6-16-2016

Using Approximate Dynamic Programming to Solve the Stochastic Demand Military Inventory Routing Problem with Direct Delivery

Ethan L. Salgado

Using Approximate Dynamic Programming to Solve the Stochastic Demand Military Inventory Routing Problem with Direct Delivery

THESIS

JUNE 2016

Ethan L. Salgado, Second Lieutenant, USAF

AFIT-ENS-MS-16-J-031

DEPARTMENT OF THE AIR FORCE
AIR UNIVERSITY

AIR FORCE INSTITUTE OF TECHNOLOGY

Wright-Patterson Air Force Base, Ohio

AFIT-ENS-MS-16-J-031

USING APPROXIMATE DYNAMIC PROGRAMMING TO SOLVE THE STOCHASTIC DEMAND MILITARY INVENTORY ROUTING PROBLEM WITH DIRECT DELIVERY

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 Operations Research

Ethan L. Salgado, BS

Second Lieutenant, USAF

JUNE 2016

AFIT-ENS-MS-16-J-031

USING APPROXIMATE DYNAMIC PROGRAMMING TO SOLVE THE STOCHASTIC DEMAND MILITARY INVENTORY ROUTING PROBLEM WITH DIRECT DELIVERY

Ethan L. Salgado, BS
Second Lieutenant, USAF

Committee Membership:

Lt Col Matthew J. Robbins, PhD
Chair

Jeffery D. Weir, PhD
Member

AFIT-ENS-MS-16-J-031

Abstract

A brigade combat team must resupply forward operating bases (FOBs) within its area of operations from a central location, mainly via ground convoy operations, in a way that closely resembles vendor managed inventory practices. Military logisticians routinely decide when and how much inventory to distribute to each FOB. Technology currently exists that makes utilizing cargo unmanned aerial vehicles (CUAVs) for resupply an attractive alternative due to the dangers of utilizing convoy operations. However, enemy actions, austere conditions, and inclement weather pose a significant risk to a CUAV's ability to safely deliver supplies to a FOB. We develop a Markov decision process model that allows for multiple supply classes to examine the military inventory routing problem, explicitly accounting for the possible loss of CUAVs during resupply operations. The large size of the motivating problem instance renders exact dynamic programming techniques computationally intractable. To overcome this challenge, we employ approximate dynamic programming (ADP) techniques to obtain high-quality resupply policies. We employ an approximate policy iteration algorithmic strategy that utilizes least squares temporal differencing for policy evaluation. We construct a representative problem instance based on an austere combat environment in order to demonstrate the efficacy of our model formulation and solution methodology. Because our ADP algorithm has many tunable features, we perform a robust, designed computational experiment to determine the ADP policy with the best quality of solutions. Results indicate utilizing least squares temporal differences with a first-order basis function is insufficient to approximate the value function when stochastic demand and penalty functions are implemented.

Keywords: approximate dyanmic programming, Markov decision process, ven–

dor managed inventory, vehicle routing, military inventory routing (MILIRP), least squares temporal differences

For my wife. Her continual support strengthened me daily.

Acknowledgements

I would like to express my sincere gratitude to my advisor, Dr. J.D. Robbins, for all his help. This would not have been possible without his support.

Ethan L. Salgado

Table of Contents

List of Figures

List of Tables

USING APPROXIMATE DYNAMIC PROGRAMMING TO SOLVE THE STOCHASTIC DEMAND MILITARY INVENTORY ROUTING PROBLEM WITH DIRECT DELIVERY

I. Introduction

United States military logistical planners must consider the timing, routing, and supply configuration of distribution assets when preparing and executing routine re–supply missions (i.e., distribution, replenishment, or sustainment operations) in sup–port of brigade combat team (BCT) operations. The brigade support battalion (BSB) is the primary organization within the BCT that plans, coordinates, synchronizes, and executes sustainment operations. Sustainment operations typically involve the establishment of a brigade support area (BSA) as the distribution center from which supplies of various classes are delivered to company- and platoon-sized units located at forward operating bases (FOBs) geographically dispersed throughout the BCT's area of operation [10]. Logistical planners at the BSB monitor the supply levels of the FOBs utilizing logistics situation reports and automated sustainment data-gathering systems such as the Battle Command Sustainment Support System, and the Force XXI Battle Command Brigade and Below logistical support system [10]. As such, the BSB knows the inventory level at all of the FOBs when making inventory routing decisions. At the beginning of each day the BSB must decide which FOBs to resupply, how much of each supply class to deliver to each FOB, how to combine FOBs (i.e., customers) into routes, and which routes to assign to each of the available delivery assets.

The BCT is the primary combined arms force that executes decisive actions for

the Army. The BCT supports offensive, defensive, stability and Defense Support of Civil Authorities tasks [10]. The BSB operations are accomplished by planning and executing missions within the context of the sustainment warfighting function and by applying the principles of sustainment when executing the support of decisive actions. The objective of sustainment in a wartime environment is to provide sufficient support to enable the BCT to conduct its four primary tasks when necessary: movement to contact, attack, exploitation, and pursuit [10].

Distribution assets that move supplies from the BSA to the FOBs include both ground assets (e.g., medium- and heavy-capacity cargo trucks and tanker trucks) and aerial assets (e.g., the CH-47 Chinook helicopter). While distribution via ground asset accounts for the majority of tonnage delivered, aerial delivery distribution provides an effective means of conducting distribution operations because it bypasses casualty-inducing enemy activities and reduces the need for route clearance of ground lines of communications.

Aerial resupply does have its own risks that must be independently considered. North Atlantic Treaty Organization military forces must account for adversaries with the capability and intent to oppose and disrupt allied aerial assets [14]. Threat levels for aerial assets are classified based on the availability, accessibility, and probability of attack. Among the threats, man-portable air-defense systems (MANPADS) are already highly proliferated with an estimated $500,000$ to $750,000$ licensed units worldwide [14]. MANPADS are particularly effective against low or slow aircraft which makes rotary wing assets particularly vulnerable during take-off and landing.

Military logistical planners face many important challenges when making daily inventory routing decisions in a combat environment. Poorly developed transportation infrastructure, adverse weather conditions, terrain, enemy threat and actions, and the availability of distribution assets all inhibit successful distribution of sup–

plies from the BSA to the FOBs. Moreover, insurgent use of improvised explosive devices (IEDs) greatly affects truck mobility throughout the operational environment and has been successful in disrupting replenishment procedures [24]. This is concerning because current resupply efforts operate mainly via convoys. The probability of successful distribution to the FOBs must be considered before a resupply decision can be made. Logistical planners at the BSB must consider what supplies (e.g., water, food, fuel, ammunition) should be sent and how much is required. Limiting factors may include distribution asset availability, convoy maintenance requirements, and current threat locations. Constantly evolving socio-political factors may cause a rapid change in current threat areas in the operational environment. Moreover, wartime logistics often do not have a short-term horizon, so logisticians must plan for sustainable resupply over an indefinite horizon.

The United States Department of Defense is interested in the design, development, and utilization of cargo unmanned aerial systems (CUAS) for resupply operations. A CUAS is the collection of all components required to allow the operation of a CUAV. A CUAS includes the operating crew (maintenance crew and pilots), required software, ground station, and the CUAVs. The United States Army intends to increase the utilization of CUAVs as an integral component of integrated logistics aerial resupply. As such, examination of inventory routing decisions for CUAS across an austere combat environment is needed.

In this thesis, we consider the military inventory routing problem (MILIRP) wherein the BSB must simultaneously decide how to route and configure CUAVs to fulfill FOB supply needs. We develop a Markov decision process (MDP) model of the MILIRP. The high-dimensionality of the state and action space renders classical dynamic programming methods computationally intractable. Thus, we apply approximate dynamic programming (ADP) techniques to obtain high-quality inven–

tory routing policies. We construct an approximate policy iteration (API) algorithm that utilizes least-squares temporal difference (LSTD) learning for policy evaluation. To demonstrate the efficacy of our proposed solution methodology, we construct a notional, representative planning scenario based on an austere combat environment like that of Afghanistan. Because our ADP algorithm has many tunable features, we perform a robust designed computational experiment to identify the ADP policy with the best quality of solutions.

The unique military aspect of the MILIRP warrants further discussion. In contrast to much of the previous work on the inventory routing problem (IRP), we explicitly account for the possible destruction of our delivery vehicles. We must model the evolution of threat and weather and their attendant impact on the likelihood of CUAV delivery success. We must also model the permanent impact CUAV destruction has on the resupply operations over an indefinitely long horizon. Moreover, in a combat environment the military does not take into account various external costs commonly associated with IRPs. Thus, the MILIRP objective function focuses on total amount of supplies delivered over the life of the system and does not consider holding, ordering, or transportation costs.

The remainder of this thesis is organized as follows. Chapter II presents a review of relevant literature concerning vendor managed inventory practices and the inventory routing problem. We also review several ADP papers to inform the development of our solution methodology. Chapter III provides a description of the MILIRP and introduces our methodology, including our MDP formulation and ADP solution method. Chapter IV presents our computational results and analysis. We perform a designed experiment on problem and algorithmic specific features to obtain the best quality solution. Chapter V provides conclusions and directions for future research.

II. Literature Review

Our literature review focuses on two areas of research pertinent to our problem formulation and solution methodology. The first is the inventory routing problem (IRP) which has been widely researched. The second area of interest is approximate dynamic programming (ADP).

2.1 Inventory Routing Problem:

The IRP is an optimization problem wherein inventory is sent from a supplier to a customer across a set of locations. The IRP is a natural evolution from the vehicle routing problem (VRP) and is an area of research that has been throughly studied in the operations research field because of the constant need to improve supply chain logistics. The IRP integrates inventory management, vehicle routing, and delivery scheduling decisions. Inventory routing has been a topic of research in the operational research field for over 30 years [7]. The IRP arises from the idea of vendor managed inventory (VMI) replenishment, a centralized approach to inventory management used to reduce overall costs.

VMI replenishment is a business practice in which the vendor monitors inventory levels of the customers. Conversely, in traditional inventory management, customers keep track of their own inventory and determine when and how much to order from the supplier; The vendor (i.e., supplier) receives orders and uses its vehicles to fill the demand. VMI replenishment is an attractive alternative because it is a mutually beneficial relationship between the supplier and the customer; the supplier reduces transportation costs by deciding when and how much inventory to distribute to each customer and the customer reduces costs by not allocating resources to monitoring

inventory scheduling. There are three main advantages to utilizing VMI practices [15]. First, VMI may lead to reduced production and inventory costs by reducing variation in orders and obtaining a more uniform utilization of resources for both the supplier and the customer. Second, proactive rather than reactive planning used in VMI may reduce transportation costs beyond that of more uniform utilization alone. It may be possible to increase low-cost full truckload shipments and decrease the frequency of high-cost less than full truckload shipments. Moreover, it may be possible to use more efficient routes by coordination of replenishment at different customers close to each other. Third, VMI may increase service levels, measured in terms of reliability of product availability.

There are two requirements necessary to obtain the benefits of VMI: the avail–ability of relevant, accurate, and timely data for the decision maker and the ability of the central decision maker to use increased amount of information to make good decisions [15]. To succeed in VMI, an organization must not only have access to rele–vant information such as current and past inventory levels at all customers, customer demand behavior, and customer location relative to vendor and each other, but they must also have the ability to utilize that data in the construction of relevant and use–ful distribution policies. This is a very complex task and many failures to implement VMI are a direct result of failing to meet one or both of the above requirements [15]. While a responsible vendor implementing VMI can save both time and money, misuse of VMI business practices can result in lost sales and revenue. Understanding of VMI practices builds the knowledge base necessary to understand the IRP.

The IRP falls into a class of problems called NP-hard, meaning they are at least as hard as the hardest non-deterministic polynomial-time problems [17]. The IRP is inherently difficult because it subsumes the classical VRP. A supplier must make three simultaneous decisions [7]: 1) when to serve a given customer, 2) how much

to deliver to this customer when it is served, and 3) how to combine customers into vehicle routes. A basic IRP seeks to minimize total inventory-distribution costs while meeting demand of each customer subject to the following constraints: inventory at each customer can never exceed its maximum capacity, inventory levels are not allowed to be negative, the supplier's vehicles can perform at most one route per time period with each route starting and ending at the supplier, and vehicle capacities cannot be exceeded.

Common problem features that describe IRPs include [7]: time horizon, structure, routing, inventory policy, inventory decisions, fleet composition, and fleet size. Within the IRP framework, time horizon is a problem dependent feature that can either be finite or infinite. With respect to structure, the number of suppliers and customers can be categorized as follows: *one-to-one* when there is only one supplier and one customer, *one-to-many* when there are many customers, or more rarely, *many-to-many*. Routing can be *direct* when there is only one customer per route, *multiple* when there are several customers in the same route, or *continuous* when there is no central depot (e.g., as seen in maritime applications). Direct delivery greatly simplifies the IRP by removing the optimization of the routing portion of the problem. Direct delivery involves the vehicle moving directly from the supplier to the customer and returning to the vendor immediately after delivery. Direct delivery is appropriate for our application of the MILIRP due to the current CUAV maximum loaded range of under 400km [18]. The two most common inventory polices are the *maximum level* and *order-up-to level* (OL) policies. Maximum level policies allow flexibility in deciding the amount to refill whereas in OL policies, the supplier always replenishes a customer to full capacity each time the customer is visited. Inventory decisions can include lost sales when excess demand becomes lost revenue, or back-orders when demand can be filled at a later date. Fleet composition can either be *homogeneous*

or *heterogeneous* while fleet size can be single, limited, or unconstrained.

Coelho *et al.* [7] and Toth & Vigo [29] give a basic introduction to the stochastic variant of the basic IRP. In the stochastic inventory routing problem (SIRP), the sup–plier knows the customer demand only in a probabilistic sense. Demand stochasticity means shortages may occur. In order to discourage shortages, a penalty function is imposed whenever a customer runs out of stock and is usually modeled as unsatis–fied demand. With no backlogging, unsatisfied demand is considered lost. There are several methods employed to solve IRP which include but are not limited to heuristic algorithms, link optimization, simulation, and dynamic programming. See Table 1 for a summary of relevant papers on the SIRP.

Campbell *et al.* [6] and Minkoff [22] formulate their SIRP in a similar fashion. They both model the use of an unconstrained fleet (in terms of size) to meet demand across their network, additionally allowing for multiple routing. While Campbell *et al.* [6] did not present specific analysis for their SIRP formulation, they provided challenging test instances of the IRP. Minkoff [22] applied a heuristic approach to solving the SIRP based on a decomposition of the problem by customer. The solution to the customer subproblems generated the penalty functions applied to their master dispatching problem.

Adelman [1] and Kleywegt *et al.* [16] provide very similar SIRP formulations. They both solve infinite horizon problems with a one-to-many structure. While their solution methodologies differ, they both focused on multiple routing and maximum level inventory policies. They both employ homogeneous fleet composition without backlogging and with a fixed, limited fleet. Adelman [1] differs from Kleywegt *et al.* [16] in that he uses linear programming techniques to obtain his solution whereas Kleywegt *et al.* [16] use ADP.

Two papers deserve more in-depth discussion because they both greatly informed

our research, and closely resemble our work in both problem formulation and solution methodology.

Kleywegt *et al.* [15] model their IRP as a direct delivery, limited fleet application with stochastic demand and an unrestricted vehicle supply. Similarly, Kleywegt *et al.* [16] model their IRP with multiple routing, limited fleet with stochastic demand, and deterministic vehicle supply. Our formulation is similar to both of these models in that we use infinite time horizon, one-to-many structure, maximum level inventory policy, no backlogging, and a homogeneous fleet composition. While we are similiar to Kleywegt *et al.* [15] in that we employ direct delivery, we are more similar to Kleywegt *et al.* [16] in that our fleet size is constrained. Our formulation differs from either paper in that the distinct military nature of our formulation yields a stochastic vehicle supply. The stochastic nature of our vehicle supply is discussed in more detail in Section 3.1.

Kleywegt *et al.* [15] and Kleywegt *et al.* [16] both employ ADP as a solution tech–nique. Kleywegt *et al.* [15] employ an approximate policy iteration (API) algorithmic strategy utilizing a parametric value function approximation. They construct a set of basis functions to create a linear architecture around the pre-decision state. Kley–wegt *et al.* [16] adopt the same ADP approach for the first part of their optimization problem, before considering multiple delivery, then use a heuristic search method to determine additional delivery opportunities afterwards, if possible. Several differences exist that distinguish our problem from theirs. First, we focus our value function ap–proximation around the post-decision state. Second, the stochastic nature of vehicle supply is a distinguishing feature unlike other IRP in the current literature.

Table 1. Classification of Relevant Stochastic IRP Papers

Reference	Routing	Fleet Size
Adelman [1]	Multiple	Multiple
Berman & Larson [2]	Multiple	Single
Campbell *et al.* [6]	Multiple	Unconstrained
Kleywegt *et al.* [15]	Direct	Unconstrained
Kleywegt *et al.* [16]	Multiple	Multiple
Minkoff [22]	Multiple	Unconstrained

2.2 Approximate Dynamic Programming

Inventory routing decisions in a combat environment involves sequential decision making under uncertain conditions. Due to enemy threats, the routing of a cargo unmanned aerial vehicle (CUAV) to replenish supplies has an uncertain outcome. The loss of a CUAV impacts the ability of the Brigade Supply Battalion (BSBs) to replenish supplies in the future. Thus, we must account for the safety of our CUAV in our formulation. We formulate the military inventory routing problem (MILIRP) as a Markov decision process (MDP). However, due to the high dimensionality of this problem, it is unable to be solved exactly using dynamic programming techniques. To overcome the curse of dimensionality, we implement an ADP methodology to solve the MILIRP. ADP is being concurrently developed by multiple different communities to include engineering controls, computer science (artificial intelligence), and operations research. For a more detailed introduction to ADP from an operations research perspective we refer the reader to Powell [25, 26, 27]. For a different ADP outlook, we refer the reader to Bertsekas & Tsitsiklis [4] (engineering control theory) or Sutton & Barto [28] (artificial intelligence).

Two ADP algorithmic strategies exist for obtaining approximate solutions to our stochastic optimization problem: approximate value iteration (AVI) and API. Al–though we will not spend time introducing API here, the interested reader may read Bertsekas [3] for a more detailed discussion. We chose API as our algorithmic strat–

egy to obtain our mapping of a state to an action where our state includes inventory levels and our actions include when and how much inventory to send to each location. In general, there exists four classes of policies: myopic cost function approximation, lookahead policies, policy function approximations, and value function approximation policies [26]. Our approximation strategy involves using the post-decision state to construct a linear architecture based on an appropriate set of basis functions. Van Roy *et al.* [30] was the first to introduce post-decision state approximation as a way to modify Bellman's equation to obtain an equivalent, deterministic expression. Using the post-decision state is useful because it reduces what Powell [26] refers to as the the outcome state portion of the curse of dimensionality. API consists of two basic steps: policy improvement and policy evaluation. Within the policy improvement step of our API algorithm, we update the value function approximation for a fixed policy using least squares temporal differencing (LSTD). Bradtke & Barto [5] introduced LSTD as a computationally efficient method for estimating the adjustable parameters when using a linear architecture with fixed basis functions to approximate the value function for a fixed policy. LSTD updates its estimate of the expected contribution and projects this over the infinite horizon [26]. We implement a variant of the LSTD algorithm that utilizes post-decision state value function approximations.

III. Methodology

This section describes the Markov decision process (MDP) model formulation of the military inventory routing problem (MILIRP) with direct delivery. We also present the approximate dynamic programming (ADP) methodology utilized to ob–tain high quality solutions to the MILIRP.

3.1 Problem Description

A brief discussion of the U.S. Army replenishment structure provides context for the MILIRP. The brigade combat team (BCT) is the highest echelon organization able to act independently in regional combat operations. The BCT is responsible for the forward operating bases (FOBs) within its area of operation. A sub-organization within the BCT, called a brigade support battalion (BSB), is responsible for the replenishment of the FOBs. The interaction between the BSB and the FOBs parallel the supplier-to-customer relationship seen in vendor managed inventory practices.

The BSB plans, coordinates, synchronizes, and executes replenishment operations in support of brigade combat teams operations [11]. The BSB is the organization within the BCT that establishes and operates the brigade support area (BSA), a central location utilized to resupply its customers (i.e., FOBs) at locations of varying distances. The BSB is responsible for the periodic resupply of the BCT's subordinate units, which closely mirrors vendor managed inventory (VMI) practices used in the private sector. To accomplish its resupply missions, the BSB is kept informed of inventory levels at the FOBs through regular reporting and automated data systems. VMI practices allow the BSB to decide when, where, and how much supplies to send to FOBs. The routing and resupply operations of the BSB can be formulated as a variant of the inventory routing problem (IRP).

Table 2. U.S. Army supply classes (*denotes classes delivered by the BSB in combat)

U.S. Army Supply Classes	
I.	*Subsistence
II.	*Clothing, individual equipment
III.	*Fuels, lubricants/fluids
IV.	*Construction materials
V.	*Ammunition
VI.	Personal demand items
VII.	*Major end items (tanks, vehicles etc.)
VIII.	*Medical supplies
IX.	*Repair parts
X.	Non-military programs material

Our formulation of the MILIRP includes multiple supply classes to improve model realism, as compared to previous efforts (e.g. McCormack [19], McKenna [21], McKenna *et al.* [20]). Thus, the Army's definition of supply classes deserves more attention. The U.S. Army defines ten different supply classes as indicated in Table 2 [8].

In a combat environment, however, the BSB cannot provide all ten classes; the BSB is restricted to only provide classes I, II, III, IV, V, VII, VIII, and IX to a FOB [10]. Replenishment during combat operations includes difficult, deliberate, and time-sensitive resupply missions conducted to provide forward companies with essential supplies to sustain the pace of operations [11]. The U.S. Army employs trucks, manned air assets, and now cargo unmanned aerial vehicles (CUAVs) to perform replenishment (i.e., distribution) operations. Utilization of these distribution assets requires deliberate logistical planning.

Similar to the number of drivers available to operate a fleet of trucks, the cargo unmanned aerial system (CUAS) has additional limiting factors that must be considered. The CUAS is a complex system because of the many factors required for successful operation: remote pilot (for emergency or combat purposes), maintenance requirements, maintenance crew, aircraft fuel, required software, and the CUAVs. In this thesis, we consider two features of the CUAS, the number of CUAVs and the size

of the crew. We refer to *crew* as all other factors required for CUAS operation other than the CUAVs themselves.

One important complicating feature of the MILIRP that has yet to be discussed is vehicle destruction. Resupply efforts pose a substantial risk to personnel in combat environments due to the harsh and rugged environments in which the Army typically operates. The CUAV is most vulnerable to man-portable air-defense systems (MAN–PADS) and small arms fire during takeoff and landing operations at FOBs. As such, successful delivery of supplies is not guaranteed when the BSB makes a CUAV resup–ply routing decision. For this reason, the MILIRP necessarily takes into account the stochasticity of routing decisions. Moreover, CUAV routing decisions are influenced by the current threat conditions because a destroyed CUAV cannot be replaced; the loss of a CUAV has a permanent impact on the ability of the BSB to deliver supplies in the future. We assume no CUAV replacement because of the logistical cost required to replace a CUAV. Each BCT deploys with its own organic CUAS contingent. Once lost, the CUAV is not replaced until the next BCT arrives in theater.

Lack of transportation infrastructure within the BCT's area of operations and enemy aggression make resupply via ground transport inherently dangerous. Impro–vised explosive devices caused 18% of all deployed fatalities between November 2002 and March 2009, all occurring during sustainment operations [12]. If CUAV resup–ply is unable to meet supply requirements, FOBs must be supplied through ground convoy operations. Due to the human capital exposure to risk necessary to resupply FOBs via ground convoy, we impose a penalty on the system if the CUAS is unable to fulfill FOB supply requirements.

3.2 MDP Formulation

The objective of the MILIRP is to determine the optimal resupply of forward operating bases (FOBs) via inventory routing and cargo configuration decisions in order to maximize expected total discounted reward over an infinite horizon. The reward function maintains increasing monotonicity with respect to supplies delivered to FOBs until it reaches the FOB's maximum holding capacity after which additional supplies delivered yield no reward. We assume all inventory levels at each FOB are known at the start of each period and that demand for each supply class has a known historical average with some variability modeled as an independent and identically distributed error term. Inherent in this formulation is the assumption that no other external event (e.g., enemy action, fire, expiration of supplies) other than demand causes a loss of inventory.

In the MILIRP, a brigade combat team (BCT) is responsible for B FOBs within its area of operation. The BCT contains a brigade support battalion (BSB) which manages resupply efforts for N supply classes for each FOB. The BSB distributes supplies to the B FOBs utilizing U identical cargo unmanned aerial vehicles (CUAVs). Each CUAV has an identical load capacity of H tons. FOB $i = 1, 2, ..., B$ requires $\hat{D}_{in}$ tons of supplies per time period for supply class $n = 1, 2, ..., N$, a stochastic demand with a mean demand $\bar{d}_{in}$ and an independent and identically distributed exogenous error term $\hat{\epsilon}_{in}$. Each FOB also has a finite maximum holding quantity Q_{in} for each supply class. A total of M threat maps captures the threat conditions in the BCT's area of operation.

Given an austere combat environment, there is potential for delivery failure due to extrinsic uncontrollable factors (e.g., enemy action, mechanical failure, extreme weather conditions). We propose a tessellation of the area of operations with each hexagonal cell identified as a high or low threat area as done in other work [19].

The probability of a CUAV being destroyed depends on the FOB being supplied and the current threat conditions. The set of M threat maps models the periodic changes in risk throughout the area of operation. Dijkstra's algorithm is applied to determine an optimal path that minimizes risk when traveling from the BSB to each FOB $i = 1, 2, ..., B$ for each tessellated threat map $m = 1, 2, ..., M$. Under threat map m, the parameter ψ_{im} denotes the probability of a one-way successful trip to FOB i (and back again). A CUAV may be destroyed either on its way to a FOB or after delivering supplies on the return route back to the depot at the brigade support area (BSA).

We proceed by describing the MDP model formulation of the MILIRP. With respect to a conventional inventory routing formulation, CUAVs are vehicles, FOBs are customers, and the centralized BSB is the supplier. Table 3 located at the end of this chapter provides a summary of notation.

The MILIRP is formulated as an infinite time horizon problem where $t \in \mathcal{T} = \{1, 2, ...\}$. During each time period a CUAV is fueled, supplied, maintained, travels to the assigned FOB, unloads, and returns to the BSB. It is assumed that all FOBs are within the CUAV's range when fully loaded and that this route is serviceable in one time period. Current CUAV limitations validate this assumption [18].

The state space includes three components: the inventory level at each FOB, the number of operational CUAVs, and the threat map index number. The inventory at each FOB is defined as $R_t = (R_{ti})_{i \in \mathcal{B}} \equiv (R_{t1}, R_{t2}, ...R_{tB})$, where $R_{ti} = (R_{tin})_{n \in \mathcal{N}} \equiv (R_{ti1}, R_{ti2}, ..., R_{tiN})$. We define $\mathcal{B} = \{1, 2, ..., B\}$ as the set of all FOBs, $\mathcal{N} = \{1, 2, ..., N\}$ as the set of all supply classes, and $R_{tin} \in (0, r_{in}]$ as the number of tons of supplies for each supply class in $\mathcal{N}$ and at each FOB in $\mathcal{B}$ at time t. Moreover, r_{in} is the inventory capacity of each supply class $n \in \mathcal{N}$ at each FOB $i \in \mathcal{B}$. In the remainder of this thesis, we assume a single aggregate supply class (i.e., $N = 1$) and

therefore drop the supply class dimension for the state space. This simplification of the MILIRP will be addressed in subsequent research efforts. The number of operational CUAVs able to perform resupply operations at time t is defined as v_t. The threat map index number at time t is defined as $\hat{M}_t \in \{1, 2, ..., M\}$ where M is the number of threat maps utilized to model the security in the BCT's area of operation. The threat map impacts the flight risk associated with successfully completing sorties between FOBs and the brigade support area (BSA). This threat information, $\hat{M}_t$ is available at time t. The threat information for time $t+1$, $\hat{M}_{t+1}$, is conditioned on $\hat{M}_t$ and is unknown at time t. Utilizing these components, we define $S_t = \left(R_t, v_t, \hat{M}_t\right) \in \mathcal{S}$ as the state of the system at time t, where $\mathcal{S}$ is the set of all possible states.

We let $\mathcal{X}(S_t)$ be the set of all feasible actions when the system is in state S_t. Let $x_t = (x_{t11}, x_{t12}, ..., x_{t1v_t}, x_{t21}, x_{t22}, ..., x_{tij}, ..., x_{tBv_t}) \in \mathcal{X}(S_t)$ denote an inventory routing decision, where $x_{tij} \in \{0, 1\}$ is 1 if CUAV j is to resupply FOB i at time t and 0 otherwise. There are four restrictions placed on CUAV routing in a time period: first, the number of CUAVs deployed cannot exceed the number of operational CUAVs, v_t; second, the total number of CUAVs deployed cannot exceed the number of crews available K; third, a CUAV cannot deliver more than its maximum capacity, H; fourth, a CUAV can only deliver to one FOB per time period. Finally, our policy (i.e., decision function) is defined in Equation 1.

$$\mathcal{X}(S_t) \ni x_t = X^{\pi}(S_t) \tag{1}$$

Transition probabilities are defined for each dimension of our state space to include inventory levels at FOBs, current CUAV count, and threat map condition. Inventory transitions are based on routing decisions each time period, x_t, and the current state of the system S_t. When CUAVs are routed to FOBs there are three possible outcomes governed by a trinomial distribution: first, a CUAV may successfully travel to and

from a FOB; second, a CUAV may successfully deliver supplies and be destroyed upon the return to the BSB; third, a CUAV may be destroyed before successfully delivering the routed supplies. Let ψ_{im}^2, $\psi_{im}(1-\psi_{im})$, and $(1-\psi_{im})$ denote the probabilities of a successful two-way delivery (SS), successful one-way delivery (SF), and failure (F) for a single CUAV routed to resupply FOB i during threat condition map m. Since we are interested in a particular outcome of a routing decision, we proceed by defining the binomial marginal distributions for each outcome type (i.e., SS, SF, F). With the assumption that each outcome of a resupply mission to a FOB is independent of other missions and defining x_{tij} as the decision to route CUAV j to FOB i with any supply load, we let $\hat{Z}_{t+1,i}^{SS}(\psi_{im}^2, \sum_{j=1}^{v_t} x_{tij})$ denote the binomial random variable with parameters ψ_{im}^2 and $\sum_{v=1}^{v_t} x_{tiv}$ governing the number of two-way successful deliveries of CUAVs routed to FOB i during time interval $[t, t+1)$, on map m. Let $\hat{Z}_{t+1,i}^{SF}(\psi_{im}(1-\psi_{im}), \sum_{v=1}^{v_t} x_{tiv})$, and $\hat{Z}_{t+1,i}^{F}((1-\psi_{im}), \sum_{v=1}^{v_t} x_{tiv})$ be similarly defined. For simplicity, we refer to the quantity as:

$$\hat{Z}_{t+1} = \big((\hat{Z}_{t+1,i}^{SS})_{i\in\mathcal{B}}, (\hat{Z}_{t+1,i}^{SF})_{i\in\mathcal{B}}, (\hat{Z}_{t+1,i}^{F})_{i\in\mathcal{B}}\big). \tag{2}$$

Inventory levels at each FOB are limited by the maximum holding quantity Q_{in}. Moreover, if the FOB supply level falls below a certain threshold the FOB is imme–diately resupplied via ground convoy. Although we model this threshold as zero we note that falling to zero may represent falling to a preallocated safety stock at which the convoy is required for resupply. Equation 3 is the inventory transition function for FOB i.

$$R_{t+1,i} = \begin{cases} r_i & \text{if } R_{ti} + H(\hat{Z}_{t+1,i}^{SS} + \hat{Z}_{t+1,i}^{SF}) - \hat{D}_{t+1,i} \leq 0, \\ \min\left(R_{ti} + H(\hat{Z}_{t+1,i}^{SS} + \hat{Z}_{t+1,i}^{SF}) - \hat{D}_{t+1,i}, r_i\right) & \text{otherwise.} \end{cases} \tag{3}$$

In the first case, convoy resupply is necessary so the FOB is resupplied to capacity. In the second case, the FOB supply changes according to supplies received and realized demand. The minimization in the second case enforces the FOB capacity constraint.

CUAV transition is contingent on the probability of two-way successful delivery between the BSA and the FOB. The number of available vehicles is the minimum of CUAVs at time t, v_t, and the number of crews K. Thus, the number of CUAVs transition according to Equation 4.

$$v_{t+1} = v_t - \sum_{i=1}^{B} (\hat{Z}^{SF}_{t+1,i} + \hat{Z}^{F}_{t+1,i}) \tag{4}$$

The map transition functions is a representation of the uncontrolled stochastic aspect of the combat environment. The set of all maps captures the threat level of the operational environment via the tessellation of a geographic region. Each region within the threat map represents either a high or low threat level; maps are classified as either high or low threat maps depending on the number of high threat conditions associated with each map. Larger numbers of high threat condition cells help capture the increased risk of sending a CUAV to a particular FOB. Different CUAV routes are applied for each threat map; recall we apply Dijkstra's algorithm to each FOB for each map prior to solving the MDP. This allows us to know the best route with the highest one-way probability of survival depending on the location of the BSA, FOB, and the high threat regions. The map transitions are representative of the changing environment; for relatively static combat conditions, the map transition probability would be relatively low. More dynamic combat environments yield a relatively higher map transition probability. The BCT intelligence teams gather information on threat conditions and may label the tessellated region based on information such as enemy action, season, historical trends, and weather.

The contribution function is defined by the total amount of supplies delivered to

FOBs during each time period. The amount of supplies delivered is bounded by the maximum inventory quantity at each FOB, constraining any excess supplies delivered from affecting the system behavior. An immediate penalty is applied when stock out occurs due to the human risk associated with convoy resupply. Letting τ_i be the stock out penalty for FOB i allows us to apply different penalties that can capture the difficulty of resupplying a particular FOB via convoy; FOBs with higher penalty would receive more weight when routing decisions are considered. We present our contribution function in Equation 5

$$C(S_t, x_t) = \mathbb{E}\Big\{ \Big[\sum_{i=1}^{B} \min \Big(r_i - R_{ti} + \hat{D}_{t+1,i}, H(\hat{Z}^{SS}_{t+1,i} + \hat{Z}^{SF}_{t+1,i}) \Big), - \sum_{i=1}^{B} \tau_i I_{\{i\}} \Big] \Big| S_t, x_t \Big\} \quad (5)$$

where $I_{\{i\}}$ is 1 when the system is in a state of depleted supply (i.e., $R_{ti} + H(\hat{Z}^{SS}_{t+1,i} + \hat{Z}^{SF}_{t+1,i}) - \hat{D}_{t+1,i} \leq 0$) and 0 otherwise. This applies a penalty for risking lives for resupply via ground convoy. The amount of rewardable supplies is determined by the minimum of available capacity at FOB i and the number of supplies successfully delivered to FOB i.

The objective of this MDP is to maximize the expected total discounted value over an infinite horizon. By definition the transitions are Markovian, thus all decisions made at time t depend only on the current state of the system. To obtain the policy that maximizes the expected total discounted reward, Bellman's optimality equation is used:

$$V_t(S_t) = \max_{x \in \mathcal{X}(S_t)} \Big(C(S_t, x) + \lambda \mathbb{E}\{ V_{t+1}(S_{t+1}) | S_t, x \} \Big). \quad (6)$$

The value of being in state S_t results from choosing the action that maximizes the sum of the immediate expected contribution and the discounted expected total value

of the state of the system at time $t+1$. Using this MDP formulation, an approximate dynamic programming algorithm is developed to obtain policies for resupplying FOBs via CUAVs.

3.3 ADP Formulation

We implement an approximate policy iteration (API) algorithmic strategy using least squares temporal differences (LSTD). API mirrors the exact policy iteration algorithm closely. Instead of using the one-step transition matrix that is difficult to utilize for problem instances with high dimensionality, our API implementation approximates and updates the value function. We use the post-decision state which is the state of the system immediately after a decision is made but before the exogenous information processes are realized. This convention allows the expectation to be moved outside the maximization operator, altering our value function to the form of

$$V_t^x(S_t^x) = \mathbb{E}\Big\{ \max_{x \in \mathcal{X}(S_{t+1})} \big(C(S_{t+1}, x) + \lambda V_{t+1}^x(S_{t+1}^x)\big) | S_t^x \Big\}. \tag{7}$$

LSTD utilizes a set of basis functions that captures relevant information in the sys–tem, thus reducing the dimensionality of the state space and providing an adequate solution [26]. Letting $(\phi_f(s))_{f \in \mathcal{F}}$ be a basis function where $\mathcal{F}$ is a set of features, the value function approximation is given by: $\bar{V}^x(S_t^x|\theta) = \sum_{f \in \mathcal{F}} \theta_f \phi_f(S_t^x)$ wherein $(\theta_f)_{f \in \mathcal{F}}$ is a vector of weights with one coefficient for each basis function. Because we choose the number of features to be less than the dimensionality of the state space, it is com–putationally efficient to estimate the value function using basis functions. Although classical linear regression methods can be used to estimate θ, choosing an appropriate set of basis functions can be challenging. LSTD updates θ iteratively throughout the algorithm.

Algorithm 1 Approximate Policy Iteration Using Least Square Temporal Differences

Step 0. Initilize θ^0
Step 1. **For** a=1 to A **(Policy Improvement Loop)**
  Step 2. **For** q=1 to Q **(Policy Evaluation Loop)**
    a. Generate a random post-decision state, $S^x_{t-1,q}$
    b. Record $\phi(S^x_{t-1,q})$
    c. Simulate transition to next event, obtain a pre-decision state, $S_{t,q}$
    d. Determine decision $x_t = X^\pi(S_{t,q}|\theta^{a-1})$
    e. Record contribution $C(S_{t,q}, x_t)$
    f. Record basis function evaluation $\phi(S^x_{t,q})$
    End
  Step 3. Compute θ^a using smoothing rule
  End

LSTD iteratively updates the discounted total value function approximation for a fixed policy and projects it over an infinite horizon. LSTD derives its name from comparing the differences between the current value of being in a state with the up–dated value of being in a state at the following iteration. Alternatively, this can be viewed as a batch algorithm that operates by collecting samples of temporal differ–ences and then using least squares regression to find the best linear fit [26]. LSTD performs least squares regression so that the sum of the temporal differences over the simulation is equal to zero. The LSTD pseudo code is summarized in Algorithm 1.

$$\Phi_{t-1} \triangleq \begin{bmatrix} \phi(S^x_{t-1,1})^\top \\ \vdots \\ \phi(S^x_{t-1,Q})^\top \end{bmatrix}, \Phi_t \triangleq \begin{bmatrix} \phi(S^x_{t,1})^\top \\ \vdots \\ \phi(S^x_{t,Q})^\top \end{bmatrix}, C_t \triangleq \begin{bmatrix} C(S_{t,1}, x_t) \\ \vdots \\ C(S_{t,Q}, x_t) \end{bmatrix} \tag{8}$$

A total of Q temporal difference sample realizations are collected in each policy evaluation loop where the q^{th} temporal difference is denoted $C(S_{t,q}, X^\pi(S_{t,q}|\theta)) + \gamma\theta^\top\phi(S^x_{t,q}) - \theta^\top\phi(S^x_{t-1,q})$. Let Φ_{t-1} and Φ_t consist of rows of basis function evaluations of the sampled post-decision states and C_t as the contribution vector for the sampled states as indicated in Equation 8. The sample realization $\hat{\theta}$ is an estimation of θ and is defined in Equation 9.

$$\hat{\theta} = \Big[(\Phi_{t-1} - \gamma\Phi_t)^\top(\Phi_{t-1} - \gamma\Phi_t)\Big]^{-1}(\Phi_{t-1} - \gamma\Phi_t)^\top C_t \tag{9}$$

For comparison, we use both LSTD and instrumental variables (IV) LSTD. Instru–mental variables as introduced by Bradtke & Barto [5], are correlated with regressors, but uncorrelated with the errors in the regressors and the observations. An instru–mental variables method makes it possible to generate consistent estimators of the θ-vector. The application of IV to obtain $\hat{\theta}$ is shown below.

$$\hat{\theta} = \Big[(\Phi_{t-1}^\top)(\Phi_{t-1} - \gamma\Phi_t)\Big]^{-1}(\Phi_{t-1}^\top C_t) \tag{10}$$

We then apply a harmonic stepsize rule to smooth in the new observation $\hat{\theta}$ with the previous estimate θ during implementation. The stepsize rule α_a is a function of the outer loop iteration count and is defined below.

$$\alpha_a = \frac{1}{a} \tag{11}$$

The stepsize rule α_a greatly influences the rate at which the API algorithm con–verges thus impacting the attendant solutions. Utilizing the harmonic stepsize rule, we update our θ in the following way:

$$\theta \leftarrow \theta(1 - \alpha_a) + \hat{\theta}(\alpha_a). \tag{12}$$

Equation 12 shows that the updated θ is weighted most heavily by our current estimate of θ and then moved toward our new estimate, $\hat{\theta}$, by an incremental amount proportional to α_a. Initially, greater emphasis is placed on $\hat{\theta}$, but as the number of iterations increases the incremental effect of $\hat{\theta}$ is lessened. Moreover, as the number of iterations increases, any singe $\hat{\theta}$ has less influence than the estimate based on information from the first $a - 1$ iterations.

Upon obtaining an updated parameter vector θ, we have completed one policy improvement iteration of the algorithm. The parameters A and Q are tunable, where A is the number of policy improvement iterations completed and Q is the number of policy evaluation iterations completed.

Table 3. Table of Notation

A	-	Number of outer loops
B	-	Number of FOBs
C	-	Contribution function
C_t	-	Contribution vector for sampled states
$\hat{D}$	-	Total daily FOB demand
$\bar{d}$	-	Mean FOB demand
H	-	CUAV carrying (holding) capacity
I	-	Indicator variable for stockout penalty
K	-	Number of crews
M	-	Number of threat maps
N	-	Number of supply classes
Q	-	Number of inner loops
r	-	Maximum FOB quantity
R	-	Inventory at FOB
S	-	State of the system
t	-	Time epoch
U	-	Initial number of CUAVs (unmanned)
V	-	Total expected value
v_t	-	Number of CUAVs available at time t
X^π	-	Policy function
x	-	Action, sending a CUAV a FOB
$\hat{Z}$	-	CUAV delivery outcome
$\mathcal{B}$	-	Set of all FOBs
$\mathcal{F}$	-	Set of basis functions
$\mathcal{N}$	-	Set of supply classes
$\mathcal{S}$	-	State space
$\mathcal{T}$	-	Set of time epochs
$\mathcal{X}$	-	Action space
γ	-	Algorithmic discount factor
$\hat{\epsilon}$	-	FOB demand error term
θ	-	Vector of weights
$\hat{\theta}$	-	Vector of sample realized weights
λ	-	Time discount factor
π	-	Policy
τ	-	Stock out penalty
Φ	-	Matrix of fixed basis functions
ϕ	-	basis function
ψ	-	One-way probability of CUAV success

IV. Analysis

Utilizing the Markov decision process (MDP) formulation discussed in chapter III, we can find a policy for a 9-forward operating base (FOB) problem instance. As a baseline, basis functions for the approximate dynamic programming (ADP) algorithm are explored to find the ADP's optimal parameters. Finally, we run an experimental design to find the algorithmic and model parameters that yield the best results for our ADP algorithm.

4.1 MDP Parameterization

The military inventory routing problem (MILIRP) is formulated as an infinite horizon MDP where days are divided into four epochs of equal time. We assume that during each epoch the cargo unmanned aerial vehicle (CUAV) can complete all mission preparation tasks and perform the assigned mission. For this thesis, we assume an aggregate supply class and stochastic demand. While supply consumption is continuous, supply delivery is integer because a CUAV always delivers its maximum capacity, and clearly only an integer number of CUAVs may be sent.

For our problem instance we chose a 9-FOB design to represent an average sized battalion. We chose 9 because there are three platoons in a company and three companies in an battalion. A platoon will typically man a FOB within the battalion's area of responsibility. We test our ADP at a battalion's average operating conditions.

Each FOB has a consumption rate and storage capacity based on the number of personnel on site. Based on a General Dynamics report [12], the expected daily consumption requirements of a platoon is $7,482$ pounds. We round up as a conservative estimate to 8,000 pound daily average consumption per FOB. With four epochs in one day, about one ton of supplies per period is required for sustainment. For our

testing, we model the stochastic demand using this known historical average $\bar{d}$ and an independent and identically distributed error term, $\hat{\epsilon}$, normally distributed with a mean of 0 and standard deviation of 0.5. We also make the conservative assump–tion that a FOB has a maximum holding capacity of three times the daily average requirement totaling 12 tons. We assume that there are no logistical failures limiting the amount of supplies available at the centralized brigade support battalion (BSB). This assumption is reasonable since the BSB is supplied via fixed wing aircraft from outside the theater of operations.

Although technology is quickly progressing, Lockheed Martin's K-MAX has deliv–ered two tons at $15,000$ feet above ground level (AGL) with more tonnage delivered at lower altitudes [18]. Thus, we chose a conservative two ton carrying capacity for CUAV resupply for our 9-FOB design. We also chose the number of CUAVs and crews to be four and two respectively which mirrors the tactical unmanned aircraft system (TUAS) platoon [9]. As the requirements for CUAV resupply increase, we expect to see the number of CUAVs and crews the BSB utilizes to increase. As such, we parameterize the CUAVs and crews as multiples of TUAS platoon ratios. For example, if three TUAS platoons are deployed at the BSB, the number of CUAVs would be 12 and the number of crews 6.

We define ψ_{im} as the probability a CUAV will successfully travel to and from FOB i on map m. An intelligence team would ideally assign risk values to each zone in the tessellated region. This number would account for threats to include but not limited to: weather, enemy action, and mechanical breakdown. Transition between maps can be created by observed trends specific to the region of interest. These problems influence threat levels which might include time of the year. Thus, the optimal path can be found that minimizes the risk to the aircraft while in transit via a shortest path algorithm. For our example, we chose to use $M = 2$ threat maps. In the absence

of an intelligence team and a specific pattern of enemy activity, we use 0.40 as the probability of staying on any given threat map.

When a FOB's supply level falls below a predetermined minimum threshold, the FOB is immediately resupplied via ground convoy to full capacity. When a convoy is sent, a penalty is applied proportional to how far below the threshold the FOBs supplies falls. The penalty represents the increased human capital risk inherent in ground convoy operations along with the risk of the FOBs. The penalty associated with resupplying a specific FOB would ideally be supplied by a subject matter expert who knows the terrain and enemy activity levels associated with each FOB. For example, FOBs further away across rough terrain would have a higher penalty than a closer and more readily accessible FOB. This penalty creates a strong incentive to ensure FOBs are resupplied by CUAV when possible.

We chose $\lambda = 0.98$ to be a discount factor that balances future needs with current needs. We utilized the above described MDP parameterization to create both myopic and ADP policies for comparison.

4.2 Myopic Policy

The myopic policy ignores the future needs of the system and chooses actions based on current needs in the system. This is accomplished by setting $\lambda = 0$ which means there is no value from the future outcome of the current decision. We use the myopic policy as a benchmark policy to compare our ADP's algorithmic performance. Because of the size of our problem, there is no optimal value function to compare our ADP solution. Thus, the myopic policy serves as a stable benchmark to compare the ADP value.

4.3 ADP Policy

The ADP policy is obtained from our approximate policy iteration algorithm using least squares temporal differences (API-LSTD). We compare API-LSTD to approx–imate policy iteration with instrumental variables (IV) bellman error minimization (IVAPI) algorithm. The challenge with both these algorithms is developing basis functions that accurately approximate the unknown optimal value function. The API algorithms are employed with the system initialized at full capacity for each FOB.

To solve the approximate dynamic program (ADP), we need to solve the inner maximization problem. The inner maximization problem can be solved exactly using complete enumeration for smaller problem instances. However, for the size of our problem instance, this is intractable. We chose to formulate the inner maximization problem as an integer program (IP) because only an integer number of CUAVs can be sent for resupply. We define our IP as follows:

Decision Variables:

x_{ij}, binary. 1 when CUAV j sent to resupply FOB i, 0 otherwise.

y_i, slack variable at FOB i corresponding to the amount below stock-out.

Parameters:

θ_{ij}, coefficient value corresponding to action taken at FOB i.

θ_0, coefficient value corresponding to the number of CUAVs available.

τ_i, penalty associated with stock-out at FOB i.

H, CUAV holding capacity.

K, number of crews available.

v_t, number of CUAVs available at time t.

λ, time discount factor.

IP:

$$\sum_{j=1}^{v_t} \sum_{i=1}^{B} x_{ij}(\psi_i H + \lambda(\theta_{ij} - \theta_0)) - \tau_i y_i \tag{13}$$

st:

$$\sum_{j=1}^{v_t} \sum_{i=1}^{B} x_{ij} \leq \min(K, v_t) \tag{14}$$

$$\psi_i H \sum_{j=1}^{v_t} x_{ij} + R_i - \bar{d}_i \leq r_i, \forall i \in \mathcal{B} \tag{15}$$

$$R_i + \psi_i H \sum_{j=1}^{v_t} x_{ij} - \bar{d}_i + y_i \geq 0, \forall i \in \mathcal{B} \tag{16}$$

$$x_{ij} \in \{0, 1\} \tag{17}$$

The objective function balances current rewards, future expected rewards, and stock-out penalties at each FOB. The first constraint limits our actions to utilizing at most the total number of CUAVs available as dictated by the crew limitations of CUAS. The second constraint limits the expected amount of supplies delivered to be no greater than the FOB capacity. The third constraint penalizes our objective function by the amount the system drops below the supply requirement (which causes ground convoy resupply). The final constraint enforces our assumption that CUAVs are only sent in integer numbers.

We develop ADP policies using our IP in our API-LSTD algorithm. We chose the first order model excluding bilinear interaction terms because it both performed better in preliminary testing and allows us to use a linear IP rather than non-linear IP. The simpler inner maximization problem allows us to perform an experimental design with more breadth in a reasonable amount of time.

4.4 Experimental design

We created an experimental design to test the robustness of our design parameters and find the parameter settings that allow our explored algorithmic approach to achieve the best performance. We focused our response variable on the total number of supplies delivered via CUAV in the system. The total number of supplies delivered via CUAV is reported in tons. In each experimental run we simultaneously assess four problem features and four algorithmic features. The four problem features of interest were chosen based on what we thought might have the most effect on the system performance. The four problem features we chose to investigate are number of CUAVs initially in the system (U), number of crews available (K), probability of staying in a low threat map (α), probability of staying in a high threat map (β). The four algorithmic features we chose to experiment on are inner loop iteration count (Q), outer loop iteration count (A), a categorical variable where -1 denotes API-LSTD using Bellman's error minimization and 1 denotes IVAPI, and another categorical variable where 1 denotes using smoothing in the specified algorithm and -1 denotes no soothing. We recorded the response variable over a three month simulation with 100 replications per treatment.

Each of the four problem features are considered to be continuous. We chose the CUAV and crew level to be levels associated with deploying one, two, and three TUAS platoons at the BSB. This is done under the assumption that as commanders increasingly value CUAV resupply, TUAS platoons will be sent in greater numbers to support the brigade operations. The transitions probabilities, α and β, are pa–rameterized similar to earlier work by McKenna [21]. The lower value, 0.2, denotes a low chance of transitioning to a different threat map condition. The upper bound explored, 0.8, denotes a high probabilty of transitioning to a high threat map condi–tion.

Table 4. Factorial Design Settings

Description	Factor	Low	Center	High
Initial number of CUAVs	U	4	8	12
Number of crews	K	2	4	6
Probability of remaining low threat	α	0.2	0.4	0.8
Probability of remaining high threat	β	0.2	0.4	0.8
Number of inner loops	Q	5000	20000	35000
Number of outer loops	A	10	30	50
Using instrumenal variable	IV	-1	-	1
Smoothing	SM	-1	-	1

The four algorithmic features were also chosen to best explore the experimental space. The inner loop count was set to a low of $5,000$ and a high of $35,000$ based on initial testing. The center run is the midpoint of the upper and lower bounds and allows us to check to see if our response variable demonstrates nonlinearity. The outer loop iteration counter was similarly chosen allowing for a large upper bound to achieve the most accurate value function approximation for the basis function we chose. Table 4 shows the problem and algorithmic settings for our experimental design.

We implemented a 2^{8-2} resolution V fractional factorial design with three center runs totaling 67 runs. In a resolution V design, all first- and second-order effects are free from being aliased with other first- or second-order interactions. Second-order interactions are, however, aliased with three factor interactions. Our ADP policy utilizes the θ coefficients for the selected basis functions. After the θ coefficients are calculated, we simulate the myopic and ADP policies to attain our response variable statistics.

All treatments within the experiment are conduced in MATLAB R2015a calling CPLEX to solve the inner maximization problem. The experimental design was run on an Intel(R) Xeon(R) E3-1226 v3 3.30 GHz processor with 32.0 GB memory. The time reported for computational effort is only the run time for the ADP algorithm;

preprocessing operations and simulation times are not included. We conduct two simulations per experimental treatment over 100 replications. These two simulations account for determining the ADP policy and utilizing the myopic policy. Moreover, we use the common random numbers variance reduction technique for both the ADP algorithm and simulation.

4.5 Results

Tables 7 and 8 at the end of this section show the results from the experiment. The ADP algorithm did not perform well compared to the myopic strategy overall however there are instances where the ADP algorithm did perform well. The ADP policy significantly outperformed the myopic policy when a high number of CUAVs were deployed (at the 95% confidence level). The settings where our ADP algorithm performed best are when CUAV number, crew number, inner loop count, and outer loop count were at their high levels, both threat transition probabilities (probability of staying in current map) were at their low levels, and smoothing was not used. This treatment delivered on average of 85.6 more tons than the myopic policy. This analysis provides indication of which variables are influential in the design structure of this problem however a metamodel is necessary to draw direct conclusions.

We next created a regression metamodel to analyze the effects with more statistical rigor. Using a factor screening method yields factors which produce a significant relationship that passes the lack of fit test, however, the residual by predicted plot exhibits a strong funnel pattern that may be problematic. This outward-opening funnel pattern shown in Figure 1 implies the variance is an increasing function of the response variable [23].

To overcome the unequal variance in our response variable, we perform a Box-–Cox Y transformation test and find the logarithmic transformation is the best fit to

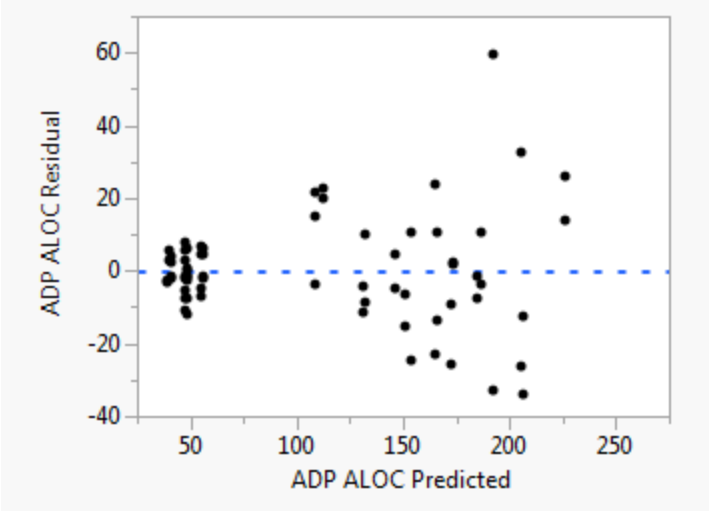

Figure 1. Residual by Predicted Plot Demonstrating Funnel Pattern

Table 5. Factors Influencing CUAV Resupply Amount

Variable	Sum of Squares	F Test	% Contribution
U	25.355997	$< .0001$	93.76
β	0.610391	$< .0001$	2.26
α	0.539198	$< .0001$	1.99
Q	0.08314	0.012	0.28
$U\ U$	0.267186	$< .0001$	1.01
$Q\ (SM)$	0.067605	0.0228	0.25
$(SM)\ K$	0.060227	0.0312	0.24
$\beta\ K\ (IV)$	0.057024	0.0359	0.21

reduce the sum of squared errors. After applying the logarithmic transformation and performing an additional screening process, we again find a significant model that passes the lack-of-fit test. After transformation, this model no longer exhibits the same problems with the residuals. This model eliminates most of the second- and third-order interactions and only leaves four first-order interaction terms significant. Table 5 summarized the significant variables.

We consider variables within the 95% confidence interval significant. With this cri–teria, number of CUAVs, probability of staying in a high threat condition, probability of staying in a low treat condition, and inner loop count are all significant first-order terms. Both smoothing and crew only become significant in the second-order terms

Table 6. Coefficient Estimates

Variable	Estimate	Percentage Increase	P-Value	Lower 95%	Upper 95%
U	0.629	87.65	$< .0001$	0.602	0.657
β	-0.098	-9.30	$< .0001$	-0.125	-0.070
α	0.092	9.61	$< .0001$	0.064	0.120
Q	0.035	3.52	0.0154	0.007	0.062
$U\ U$	-0.309	-26.57	$< .0001$	-0.440	-0.178
$Q\ (SM)$	0.033	3.30	0.0225	0.005	0.060
$(SM)\ K$	0.032	3.22	0.0259	0.004	0.059
$\beta\ K\ (IV)$	0.029	2.99	0.0379	0.002	0.057

while using instrumental variables only shows significance as a third-order effect. One area of note is that outer iteration count is the only effect that did not prove to be significant in our testing or previous research done to date on this problem [21]. This could be of significance because in future runs we could drop the outer iteration count to 10 iterations without having deleterious effects on the performance of the ADP policy in terms of tons of supplies delivered.

Using the values proved to be significant, we created a regression model. Table 6 provides the coefficient estimates of the significant terms, and the p-values for their associated t-statistic. Because of the logarithmic transformation on our response variable, we can no longer directly interpret our results; we can, however, interpret a unit change in the independent variable as a percentage per unit increase by taking any coefficient ω and performing the following operation, $100(e^{\omega} - 1)$. These values have been generated and added to Table 6 for ease of interpretation.

According to our model, the amount of supplies delivered is maximized with 12 CUAVs, probability of staying in a high threat map of 0.2, probability of staying in a low threat map of 0.8, $35,000$ inner loops, no smoothing, 6 crews, and using IVAPI. The fact that no smoothing appear significant indicates that the smoothing rule we chose was a poor value for this problem structure. We suspect that choosing a better smoothing rule will show the smoothing significant in future research.

Table 7. Experimental Results Over 3 Month Horizon

			ADP Policy		Myopic Policy		
Run	Coded Factor Levels	Comp time (sec)	ALOC (tons)	Convoy Incidents	ALOC (tons)	Convoy Incidents	ALOC Diff. (tons)
1	− − − − − − −+	188.6	55.48 ± 5.11	2.40 ± 0.48	68.86 ± 5.75	2.55 ± 0.51	-13.4 ± 7.6
2	− − − − − + +−	948.4	40.09 ± 4.33	2.32 ± 0.46	68.86 ± 5.75	2.55 ± 0.51	-28.8 ± 7.2
3	− − − − + − +−	1368.3	41.34 ± 4.80	2.36 ± 0.47	68.86 ± 5.75	2.55 ± 0.51	-27.5 ± 7.4
4	− − − − + + −+	6372.6	49.39 ± 4.77	2.46 ± 0.49	68.86 ± 5.75	2.55 ± 0.51	-19.5 ± 7.4
5	− − − + − − +−	184.1	42.83 ± 3.42	2.04 ± 0.40	54.58 ± 4.45	2.56 ± 0.51	-11.7 ± 5.6
6	− − − + − + −+	919.3	45.52 ± 4.30	2.44 ± 0.48	54.58 ± 4.45	2.56 ± 0.51	-9.1 ± 6.1
7	− − − + + − −+	1331.5	43.31 ± 4.24	2.34 ± 0.46	54.58 ± 4.45	2.56 ± 0.51	-11.3 ± 6.1
8	− − − + + + +−	6851.1	39.49 ± 5.28	2.86 ± 0.57	54.58 ± 4.45	2.56 ± 0.51	-15.1 ± 6.9
9	− + − − − + +	188.6	61.77 ± 6.42	2.97 ± 0.59	80.41 ± 6.14	2.73 ± 0.54	-18.6 ± 8.8
10	− + − − + − −	911.0	48.30 ± 6.00	2.78 ± 0.55	80.41 ± 6.14	2.73 ± 0.54	-32.1 ± 8.5
11	− − + − + − −−	1333.8	54.63 ± 6.33	3.15 ± 0.63	80.41 ± 6.14	2.73 ± 0.54	-25.8 ± 8.8
12	− − + − + + ++	7178.5	60.92 ± 6.31	2.91 ± 0.58	80.41 ± 6.14	2.73 ± 0.54	-19.5 ± 8.7
13	− − + + − − −−	183.0	36.62 ± 4.17	2.33 ± 0.46	66.45 ± 5.70	2.75 ± 0.55	-29.8 ± 7.0
14	− − + + − + ++	971.0	50.61 ± 5.33	2.37 ± 0.47	66.45 ± 5.70	2.75 ± 0.55	-15.8 ± 7.8
15	− − + + + − ++	1385.9	54.71 ± 5.55	2.76 ± 0.55	66.45 ± 5.70	2.75 ± 0.55	-11.7 ± 7.9
16	− − + + + + −−	6743.6	36.45 ± 5.11	2.70 ± 0.54	66.45 ± 5.70	2.75 ± 0.55	-30.0 ± 7.6
17	− + − − − − ++	192.6	45.89 ± 5.38	2.77 ± 0.55	67.65 ± 6.35	3.03 ± 0.60	-21.8 ± 8.3
18	− + − − − + −−	914.0	42.06 ± 4.75	2.69 ± 0.53	67.65 ± 6.35	3.03 ± 0.60	-25.6 ± 7.9
19	− + − − + − −−	1347.0	46.72 ± 4.82	2.25 ± 0.45	67.65 ± 6.35	3.03 ± 0.60	-20.9 ± 7.9
20	− + − − + + ++	7119.1	45.89 ± 4.77	2.56 ± 0.51	67.65 ± 6.35	3.03 ± 0.60	-21.8 ± 7.9
21	− + − + − − −−	182.3	37.16 ± 4.38	2.48 ± 0.49	48.59 ± 4.70	2.49 ± 0.49	-11.4 ± 6.4
22	− + − + − + ++	998.3	36.56 ± 4.19	2.33 ± 0.46	48.59 ± 4.70	2.49 ± 0.49	-12.0 ± 6.3
23	− + − + + − ++	1430.6	38.50 ± 3.73	2.28 ± 0.45	48.59 ± 4.70	2.49 ± 0.49	-10.1 ± 6.0
24	− + − + + + −−	6783.5	44.73 ± 4.54	2.50 ± 0.50	84.36 ± 7.80	2.49 ± 0.49	-3.9 ± 6.5
25	− + + − − − +	197.0	59.44 ± 6.41	3.02 ± 0.60	84.36 ± 7.80	3.48 ± 0.69	-24.9 ± 10.0
26	− + + − − + +−	907.7	50.05 ± 5.70	2.88 ± 0.57	48.59 ± 4.70	3.48 ± 0.69	-34.3 ± 9.6
27	− + + − + − +−	1341.9	62.34 ± 6.66	3.04 ± 0.60	84.36 ± 7.80	3.48 ± 0.69	-22.0 ± 10.2
28	− + + − + + −+	6876.5	54.24 ± 6.23	2.89 ± 0.57	84.36 ± 7.80	3.48 ± 0.69	-30.1 ± 9.9
29	− + + + − − +−	182.1	44.94 ± 6.23	2.96 ± 0.59	62.34 ± 6.84	3.27 ± 0.65	-17.4 ± 9.2
30	− + + + − + −+	970.5	52.51 ± 5.30	2.60 ± 0.52	62.34 ± 6.84	3.27 ± 0.65	-9.8 ± 8.6
31	− + + + + − −+	1380.3	48.50 ± 4.78	2.59 ± 0.51	62.34 ± 6.84	3.27 ± 0.65	-13.8 ± 8.3
32	− + + + + + +−	6776.9	46.59 ± 6.77	3.01 ± 0.60	62.34 ± 6.84	3.27 ± 0.65	-15.8 ± 9.6
33	+ − − − − − −−	191.6	128.99 ± 6.63	3.27 ± 0.65	183.05 ± 8.64	3.66 ± 0.73	-54.1 ± 10.8
34	+ − − − − + ++	1011.6	164.23 ± 9.67	4.25 ± 0.84	183.05 ± 8.64	3.66 ± 0.73	-18.8 ± 12.9

Table 8. Experimental Results Over 3 Month Horizon Continued

			ADP Policy		Myopic Policy		
Run	Coded Factor Levels	Comp time (sec)	ALOC (tons)	Convoy Incidents	ALOC (tons)	Convoy Incidents	ALOC Diff. (tons)
35	+ − − − + − ++	1493.7	175.81 ± 10.16	4.64 ± 0.92	183.05 ± 8.64	3.66 ± 0.73	-7.2 ± 13.3
36	+ − − − + + −−	7248.0	174.86 ± 9.56	4.13 ± 0.82	183.05 ± 8.64	3.66 ± 0.73	-8.2 ± 12.8
37	+ − − + − − ++	206.2	134.81 ± 7.55	3.47 ± 0.69	141.71 ± 7.84	3.60 ± 0.71	-6.9 ± 10.8
38	+ − − + − + −−	955.6	132.42 ± 7.17	3.38 ± 0.67	141.71 ± 7.84	3.60 ± 0.71	-9.3 ± 10.6
39	+ − − + + − −−	1405.9	123.62 ± 6.93	3.23 ± 0.64	141.71 ± 7.84	3.60 ± 0.71	-18.1 ± 10.4
40	+ − − + + + ++	7559.3	141.99 ± 7.30	3.35 ± 0.66	141.71 ± 7.84	3.60 ± 0.71	0.3 ± 10.6
41	+ − + − − − +−	193.8	183.35 ± 10.67	4.55 ± 0.90	210.59 ± 11.91	4.90 ± 0.97	-27.2 ± 15.9
42	+ − + − + − +	999.6	197.47 ± 12.31	4.93 ± 0.98	210.59 ± 11.91	4.90 ± 0.97	-13.1 ± 17.0
43	+ − + − + − −+	1546.4	194.07 ± 10.22	4.45 ± 0.88	210.59 ± 11.91	4.90 ± 0.97	-16.5 ± 15.6
44	+ − + − + + +−	7108.8	173.06 ± 9.33	3.75 ± 0.74	210.59 ± 11.91	4.90 ± 0.97	-37.5 ± 15.0
45	+ − + + − − −+	201.4	150.37 ± 9.24	3.76 ± 0.75	171.82 ± 9.41	4.25 ± 0.84	-21.5 ± 13.1
46	+ − + + − + +−	943.4	141.37 ± 8.29	3.87 ± 0.77	171.82 ± 9.41	4.25 ± 0.84	-30.4 ± 12.5
47	+ − + + + − +−	1441.3	176.35 ± 10.66	4.67 ± 0.93	171.82 ± 9.41	4.25 ± 0.84	4.5 ± 14.1
48	+ − + + + + −+	6933.4	152.09 ± 9.47	3.98 ± 0.79	171.82 ± 9.41	4.25 ± 0.84	-19.7 ± 13.3
49	+ + − − − − +−	199.9	162.93 ± 10.50	4.71 ± 0.93	166.30 ± 8.14	3.76 ± 0.75	-3.4 ± 13.2
50	+ + − − − + −+	1188.1	146.74 ± 9.79	3.98 ± 0.79	166.30 ± 8.14	3.76 ± 0.75	-19.6 ± 12.7
51	+ + − − + − −+	1676.1	159.15 ± 8.51	3.96 ± 0.79	166.30 ± 8.14	3.76 ± 0.75	-7.2 ± 11.7
52	+ + − − + + +−	7037.1	251.91 ± 15.46	6.04 ± 1.20	166.30 ± 8.14	3.76 ± 0.75	85.6 ± 17.4 *
53	+ + − + − − −+	235.3	127.03 ± 7.49	3.42 ± 0.68	133.48 ± 6.74	3.19 ± 0.63	-6.4 ± 10.0
54	+ + − + − + +−	1017.6	119.61 ± 7.95	3.64 ± 0.72	133.48 ± 6.74	3.19 ± 0.63	-13.9 ± 10.4
55	+ + − + + − +−	1458.9	144.45 ± 8.18	3.81 ± 0.76	133.48 ± 6.74	3.19 ± 0.63	11.0 ± 10.5 *
56	+ + − + + + −+	8240.2	135.96 ± 8.22	3.56 ± 0.71	133.48 ± 6.74	3.19 ± 0.63	2.5 ± 10.6
57	+ + + − − − −−	200.0	179.41 ± 11.00	4.35 ± 0.86	187.53 ± 9.41	4.24 ± 0.84	-8.1 ± 14.4
58	+ + + − − + ++	1049.1	238.20 ± 13.78	5.47 ± 1.09	187.53 ± 9.41	4.24 ± 0.84	50.7 ± 16.6 *
59	+ + + − + − ++	1435.9	251.60 ± 15.42	6.50 ± 1.29	187.53 ± 9.41	4.24 ± 0.84	64.1 ± 18.0 *
60	+ + + − + + −−	7231.1	239.71 ± 14.54	5.97 ± 1.18	187.53 ± 9.41	4.24 ± 0.84	52.2 ± 17.2 *
61	+ + + + − − ++	208.2	188.38 ± 11.85	5.06 ± 1.01	155.88 ± 8.87	3.82 ± 0.76	32.5 ± 14.7 *
62	+ + + + − + −−	1021.0	141.88 ± 9.39	4.12 ± 0.82	155.88 ± 8.87	3.82 ± 0.76	-14.0 ± 12.8
63	+ + + + + − −−	1459.7	176.91 ± 11.45	5.07 ± 1.01	155.88 ± 8.87	3.82 ± 0.76	21.0 ± 14.4 *
64	+ + + + + + ++	7553.9	182.99 ± 13.17	5.56 ± 1.10	155.88 ± 8.87	3.82 ± 0.76	27.1 ± 15.8 *
65	0 0 0 0 0 0 − −	2422.4	124.18 ± 10.14	4.30 ± 0.85	128.62 ± 8.82	3.59 ± 0.71	-4.4 ± 13.4
66	0 0 0 0 0 0 − +	2409.0	130.73 ± 8.46	3.68 ± 0.73	128.62 ± 8.82	3.59 ± 0.71	2.1 ± 12.1
67	0 0 0 0 0 0 + −	2386.5	105.66 ± 7.86	3.46 ± 0.69	128.62 ± 8.82	3.59 ± 0.71	-23.0 ± 11.7

V. Conclusions and Recommendations

5.1 Conclusions

Management of cargo unmanned aerial vehicle (CUAV) assets for resupply is an important issue to the United States military. Poorly developed transportation infrastructure, adverse weather conditions, terrain, enemy threat and actions, and the availability of distribution assets all inhibit successful distribution of supplies from the brigade support area (BSA) to the forward operating bases (FOBs). Moreover, insurgent use of improvised explosive devices (IEDs) greatly affects truck mobility throughout the operational environment and has been successful in disrupting replenishment procedures [24]. Since 2012 when the K-MAX successfully deployed to Afghanistan [13], CUAVs have been of increasing interest both to the United States and worldwide [14]. This thesis provides insight into using cargo unmanned aerial vehicles (CUAVs) in combat environments for resupply. High casualty rates for convoy resupply mission has highlighted the importance of CUAV aerial resupply. CUAV benefits include: better performance in adverse weather conditions, higher flight ceilings, and no escort requirement restrictions. All these yield a lower probability of vehicle destruction via man portable air defense systems (MANPADS) and small arms fire. The most important benefit of CUAVs is their ability to save lives by alleviating ground convoy resupply requirements. Although CUAVs do not yet have the ability to completely handle FOB supply requirements, each successful CUAV delivery means less men and women exposed to enemy threats to include IEDs.

We formulated an Markov decision process (MDP) of the military inventory routing problem (MILIRP). We expanded previous research by adding model realism to include stochastic demand and a penalty function while developing the general model to introduce supply classes. We utilized approximate dynamic programming (ADP)

to determine approximate solutions. We tested our approach and obtained mixed results.

Although our results are situationally better than the myopic policy, we cannot conclude that our methodology properly captures the nuances of the problem structure. Ultimately this thesis shows utilizing least squares temporal differences with first order terms is an insufficient basis function approximation technique to approximate the value function when stochastic demand and penalty functions are implemented. We suspect that the addition of the penalty function creates non-linearities in the value function. It would be advantageous to increase the size of the basis function to include bilinear interaction and nonlinear terms; this addition would require the inner maximization problem to be solved using nonlinear integer solution techniques which would increase the computational intensity of the problem. A better understanding of how the penalty function affects the value function is necessary to model the value function properly.

5.2 Future Research

There are a plethora of areas for future research into the military inventory routing problem (MILIRP). The first way this research can be extended is to improve model realism. This thesis did not consider time limiting restrictions that may be a necessary consideration in combat conditions. Time windows for delivery and relaxing the direct delivery constraints also could provide more model realism as well. Additionally, the addition of supply classes would add model realism that has previously not been researched. Multiple supply classes greatly increase the computational complexity that the model must handle however it would be a great stride in modeling the MILIRP.

Moreover, applying a greater variety of ADP algorithms to this problem would also

be useful. This thesis examined policies generated from two different ADP algorithms. Thus, applying additional ADP algorithms to obtain solutions to the MILIRP is another avenue for future research.

Finally, additional research into special problem structure must be explored. A theoretical result has yet to be proven to show how a penalty function affects the value function. A greater understanding of the penalty induced value function can allow for customized algorithms to better approximate the value function.

Appendix A. Acronyms

ADP = approximate dynamic programming

BCT = brigade combat team

BSA = brigade supply area

BSB = brigade supply battalion

CUAV = cargo unmanned aerial vehicle

F = CUAV does not successfully deliver supplies

FOB = forward outpost

IED= improvised explosive device

IRP = inventory routing problem

IV = instrumental variables

MANPADS = man-portable air-defense system

MDP = Markov decision process

MILIRP = military inventory routing problem

SF = CUAV delivers supplies to COP, but does not successfully return

SIRP = stochastic inventory routing problem

SS = CUAV completes both legs of the journey

VMI = vendor managed inventory

VRP = vehicle routing problem

Bibliography

1. Adelman, Daniel. 2004. A price-directed approach to stochastic inventory/routing. *Operations Research*, **52**(4), 499–514.

2. Berman, Oded, & Larson, Richard C. 2001. Deliveries in an inventory/routing problem using stochastic dynamic programming. *Transportation Science*, **35**(2), 192–213.

3. Bertsekas, Dimitri P. 2011. Approximate policy iteration: A survey and some new methods. *Journal of Control Theory and Applications*, **9**(3), 310–335.

4. Bertsekas, Dimitri P, & Tsitsiklis, John N. 1996. Neuro-dynamic programming (optimization and neural computation series, 3). *Athena Scientific*, **7**, 15–23.

5. Bradtke, Steven J, & Barto, Andrew G. 1996. Linear least-squares algorithms for temporal difference learning. *Machine Learning*, **22**(1-3), 33–57.

6. Campbell, Ann, Clarke, Lloyd, Kleywegt, Anton, & Savelsbergh, Martin. 1998. The inventory routing problem. *Pages 95–113 of: Fleet management and logistics*. Springer.

7. Coelho, Leandro C., Cordeau, Jean-Franois, & Laporte, Gilbert. 2014. Thirty Years of Inventory Routing. *Transportation Science*, **48**(1), 1–19.

8. Department of the Army. 1995. Army Field Manual: Army Operational Support No. 100-16.

9. Department of the Army. 2010. Army Field Manual: Brigade Combat Team No. 3-90.6.

10. Department of the Army. 2014a. Army Field Manual: Brigade Support Battalion No. 4-90.

11. Department of the Army. 2014b. General Supply and Field Services Operations No. 4-42. *Army Techniques Publication*.

12. General Dynamics Information Technology. 2010. *Future Modular Force Resupply Mission for Unmanned Aircraft Systems (UAS)*. General Dynamics Information Technology.

13. Hoffman, Michael. 2012. *K-Max cargo UAS exceeds expectations in Afghanistan test*.

14. JAPCC, Joint Air Power Competence Centre. 2014. Remotely Piloted Aircraft Systems in Contested Environments.

15. Kleywegt, Anton J., Nori, Vijay S., & Savelsbergh, Martin W. P. 2002. The Stochastic Inventory Routing Problem with Direct Deliveries. *Transportation Science*, **36**(1), 94.

16. Kleywegt, Anton J., Nori, Vijay S., & Savelsbergh, Martin W. P. 2004. Dynamic Programming Approximations for a Stochastic Inventory Routing Problem. *Transportation Science*, **38**(1), 42 – 70.

17. Lenstra, Jan Karel, & Kan, AHG. 1981. Complexity of vehicle routing and scheduling problems. *Networks*, **11**(2), 221–227.

18. Lockheed Martin. 2010. *K-MAX Unmanned Aircraft System.* `http://www.lockheedmartin.com/content/dam/lockheed/data/ms2/documents/K-MAX-brochure.pdf`. Accessed: 2016-3-1.

19. McCormack, Ian. 2014. *The Military Inventory Routing Problem with Direct Delivery.* M.Sci., Air Force Institute of Technology.

20. McKenna, R. S., Robbins, M. J., Lunday B. J., & McCormack, I. M. 2016. *Approximate Dynamic Programming for the The Military Inventory Routing Problem with Direct Delivery.* Tech. rept. Air Force Institute of Technology.

21. McKenna, Rebekah. 2015. *Using Approximate Dynamic Programming to Solve the The Military Inventory Routing Problem with Direct Delivery.* M.Sci., Air Force Institute of Technology.

22. Minkoff, Alan S. 1993. A Markov decision model and decomposition heuristic for dynamic vehicle dispatching. *Operations Research*, **41**(1), 77–90.

23. Montgomery, Douglas C, Peck, Elizabeth A, & Vining, G Geoffrey. 2015. *Introduction to linear regression analysis.* John Wiley & Sons.

24. Peterson, Troy M., & Staley, Jason R. 2011. Business Case Analysis of Cargo Unmanned Aircraft Systems (UAS) Capability in Support of Forward Deployed Logistics in Operation Enduring Freedom (OEF). 1–15.

25. Powell, Warren B. 2009. What you should know about approximate dynamic programming. *Naval Research Logistics (NRL)*, **56**(3), 239–249.

26. Powell, Warren B. 2011. *Approximate Dynamic Programming: Solving the Curses of Dimensionality.* 2 edn. John Wiley & Sons, Inc.

27. Powell, Warren B. 2012. Perspectives of approximate dynamic programming. *Annals of Operations Research*, 1–38.

28. Sutton, Richard S, & Barto, Andrew G. 1998. *Reinforcement learning: An introduction.* MIT press.

29. Toth, Paolo, & Vigo, Daniele. 2001. *The Vehicle Routing Problem*. Vol. 18. SIAM.

30. Van Roy, Benjamin, Bertsekas, Dimitri P, Lee, Yuchun, & Tsitsiklis, John N. 1997. A neuro-dynamic programming approach to retailer inventory management. *Pages 4052–4057 of: Decision and Control, 1997., Proceedings of the 36th IEEE Conference on*, vol. 4. IEEE.

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704–0188

1. REPORT DATE *(DD–MM–YYYY)*	2. REPORT TYPE	3. DATES COVERED *(From — To)*
16–06–2016	Master's Thesis	SEP 2014 — JUN 2016

4. TITLE AND SUBTITLE

Using Approximate Dynamic Programming to Solve the Stochastic Demand Military Inventory Routing Problem with Direct Delivery

5a. CONTRACT NUMBER

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)

Salgado, Ethan L., Second Lieutenant, USAF

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-16-J-031

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)

TRADOC Capability Manager for Unmanned Aircraft Systems
Deputy, TCM-UAS Mr. Glenn A. Rizzi
453 Novosel Street
Fort Rucker, AL 36362
glenn.a.rizzi.civ@mail.mil

10. SPONSOR/MONITOR'S ACRONYM(S)

TCM-UAS

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION / AVAILABILITY STATEMENT

Distribution Statement A.
Approved for Public Release; distribution unlimited.

13. SUPPLEMENTARY NOTES

14. ABSTRACT

A brigade combat team must resupply forward operating bases (FOBs) within its area of operations from a central location, mainly via ground convoy operations, in a way that closely resembles vendor managed inventory practices. Military logisticians routinely decide when and how much inventory to distribute to each FOB. Technology currently exists that makes utilizing cargo unmanned aerial vehicles (CUAVs) for resupply an attractive alternative due to the dangers of utilizing convoy operations. However, enemy actions, austere conditions, and inclement weather pose a significant risk to a CUAV's ability to safely deliver supplies to a FOB. We develop a Markov decision process model that allows for multiple supply classes to examine the military inventory routing problem, explicitly accounting for the possible loss of CUAVs during resupply operations. The large size of the motivating problem instance renders exact dynamic programming techniques computationally intractable. To overcome this challenge, we employ approximate dynamic programming (ADP) techniques to obtain high-quality resupply policies. We employ an approximate policy iteration algorithmic strategy that utilizes least squares temporal differencing for policy evaluation. We construct a representative problem instance based on an austere combat environment in order to demonstrate the efficacy of our model formulation and solution methodology. Because our ADP algorithm has many tunable features, we perform a robust, designed computational experiment to determine the ADP policy with the best quality of solutions. Results indicate utilizing least squares temporal differences with a first-order basis function is insufficient to approximate the value function when stochastic demand and penalty functions are implemented.

15. SUBJECT TERMS

Approximate dynamic programming, Markov decision process, Vendor managed inventory, Vehicle routing, Military inventory routing (MILIRP), Least squares temporal differences

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			Lt Col Matthew J. Robbins, AFIT/ENS
U	U	U	U	56	19b. TELEPHONE NUMBER *(include area code)* (937) 255-3636, x4539 matthew.robbins@afit.edu