO-Operations Research*, 49(1):161–194, 2015.
6. Coelho, Leandro C. and Gilbert Laporte. The Exact Solution of Several Classes of Inventory Routing Problems. *Computers & Operations Research*, 40(2):558–565, 2013.
7. Coelho, Leandro C., Jean-François Cordeau, and Gilbert Laporte. Thirty Years of Inventory Routing. *Transportation Science*, 48(1):1–19, 2014.
8. Cordeau, Jean-François, Michel Gendreau, Gilbert Laporte, Jean-Yves Potvin, and Frédéric Semet. A Guide to Vehicle Routing Heuristics. *Journal of the Operational Research Society*, 53(5):512–522, 2002.
9. Couch, Mark and Dennis Lindell. Study on Rotorcraft Survivability. *Joint Aircraft Survivability Program Office*, Summer, 2010.
10. Dantzig, George B. and John H. Ramser. The Truck Dispatching Problem. *Management Science*, 6(1):80–91, 1959.
11. Davies, Alex. The Marines’ Self-Flying Chopper Survives a Three-Year Tour. *Wired*, 2014.
12. Dennard, Kerry, Christine Haffey, and Ray Ferguson. 45th Sustainment Brigade: Supply Distribution in Afghanistan. *Army Sustainment*, 42(6):12–14, 2010.
13. Department of Defense. *Unmanned Systems Integrated Roadmap*. Office of the Under Secretary of Defense for Acquisition and Sustainment, Washington, D.C., 2017.

14. Department of the Army. *Brigade Combat Team*. FM 3-96. Washington, D.C., 2015.
15. Department of the Army. *Sustainment Operations*. FM 4-0. Washington, D.C., 2019.
16. Department of the Army. *Army Geospatial Guide for Commanders and Planners*. TC 3-34.80. Washington, D.C., 2019.
17. Department of the Army. *Brigade Support Battalion*. ATP 4-90. Washington, D.C., 2020.
18. Gehring, Clement, Yangchen Pan, and Martha White. Incremental Truncated LSTD. 2015. URL <https://arxiv.org/pdf/1511.08495.pdf>.
19. Ho, Yu-Chi, Ramavarapu Sreenivas, and Pirooz Vakili. Ordinal Optimization of DEDS. *Discrete Event Dynamic Systems*, 2(1):61–88, 1992.
20. Johnson, Donald B. A Note On Dijkstra’s Shortest Path Algorithm. *Journal of the ACM (JACM)*, 20(3):385–388, 1973.
21. Kleywegt, Anton J., Vijay S. Nori, and Martin W. Savelsbergh. The Stochastic Inventory Routing Problem With Direct Deliveries. *Transportation Science*, 36(1):94–118, 2002.
22. Kleywegt, Anton J., Vijay S. Nori, and Martin W. Savelsbergh. Dynamic Programming Approximations for a Stochastic Inventory Routing Problem. *Transportation Science*, 38(1):42–70, 2004.
23. Laporte, Gilbert. The Vehicle Routing Problem: An Overview of Exact and Approximate Algorithms. *European Journal of Operational Research*, 59(3):345–358, 1992.
24. McCormack, Ian M. The Military Inventory Routing Problem with Direct Delivery. Master’s thesis, Air Force Institute of Technology, Wright Patterson Air Force Base, Ohio, March 2014.
25. McCullough, Amy. Saving Outpost Keating. *Air Force Magazine*, 93(10), 2010.
26. McKenna, Rebekah S. Using Approximate Dynamic Programming to Solve the Military Inventory Routing Problem with Direct Delivery. Master’s thesis, Air Force Institute of Technology, Wright Patterson Air Force Base, Ohio, March 2015.
27. McKenna, Rebekah S., Matthew J. Robbins, Brian J. Lunday, and Ian M. McCormack. Approximate Dynamic Programming for the Military Inventory Routing Problem. *Annals of Operations Research*, 288(1):391–416, 2020.

28. Minkoff, Alan S. A Markov Decision Model and Decomposition Heuristic for Dynamic Vehicle Dispatching. *Operations Research*, 41(1):77–90, 1993.
29. Montgomery, Douglas C. *Design and Analysis of Experiments (9th edition)*. John Wiley & Sons, Incorporated, New Jersey, 2017.
30. Mu, Qianxin, Zhuo Fu, Jens Lygaard, and Richard Eglese. Disruption Management of the Vehicle Routing Problem With Vehicle Breakdown. *Journal of the Operational Research Society*, 62(4):742–749, 2011.
31. O’Brien, Henry. US Army Futures Command Sustainment (CDID). Personal Correspondence. June 30, 2020.
32. Parson, Carl R. Approximate Dynamic Programming for Military Resource Allocation. Master’s thesis, Air Force Institute of Technology, Wright Patterson Air Force Base, Ohio, March 2014.
33. Peterson, Troy M. and Jason R. Staley. Business Case Analysis of Cargo Unmanned Aircraft System (UAS) Capability in Support of Forward Deployed Logistics in Operation Enduring Freedom (OEF). Master’s thesis, Naval Postgraduate School, Monterey, CA, December 2011.
34. Pillac, Victor, Michel Gendreau, Christelle Guéret, and Andrés L. Medaglia. A Review of Dynamic Vehicle Routing Problems. *European Journal of Operational Research*, 225(1):1–11, 2013.
35. Puterman, Martin L. *Markov Decision Processes: Discrete Stochastic Dynamic Programming*. Wiley-Interscience, New Jersey, 2005.
36. Ruszczyński, Andrzej. Commentary—Post-Decision States and Separable Approximations Are Powerful Tools of Approximate Dynamic Programming. *INFORMS Journal on Computing*, 22(1):20–22, 2010.
37. Salgado, Ehtan L. Using Approximate Dynamic Programming to Solve the Stochastic Demand Military Inventory Routing Problem with Direct Delivery. Master’s thesis, Air Force Institute of Technology, Wright Patterson Air Force Base, Ohio, March 2016.
38. Solyalı, Oğuz and Haldun Sürat. The One-Warehouse Multi-Retailer Problem: Reformulation, Classification, and Computational Results. *Annals of Operations Research*, 196(1):517–541, 2012.
39. Toner, Ellen. Profile of the United States Army. *The Institute of Land Warfare at the Association of the United States Army*, 2018.
40. Toth, Paolo, Daniele Vigo, Paolo Toth, and Daniele Vigo. *Vehicle Routing: Problems, Methods, and Applications (2nd edition)*. Society for Industrial and Applied Mathematics, Philadelphia, PA, 2014.

- 41. Trudeau, Pierre and Moshe Dror. Stochastic Inventory Routing: Route Design with Stockouts and Route Failures. *Transportation Science*, 26(3):171–184, 1992.
- 42. Tu, Stephen and Benjamin Recht. Least-Squares Temporal Difference Learning for the Linear Quadratic Regulator. In *International Conference on Machine Learning*, pages 5005–5014. PMLR, 2018.
- 43. Ulmer, Marlin W. *Approximate Dynamic Programming for Dynamic Vehicle Routing*, volume 61 of *Operations Research/Computer Science Interface Series*. Springer International Publishing, Cham, Switzerland, 2017.
- 44. Wikipedia.Org. Combat Outpost Keating, 30 September 2010. URL https://en.wikipedia.org/wiki/Combat_Outpost_Keating.
- 45. Xu, Xin, Tao Xie, Dewen Hu, and Xicheng Lu. Kernel Least-Squares Temporal Difference Learning. *International Journal of Information Technology*, 11(9):54–63, 2005.

REPORT DOCUMENTATION PAGE					Form Approved OMB No. 0704-0188	
The public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden to Department of Defense, Washington Headquarters Services, Directorate for Information Operations and Reports (0704-0188), 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to any penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number. PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ADDRESS.						
1. REPORT DATE (DD-MM-YYYY)		2. REPORT TYPE		3. DATES COVERED (From — To)		
25-03-2021		Master's Thesis		August 2019 — March 2021		
4. TITLE AND SUBTITLE Resupply Operations of a Dispersed Infantry Brigade Combat Team using Approximate Dynamic Programming				5a. CONTRACT NUMBER		
				5b. GRANT NUMBER		
				5c. PROGRAM ELEMENT NUMBER		
6. AUTHOR(S) Song, Camero K., MAJ, USA				5d. PROJECT NUMBER		
				5e. TASK NUMBER		
				5f. WORK UNIT NUMBER		
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Air Force Institute of Technology Graduate School of Engineering and Management (AFIT/EN) 2950 Hobson Way WPAFB OH 45433-7765				8. PERFORMING ORGANIZATION REPORT NUMBER AFIT-ENS-MS-21-M-184		
9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES) Mr. David M. Panson Strategic Development Planning & Experimentation (SDPE) Office 1864 4th Street Wright-Patterson AFB, OH 45433 (937) 904-6539				10. SPONSOR/MONITOR'S ACRONYM(S) SDPE		
				11. SPONSOR/MONITOR'S REPORT NUMBER(S)		
12. DISTRIBUTION / AVAILABILITY STATEMENT DISTRIBUTION STATEMENT A: APPROVED FOR PUBLIC RELEASE: DISTRIBUTION UNLIMITED.						
13. SUPPLEMENTARY NOTES This work is declared a work of the U.S. Government and is not subject to copyright protection in the United States.						
14. ABSTRACT Military sustainment planners must consider the effective employment of Cargo Unmanned Aerial Vehicles (CUAVs) into resupply operations under austere threat conditions. Viewed as an inventory routing problem, the complexities of resupplying dislocated Forward Operating Bases (FOBs) under dangerous and dynamic conditions warrant additional features to the baseline problem. This research models a military stochastic inventory routing problem with multiple routing (MIL SIRP- MR) as a Markov decision process (MDP) model and provides high-quality policy solutions using a combination of approximate dynamic programming and ordinal optimization techniques.						
15. SUBJECT TERMS inventory routing problem, approximate dynamic programming						
16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON	
a. REPORT	b. ABSTRACT	c. THIS PAGE			Dr. Matthew J. Robbins, AFIT/ENS	
U	U	U	UU	56	19b. TELEPHONE NUMBER (include area code) (937) 255-3636, x4606; matthew.j.robbs@afit.edu	