



**United States Air Force Officer  
Manpower Planning Problem  
via Approximate Dynamic Programming**

THESIS

MARCH 2016

Amelia E. Bradshaw, 2d Lt, USAF  
AFIT-ENS-MS-16-M-092

**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 Air Force, 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-16-M-092

UNITED STATES AIR FORCE OFFICER  
MANPOWER PLANNING PROBLEM via  
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

Amelia E. Bradshaw, BS, MS

2d Lt, USAF

MARCH 2016

DISTRIBUTION STATEMENT A  
APPROVED FOR PUBLIC RELEASE; DISTRIBUTION UNLIMITED.

AFIT-ENS-MS-16-M-092

UNITED STATES AIR FORCE OFFICER  
MANPOWER PLANNING PROBLEM via  
APPROXIMATE DYNAMIC PROGRAMMING

THESIS

Amelia E. Bradshaw, BS, MS  
2d Lt, USAF

Committee Membership:

Lt Col Matthew J. Robbins, Ph.D.  
Chair

Raymond R. Hill, Ph.D.  
Member

## **Abstract**

The United States Air Force (USAF) is concerned with managing its officer corps to ensure sufficient personnel for mission readiness. Manpower planning for the USAF is a complex process which requires making decisions about accessions. Uncertainty about officer retention complicates such decisions. We formulate a Markov decision process model of the Air Force officer manpower planning problem (AFO-MPP) and utilize a least squares approximate policy iteration algorithm as an approximate dynamic programming (ADP) technique to attain solutions. Computational experiments are conducted on two AFO-MPP instances to compare the performance of the policy determined with the ADP algorithm to a benchmark policy. We find that the ADP algorithm performs well for the basis functions selected, providing policies which reduce soft” costs associated with shortages and surpluses in the force.

## Acknowledgements

I would like to thank Dr. Robbins for being a most instructive advisor. I have learned a lot about the research process, Markov decision processes, and being an effective analyst under your tutelage. I appreciate your patience throughout this thesis effort. Special thanks to Captain Mike Davis for his assistance throughout the coursework for this thesis and assistance with MATLAB.

Amelia E. Bradshaw

# Table of Contents

	Page
Abstract .....	iv
Acknowledgements .....	v
List of Figures .....	viii
List of Tables .....	ix
I. Introduction .....	1
1.1 Problem Background .....	1
1.2 Thesis Outline .....	5
II. Literature Review .....	6
2.1 Simulation .....	6
2.2 Regression .....	7
2.3 Linear Programming .....	8
2.4 Stochastic Modeling .....	9
2.5 Markov Decision Processes .....	10
2.6 Heuristics .....	14
2.7 Approximate Dynamic Programming .....	14
III. Methodology .....	19
3.1 Problem Statment .....	19
3.2 MDP Formulation .....	21
3.3 Approximate Dynamic Programming Algorithms .....	26
Least Squares Temporal Differences .....	26
IV. Computational Results .....	30
4.1 Small Problem Instance .....	30
Experimental Design .....	31
Experimental Results .....	34
4.2 Full Problem Instance .....	35
Experimental Design .....	37
Over-manned Scenario .....	40
Under-manned Scenario .....	42
Target Level Scenario .....	44

	Page
V. Conclusions .....	47
5.1 Conclusions.....	47
5.2 Future Work.....	48
VI. Appendix .....	50
Bibliography .....	51

# List of Figures

Figure		Page
1	Event Timing Diagram for AFO-MPP .....	20
2	Officer Progression Throughout AFO-MPP .....	24
3	Retention Factor Levels .....	39

## List of Tables

Table		Page
1	Factor Settings for Small Problem Experimental Design .....	33
2	Small Problem Experimental Results .....	34
3	Small Problem Parameter Estimates .....	35
4	Factor Settings for Full Problem Experimental Design .....	40
5	Over-manned Experimental Results .....	41
6	Over-manned Scenario Parameter Estimates .....	42
7	Under-manned Experimental Results .....	43
8	Under-manned Scenario Parameter Estimates .....	44
9	Target Level Experimental Results .....	45
10	Target Level Scenario Parameter Estimates .....	46

UNITED STATES AIR FORCE OFFICER  
MANPOWER PLANNING PROBLEM via  
APPROXIMATE DYNAMIC PROGRAMMING

## I. Introduction

### 1.1 Problem Background

The United States Air Force (USAF) is a military service organization within the Department of Defense (DoD) that provides air and space national security capabilities to the country. The USAF, alongside the other branches of the United States military, safeguards national security to reinforce America's position as a leading world power. The United States military leads the world in defense technology and in the superior training and discipline of its personnel. In the midst of the changing landscape of warfare and economic difficulties, the USAF is concerned with maintaining its war-fighting capabilities (Conley & Robbert, 2009). In particular, senior leadership seeks to recruit and maintain the optimal number of personnel to carry out the mission of the USAF. This research studies how to best manage the USAF officer corps to ensure sufficient personnel for mission readiness.

Manpower planning in an organization involves recruiting and retaining enough employees to carry out the functions of the business, subject to the budgetary considerations for employees salaries. It is a dynamic issue on which all organizations focus in the ebb and flow of business booms and economic downturns. Such fluctuations influence USAF manpower policy. The structure of the force within the military involves the hierarchy of an officer corps who are responsible for leading the enlisted

force of personnel. The government mandates the numbers of officers and enlisted members in the military. Depending on the country's financial state and the level of threat to national security, the number of officer and enlisted personnel allowed for the fiscal year varies.

The USAF classifies its officers by one of ten ranks: from Second Lieutenant through General. The ranks have corresponding pay grade labels: O-1 through O-10. The hierarchical nature of the military dictates that all officers enter the system as an O-1 and progress through the ranks sequentially (Brauner *et al.*, 2009). Company grade officers are those in grade labels O-1 through O-3, field grade officers in ranks O-4 through O-6 and general officers in O-7 through O-10 (Conley & Robbert, 2009). As the officer progresses through the ranks, the officer's authority in the organization increases. A field grade or general officer is not only responsible for the enlisted troops in their command, but also the company grade officers. The company grade officer ranks are highly populated but in the higher ranks, the number of officers is lower.

An important aspect of manpower planning is the natural turnover as employees retire or move on to other opportunities. These employees are replaced by promoting within the organization or hiring new employees to train. The issue of manpower planning in the DoD and USAF is unique in comparison to other organizations due to the hierarchical, "closed" nature of the military. In civilian organizations, theoretically, employees of any skill level can be hired from outside of the organization to replace an employee of a higher status. This is not the case for the military. Any replacement of higher-ranking officers comes from within the military. All recruiting and admittance of new officers into the personnel system occurs at the entry level only (Wang, 2005). This aspect of the military personnel system presents a challenge when there is a shortfall in manpower in higher ranks because there is no immediate solution; the officers must be recruited and trained to the appropriate level within

the system.

The USAF categorizes its force further by the career field in which personnel work. The Air Force Specialty Code (AFSC) designates the job of the individual. Each AFSC has a specific mission (i.e., service requirement) it carries out for the Air Force. Individuals within the AFSC acquire mission-specific skill sets via specialized training and education programs. Over one hundred career fields exist within the Air Force. As warfare evolves over time, certain career fields become obsolete while others emerge due to the development of new technology (Conley & Robbert, 2009). This situation contributes to the uncertainty involved with manpower planning for the USAF.

The current method utilized by the USAF to manage the size of officer and enlisted personnel for the AFSCs is based on the sustainment line concept. Constructed from historical personnel data and current requirement levels for the AFSC, the sustainment line enables the estimation of the number of personnel that should be hired and leave the force in a particular year group to maintain optimal mission effectiveness over a thirty year time period. The sustainment line for officers and enlisted are treated separately because the two personnel groups have different functionalities within AFSCs. The sustainment line method provides the Headquarters Air Force with information concerning how many officers to commission to meet future end-strength numbers (Conley & Robbert, 2009). The sustainment line provides insight as to which career fields may be over-manned and able to withstand cuts in times of force reduction or may be understaffed and need additional personnel. The sustainment line has proved effective for senior decision makers, but there is much uncertainty involved with manpower planning in the military. The levels of personnel are trimmed by congressional mandate in times of peace and financial uncertainty, as is the current state of affairs. However, if threats to national security develop, the

size of the military increases to provide government leaders the ability to respond with appropriate force, if required.

Manpower planning for the USAF is a complex process which requires making decisions about accessions and promotions, with the uncertainty about factors which could affect employee turnover. In this study, we formulate a model which incorporates the uncertainty involved in the decision making process to provide insight into optimal policies for USAF officer accessions. We develop the Air Force Officer Manpower Planning Problem (AFO-MPP) to study this process.

We formulate a Markov decision process model of the AFO-MPP. Markov decision processes have evolved and grown in popularity in recent years as a methodology for solving manpower planning problems. A Markov decision process (MDP) involves a set of states, actions, time periods, transition probabilities, and rewards (Puterman, 2005). The objective of an MDP is to determine a policy, or set of decisions, so that the corresponding random reward sequence is as large as possible if a maximization problem is considered. If the objective is minimization, the optimal policy is the sequence of actions that makes the random cost sequence as small as possible. The state space reflects the elements of interest for the problem. These action sets or state spaces can potentially be infinite sets. The system then evolves deterministically from state to state as a function of the actions taken. There is a probabilistic element with transitions from one state to another depending on uncertainty involved in the system.

The MDP for the AFO-MPP makes use of yearly decision epochs to reflect the yearly manpower authorizations decisions made by Congress within the DoD's Future Years Defense Program (Conley, 2006). The state of the model is the number of officers within the AFSC at a specific year group and rank. The decision to be made at each time epoch is how many officers to enter the system at the lowest level, reflecting how military accessions occur at the O-1 rank. The cost function considers

the “soft” costs of a potentially under-manned or over-manned force. The transition probabilities reflect the turnover decisions within the organization from skill level to skill level. The objective of the MDP is to determine a policy that minimizes the expected total cost criterion. The optimal policy is a set of decision rules for each time period which will give the decision maker the least costly accession program. For the AFO-MPP, a decision rule specifies the number of second lieutenants to commission at a given point in time considering the current inventory of officers in the system.

In this thesis, we seek to solve the AFO-MPP by finding the optimal number of officers to assess that minimizes the costs of maintaining force. We utilize approximate dynamic programming (ADP) techniques to solve this large-scale dynamic programming problem within the framework of our MDP formulation. We test various parameters of our ADP algorithm to see which settings produce the highest solution quality for a small scale AFO-MPP problem instance. We then perform three design of experiments on realistic MP scenarios to see how our ADP algorithm performs. We intend to show the advantages of the sequential decision making techniques of MDP and ADP when applied to the AFO-MPP.

## **1.2 Thesis Outline**

Chapter 2 reviews published works concerning the manpower planning problem, making note of the many methodologies employed. Chapter 3 explains the MDP framework and ADP algorithm employed in this study. Chapter 4 provides the results of our findings after testing our ADP algorithm. Chapter 5 summarizes the results found and gives recommendations for future research.

## II. Literature Review

Manpower planning (MP) is extensively researched in the academic community. The main idea of manpower planning is to utilize employees in the most effective manner, given a range of skill levels (Khoong, 1996). Many different aspects of manpower planning can be examined with operations research techniques. The diversity of demands for analysis results from the broad applications in MP. In addition to the many aspects of manpower planning to examine, there also is a wide variety of methodologies used to solve MP problems. This literature review provides a background of the applications and methods used for MP to better inform our research of the topic.

### 2.1 Simulation

A popular method for gaining insight into MP problems involves computer simulation techniques. Analysts build simulation models of MP systems utilizing a variety of software products. Simulations are run to mimic system behavior and to collect statistics concerning the behavior of system entities under the conditions set in the model (Banks *et al.*, 2010). Wang (2005) utilizes a systems dynamics modeling approach to simulate officer training in the Australian Army and to examine how the training regimen affects the number of officers across the different manpower categories. Wang (2005) highlights that because of the closed structure of the military, increasing demand in a higher rank increases the demand in the lower ranks.

A Decision Support System (DSS) can be a particular type of computer simulation made for use by decision makers to give a comprehensive look at their organization. RAND created a DSS while studying the competency requirements in the career progression of a General Officer (GO) in the Air Force (Robbert *et al.*, 2005). The

RAND study surveys a set of GOs concerning factors they deemed most important in execution of their duties. The authors utilize the survey results to create a career progression model for a potential future GO from the entry level position onward. The authors leverage their proposed progression model to identify potential disqualifying factors for consideration. In the health career field, Hagopian *et al.* (2008) use a DSS to predict the demand for health care providers for people suffering from AIDS in the low-income country of Mozambique. Their staffing model allows the decision maker to input the demands based on the expected numbers of patients and receive as output the optimal schedule for each of the types of health care professionals needed to care for those patients.

## 2.2 Regression

Regression analysis is another technique applied to the study of MP problems. A topic of particular interest concerning military models is retention rates since the US military is volunteer-based. The regression analysis performed by Castro & Huffman (2002) explores the factors that influence military members' decisions to remain in the service. The study identifies five important topical areas relevant to retention: operations tempo (OPTEMPO), leadership, personal factors, work climate, and family considerations. The survey data were collected from soldiers near reenlistment and ascertained the relative importance of these factors in the soldiers' decision to leave the force or remain. The authors perform their analysis utilizing two approaches: multinomial logistic regression and a chi-squared automated interaction detection. The authors identify the most significant variables for incorporation into the model for predicting retention among military members. The important variables included deployment experience, years in the military, soldier pride, non-commissioned officer leadership, and marital status. The models created utilizing the significant factors

correctly predicted about 60% of the cases for identifying members who intended to remain in the military.

Lakhani (1988) also studies retention rates with regression analysis. The objective is to see how pay and retention bonuses affect retention in the US Army. The data on quit rates are collected from the historical data of force levels from year to year. The study examines how military members in combat career fields responded over time to pay incentives, as compared to members in non-combat career fields. The author concludes that members in the combat fields are more incentivized to remain in the force with reenlistment bonuses than their counterparts in non-combat fields. Stanley III (2012) uses regression analysis to identify factors that affect USAF pilot retention. As with most manpower planning problems, building up a force of fully trained pilots for the AF in the case of a shortage proves more difficult than relieving an over-manned force of personnel. Stanley identifies the presence of a force shaping program, acceptance of aviator continuation pay, and increased hiring from civilian airlines as significant to retention of pilots. The model formulated by the regression analysis is to be utilized to identify potential shortfalls in pilot retention in the future and to take a proactive measure to prevent such an occurrence.

### **2.3 Linear Programming**

For some MP model applications, the decision maker wants to optimize multiple objectives. Sayın & Karabatı (2007) formulate a two-stage optimization model to maximize utilization of each department in the organization and to maximize skill improvement of the individual employees. The first stage assigns the workers to departments and evolves the skill sets of the workers according to the workers' learning curve. The learning curve is a common term in the MP literature and refers to an individual's rate at which he or she picks up skills on the job (Gans & Zhou, 2002).

The evolution of the individual’s learning curve contributes to the maximization of department utility. The second stage examines worker assignments from the framework of maximizing worker skill evolution. The authors seek a balance between assigning workers so as to benefit the organization and to benefit the workers themselves. Examining both aspects benefits the overall performance of the company.

An important issue for employers is employee turnover. Sohn *et al.* (2007) refer to the loss of the departing employees’ skills as “brain-drain”. Sohn *et al.* (2007) implement a random effects Weibull regression model with an inverse gamma distribution to capture turnover. The Weibull distribution is used for its flexibility. Maximum likelihood estimates of the parameters for the Weibull distribution prove to be significant at the 1% confidence level for data provided about employee turnover in a Korean trading company.

## 2.4 Stochastic Modeling

Chattopadhyay & Gupta (2007) develop a stochastic model for a MP problem with different “class” sizes. The authors emphasize the importance of modeling MP problems as stochastic because of the amount of uncertainty involved in turnover for employees within an organization. They fit a beta distribution to the proportion of employees promoting to the next skill level, a binomial distribution to the current number of employees of a certain skill level, a Poisson distribution to the number of employees to be recruited into the system, and a geometric distribution to the time at which an employee experiences a promotion. Many of the distributions fit the data well, but the authors experience difficulty in parameterization. A simulation model using this information identifies the optimal age intervals for promotions within an airline’s hierarchy of flight attendants.

Vernez *et al.* (2006) study Air Force space and missile officers to identify ways to

maximize their utilization within the military. The 13S career field (i.e., space and missile officers) is identified as having officers with shortfalls of experience for the jobs at higher levels. After a thorough analysis of historical data on 13S officers, types of jobs the officers could hold were binned into 12 categories. The authors identify gaps for officers with respect to the necessary work experience for the job to which they are assigned and the experience they have gained so far in their career path. The optimization performed with flow analysis matches the officers with a specific type of work experience to the job that best utilizes their set of skills. The authors utilize their model to investigate how policy changes, such as requiring officers to take certain career paths to broaden their command potential, affect the overall gaps in experience in the field. Brauner *et al.* (2009) perform a similar case study on the Air Force 14N career field (i.e., intelligence officers). In the 14N career field, officers gain, on average, 35 types of experiences by the rank O-6. However, the typical O-6 only uses about four to seven of those types of experiences for his or her current job. The authors suggest that greater attention be given to the career paths of 14N officers so as to maximize the skills they gain for effective utilization in the future. As a proposal for future work, Brauner *et al.* (2009) recommend leveraging flow analysis in a similar manner as Vernez *et al.* (2006) to eliminate deficiencies in experience and decrease extraneous types of job experiences for AF intelligence officers.

## 2.5 Markov Decision Processes

Markov decision processes (MDPs) are also used as a modeling technique for MP problems. Gans & Zhou (2002) focus on finding the optimal hiring policy for a company given non-stationary service requirements. The model is formulated as a discrete-time continuous state space MDP. The authors utilize a state variable vector to represent the number of employees at different stages of the learning curve. Hiring

decisions in the model are made only at the entry skill level. The employee progresses through the skill levels stochastically as learning on the job occurs. Turnover, or loss of employees at a certain skill level, occurs at any stage of the learning curve. The authors employ the multinomial distribution to model the natural occurrence of learning and turnover because of the independence of the individual employees within the system. There are three types of costs used for the objective function: a fixed cost for hiring an employee, the wages for employees at each skill level, and the operating costs due to outsourcing and employee overtime pay when current worker levels cannot meet demand. The objective is to minimize the expected total costs for the company. The structure of the optimal policies for the system is examined and is shown to be “hire-up-to” for systems with stationary and stochastically increasing service requirements. The authors also show that myopic, or short-sighted, policies are optimal when considering systems with only one dimension for the state space.

McClellan (1991) applies a non-parametric semi-Markov model to an application of manpower planning for Northern Ireland nurses. The author examines a subsystem in which four grades of nurses are considered: student, staff nurse, sister, and higher grades. A traditional Markov model cannot be used because transitions for each time period are not equally likely. The state of the chain reflects the number of nurses by grade and duration in grade. Use of maximum likelihood estimators within the Markov framework allows the model to more accurately reflect the scale of the manpower planning system.

Guerry & De Feyter (2012) create a model that provides a trade-off between making decisions at the long-term upper-echelon management level of an organization and incorporating inputs from the tactical level in each department. The authors present a multi-level Markov manpower model and apply it to an example of a company with two departments and three worker types. Modeling both sides of the decision making

process results in decisions that address both general and limited company needs.

Manpower models are quite pertinent to the management of personnel in the military. Gass *et al.* (1988) developed the Army Manpower Long-Range Planning System (MLRPS) to help the military make long term projections of force levels. It is utilized for planning and analysis for policy, force structure, and cost. The elements of this tool include the data processing subsystem which collects and stores data relevant to force planning. The flow model subsystem utilizes Markov decision processes to take current states, or levels of personnel, and determine future force levels given current military service demands. The optimization model subsystem gives the user options concerning preferred force structure end goals and explores the necessary changes required to achieve those goals.

Škulj *et al.* (2008) apply manpower planning to the Slovenian armed forces. The authors develop an MDP model and utilize historical data to parameterize the transition probability matrix. They compare results obtained from the MDP model to a manpower projection to identify potential deficiencies in the force. Filinkov *et al.* (2011) develop a sustainability tool for modeling operational readiness requirements of the Australian Army. The model developed has two components: the units enabling components module and the personnel structure module. The first component focuses on the overall structure of the Australian Army to match units with varying levels of operational readiness to military operations. The next component models flow of personnel by recruitment and promotion for employment in various army units.

Dietz (1996) formulates an MDP model to optimize the number of officers receiving an advanced academic degree from the Air Force Institute of Technology (AFIT). The Air Force has a specific number of officers needed with certain educational backgrounds, such as Master's and PhD degrees. However, educating officers at the higher level is a significant investment. The MDP model determines the optimal number of

officers of specific ranks to educate as Master’s and PhD students to meet the requirements for officers of higher level education. Different scenarios reveal how the costs vary with different policies for admitting students to AFIT.

In the MP literature, there are two ways to characterize the flow in a military training model. Wang (2005) defines *pull flow* as when recruitment and promotion in an organization happen as gaps appear. In contrast, *push flow* does not consider vacancy restrictions and employees flow stochastically through the levels of the organization. In most manpower models, the authors consider only one flow approach depending on the nature of the problem. De Feyter (2007) creates a manpower model with both push and pull flows considered and analyzes the asymptotic nature of the model with certain assumptions. Modeling both types of flow simultaneously allows for greater precision of future manpower levels.

Knapp & Mahajan (1998) formulate a maintenance manpower model to schedule the appropriate maintenance specialist to the maintenance task. With the results from the MDP model, the authors create a simulation of the worker allocation problem and compare the results of the optimal policy to a benchmark policy. The optimal policy decreases costs because it requires fewer workers. The simulation tracked the metrics of how effectively workers are employed and how long a unit waits in a queue for maintenance for both the optimal and benchmark policies. The simulation with the optimal policy shows improvements in both metrics while utilizing fewer workers.

A limitation of MDPs, and many modeling techniques, is that the computational complexity of the problem increases exponentially as the state and action space of the problem grows. This is a problem discussed by De Bruecker *et al.* (2015) who suggest that MP problems that accurately reflect the size of the corporation require specific solution techniques. In the next sections, we discuss two methods authors have employed for solving large-scale MDPs: the use of heuristics and approximate

dynamic programming.

## 2.6 Heuristics

Heuristics are a popular method for solving problems considered computationally intractable. Gans & Zhou (2002) test an LP heuristic against the optimal policy found by the MDP method. The LP heuristic finds solutions within 1% of the optimal policy, performing very well. Within a mixed integer programming framework, Fowler *et al.* (2008) employ a genetic algorithm, a solution space partition, and linear programming heuristics to solve the problem of hiring, training, and firing employees of discrete skill levels. Genetic algorithms treat the set of current solutions as a population from which to generate new solutions. The solutions are ranked and matched with each other for reproduction of a new generation of solutions. Eventually, the solution population converges to optimality with careful selection of mating schemes for reproduction (Hill *et al.*, 2001). The solution space partition divides the workers into subsets based on their skill levels and considers cross training of employees only within subgroups. The optimal solution is found within those subgroups and combined to form the solution to the overall integer program. The linear programming heuristics fix the fractional values for the solutions to be integer. While the genetic algorithm had the longest computation time in experimentation, it provided the best results (Fowler *et al.*, 2008).

## 2.7 Approximate Dynamic Programming

Approximate dynamic programming (ADP) evolved as a methodology for solving large scale sequential decision problems. Powell (2012) describes how ADP developed in a few different academic communities such as simulation optimization, reinforcement learning, and stochastic programming in response to community-specific re-

search challenges. ADP becomes necessary when modelers run into the three “curses of dimensionality” of a large state space, action space, and/or outcome space. The methodology offers various algorithms and techniques to reduce the intractability of the problem such as cost function approximations, value function approximations, or aggregation of the state space. These algorithms employ Monte Carlo simulation to sample the random outcomes in the state and action space and determine the value of these outcomes. Some ADP algorithms that sample a sufficiently large state space are shown to converge to optimality (Powell, 2012). For example, Song & Huang (2008) employ an approximate dynamic programming technique known as the successive convex approximation method (SCAM) to obtain the optimal policy for their MDP model. The authors apply the SCAM algorithm to the MP problem by using piecewise linear approximations of contribution function. The SCAM algorithm solves a small MP problem nearly to optimality when compared to the exact solution.

Large MDP models are typically solved myopically, for some short time period, to simplify computations. Solving myopically means we do not consider future exogenous system changes when making the current period decision. Because of the structure of Bellman’s equation, an advantage of ADP is that it provides the framework to make decisions that are robust to future changes in the system. Powell *et al.* (2011) solve the aircraft fleet management problem of where to locate aircraft to serve known and random customer demands. The problem has many elements of uncertainty, to include: demand, equipment failure, and demand priority. The authors found that the ADP algorithm consistently outperformed the myopic policy when there is greater uncertainty in the problem.

Applications of ADP typically model a resource of multiple attributes (Powell, 2011). Simão *et al.* (2009) apply ADP to a fleet management problem for Schneider National trucking company. Within their ADP framework, they create a simulation

model of trucking operations with drivers of multi-dimensional attributes carrying loads across various regions. They model drivers as company resources, where the description of each driver is a vector that tracks attributes such as current location of driver, home of driver, and time at home. The authors' use of the attribute vector to describe the attributes of the truck driver informed the development of our notation in this thesis. The authors use ADP techniques such as a double pass algorithm, an optimal step size algorithm, and value function approximation to model and solve an extremely complicated, but realistically sized trucker assignment problem instance. The goal of the study was to match the performance of experienced trucking dispatchers with the performance of their ADP policy. Schneider National uses the authors' approach to test various policies for ensuring driver satisfaction in regards to minimizing time away from home, hiring drivers for specific regions, and setting up pickups for loads.

Another application of ADP by Simao & Powell (2009) models an airline's spare parts replacement network. Various types of parts and different costs and times for repair of those parts complicate the problem of ensuring expedient repair. Simao & Powell (2009) examine the network of central and distribution centers and demand locations. The states include the inventory of parts and the knowledge of the status of the part, which changes over time. Similar to Simão *et al.* (2009), the authors use attribute vectors to describe the state space, which informed the formulation of our model. The decisions are whether to buy, move, or replace parts depending on their status after random equipment failures or delays. The authors note that high-cost repair may be a rare event over the course of a simulation. Therefore, a conditional marginal value for the event of a large cost failure is explicitly developed to ensure those events are properly considered in the model. To simplify calculations for the large replacement problem, Simao & Powell (2009) utilize the post-decision state variable,

which allows actions to be chosen without having to compute the expectation of the value function (Powell, 2011). The post-decision state is utilized in the algorithms employed in this thesis.

Approximate value iteration is an ADP algorithmic strategy that employs value function approximation. When the value function is monotonic, piecewise-linear concave approximations give good estimates of the true function. Powell (2011) outlines three algorithms which approximate the true value function well: SPAR, levelling, and concave adaptive value estimation (CAVE). All three algorithms iteratively update a value function approximation while maintaining monotonicity. If an update to the approximation produces a violation of monotonicity, the leveling algorithm forces the offending value to a monotonic value (Topaloglu & Powell, 2003). SPAR averages any new approximations which violate monotonicity (Powell *et al.*, 2004). In contrast to the leveling and SPAR algorithms, CAVE handles violations of monotonicity by expanding the range of the function being approximated to incorporate the non-monotonic value. The leveling and SPAR algorithms are proved to converge to the true value function (Powell, 2011) and variants of CAVE also have convergence proofs (Topaloglu & Powell, 2003).

Topaloglu & Powell (2006) use value function approximation to solve the problem of serving customers with a fleet of business jets. The fleet includes jets of different types and the demands to be flown from location  $i$  to location  $j$  must be served. The value of serving the customer is a concave function. Therefore, the authors approximate that function as a separable, piecewise-linear, concave function. In experimentation, the authors utilize three different value function approximation strategies. The hybrid value function approximation performed best for the deterministic experiments. For the stochastic experiments, the linear and piecewise-linear approximations outperform the rolling horizon strategy, which is common for stochastic MDPs.

Approximate dynamic programming techniques have previously been applied to the AFO-MPP in the work by Hoecherl (2015). Hoecherl (2015) utilize the least-squares temporal differencing (LSTD) algorithm and the CAVE algorithm in testing. The study considers accessions and promotions decisions for multiple AFSCs, officer grades, and year groups. The study finds moderate success with the LSTD algorithm under certain conditions. However, the author suggests further study of different basis functions for the algorithm to give better solution quality. The CAVE algorithm converges to optimal solutions despite the non-linearity involved due to interactions between the accessions and promotions decisions. However, the author cites this as a limitation of the algorithm and suggests exploring alternatives that overcome this drawback.

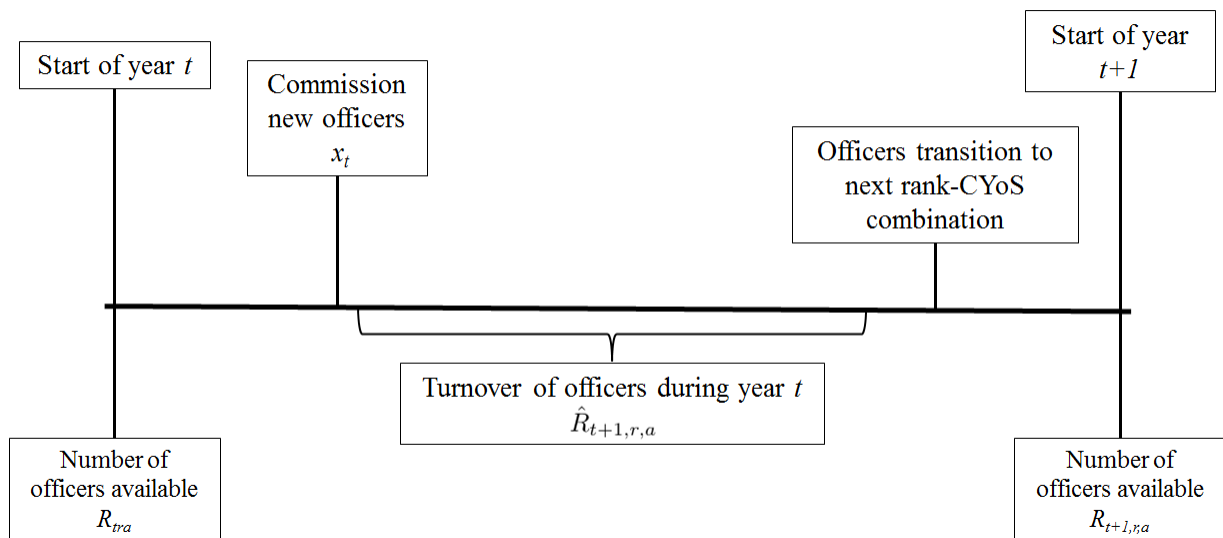
## III. Methodology

### 3.1 Problem Statment

We seek to solve the Air Force Officer Manpower Planning Problem (AFO-MPP) by assessing the optimal number of officers over an infinite time horizon. We formulate a multi-class inventory control model of the AFO-MPP. The inventory control model is a discrete-time, discrete state space Markov decision process (MDP) wherein the state is the number of officers in each level of the organization. The stochastic nature of the MDP allows us to study how the manpower capabilities of the USAF will evolve over time with specific conditions. In particular, the system varies with regards to the promotion and turnover rates of officers from one year group to the next. The optimal policy minimizes the costs of maintaining a certain number of officers in the system.

For our formulation of the AFO-MPP as an MDP, the state space of the model captures the number of officers within the AFSC at each specific year group and rank combination. The tuple of year group and rank is chosen for the state space to reflect the number of years the officer has remained in the system and the military pay grade of the officer. A previous study of the officer sustainment problem by Hoecherl (2015) includes the AFSC of the officer in the tuple, while we focus on one AFSC for this thesis. Our study considers the 61A career field, also known as an Operations Analyst. We further limit the state space by only considering ranks O-1 through O-6 in our problem, as the general officer ranks have unique structure in comparison to the lower set of ranks. The action space of the problem captures how many officers to admit into the system at the lowest level, reflecting how military accessions occur at the O-1 rank. Officers accessed at time period  $t$  are assumed ready for duty at that time. We simplify the real life accession process in that regard because officers are

accessed at points throughout the year depending on the commissioning source. We then model the transition of officers from one year to the next, taking into account promotions and turnover of officers throughout the year. The event timing diagram in Figure 1 displays this transition of officers through the AFO-MPP.



**Figure 1. Event Timing Diagram for AFO-MPP**

The AFO-MPP considers the costs as two separate elements, the cost of over-manning and cost of under-manning. These are two “soft” costs which many MP models take into account. Over-manning causes problems of redundancies in the workplace and job dissatisfaction for the perceived lack of fulfilling work (Galway *et al.*, 2005). Under-manning leads to outsourcing work that cannot be completed in the workplace, which incurs costs to hire outside organizations to complete (Gans & Zhou, 2002).

The objective of the MDP is to determine a hiring policy that minimizes the expected total discounted cost. A policy is a set of decision rules for each time period that indicates how many O-1’s to admit into the Air Force given the current state of the system. The formulation of the AFO-MPP as a discrete-time, discrete-state

Markov decision problem is informed in part by Gans & Zhou (2002).

### 3.2 MDP Formulation

The MDP model for the AFO-MPP is described as follows:

Decisions are made annually about the hiring of new officers into the system. The set of decision epochs is denoted as follows:

$$\mathcal{T} = \{1, 2, \dots, T\}, T \leq \infty.$$

The epochs are the points in time at which decisions are made. In this case, the accession decisions are made at the beginning of the year.

The full state space of the AFO-MPP system is comprised of the number of officers with certain AFSC, rank, and year group elements. For this thesis, we propose an aggregate officer replacement model where we express the rank and commissioned years of service (CYoS) of the officer. We consider just one AFSC to simplify the state space. Let

$$r \in \mathcal{R}^r = \text{rank of an officer,}$$

where  $\mathcal{R}^r = \{1, 2, \dots, 6\}$  denotes the set of ranks. We utilize six categories of rank to represent officer ranks O-1 through O-6. We do not consider the general officers in the scope of this thesis. Let

$$a \in \mathcal{R}^a = \text{the commissioned years of service of an officer,}$$

where  $\mathcal{R}^a = \{1, 2, \dots, 29\}$  denotes the set of CYoS. While some officers enter the system with enlisted years of service, we only model commissioned years of service when considering how long an officer has served in the Air Force.

The set  $\mathcal{R}$  contains the full scope of possible combinations of ranks  $r$  and CYoS  $a$ . Let

$$(r, a) \in \mathcal{R} = \text{Set of all possible officer rank-CYoS combinations,}$$

where  $\mathcal{R} = \mathcal{R}^r \times \mathcal{R}^a$ . We note that not every rank-CYoS combination is feasible. For example, an officer of rank O-3 would not be in the system with 28 CYoS.

The state of the system is determined by the number of officers of rank  $r$  and CYoS  $a$ . Let

$$R_{tra} = \text{the number of officers at time } t \text{ of rank } r \text{ and CYoS } a.$$

Note that  $R_{tra} \in \mathbb{N}^0$  cannot take on negative values because of the nature of the personnel system. The pre-decision state is a vector denoted as

$$R_t = (R_{tra})_{(r,a) \in \mathcal{R}}.$$

The action at time  $t$  indicates the number of officers to be commissioned into rank  $r = 1$ . Let

$$x_t = \text{number of officers commissioned into the AF at time } t.$$

The action reflects how many military officers enter the AF at rank O-1 (with no experience). Thus, these officers are considered to have zero CYoS.

In the military, key factors that influence the probability of turnover, or the officer leaving the military, are the rank and CYoS of the officer. Let

$\hat{R}_{tra}$  = the number of officers of rank  $r$  and CYoS  $a$   
that turnover in the time interval  $(t - 1, t)$ .

The random variable  $\hat{R}_{tra}$  follows a binomial distribution with parameters  $R_{tra}$  and  $\rho_{ra}$  because the number of officers leaving the system depends on the number of officers currently in the system. Let

$\rho_{ra}$  = the probability an officer of rank  $r$  and CYoS  $a$   
will turnover in the time interval  $(t - 1, t)$ .

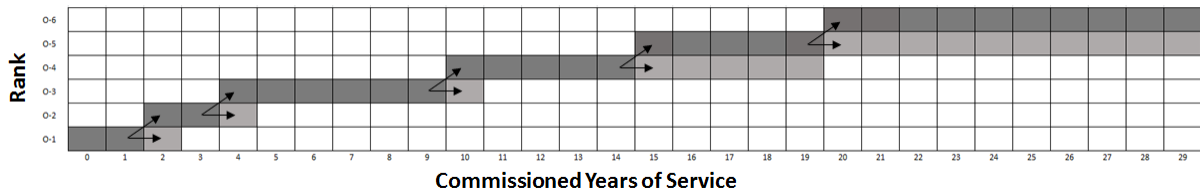
This parameter is indexed by  $(r, a)$  because the probability of an officer leaving the system changes with different rank-CYoS combinations.

The number of officers of rank  $r$  and CYoS  $a$  available at time  $t + 1$ ,  $R_{t+1,r,a}$ , results from the number of officers of rank  $r$  and age  $a - 1$  in the system at time  $t$ ,  $R_{t,r,a-1}$ , the number of new officers accessed at epoch  $t$ ,  $x_t$ , and the number of officers of type  $r$  that turnover during time interval  $(t, t + 1)$ ,  $\hat{R}_{t+1,r,a-1}$ . Moreover, we consider officer promotions. We model the turnover and promotions system via the following transition function:

$$R_{t+1,r,a} = \begin{cases} \lceil v_{ra}(R_{t,r-1,a-1} - \hat{R}_{t+1,r-1,a-1}) \rceil & \text{if } (r, a) \in \mathcal{R}^p, \\ (R_{t,r,a-1} - \hat{R}_{t+1,r,a-1}) - \lceil v_{r+1,a}(R_{t,r,a-1} - \hat{R}_{t+1,r-1,a-1}) \rceil & \text{if } (r, a) \in \mathcal{R}^n, \\ x_t & \text{if } (r, a) = (1, 1) \\ R_{t,r,a-1} - \hat{R}_{t+1,r,a-1} & \text{otherwise.} \end{cases}$$

The parameter  $v_{ra}$  is the proportion of officers of rank  $r - 1$  and CYoS  $a - 1$  selected

for promotion to rank  $r$ . The set  $\mathcal{R}^p = \{(2, 2), (3, 4), (4, 10), (5, 15), (6, 20)\}$  denotes the rank-CYoS combinations at which “in-the-zone” promotions occur. The set  $\mathcal{R}^n = \{(1, 2), (2, 4), (3, 10), (4, 15), (5, 20)\}$  denotes the rank-CYoS combinations that receive officers passed up for promotion. In the third case of the function, the officers accessed at time  $t$  transition into rank O-1 with one CYoS. The last case accounts for the transitions of any rank-CYoS combination not previously specified. Figure 2 helps to visualize these transitions.



**Figure 2. Officer Progression Throughout AFO-MPP**

The transition of officers in the AFO-MPP can be written in the following system dynamics form:

$$R_{t+1} = R^M(R_t, x_t, \hat{R}_{t+1}).$$

The cost for the AFO-MPP incorporates the “soft” cost of over-manning and under-manning in a workforce. Manpower planning models typically consider how to hire employees to fit set service requirements (Gans & Zhou, 2002). There are certain staffing requirements that the organization has prescribed over time. To reflect this requirement in the model, the decision maker sets a value  $\bar{R}_r$  for the number of officers required. These values are indexed by  $r$  because there are target levels for each officer rank  $r \in \mathcal{R}^r$ . Let  $c_r^o$  denote the cost associated with an over-manned workforce of rank  $r$  and  $c_r^u$  for that of an under-manned workforce of rank  $r$ . The cost parameters  $c_r^o$  and  $c_r^u$  take on positive values to penalize the objective function. These parameters are indexed by  $r$  to reflect that it is more costly for higher ranks to have shortages and surpluses than others. The cost function for officer surpluses is

$$O_r(R_t, x_t) = \begin{cases} c_r^o(\max\{x_t + \sum_{a \in \mathcal{A}^r} R_{t,r,a} - \bar{R}_r, 0\})^{\beta^o} & \text{if } r = 1, \\ c_r^o(\max\{(\sum_{a \in \mathcal{A}^r} R_{t,r,a}) - \bar{R}_r, 0\})^{\beta^o} & \text{if } r \neq 1. \end{cases}$$

There is a separate cost function for O-1 officer surpluses because we must include officers accessed during time period  $t$ . The cost equation for officer surpluses in ranks O-2 through O-6 are of the same form.

The cost function for officer shortages is

$$U_r(R_t, x_t) = \begin{cases} c_r^u(\max\{\bar{R}_r - (x_t + \sum_{a \in \mathcal{A}^r} R_{t,r,a}), 0\})^{\beta^u} & \text{if } r = 1, \\ c_r^u(\max\{\bar{R}_r - (\sum_{a \in \mathcal{A}^r} R_{t,r,a}), 0\})^{\beta^u} & \text{if } r \neq 1. \end{cases}$$

Again, there is a separate cost function for O-1 officer shortages because we must include officers accessed during time period  $t$ . The cost equation for officer shortages in ranks O-2 through O-6 are of the same form. Therefore, the single period cost function for the AFO-MPP is the sum of the surplus and shortage costs. The function is denoted as follows:

$$C(R_t, x_t) = \sum_{r \in \mathcal{R}^r} O_r(R_t, x_t) + U_r(R_t, x_t).$$

Having described the components of our MDP model, we can now state the objective function for the AFO-MPP. Let

$$V(R_t) = \min_{\pi \in \Pi} \mathbb{E} \left\{ \sum_{t=0}^{\infty} \gamma^t C(R_t, x_t) \right\}.$$

We seek a policy  $\pi \in \Pi$  that minimizes the expected total discounted cost criterion. The optimal policy is the set of decision rules for each time period that provides the decision maker with the least costly officer accession program. Solving Bellman's

equation provides the optimal policy:

$$V(R_t) = \min_{x_t} \left( C(R_t, x_t) + \gamma \mathbb{E} V_{t+1}(R^M(R_t, x_t, \hat{R}_{t+1})) | R_t \right).$$

### 3.3 Approximate Dynamic Programming Algorithms

#### Least Squares Temporal Differences.

The AFO-MPP is solved using an ADP technique because our model has high dimensionality. ADP provides a mechanism to approximate the value function without having to compute each state-action pair. ADP algorithms employ Monte Carlo simulation to sample the random outcomes in the state and action space and determine the value of these outcomes. Some ADP algorithms that sample a sufficiently large state space are shown to converge to optimality (Powell, 2012). The use of Monte Carlo simulation alleviates the need to solve for the value for each state-action pair. Utilization of the post-decision state convention can reduce the computational complexity in ADP algorithms (Powell, 2011). The post-decision state,  $R_t^x$ , considers the state of the system immediately prior to the revelation of exogenous changes to the system, allowing the expectation to be computed outside of the minimization operator. Bellman's equation is represented as follows when the post-decision state is incorporated:

$$V_{t-1}^x(R_{t-1}^x) = \mathbb{E} \left\{ \min_x C(R_t, x) + \gamma V_t^x(R_t^x) | R_{t-1}^x \right\}. \quad (1)$$

To simplify calculations for the large replacement problem, Simao & Powell (2009) and Simão *et al.* (2009) utilize the post-decision state variable for a noted computational cost savings.

Approximate policy iteration (API) is an ADP algorithmic strategy that evaluates

the values associated with states and outcomes for a fixed policy, or set of actions, for an MDP problem. The policy iteratively updates based on the the observed values of the fixed policy. The advantages of API are strong convergence theory and avoiding the computational burden of calculating each state-action pair (Powell, 2012). There is a focus in ADP literature on using API with parametric modeling and linear basis functions (Powell, 2011). With these mechanisms, linear regression techniques are applied to estimate a parameter vector  $\theta$  to fit a model of value functions using the selected basis functions, or features (Bertsekas & Tsitsiklis, 1995). The value function approximation utilized in our API implementation leverages the post-decision state variable convention and is denoted below:

$$\bar{V}_t^x(R_x^t) = \sum_{f \in \mathcal{F}} \theta_f \phi_f(R_t^x) = \theta^\top \phi(R_t^x) \quad (2)$$

where  $\phi_f(R_t^x)$  is the vector of basis functions for the post-decision state (Scott *et al.*, 2014). Substituting this equation into Bellman’s equation gives the foundation for least squares value function approximation:

$$\theta^\top \phi(R_{t-1}^x) = \mathbb{E}[C(R_t, X^\pi(R_t|\theta)) + \gamma \theta^\top \phi(R_t^x) | R_{t-1}^x], \quad (3)$$

where  $X^\pi$  is the policy function for the MDP model.

The least squares temporal differencing algorithm is an on-policy algorithm outlined by Bradtke & Barto (1996) that minimizes the sum of the temporal differences, or Bellman’s error, for approximating the estimation of the true value function. Minimizing Bellman’s error minimizes the difference between the approximation of the value function and the observed value of the approximations. The estimator for the

least squares Bellman error minimization is as follows:

$$\hat{\theta} = [(\Phi_{t-1} - \gamma\Phi_t)^\top (\Phi_{t-1} - \gamma\Phi_t)]^{-1} (\Phi_{t-1} - \gamma\Phi_t)^\top C_t, \quad (4)$$

where  $\Phi_{t-1}$  is a matrix of values for the basis function for the sampled post-decision states,  $\Phi_t$  is a matrix that records basis function values for the post-decision state in the next period, and  $C_t$  is a vector that records the observed contributions of the period in each iteration. The full LSTD algorithm is given by Scott *et al.* (2014) and is outlined in Algorithm 1.

---

**Algorithm 1** API Algorithm

---

- 1: **Step 0:** Initialize  $\theta^0$ .
  - 2: **Step 1:**
  - 3: **for** n=1 **do to**  $N$  (**Policy Improvement Loop**)
  - 4:     **Step 2:**
  - 5:     **for** m=1 **do to**  $M$  (**Policy Evaluation Loop**)
  - 6:         Generate a random post-decision state,  $R_{t-1,m}^x$ .
  - 7:         Record  $\phi(R_{t-1,m}^x)$ .
  - 8:         Simulate transition to next event, obtain a pre-decision state,  $R_{t,m}$ .
  - 9:         Determine decision  $x = X^\pi(R_{t,m}|\theta^{n-1})$ .
  - 10:         Record contribution  $C(R_{t,m}, x)$ .
  - 11:         Record basis function evaluation  $\phi(R_{t,m}^x)$
  - 12:     **End**
  - 13:     Compute  $\theta^n$  using LSTD or IV and smoothing rule.
  - 14: **End**
- 

When generating the random sample of post-decision states for Step 2 of Algorithm 1, we employ Latin hypercube sampling (LHS). Because we have a high-dimension state space, use of LHS allows for uniform sampling across all dimensions. This improves solution quality by allowing the  $\theta$  coefficients to more clearly identify the regressors affecting the value function (McKay *et al.*, 1979).

A variation on the least squares temporal differencing algorithm includes instrumental variables in the approximation of  $\hat{\theta}$  (Scott *et al.*, 2014). Instrumental variables

obtain consistent estimates because they are correlated with the regressors, or the basis functions, but not correlated with the errors of the regressors or the observations of the contributions (Bowden & Turkington, 1990). Bradtke & Barto (1996) and Scott *et al.* (2014) use instrumental variables within an approximate policy iteration algorithm. The estimator in the instrumental variables Bellman error minimization is as follows:

$$\hat{\theta} = [(\Phi_{t-1})^\top (\Phi_{t-1} - \gamma \Phi_t)]^{-1} (\Phi_{t-1})^\top C_t. \quad (5)$$

## IV. Computational Results

In this chapter, we apply the approximate dynamic programming (ADP) techniques outlined in Chapter 3 to two representative problem instances of the U.S. Air Force officer manpower planning problem (AFO-MPP). We investigate different features of the ADP algorithm to demonstrate the effectiveness of our proposed procedure on a small instance of the problem. To clearly investigate the impact of varying certain algorithmic features on solution quality and computational effort, we design and conduct a computational experiment. The experiments are run using MATLAB. Using information gathered from testing algorithmic feature settings on our small problem instance, we investigate three notional scenarios for the AFO-MPP. We execute experimental designs on each of these scenarios to test the performance of our ADP algorithm.

### 4.1 Small Problem Instance

The problem instance we outline in this section considers a subsection of the full AFO-MPP to maintain reasonable computational effort. As discussed in Chapter 3, we examine the Air Force Specialty Code of 61A, Operations Research Analyst. We examine a problem with two categories of officer rank  $r$ . By this convention, we divide rank into the categories of company grade officer (CGO), O-1 through O-3, and field grade officer (FGO), O-4 through O-5. We also consider 20 categories of commissioned years of service (CYoS),  $a \in \{1, 2, \dots, 19\}$  and a separate category for newly accessed officers with zero CYoS. Therefore, we model up to 40 different rank-CYoS combinations for the AFO-MPP. We note that certain rank-CYoS combinations are infeasible. For example, there would not be an FGO in the system with two CYoS. Having many dimensions to the state space makes the AFO-MPP a large-scale

dynamic programming problem. However, because we are using ADP techniques, we are not seeking exact solutions to the problem.

We set the parameters for the cost functions in this instance of the AFO-MPP based on examination of the literature on manpower planning for the USAF. The target levels of officers for each rank-CYoS combination,  $\bar{R}_r$ , are derived from information presented by analysts at HAF-A1. After extensive testing of the “soft” cost parameters for over-manning,  $c_r^o$ , and under-manning,  $c_u^r$ , we set those values to small positive values to penalize the cost function for shortages and surpluses. We employ a discount factor ( $\gamma$ ) of 0.95 in the objective function.

The parameters for the transition functions of the AFO-MPP are based on literature concerning an AF officer’s career and discussions with AF personnel at HAF-A1. The probability of promotion for an officer in each rank-CYoS category,  $v_{ra}$ , is based on the information from a typical US military officer’s career, gathered from Rostker *et al.* (1993) and Hosek (2001). Officers in the military typically promote automatically from ranks O-1 to O-2. The subsequent promotions are more competitive as the officer must meet promotion boards to advance to the next officer rank. Each officer grade has an associated promotion opportunity percentage and typical promotion timing that dictates promotions. To parameterize retention rates,  $\rho_{ra}$ , for each rank-CYoS combination, we incorporate information provided by HAF-A1 on retention in the 61A career field.

### **Experimental Design.**

We construct a set of experiments to evaluate the proposed ADP algorithm’s solution quality and computational effort by studying the impact of systematically varying different features of the ADP algorithm and certain MDP parameters (Montgomery, 2008). The policy resulting from the ADP algorithm is assessed based on its

improvement over a benchmark policy for the AFO-MPP. The benchmark policy is set based on information provided by HAF-A1 on the 61A Career Field. The response variable for our design of experiments is reported as the percentage improvement (i.e., decrease in cost) over the benchmark policy. The halfwidth for this value is reported at the 95% confidence level. The computation times for the ADP algorithm are recorded to measure the computational effort needed to perform the ADP algorithm.

We investigate five algorithmic features in our design of experiments for the AFO-MPP. The first algorithmic feature we examine is the effect of instrumental variables on the performance of the algorithm. This is a two level categorical variable, indicating whether the instrumental variables are used or if Bellman error minimization is solely employed. The next factor we investigate is a categorical variable of the type of basis functions utilized. The two options for basis function features are second order basis functions with selected interaction terms or fourth order basis functions also with selected interaction terms. These interaction terms are described as “selected” because only specific rank-CYoS combination interactions, such as those of subsequent rank-CYoS combinations, are anticipated to have an impact on algorithmic performance. For example, including the interaction between a CGO of five CYoS compared to a CGO of six CYoS in our basis function would give more information about the AFO-MPP than the interaction between a CGO of five CYoS compared to an FGO of 18 CYoS. The number of policy improvement (outer) loops (N) are set to values of 10 and 30. The number of policy evaluation (inner) loops (M) are set to values of 1,000 and 3,000. The final algorithmic feature we examine is the parameter of the generalized harmonic smoothing function,  $a$ . The formula for generalized harmonic smoothing is as follows:

$$\frac{a}{a + n - 1}.$$

We choose a value of one for the low factor setting for  $a$  to study how simple harmonic smoothing affects our response and a value of  $a = 30$  for the high setting. This allows us to slow the algorithm convergence (Powell, 2011). To summarize our discussion, the algorithmic features and their levels for the small problem design of experiments are displayed in Table 1.

**Table 1. Factor Settings for Small Problem Experimental Design**

Factor	Low (-1)	High (+1)
Instrumental Variables	Not Present	Present
Basis Functions	Second Order with Interactions	Fourth Order with Interactions
Policy Improvement Loops ( $N$ )	10	30
Policy Evaluation Loops ( $M$ )	1,000	3,000
Harmonic Smoothing ( $a$ )	1	30

A  $2^{5-1}$ , resolution V, fractional-factorial design is implemented. This design investigates the five selected features in 16 runs. A resolution V design specifies that main effects and two-factor interactions are not aliased with each other. However, some two-factor interactions may be confounded with three factor interactions (Montgomery, 2008). When applying this experimental design, we create our ADP policy by calculating the  $\theta$  coefficients for our basis functions from the implementation of our ADP algorithm. Once we have the  $\theta$  coefficients, we simulate both the ADP policy and our benchmark policy over a 40 year horizon for 100 replications per treatment to obtain the statistics of our response variables. The simulation starts out with the target level of officers in each rank-CYoS combination.

## Experimental Results.

Table 2 reports the results from the experimental design. The best performing design point (i.e., set of algorithmic feature parameters) is highlighted in Run 5. It provides a trade-off between solution quality and computational effort. In order to provide degrees of freedom for error in our analysis of the model created in this experimental design, we analyze our model with a replication of each design point. The  $R^2$  for the model created with this experimental design is 0.96936 and the  $R^2_{adj}$  is 0.94065. This shows the model provides a good fit for the response, indicating the parameters chosen explain the variation in the response. Table 3 displays the pa-

**Table 2. Small Problem Experimental Results**

Run	Instrumental Variables	Basis Functions	Outer (N)	Inner (M)	Smoothing (a)	Percent Improvement	Time (In Sec.)
1	-	-	-	-	+	37.11±1.21%	164.62
2	-	-	-	+	-	14.07±1.63%	484.90
3	-	-	+	-	-	39.55±1.47%	488.07
4	-	-	+	+	+	39.56±1.34%	1475.38
5	-	+	-	-	-	39.15±1.29%	177.18
6	-	+	-	+	+	33.01±1.40%	518.65
7	-	+	+	-	+	39.38±1.47%	508.0
8	-	+	+	+	-	24.04±1.54%	1518.60
9	+	-	-	-	-	-765.12±13.33%	168.21
10	+	-	-	+	+	-316.83±6.55%	485.97
11	+	-	+	-	+	-574.87±12.21%	487.88
12	+	-	+	+	-	-769.21±14.02%	1518.97
13	+	+	-	-	+	-412.30±9.36%	176.01
14	+	+	-	+	-	-722.79±15.51%	521.01
15	+	+	+	-	-	-675.17±12.82%	509.40
16	+	+	+	+	+	-664.06±11.70%	1582.52

parameter estimates for the model created with this experimental design. The factor with the most influence on the response is the categorical factor of instrumental variables. Our ADP algorithm performs poorly when instrumental variables are utilized, a surprising result as past research with instrumental variables has improved least squares temporal differencing performance. The responses when instrumental variables are not utilized all outperform the benchmark policy. Additionally, we note that

the algorithm generally performs better for high levels of  $a$  in generalized harmonic smoothing. The interaction term of the categorical variables of instrumental smoothing and the continuous variable for the parameter of the harmonic smoothing is also significant. The results we gathered for this experimental design guides subsequent analysis on larger AFO-MPP instances, presented later in this chapter.

**Table 3. Small Problem Parameter Estimates**

<b>Term</b>	<b>Estimate</b>	<b>Prob&gt; t </b>
IV[L1]	347.78	<0.0001
BF[L1]	-4.10	0.7979
N	-13.89	0.3910
M	-12.14	0.4518
a	37.76	0.0291
IV[L1]*BF[L1]	1.79	0.9109
IV[L1]*N	20.23	0.2174
IV[L1]*M	9.52	0.5541
IV[L1]*a	-36.64	0.0335
BF[L1]*N	10.64	0.5087
BF[L1]*M	20.53	0.2109
BF[L1]*a	10.01	0.5340
N*M	-10.97	0.4959
N*a	-17.84	0.2740
M*a	-4.22	0.7921

## 4.2 Full Problem Instance

The problem instance we outline in this section considers the full AFO-MPP. As discussed in Chapter 3, we examine the Air Force Specialty Code of 61A, Operations Research Analyst. We examine a problem with six categories of officer rank  $r$ . We also consider 30 categories of commissioned years of service (CYoS),  $a \in \{1, 2, \dots, 29\}$  and a separate category for officers accessed with zero CYoS. Therefore, we model up to 180 different rank-CYoS combinations for the AFO-MPP. We note that certain rank-CYoS combinations are infeasible. For example, there would not be an officer of rank O-5 in the system with two CYoS. Having many dimensions to the state space

makes the AFO-MPP a large-scale dynamic programming problem. However, because we are using ADP techniques, we are not seeking exact solutions to the problem.

As with the small problem instance discussed earlier in this chapter, we set the parameters for the cost and transition functions in this instance of the AFO-MPP based on examination of the literature on manpower planning for the USAF and discussions with AF personnel at HAF-A1. The target levels of officers for each rank-CYoS combination,  $\bar{R}_r$ , are derived from information presented in a 61A Career Field overview. After extensive testing of the “soft” cost parameters for over-manning,  $c_r^o$ , and under-manning,  $c_u^r$ , we set those values to small positive values to penalize the cost function for shortages and surpluses. The probability of promotion for an officer in each rank-CYoS category,  $v_{ra}$ , is based on the information on a typical US military officer’s career progression. We also employ a discount factor of  $\gamma=0.95$  in the objective function.

We now outline the testing that is conducted on this larger instance of the AFO-MPP. We perform three designed experiments to investigate how our ADP algorithm performs with different problem and algorithmic features. Each designed experiment (with one experimental design per scenario) represents a different realistic manpower planning situation that the Air Force may face. The first scenario we explore is the case where there is a high number of CGO’s in the system already. We investigate the performance of the ADP policy at adapting to this setback. Similarly, our next scenario is a case where the CGO ranks are under-manned. Our last scenario, like our policy simulation in the small problem instance, investigates the ADP policy performance when the target levels of officers are already present in the system.

To provide more insight into realistic costs that are incurred in manpower planning scenarios, we report the “hard” cost of employing officers in the AFO-MPP. This cost is the fully burdened DoD cost for employing an officer of a certain rank. The data for

the fully burdened cost is gathered from a testimony before the National Commission of the Structure of the Air Force (Jackson, 2013). Let  $w_{ra}$  denote the fully burdened DoD annual cost for an officer of rank  $r$  and CYoS  $a$ . The cost associated with the wage is denoted as follows:

$$W_r(R_t, x_t) = \begin{cases} w_{r1}x_t + \sum_{a \in \mathcal{A}^r} w_{ra}R_{tra} & \text{if } r = 1, \\ \sum_{a \in \mathcal{A}^r} w_{ra}R_{tra} & \text{if } r \neq 1. \end{cases}$$

The set  $\mathcal{A}^r$  indicates all feasible CYoS for a particular rank  $r$ . There is a separate wage function for officer rank  $r = 1$  because we must include officers accessed during time period  $t$ , as indicated by  $x_t$ . The cost equation for officer wages in ranks O-2 through O-6 are of the same form. We report the fully burdened cost with the other responses in our experimental design.

### **Experimental Design.**

We construct a set of experiments to evaluate the proposed ADP algorithm's solution quality and computational effort for each of the three scenarios outlined earlier. We employ a design of experiments to systematically study the impact of varying different algorithmic and problem features (Montgomery, 2008). The policy resulting from the ADP algorithm is assessed based on its improvement over a benchmark policy for the AFO-MPP. The benchmark policy is set based on information provided by HAF-A1 on the 61A Career Field. The response variable for our design of experiments is reported as the percentage improvement (i.e., decrease in cost) over the benchmark policy. The halfwidth for this value is reported at the 95% confidence level. The mean fully burdened DoD cost associated with maintaining officers in the system over the specified time horizon is another response. The computation times for the ADP algorithm are recorded to measure the computational effort needed to

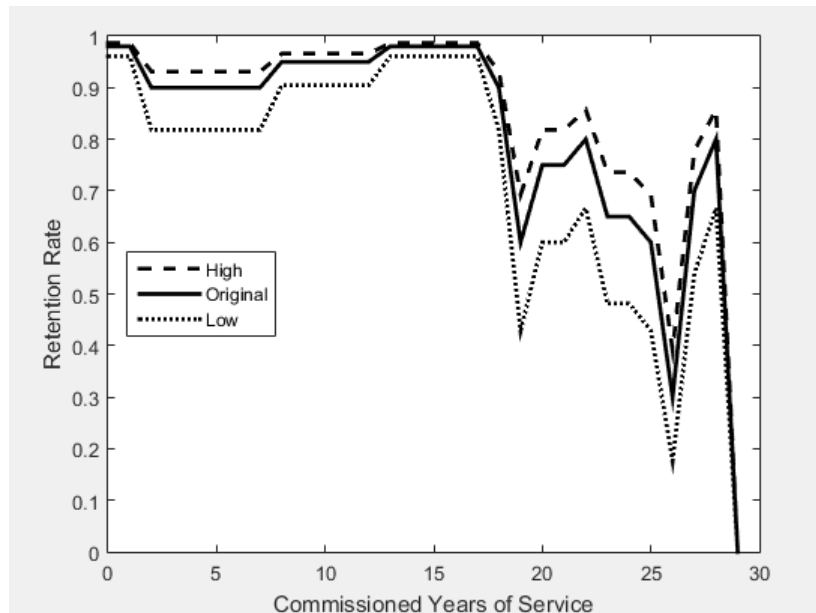
implement the ADP algorithm.

For each of the three scenarios for the AFO-MPP, we investigate three algorithmic features and one problem feature of interest to test the performance of the ADP policy relative to the benchmark policy. The first two algorithmic features we vary are the number of policy improvement loops (N) and the number of policy evaluation loops (M). In the small problem instance, we found that higher computational effort did not necessarily produce better results. However, because the full problem instance has very high dimensionality, we preliminarily investigate higher values for the policy improvement and evaluation loops to see if solution quality increases with computational effort. These longer runs were not found to provide statistically significant improvement in solution quality, so therefore, we keep the same number of policy evaluation loops as in the small problem instance. This provides a trade-off between computational effort and adequate sampling of the large state space. The last algorithmic feature we investigate is the parameter for the generalized harmonic smoothing function,  $a$ . From our results in the small problem instance and initial investigations of the full problem instance, we find that more aggressive smoothing outperforms simple harmonic smoothing. Therefore, we choose the values of  $a = 10$  and  $a = 100$  for our experimental design.

The problem feature of interest for our experimental design is the retention parameter,  $\rho_{ra}$ , for the AFO-MPP. The original  $\rho_{ra}$  incorporates information provided by HAF-A1 on retention in the 61A career field. To investigate how the benchmark and ADP policies are affected when the system has higher or lower retention rates than usual, we adjust  $\rho_{ra}$ . To accomplish this, we convert the original retention rate to odds by the following formula:

$$q_{ra} = \frac{\rho_{ra}}{1 - \rho_{ra}}.$$

To get the high level factor setting for the retention rate, we then take  $\frac{(q_{ra}+(0.5 \cdot q_{ra}))}{(q_{ra}+(0.5 \cdot q_{ra}))+1}$ . For already high retention rates, this procedure allows us to remain within an acceptable probability range. The difference in retention rates with this adjustment is more evident for  $\rho_{ra}$  values that were already low. The low level factor setting for the retention rate is found using the formula  $\frac{(q_{ra}-(0.5 \cdot q_{ra}))}{(q_{ra}-(0.5 \cdot q_{ra}))+1}$ . Figure 3 depicts the adjusted retention rates in relation to the original retention rate provided. Varying this parameter simulates the situations of higher number of officers remaining in the service or many officers leaving the service.



**Figure 3. Retention Factor Levels**

A  $2^4$  factorial design is implemented. This design investigates the four selected features in 16 runs. When applying this experimental design, we create our ADP policy by calculating the  $\theta$  coefficients for our basis functions from the execution of our ADP algorithm. In the initial investigations of the full problem instance, we test each type of basis function, first order through fourth order with interactions, to see which provides the best response. We do this because the experimental design performed on the small problem instance did not provide statistically significant insight for which

order of basis functions to utilize. Instrumental variables performed poorly in the small problem instance for the AFO-MPP so we utilize Bellman error minimization for the LSTD point estimator. Throughout testing of these three scenarios, we utilize fourth order basis functions with the selected interactions because it performed best in these initial investigations. Once we have the  $\theta$  coefficients, we simulate both the ADP policy and our benchmark policy over a 40 year horizon for 30 replications per treatment to obtain the statistics of our response variables. Table 4 displays the features we investigate and their associated levels.

**Table 4. Factor Settings for Full Problem Experimental Design**

Factor	Low (-1)	High (+1)
Number of Outer Loops (N)	10	30
Number of Inner Loops (M)	1,000	3,000
Harmonic Smoothing (a)	10	100
Retention Rate ( $\rho_{ra}$ )	Decreased by 50%	Increased by 50%

### **Over-manned Scenario.**

In this section, we discuss the results of the over-manned scenario experimental design for the full problem instance of the AFO-MPP. Evaluation of the over-manned scenario involves setting the starting level of officers in the system to higher numbers in the CGO ranks than the target level of officers for those ranks. We simulate forward from this starting state over a 40 year time horizon for both the benchmark policy and our ADP policy. Table 5 displays the results for the over-manned scenario experimental design. We display the percent improvement of the ADP policy over the benchmark policy along with the fully burdened cost of maintaining the force by following both the benchmark and ADP policies over the 40 year time horizon. The computation times for each run are shown as well.

Overall, the ADP policy outperforms the benchmark policy for the over-manned scenario. Most of the runs show at least a 20% improvement over the benchmark. We

**Table 5. Over-manned Experimental Results**

Run	N	M	$\rho_{ra}$	a	Percent Improvement	Fully Burd. Benchmark (\$M)	Fully Burd. ADP (\$M)	Time (Sec.)
1	1	1	-1	1	21.76±0.089%	628.952	561.314	2235.00
2	-1	-1	1	-1	21.93±0.053%	637.682	570.224	183.62
3	-1	1	1	-1	14.37±0.055%	634.859	414.142	669.18
4	1	1	-1	-1	21.32±0.094%	629.849	552.119	2026.82
5	1	1	1	-1	21.86±0.046%	635.010	570.956	1974.44
6	1	-1	-1	1	21.32±0.088%	630.592	551.428	673.18
7	1	-1	1	-1	21.84±0.046%	635.903	569.619	687.71
8	1	-1	-1	-1	21.34±0.086%	629.383	552.765	686.36
9	-1	1	-1	-1	-55.91±0.124%	629.905	697.798	661.41
10	-1	-1	1	1	9.86±0.053%	637.207	328.722	224.24
11	-1	-1	-1	1	10.23±0.087%	630.469	325.342	220.92
12	1	1	1	1	20.52±0.059%	637.132	536.459	2030.97
13	-1	-1	-1	-1	11.89±0.084%	630.113	358.891	226.44
14	-1	1	-1	1	9.10±0.091%	630.509	305.085	676.90
15	1	-1	1	1	23.10±0.050%	637.886	604.258	678.31
16	-1	1	1	1	8.12±0.056%	638.821	294.451	830.74

note that the fully burdened costs for the ADP policy are also generally lower than those of the benchmark policy. This signifies that the ADP policy corrects for the over-manned starting state by accessing less officers to minimize the “soft” penalty costs. The  $R^2$  for the model created with this experimental design is 0.73219 and the  $R_{adj}^2$  is 0.19656. The model does not have a high  $R_{adj}^2$  value which indicates that the factors chosen do not explain all the variation in the data. We expect that the parameters we chose would only explain some of the variation in the data because there are many variables and parameters which influence the percent improvement over benchmark policy and we only chose four to investigate. Table 6 displays the parameter estimates for this experimental design. Analyzing the factors for this experimental design reveals the most significant factor at a 90% confidence level in terms of influence of the percent improvement over benchmark is the number of policy improvement loops (N). Higher values of the outer loops generally produce higher percent improvement. Run 9 is of interest because it shows a reduction in

performance as compared to the benchmark. This run has a low level of policy improvement loops, decreased retention rate, a low level of harmonic smoothing and a high number of policy evaluation loops. Run 15 performs the best, and we note that it occurs when the number of policy improvement loops is set at the high level. Run 16 provides interesting results despite the slightly lower percent improvement by producing the lowest fully burdened cost.

**Table 6. Over-manned Scenario Parameter Estimates**

<b>Term</b>	<b>Estimate</b>	<b>Prob&gt; t </b>
N	8.96	0.0906
M	-5.02	0.2940
$\rho_{ra}$	5.03	0.2930
a	2.83	0.5375
N*M	4.75	0.3177
$N^*\rho_{ra}$	-4.83	0.3102
$N^*a$	-2.79	0.5432
$M^*\rho_{ra}$	3.54	0.4463
$M^*a$	4.39	0.3520
$\rho_{ra}^*a$	-5.13	0.2846

### **Under-manned Scenario.**

In this section, we discuss the results of the under-manned scenario experimental design for the full problem instance of the AFO-MPP. Evaluation of the under-manned scenario involves setting the starting level of officers in the system to lower numbers in the CGO ranks than the target level of officers for those ranks. We simulate forward from this starting state over a 40 year time horizon for both the benchmark policy and our ADP policy. Table 7 displays the results for the under-manned scenario experimental design. We display the percent improvement of the ADP policy over the benchmark policy along with the fully burdened cost of maintaining the force by following both the benchmark and ADP policies. The computation times for each run are shown as well.

**Table 7. Under-manned Experimental Results**

Run	N	M	$\rho_{ra}$	a	Percent Improvement	Fully Burd. Benchmark (\$M)	Fully Burd. ADP (\$M)	Time (Sec.)
1	1	1	-1	1	5.00±0.135%	582.166	978.987	1652.17
2	-1	-1	1	-1	5.56±0.071%	587.277	931.090	184.68
3	-1	1	1	-1	4.50±0.091%	587.589	974.466	549.51
4	1	1	-1	-1	4.99±0.152%	583.607	978.423	1997.95
5	1	1	1	-1	4.50±0.067%	587.672	974.755	2006.94
6	1	-1	-1	1	1.53±0.134%	582.032	661.884	557.76
7	1	-1	1	-1	5.70±0.088%	588.734	910.461	562.48
8	1	-1	-1	-1	4.99±0.140%	583.330	977.903	673.00
9	-1	1	-1	-1	5.01±0.118%	583.639	979.135	549.68
10	-1	-1	1	1	5.60±0.082%	587.826	908.625	183.98
11	-1	-1	-1	1	5.29±0.151%	582.358	864.510	185.44
12	1	1	1	1	4.40±0.098%	589.300	974.808	2001.09
13	-1	-1	-1	-1	2.09±0.145%	583.227	694.500	185.34
14	-1	1	-1	1	5.00±0.140%	581.637	978.740	687.67
15	1	-1	1	1	3.02±0.103%	587.808	752.460	555.01
16	-1	1	1	1	5.56±0.072%	588.334	929.531	663.21

Overall, the ADP policy outperforms the benchmark policy for the under-manned scenario. We do not observe as much improvement over the benchmark as we do for the over-manned scenario. Most of the runs produce about five percent improvement over the benchmark policy. The  $R^2$  for the model created with this experimental design is 0.70026 and the  $R_{adj}^2$  is 0.10078. The model does not have a high  $R_{adj}^2$  value which indicates that the factors chosen do not explain all the variation in the data. We expect that the parameters we chose would only explain some of the variation in the data because there are many variables and parameters which influence the percent improvement over benchmark policy and we only chose four to investigate. Table 8 displays the parameter estimates for this experimental design. At a 90% confidence level, only the interaction term between the number of policy improvement loops and parameter for harmonic smoothing affects the response. We note that the fully burdened costs for the ADP policy are generally higher than those of the benchmark policy. This signifies that the ADP policy corrects for the under-manned starting state

by accessing more officers to minimize the “soft” penalty costs. Run 10 performs the best, and we note that the levels of policy improvement and evaluation loops are set to the low level and those of the retention rate and harmonic smoothing parameters are set to the high level. It is also interesting to note that while Run 6 and 13 produce lower percent improvements over benchmark, they also produce lower fully burdened costs.

**Table 8. Under-manned Scenario Parameter Estimates**

<b>Term</b>	<b>Estimate</b>	<b>Prob&gt; t </b>
N	-0.28	0.3887
M	0.32	0.3258
$\rho_{ra}$	0.31	0.347
a	-0.12	0.7000
N*M	0.13	0.6751
$N^*\rho_{ra}$	-0.16	0.5918
$N^*a$	-0.65	0.0779
$M^*\rho_{ra}$	-0.43	0.1989
$M^*a$	0.24	0.4528
$\rho_{ra}^*a$	-0.09	0.7775

### **Target Level Scenario.**

In this section, we discuss the results of the target level scenario experimental design for the full problem instance of the AFO-MPP. Evaluation of the target level scenario involves setting the starting level of officers in the system to the target level of officers for the CGO ranks. We simulate forward from this starting state over a 40 year time horizon for both the benchmark policy and our ADP policy. Table 9 displays the results for the target level scenario experimental design. We display the percent improvement of the ADP policy over the benchmark policy along with the fully burdened cost of maintaining the force by following both the benchmark and ADP policies. The computation times for each run are shown as well.

**Table 9. Target Level Experimental Results**

Run	N	M	$\rho_{ra}$	a	Percent Improvement	Fully Burd. Benchmark (\$M)	Fully Burd. ADP (\$M)	Time (Sec.)
1	1	1	-1	1	4.93±0.134%	605.110	1000.297	1654.15
2	-1	-1	1	-1	3.75±0.090%	610.665	820.266	223.55
3	-1	1	1	-1	4.44±0.100%	610.670	999.63	543.01
4	1	1	-1	-1	5.00±0.118%	604.452	999.74	2053.37
5	1	1	1	-1	4.48±0.101%	611.104	1001.29	2012.25
6	1	-1	-1	1	5.63±0.141%	605.905	909.11	882.53
7	1	-1	1	-1	5.07±0.104%	611.665	898.75	683.26
8	1	-1	-1	-1	5.01±0.123%	604.418	999.80	679.37
9	-1	1	-1	-1	5.11±0.155%	604.111	1000.02	669.13
10	-1	-1	1	1	5.79±0.102%	610.280	939.96	218.14
11	-1	-1	-1	1	5.75±0.150%	604.765	929.28	227.95
12	1	1	1	1	4.46±0.092%	612.781	999.48	1966.10
13	-1	-1	-1	-1	3.01±0.176%	607.592	774.15	222.95
14	-1	1	-1	1	4.60±0.162%	604.686	851.56	864.95
15	1	-1	1	1	5.23±0.100%	613.723	910.61	554.93
16	-1	1	1	1	5.58±0.085%	610.026	955.58	721.56

Again, we see that the ADP policy outperforms the benchmark policy for the target level scenario. Most of the runs produce about a five percent improvement over the benchmark policy. The  $R^2$  for the model created with this experimental design is 0.81593 and the  $R^2_{adj}$  is 0.447798. The model does not have a high  $R^2_{adj}$  value which indicates that the factors chosen do not explain all the variation in the data. We expect that the parameters we chose would only explain some of the variation in the data because there are many variables and parameters which influence the percent improvement over benchmark policy and we only chose four to investigate. Table 10 displays the parameter estimates for this experimental design. In this experimental design, we see at a 95 % confidence level that the parameter for harmonic smoothing has the most significant effect on the response. High levels of the parameter for harmonic smoothing produce higher percent improvements. Run 10 performs the best, and we note that the levels of policy improvement and evaluation loops are set to the low level and those of the retention rate and harmonic smoothing parameters are

set to the high level. Studying the fully burdened costs of the benchmark policy versus the ADP policy, we note that the costs are higher for the ADP policy. This signifies that the ADP policy is accessing more officers in the system than the benchmark policy. We make note of Run 13, which has a lower percentage improvement over benchmark, but also has a lower fully burdened cost.

**Table 10. Target Level Scenario Parameter Estimates**

<b>Term</b>	<b>Estimate</b>	<b>Prob&gt; t </b>
N	0.11	0.4536
M	-0.04	0.7883
$\rho_{ra}$	-0.01	0.9192
a	0.38	0.0394
N*M	-0.22	0.1717
$N^*\rho_{ra}$	-0.15	0.3218
$N^*a$	-0.29	0.0867
$M^*\rho_{ra}$	-0.07	0.6320
$M^*a$	-0.31	0.0727
$\rho_{ra}^*a$	0.03	0.8140

## V. Conclusions

### 5.1 Conclusions

The research completed in this thesis provides a technique to solve the large-scale United States Air Force Officer Manpower Planning Problem (AFO-MPP). With our Markov decision process framework, we utilize the approximate policy iteration least squares temporal differencing algorithm to solve two problem instances of the AFO-MPP. The small problem instance investigates the impact of five algorithmic features on solution quality of the algorithm. Contrary to previous work done with instrumental variables on the AFO-MPP (Hoecherl, 2015), our experimental design shows that instrumental variables produce increases in costs when compared to the benchmark policy. Therefore, analysis on the full problem instance of the AFO-MPP employs Bellman’s error minimization.

In testing of the full problem instance, we employ three separate experimental designs for the AFO-MPP. Each experimental design tests the impact of three algorithmic features and the retention rate parameter on solution quality. In testing of the experimental design, the simulations begin with one of three notional scenarios: an over-manned force, an under-manned force and a force at the target level of officers. Our ADP algorithm shows moderate percent improvement over the benchmark policy in the under-manned and target level scenarios and slightly higher percent improvement in the over-manned scenario.

The work of Hoecherl (2015) suggests that alternative sets of basis functions should be investigated for the AFO-MPP. We investigate the use of high order basis functions and basis functions with interaction terms in this thesis. The improvements we see in percent improvement over benchmark for the full problem instance indicates the basis functions selected for the problem are acceptable. A drawback of the LSTD

algorithm is the potential for poor sampling within a large state space. We attempt to alleviate this potential for poor solutions by use of Latin hypercube sampling.

## 5.2 Future Work

The work of this thesis is a preliminary analysis of the AFO-MPP. Future work using ADP techniques may apply alternative algorithms such as leveling, SPAR, or CAVE. Use of these value function approximation techniques could provide good solution quality for this large-scale problem. The CAVE algorithm was applied with success by Hoecherl (2015). The method of applying kernel regression as a value function approximation technique was preliminarily explored in our research. However, Powell (2011) cautions against extending this methodology to problems of more than five dimensions. Since even the small problem instance investigated in this work has 40 dimensions, kernel regression was not utilized.

Further extensions of this work may expand the action space of the algorithm to include promotion decisions. The additional control mechanism provides the AFO-MPP with more elements of realism as senior leaders generally have power to exit officers from the system in times of reduction in force. Another interesting extension of the problem would be to investigate multiple AFSCs. This provides opportunity to examine the effects of cross-flow of officers between AFSCs, which is a common occurrence in the services.

The parameterization of our model could also be refined with additional consultation with personnel managers and recent work completed on factors which influence retention of officers in the Air Force. Consultation with personnel managers would better inform the cost parameters included in the model, the notion of target levels of officers and the benchmark policy against which we compare our ADP algorithm. The recent work completed by Schofield (2015) and Zens (2016) on factors which

influence retention could be applied to the AFO-MPP to parameterize retention for specific AFSCs.

# United States Air Force Officer Manpower Planning Problem via Approximate Dynamic Programming



## Objective:

The objective of the United States Air Force Officer Manpower Planning Problem (AFO-MPP) is to find the optimal number of officers to access to maintain mission effectiveness. Using the Markov decision process framework, we incorporate the uncertainty involved in officer turnover.

## Markov Decision Process Formulation:

Decision spaces:

State space:  $T = \{1, 2, \dots, T\}, T \in \mathbb{N}$

Rank:  $r \in \mathbb{R}^r$  where  $\mathbb{R}^r = \{1, 2, \dots, 6\}$

Commissioned Years of Service (CWS):  $a \in \mathbb{R}^a$  where  $\mathbb{R}^a = \{1, 2, \dots, 20\}$

State space:  $R_{t,r,a} = \text{the number of officers at time } t \text{ of rank } r \text{ and CWS } a$

Action space:  $\pi_t = \text{number of officers commissioned into the AF at time } t$

Transition function:

$$R_{t,r,a} = \begin{cases} \text{Trans}(R_{t-1,r,a}) & \text{if } (r,a) \in \mathbb{R}^r \\ \text{Trans}(R_{t-1,r,a}) + \text{Trans}(R_{t-1,r,a-1}) & \text{if } (r,a) \in \mathbb{R}^a \\ \text{Trans}(R_{t-1,r,a}) + \text{Trans}(R_{t-1,r,a-1}) + \text{Trans}(R_{t-1,r,a-2}) & \text{if } (r,a) \in \{1, 2\} \end{cases}$$

Cost function for officer advantages:

$$C_t(R_t, \pi_t) = \begin{cases} C_1(R_t, \pi_t) & \text{if } r = 1 \\ C_2(R_t, \pi_t) & \text{if } r \neq 1 \end{cases}$$

Cost function for officer shortages:

$$U_t(R_t, \pi_t) = \begin{cases} U_1(R_t, \pi_t) & \text{if } r = 1 \\ U_2(R_t, \pi_t) & \text{if } r \neq 1 \end{cases}$$

Single period cost function for the AFO-MPP:

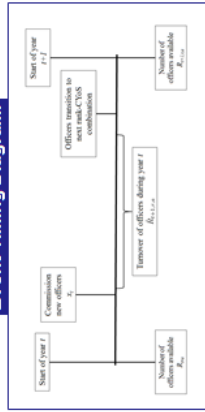
$$C_t(R_t, \pi_t) = \sum_{r,a} C_1(R_t, \pi_t) + U_t(R_t, \pi_t)$$

Objective function:

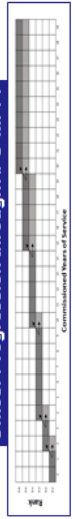
$$V^*(R_0) = \min_{\pi} \mathbb{E} \left\{ \sum_{t=0}^{T-1} C_t(R_t, \pi_t) \right\}$$

2d Lt Amelia E. Bradshaw  
 Advisor: Lt Col Matthew J. Robbins, PhD  
 Reader: Raymond R. Hill, PhD  
 Department of Operational Sciences (ENS)  
 Air Force Institute of Technology

## Event Timing Diagram:



## Officer Progression through AFO-MPP:



## Approximate Dynamic Programming:

ADP provides solutions to large scale MDPs which cannot be solved exactly. We utilize the approximate policy iteration algorithm which computes values associated with a fixed policy. Utilizing regression techniques, the value function is approximated by fitting a parameter vector to a set of basis functions.

```

Algorithm 1 ADP Algorithm
1: Step 0: Initialize  $\hat{V}^*$ 
2: for  $m=1$  do to  $N$  (Policy Improvement Loop)
3:   for  $m=1$  do to  $M$  (Policy Evaluation Loop)
4:     Record  $\hat{R}_m^*$ 
5:     Simulate transition to next event, obtain a prediction state,  $R_{m+1}^*$ 
6:     Determine decision  $\pi = \pi^*(R_m, \hat{V}_m^*)$ 
7:     Record basis function evaluations  $\phi(R_m)$ 
8:   End
9:   Compute  $\hat{V}^*$  using LSTD or IV and smoothing rule.
10: End
    
```

## VI. Appendix

### Small Problem Instance:

A  $2^{24}$  fractional factorial experimental design tests API algorithmic features on the AFO-MPP with two rank categories and 20 year group categories.

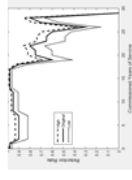
Instrumental Variables	Low (L)	High (H)
Rank Function	Not Present	Present
Second Order with Interactions	10	30
Policy Improvement Loops (N)	1,000	3,000
Policy Evaluation Loops (M)	10	30

Instrumental variables are the most significant factor for the experimental design. API performs poorly when IVs are included.

### Full Problem Instance:

Three  $2^4$  full factorial experimental designs test three algorithmic features and one problem feature on the full AFO-MPP for three notional scenarios: an over-manned force, under-manned force, and a force at the target level of officers.

Factor	Low (L)	High (H)
Number of Outer Loops (N)	10	30
Number of Inner Loops (M)	1,000	3,000
Instrumental Variables	Decreased by 50%	Increased by 50%
Rotation Rate (rho)	Decreased by 50%	Increased by 50%



The policy created by the ADP algorithm outperformed the benchmark policy in all scenarios, with the best performance for the over-manned scenario.

### Results:

Overall, API provides officer accession policies to the AFO-MPP which improve over the benchmark policy. Instrumental variables did not improve solution quality when included in the algorithm.



## Bibliography

- Banks, Jerry, Carson, John S., Nelson, Barry L., & Nicol, David M. 2010. *Discrete-event system simulation*. Prentice Hall, Upper Saddle River, NJ.
- Bertsekas, DP, & Tsitsiklis, JN. 1995. Neuro-dynamic programming: an overview. *Pages 560–564 of: Proceedings of the 34th IEEE Conference on Decision and Control, 1995.*, vol. 1. IEEE.
- Bowden, Roger J, & Turkington, Darrell A. 1990. *Instrumental variables*. Vol. 8. Cambridge University Press.
- Bradtke, Steven J, & Barto, Andrew G. 1996. Linear least-squares algorithms for temporal difference learning. *Machine Learning*, **22**(1-3), 33–57.
- Brauner, Marygail K, Massey, Hugh G, Moore, SG, & Medlin, Darren D. 2009. *Improving development and utilization of US Air Force intelligence officers*. Tech. rept. RAND Corporation, Santa Monica, CA.
- Castro, Carl A, & Huffman, Ann H. 2002. Predicting retention rates of US Soldiers stationed in Europe. *Army Medical Research Unit*.
- Chattopadhyay, Asis Kumar, & Gupta, Arindam. 2007. A stochastic manpower planning model under varying class sizes. *Annals of Operations Research*, **155**(1), 41–49.
- Conley, Raymond E. 2006. *Maintaining the Balance Between Manpower, Skill Levels, and PERSTEMPO*. Vol. 492. RAND Corporation, Santa Monica, CA.
- Conley, Raymond E, & Robbert, Albert A. 2009. *Air Force Officer Specialty Structure: Reviewing the Fundamentals*. RAND Corporation, Santa Monica, CA.
- De Bruecker, Philippe, Van den Bergh, Jorne, Beliën, Jeroen, & Demeulemeester, Erik. 2015. Workforce planning incorporating skills: State of the art. *European Journal of Operational Research*, **243**(1), 1–16.
- De Feyter, Tim. 2007. Modeling mixed push and pull promotion flows in manpower planning. *Annals of Operations Research*, **155**(1), 25–39.
- Dietz, Dennis C. 1996. Satisfying Advanced Degree Requirements for US Air Force Officers. *Military Operations Research*, **2**(4), 73–80.
- Filinkov, A, Richmond, M, Nicholson, R, Alshansky, M, & Stewien, J. 2011. Modelling personnel sustainability: a tool for military force structure analysis. *Journal of the Operational Research Society*, **62**(8), 1485–1497.
- Fowler, John W, Wirojanagud, Pornsarun, & Gel, Esma S. 2008. Heuristics for workforce planning with worker differences. *European Journal of Operational Research*, **190**(3), 724–740.

- Galway, Lionel A, Buddin, Richard J, Thirtle, Michael R, Ellis, Peter S, & Mele, Judith D. 2005. *Understrength Air Force Officer Career Fields. A Force Management Approach*. RAND Corporation, Santa Monica, CA.
- Gans, Noah, & Zhou, Yong-Pin. 2002. Managing learning and turnover in employee staffing. *Operations Research*, **50**(6), 991–1006.
- Gass, Saul I, Collins, Roger W, Meinhardt, Craig W, Lemon, Douglas M, & Gillette, Marcia D. 1988. OR Practice The Army Manpower Long-Range Planning System. *Operations Research*, **36**(1), 5–17.
- Guerry, Marie-Anne, & De Feyter, Tim. 2012. Optimal recruitment strategies in a multi-level manpower planning model. *Journal of the Operational Research Society*, **63**(7), 931–940.
- Hagopian, Amy, Micek, Mark A, Vio, Ferruccio, Gimbel-Sherr, Kenneth, & Montoya, Pablo. 2008. What if we decided to take care of everyone who needed treatment? Workforce planning in Mozambique using simulation of demand for HIV/AIDS care. *Human Resources for Health*, **6**(1), 3.
- Hill, Raymond R, McIntyre, Gregory A, & Narayanan, S. 2001. Genetic Algorithms for Model Optimization. In: *Proceedings of Simulation Technology and Training Conference (SimTechT)*. Citeseer.
- Hoecherl, Joseph C. 2015. *Approximate Dynamic Programming Algorithms for United States Air Force Officer Sustainment*. Master's Thesis, Air Force Institute of Technology, Wright-Patterson AFB, OH, USA.
- Hosek, Susan D. 2001. *Minority and gender differences in officer career progression*. Vol. 1184. Rand Corporation.
- Jackson, James F. 2013. *Testimony before the National Commission on the Structure of the Air Force*. <http://afcommission.whs.mil/public/docs/meetings/20130827/Testimony%20of%20Lieutenant%20General%20James%20F.%20Jackson.pdf>.
- Khoong, CM. 1996. An integrated system framework and analysis methodology for manpower planning. *International Journal of Manpower*, **17**(1), 26–46.
- Knapp, Gerald M, & Mahajan, Milind. 1998. Optimization of maintenance organization and manpower in process industries. *Journal of Quality in Maintenance Engineering*, **4**(3), 168–183.
- Lakhani, Hyder. 1988. The effect of pay and retention bonuses on quit rates in the US Army. *Industrial & Labor Relations Review*, **41**(3), 430–438.
- McClellan, Sally. 1991. Manpower planning models and their estimation. *European Journal of Operational Research*, **51**(2), 179–187.

- McKay, Michael D, Beckman, Richard J, & Conover, William J. 1979. Comparison of three methods for selecting values of input variables in the analysis of output from a computer code. *Technometrics*, **21**(2), 239–245.
- Montgomery, Douglas C. 2008. *Design and analysis of experiments*. Hoboken, New Jersey: John Wiley & Sons.
- Powell, Warren, Ruszczyński, Andrzej, & Topaloglu, Huseyin. 2004. Learning algorithms for separable approximations of discrete stochastic optimization problems. *Mathematics of Operations Research*, **29**(4), 814–836.
- Powell, Warren B. 2011. *Approximate Dynamic Programming: Solving the curses of dimensionality*. John Wiley & Sons.
- Powell, Warren B. 2012. Perspectives of approximate dynamic programming. *Annals of Operations Research*, 1–38.
- Powell, Warren B, Bouzaiene-Ayari, Belgacem, Berger, Jean, Boukhtouta, Abdeslem, & George, Abraham P. 2011. The Effect of Robust Decisions on the Cost of Uncertainty in Military Airlift Operations. *ACM Transactions on Modeling and Computer Simulation*, **22**(1), 1–19.
- Puterman, Martin L. 2005. *Markov Decision Processes, Discrete Stochastic Dynamic Programming*. Hoboken, New Jersey: John Wiley & Sons, Inc.
- Robbert, Albert A, Drezner, Stephen M, Boon, John, Hanser, Larry, Moore, Craig, Scott, Lynn, & Shukiar, Herbert J. 2005. *Integrated Planning for the Air Force Senior Leader Workforce. Background and Methods*. RAND Corporation, Santa Monica, CA.
- Rostker, Bernard D, Thie, Harry J, Lacy, James L, Kawata, Jennifer H, & Purnell, Susanna W. 1993. *The Defense Officer Personnel Management Act of 1980*. RAND Corporation.
- Sayın, Serpil, & Karabatı, Selçuk. 2007. Assigning cross-trained workers to departments: A two-stage optimization model to maximize utility and skill improvement. *European Journal of Operational Research*, **176**(3), 1643–1658.
- Schofield, Jill A. 2015. *Non-Rated Air Force Line Officer Attrition Rates Using Survival Analysis*. Master’s Thesis, Air Force Institute of Technology, Wright-Patterson AFB, OH, USA.
- Scott, Warren R, Powell, Warren B, & Moazehi, Somayeh. 2014. Least squares policy iteration with instrumental variables vs. direct policy search: Comparison against optimal benchmarks using energy storage. *Technical report, Princeton*.

- Simao, Hugo, & Powell, Warren. 2009. Approximate dynamic programming for management of high-value spare parts. *Journal of Manufacturing Technology Management*, **20**(2), 147–160.
- Simão, Hugo P, Day, Jeff, George, Abraham P, Gifford, Ted, Nienow, John, & Powell, Warren B. 2009. An approximate dynamic programming algorithm for large-scale fleet management: A case application. *Transportation Science*, **43**(2), 178–197.
- Škulj, Damjan, Vehovar, Vasja, & Štamfelj, Darko. 2008. The modelling of manpower by Markov chains-a case study of the Slovenian armed forces. *Informatica*, **32**(3), 289–297.
- Sohn, So Young, Chang, In Sang, & Moon, Tae Hee. 2007. Random effects Weibull regression model for occupational lifetime. *European Journal of Operational Research*, **179**(1), 124–131.
- Song, Haiqing, & Huang, Huei-Chuen. 2008. A successive convex approximation method for multistage workforce capacity planning problem with turnover. *European Journal of Operational Research*, **188**(1), 29–48.
- Stanley III, Dale W. 2012. *Predicting Pilot Retention*. Master’s Thesis, Air Force Institute of Technology, Wright-Patterson AFB, OH, USA.
- Topaloglu, Huseyin, & Powell, Warren B. 2003. An algorithm for approximating piecewise linear concave functions from sample gradients. *Operations Research Letters*, **31**(1), 66–76.
- Topaloglu, Huseyin, & Powell, Warren B. 2006. Dynamic-Programming Approximations for Stochastic Time-Staged Integer Multicommodity-Flow Problems. *INFORMS Journal on Computing*, **18**(1), 31–42.
- Vernez, Georges, Moore, Craig, Martino, Steven, & Yuen, Jeffrey. 2006. *Improving the Development and Utilization of Air Force Space and Missile Officers*. RAND Corporation, Santa Monica, CA.
- Wang, Jun. 2005. *A Review of Operations Research Applications in Workforce Planning and Potential Modeling of Military Training*. DSTO Systems Sciences Laboratory.
- Zens, Christine. 2016. *Application of Non-Rated Line Officer Attrition Levels and Career Field Stability*. Master’s Thesis, Air Force Institute of Technology, Wright-Patterson AFB, OH, USA.

# 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.**

<b>1. REPORT DATE (DD-MM-YYYY)</b> 24-03-2016		<b>2. REPORT TYPE</b> Master's Thesis		<b>3. DATES COVERED (From — To)</b> Aug 2014 – Mar 2016		
<b>4. TITLE AND SUBTITLE</b>  United States Air Force Officer Manpower Planning Problem via Approximate Dynamic Programming				<b>5a. CONTRACT NUMBER</b>		
				<b>5b. GRANT NUMBER</b>		
				<b>5c. PROGRAM ELEMENT NUMBER</b>		
				<b>5d. PROJECT NUMBER</b>		
				<b>5e. TASK NUMBER</b>		
<b>6. AUTHOR(S)</b>  Bradshaw, Amelia E., 2d Lt, USAF				<b>5f. WORK UNIT NUMBER</b>		
<b>7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)</b> Air Force Institute of Technology Graduate School of Engineering and Management (AFIT/EN) 2950 Hobson Way WPAFB OH 45433-7765			<b>8. PERFORMING ORGANIZATION REPORT NUMBER</b>  AFIT-ENS-MS-16-M-092			
<b>9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)</b> Gerald Diaz, Ph.D Headquarters Air Force/A1 1550 W. Perimeter Rd, Rm 4710 Joint Base Andrews NAF Washington, MD 20762-5000 gerald.diaz.civ@mail.mil			<b>10. SPONSOR/MONITOR'S ACRONYM(S)</b>  HAF/A1			
			<b>11. SPONSOR/MONITOR'S REPORT NUMBER(S)</b>			
<b>12. DISTRIBUTION / AVAILABILITY STATEMENT</b>  Distribution Statement A. Approved for Public Release; distribution unlimited.						
<b>13. SUPPLEMENTARY NOTES</b>  This material is declared a work of the U.S. Government and is not subject to copyright protection in the United States.						
<b>14. ABSTRACT</b>  The United States Air Force (USAF) is concerned with managing its officer corps to ensure sufficient personnel for mission readiness. Manpower planning for the USAF is a complex process which requires making decisions about accessions, with the uncertainty about factors which could affect employee turnover. We formulate a Markov decision process model of the Air Force officer manpower planning problem (AFO-MPP). We utilize the least squares approximate policy iteration algorithm as an approximate dynamic programming (ADP) technique to solve this large-scale problem. We test the performance of the policy determined with the ADP algorithm compared to a benchmark policy on two problem instances of the AFO-MPP. We find that the algorithm performs well for the basis functions selected, providing policies which reduce "soft" costs associated with shortages and surpluses in the force.						
<b>15. SUBJECT TERMS</b>  manpower planning, officer sustainment, workforce planning, approximate dynamic programming, ADP, approximate policy iteration, least squares temporal differences, LSTD						
<b>16. SECURITY CLASSIFICATION OF:</b>			<b>17. LIMITATION OF ABSTRACT</b>	<b>18. NUMBER OF PAGES</b>	<b>19a. NAME OF RESPONSIBLE PERSON</b> Lt Col Matthew J.D. Robbins, Ph.D.	
a. REPORT	b. ABSTRACT	c. THIS PAGE			<b>19b. TELEPHONE NUMBER (include area code)</b> (937) 255-3636, x4539;matthew.robbins@afit.edu	
U	U	U	UU	65		