MODELING HUMAN SUPERVISORY CONTROL IN HETEROGENEOUS UNMANNED VEHICLE SYSTEMS

by

CARL E. NEHME

Bachelors Computer Engineering, University of Toronto, 2002

Master of Science Aeronautics & Astronautics, Massachusetts Institute of Technology, 2004

Submitted to the Department of Aeronautics and Astronautics
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

February 2009

Author: ______________________________
Carl E. Nehme, Dept. of Aeronautics and Astronautics
December 19, 2008

Certified by: ______________________________
Mary L. Cummings, Thesis Supervisor
Associate Professor of Aeronautics and Astronautics
Director, Humans and Automation Laboratory

Certified by: ______________________________
R. John Hansman
Professor of Aeronautics and Astronautics
Director, International Center for Air Transportation

Certified by: ______________________________
Olivier de Weck
Associate Professor of Aeronautics and Astronautics
Associate Director, Engineering Systems Division

Accepted by: ______________________________
David L. Darmofal
Associate Head, Department of Aeronautics and Astronautics
Chair, Committee on Graduate Students

MODELING HUMAN SUPERVISORY CONTROL IN HETEROGENEOUS UNMANNED VEHICLE SYSTEMS

by

Carl E. Nehme

Submitted to the Department of Aeronautics and Astronautics in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Aeronautics and Astronautics

ABSTRACT

Given advanced technology that relieves the human operator of low-level tasking and the future vision for network-centric operations, operator supervisory control of Unmanned Vehicle (UV) teams is likely to be a focal point of future research and development. Due to requirements for interoperability among UVs of varying attributes, heterogeneity in vehicle capabilities and tasks is likely to exist in future UV systems. This will lead to a large design space for these systems, which will cause design validation to require lengthy and expensive human-in-the-loop experimentation.

This problem is addressed in this thesis through the following: First, identification of human-UV interaction attributes and associated variables that should be captured when modeling supervisory control of heterogeneous UV systems. Second, the derivation of a queuing-based multi-UV discrete event simulation (MUV-DES) model that captures both vehicle-team variables (including team composition and level of autonomy) and operator variables (including attention allocation strategies and situational awareness). The MUV-DES model supports design validation by simulating the impact of alternate designs on vehicle, operator, and system performance.

To determine the accuracy and robustness of the MUV-DES model, an Internet-based test bed was developed to support extensive and rapid data collection for supervisory control of multiple heterogeneous UVs. Using data accumulated from online experiments, a multi-stage validation process was applied. The validation process resulted in achieving confidence in the model's accuracy and determination of the model's robustness under different input settings. Following the validation process, the MUV-DES model's ability to aid in the design and assessment of heterogeneous UV teams and related technologies was evaluated. More specifically, the MUV-DES model generated design recommendations addressing three underlying research objectives: a) indicating how potential operational/developmental design modifications could lead to performance improvements including 30% reductions in average vehicle wait times, b) identifying potential capabilities and limitations of future designs, including the detrimental impact of service time heterogeneity greater than 40% on average vehicle wait times, and c) replicating observed behavior in an existing system as a means of diagnosing the causes of vehicle-performance inefficiency. A subset of the MUV-DES model design recommendations was then implemented and the predicted benefit was validated using an additional set of experiments.

Thesis supervisor: M. L. Cummings
Title: Associate Professor of Aeronautics and Astronautics

ACKNOWLEDGMENTS

First and foremost, I would like to thank my father and mother to whom I am eternally grateful. You have always stood by me and your undying encouragement consistently pushed me to strive for the best. It is because of you that I was able to make it this far and I dedicate this thesis to you. Thank you to my sisters, Fani, Maya, and Tara for making sure I always moved in the right direction. George, Elie, and Ralph, I thank you as well. Every Christmas, you made me look forward to a year of accomplishments.

My parents and family were the driving force behind my achievements, but without a teacher and a mentor, I would not have been able to move forward. Missy Cummings, I thank you for being that teacher, for teaching me how to think, how to write, and most importantly how to achieve. I am honored to have been in your team and I look forward to a lifetime of friendship and cooperation.

There are many other individuals I need and wish to thank:

John Hansman: Thank you for the many insightful comments you have given me. I enjoyed taking air traffic control with you and that is where I first experienced your thoroughness and attention to detail. Thank you for being on my committee.

Olivier de Weck: Thank you for being on my committee. I also thank you for teaching me how to not just solve problems, but to reach optimality. I very much enjoyed 16.888 and I thank you and Karen Wilcox for teaching me the so many things I learned.

Dave Darmofal: I thank you for being there at most of my doctoral events (everything except my qualifiers!). It was always great to receive your insightful and patient feedback.

Kristina Lundqvist: Thank you for your support during my qualifying exams and for being a great initial advisor.

Wesley Harris: Thank you for your continued support throughout my time at MIT.

Duncan Campbell: It was extremely rewarding having you listen to my ideas and thoughts. Thank you for always smiling and being patient. Also, thank you for making the effort to read my thesis especially with the time difference!

Ryan Kilgore: Where do I start? First and foremost, thank you for your friendship. It was great to a have an industry partner that was intelligent, patient and appreciative. Thank you for reading my thesis as well.

Jake Crandall: It is because of your founding work that I was able to reach this far. I very much cherish the intelligent and non-intelligent conversations we would have by the 77 Mass Ave. window. I look forward to our continued friendship.

Emilio Frazzoli: Thank you for always having the time to meet with me. I also thank you for providing me with the opportunity to collaborate with your CSAIL team.

Debbie Nightingale: Integrating the Lean Enterprise was extremely rewarding. Thank you for always being a supportive figure.

Sally Chapman: Thank you for being there. Thank you for your wisdom, for helping me fly to conferences, for helping me arrange for rooms to spread my ideas in, and for giving me good advice when I needed it. I hope we can remain friends for a long time to come.

Kathryn Fisher: Thank you for your constant care and support.

Marie Stuppard: I don't know how I would have reached this point had you not been there the last couple of years. I am grateful for your help, but more importantly, for your encouragement and support.

Barbara Lechner: Thank you for believing in me and for taking the time to listen to my questions. I was always motivated to work hard knowing how hard and late you worked for the department.

Gilles Coppin: Thank you for all the advice you have provided and for being a source of encouragement and support.

Ketan Savla and Tom Temple: I thank you, "the CSAIL group", for collaborating with me and for being there when I needed a listening ear.

Birsen Donmez: Shewirme! Thank you for your friendship and thank you for helping me to further develop the WTSA/UT work. Thank you for your input on the thesis.

Luca Bertuccelli: Thank you for making time to give me feedback. Your dedication to your fellow students is commendable.

Mark Ashdown: Thank you for your software engineering insight when we were building RESCHU and thank you for your input on the thesis.

Stacey Scott: Thank you for being a supporting figure when I first got to HAL and for teaching me the many things I needed to know.

Brian Mekdeci and Tuco: Dr. Deece, thanks for being a great friend. I enjoyed bouncing ideas off you and always appreciated your insightful input. Take care of Tuco.

Madhu Raman: Thank you my old friend for being there to support me throughout my dissertation writing process. It would have been extremely hard for me to write without the late night phone conversation breaks!

Yale Song: Thank you for working tirelessly with me on RESCHU and for believing in the project. I believe in your continued success.

Mauro della Penna: Thank you for being a great friend. It was great talking to you about my work and listening to your insightful comments.

Paul de Jong: Thank you for working with me. It was great to have you as a teammate and it was refreshing to listen to your thoughts. All the best in your Ph.D. endeavors.

Georges Aoude: Thank you for being a great friend. I very much enjoyed the classes we worked on together.

Yves Boussemart: Thank you for being a great office mate. I hope we bump into each other on one or more dives in the future.

Sylvain Bruni: It was great working with you on 16.888 and I thank you for all the input you have provided. Looking forward to us both finishing.

Tony McDonald and Caryn Krakauer: Thank you both for working on my projects under the UROP program.

I would also like to thank the rest of HAL (Hudson Graham, Geoff Carrigan, Christin Hart, Anna Massie, Andrew Claire, Chris Tsonis, Farzan Sasangohar…) for being great friends and making the lab experience amazing!

I would also like to thank my non-HAL friends for always being there. Rachael, Gaylee, Tarek, Naeem, Jean-Pierre, Ayah, and Raed, thank you for supporting me the past couple of years!

Finally, I would like to thank Charles River Analytics and the Office of Naval Research (ONR) for sponsoring my research.

TABLE OF CONTENTS

LIST OF ACRONYMS AND ABBREVIATIONS

DES	Discrete Event Simulation
FDS	Full Decision Support
FO	Fan-out
HALE UAV	High Altitude Long Endurance Unmanned Aerial Vehicle
IT	Interaction Time
LOA	Level of Automation
MALE UAV	Medium Altitude Long Endurance Unmanned Aerial Vehicle
MUV-DES	Multi Unmanned Vehicle Discrete Event Simulation
NDS	No Decision Support
NT	Neglect Time
R&D	Research & Development
SA	Situational Awareness
RESCHU	Research Environment for Supervisory Control of Heterogeneous Unmanned-vehicles
UAV	Unmanned Aerial Vehicle
UGV	Unmanned Ground Vehicle
USV	Unmanned Surface Vehicle
UUV	Unmanned Undersea Vehicle

UV	Unmanned Vehicle
WTI	Wait Times due to Interaction
WTQ	Wait Times due to Queuing
WTSA	Wait Times due to loss of Situational Awareness
WTSA/UT	Wait Times due to loss of Situational Awareness/Utilization

LIST OF FIGURES

LIST OF TABLES

1 INTRODUCTION

1.1 Motivation

In recent years, the use of unmanned vehicles (UVs) has become increasingly prominent. Unmanned aerial vehicles (UAVs), ground vehicles (UGVs), surface vehicles (USVs), and undersea vehicles (UUVs) have been used in applications ranging from military operations to border security (Figure 1.1). The ability of these vehicles to excel in dangerous missions has encouraged the military to embrace this technology. In addition, successes of UVs in military applications have encouraged the civil sector to look towards such technology in their future vision. For example, tasks such as fighting forest fires, monitoring wildlife and supporting police squads are ideal applications for UAVs (Ollero & Maza, 2007). Similarly, in the under-sea domain, discovering hydrothermal vents, under-sea exploration, oil-rig repairs, and subsurface wellhead maintenance are suitable applications for non-military UUVs.

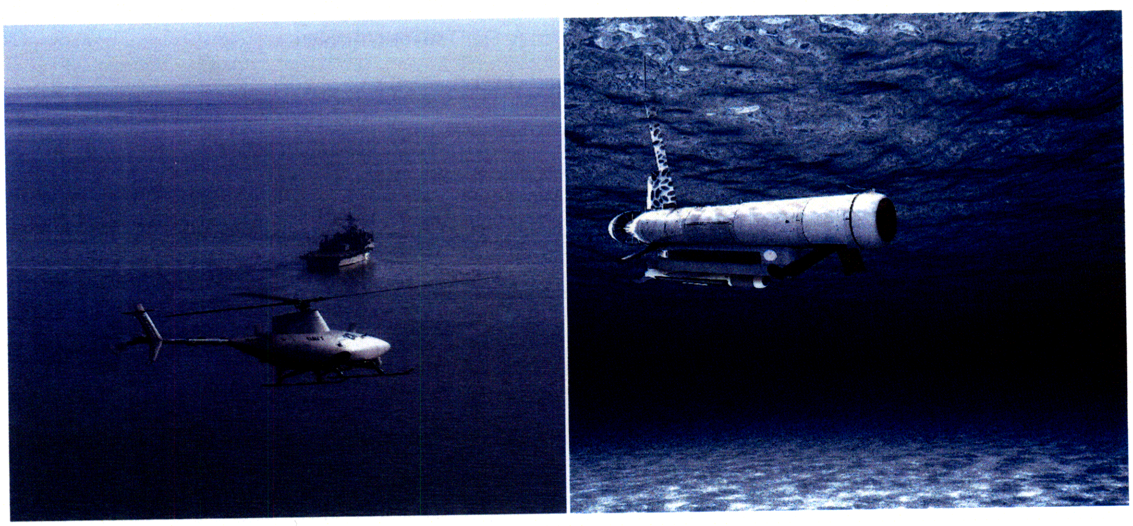


Figure 1.1 Department of Defense images of an unmanned air (image 060117-N-4935L-001) and undersea vehicle (image 021213-N-0000X-001)

1.1.1 Market for Unmanned Vehicle Systems

The desirability of unmanned vehicles is reflected in the size of the unmanned systems market and its projected growth trends. According to a study by the Teal Group, the UAV sector is the most dynamic growth sector of the aerospace industry today (Zaloga, Rockwell, & Finnegan, 2008). As shown in Figure 1.2, the UAV market is projected to more than double over the next decade from a market size of US$ 3.4 billion to US$ 7.3 billion (Zaloga, Rockwell, & Finnegan, 2008). This is expected to be divided equally between procurement and research & development (R&D). It is also forecasted that the global markets for both UUVs and UGVs will experience significant growth in the period 2007-2020, driven by both military and commercial interests (Bharat, 2007). The global UUV market is projected to see total revenues of US$ 8.83 billion by 2020, while the global UGV market could reach US$ 1.35 billion in the same period (Bharat, 2007). Such projected trends are reinforced by U.S. military roadmaps such as the fiscal year 2001 Defense Authorization Act, which stipulates that one third of U.S. Army vehicles must be unmanned by the year 2015 (DoD, 2007).

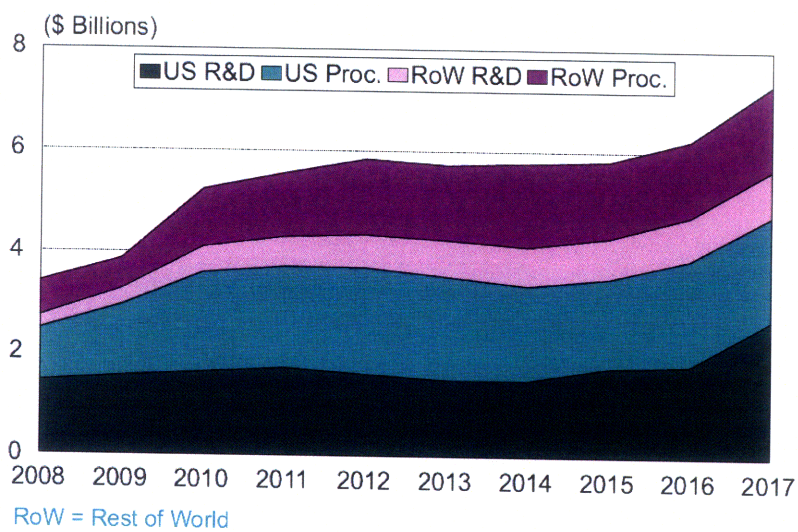


Figure 1.2 UAV market size and projected growth trends (Zaloga, Rockwell, & Finnegan, 2008)

The demand for UVs has mainly been fueled by the advantages that such systems offer over their manned counterparts such as lower acquisition costs and longer endurance (Bolkcom, 2005). Nonetheless, UV systems still compare with manned systems in terms of manning requirements. On average, two operators are required to operate a UAV. This excludes any maintenance or launch and recovery personnel (McCarley & Wickens, 2005). Examples of current systems with multiple operators include the Predator UAV, in use by the U.S. Air Force, and the Shadow UAV, in use by the U.S. Army. In these systems, both a payload operator and a navigator are required to fulfill the functions of navigation and payload control simultaneously. Similarly, UUVs, UGVs, and USVs, all currently require at least one operator per vehicle (Murphy, 2004).

1.1.2 Future Vision

As technology and computer processing power improves, several UV operator tasks are becoming more automated (Cummings, Bruni et al., 2007). The effects of increasing levels of automation can be seen in the autonomy for low level control of unmanned vehicles. For example, the angle by which the actuators have to move the ailerons on a UAV is something that is completely hidden from the human operator. Similarly, research has shown that it is possible for automation to relieve the human operator from navigation-based tasks through better path planning algorithms and technologies such as sense-and-avoid (Utt, McCalmont, & Deschenes, 2005), the unmanned vehicle equivalent of collision avoidance (Harman, 1989).

Increasing use of automation in unmanned vehicle systems has shifted the human operator's responsibility from manually controlling vehicles to managing vehicles at the supervisory control level (Cummings, Bruni et al., 2007). At the supervisory control level, implementation details of higher-level tasking initiated by the human are delegated to the automation onboard these vehicles (Sheridan, 1992). The reduced workload afforded by supervisory control has several implications. One such ramification is an increase in operator idle time, which can be used as a force multiplier that allows operators to supervise multiple vehicles simultaneously, hence inverting the current many-to-one ratio of operators to vehicles (Cummings, Bruni et al., 2007; Mitchell & Cummings, 2005). Figure 1.3 displays this in terms of a notional systems architecture with the operator interacting at the payload/mission management level while delegating routine piloting and navigation based tasks to the automation (Cummings, Bruni et al., 2007).

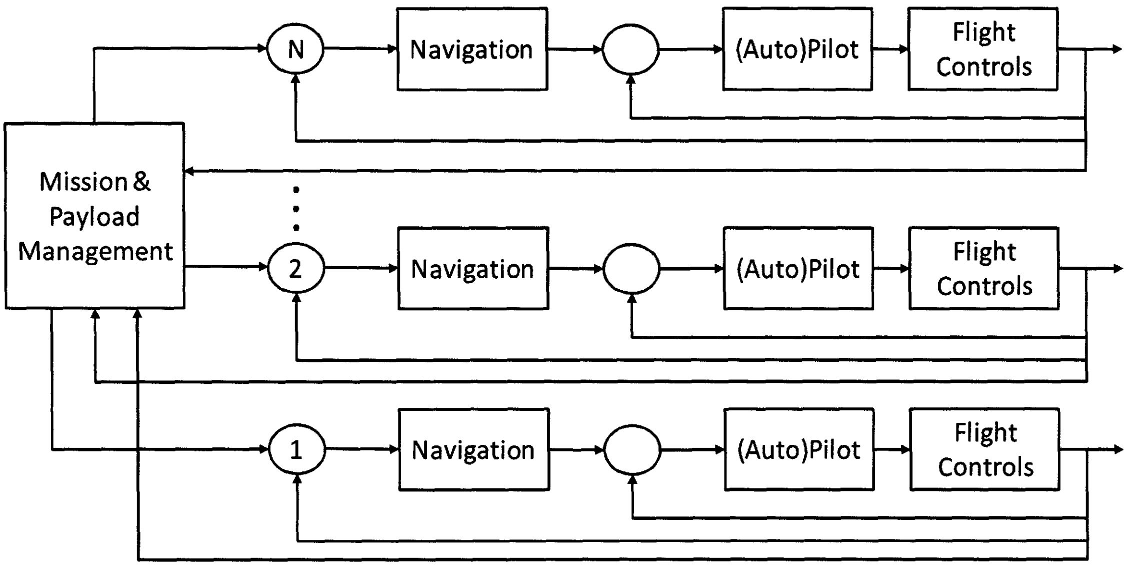


Figure 1.3 Hierarchical control for multiple unmanned vehicles (Cummings, Bruni et al., 2007)

Although the single-operator-multiple-UVs paradigm of control might not be suitable when dedicated operator attention is needed or for vehicles that require a human to serve as a backup pilot, it may serve as an important element in satisfying the needs of futuristic concepts of operation. Inverting the operator-to-vehicle ratio can be used to reduce manning in situations where the number of vehicles needed to accomplish missions exceeds that of available operators, which is currently a significant problem in the Air Force's Predator community. In addition, more complex problems similar to those that present themselves through concepts such as network-centric operations (Alberts, Garstka, & Stein, 1999) and the Future Combat System (Feickert, 2005) will require a transition from the architecture of today, where vehicles are supervised on an individual basis, to an architecture where operators integrate information retrieved from multiple platforms. Allowing for systems where multiple vehicles can be supervised by a single operator simultaneously will facilitate such operational architectures. In general, unmanned system development efforts by both the military and the general industry are guided by two roadmaps (Figure 1.4); one addressing technology-driven capabilities and the other, operations-driven missions (Cambone, Krieg et al., 2005). Given the technology which is relieving the human operator of low-level tasking, and the military's future vision for network centric operations, the operator manning paradigm inversion is likely to be a focal point of future research and development.

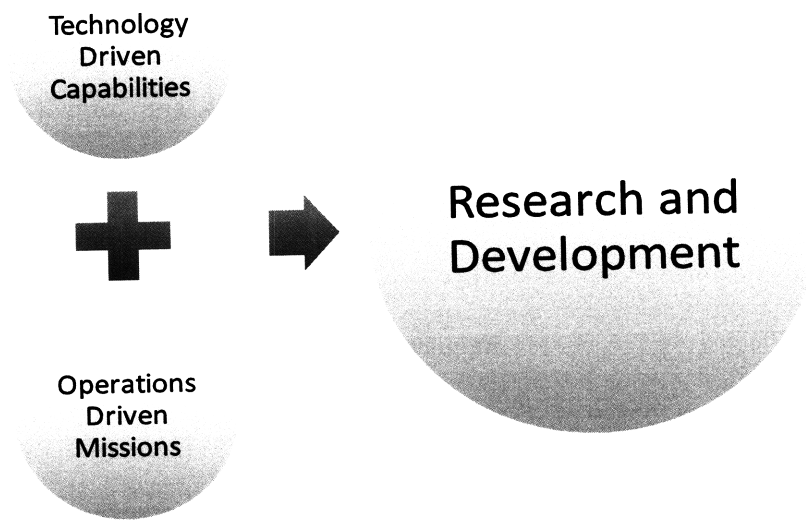


Figure 1.4 Forces driving research and development in military and industry UV settings

1.1.3 UV System Designer Needs

An important process in the design of any engineering system is the requirements, design and evaluation loop (Figure 1.5). Desired operational architectures that are driven by end user demands shape requirements generation for future engineering systems. Technology-based enablers are then used to satisfy these requirements through engineering designs. The ability of these designs to satisfy the requirements are then validated using human-in-the-loop evaluation. The fidelity of such human-in-the-loop evaluation will depend on the extent to which the evaluation reflects real world scenarios. The iterative cycle is then completed by using the performance observed in the evaluation stage to refine the original requirements.

Given that network-centric operations is a major driving force behind the possible transition to a single-operator-multiple-UVs paradigm, and these operational concepts require interoperability among UVs of varying attributes, heterogeneity is likely to exist in the requirements of future UV supervision. As unmanned vehicle system mission goals become increasingly demanding, the composition of UV teams is likely to involve vehicles of varying capabilities and levels of autonomy. In addition to heterogeneity across vehicle types, even a single UV can have multiple payloads which will ultimately lead to heterogeneity in operator tasks. It is therefore plausible that a team of unmanned vehicles could be composed of vehicles that vary in their capabilities or their assigned tasks, and form a heterogeneous system.

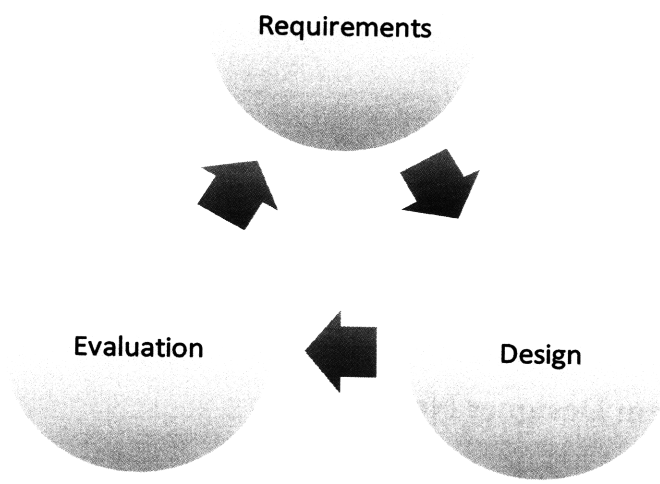


Figure 1.5 Requirements, design and evaluation loop for systems engineering

This heterogeneity in requirements is likely to cause a diversification in the choices presented to designers of the vehicles, as well as the interfaces for controlling the vehicles. In terms of vehicle team design, alternate designs can result in different levels of heterogeneity in vehicle capabilities and/or assigned tasks. In terms of interface and decision support design, disparate vehicles and associated tasks will result in more diverse methods for operators to allocate their attention to vehicle control tasks. Since operator actions are largely motivated by the type of decision support available, alternate interface designs are possible (Hanson, Roth et al., 2004a; Hanson, Roth et al., 2004b; Linegang, Haimson et al., 2003; Steinberg, 2006; Weil, Freeman et al., 2006).

Although heterogeneity is a by-product of the desired attributes of interoperability and network-centric operations, the varied design choices that it induces could pose a threat to the effectiveness of UV supervision and mission performance. Due to the disparity in the type of interaction required from the human operator for supervising heterogeneous tasks, different design choices can influence the ability of operators to supervise the teams effectively. For example, a UAV might require more frequent attention from the human operator than a UUV that spends more of its time underwater and out of communication range. Moreover, since design choices are not mutually exclusive, it is likely that design tradeoffs will exist. For example, it is likely that a tradeoff will exist between the desire to have the flexibility of varying the size and composition of teams, and the ability of operators to effectively supervise such teams. The importance of ensuring operator effectiveness is most evident in the number of catastrophic incidents exhibited in UAV operations. Of the 139 Predators delivered to the U.S. Air Force through 2007, records show that 53 have been lost, almost 40% of the inventory (Vanden Brook, 2007). Moreover, different studies have shown that the majority of accidents have been due to human error. Such studies have shown that tracing the source of the errors shifts the blame from pilot error to organizational and latent errors that may be due to faulty designs (Carrigan, Long et al., 2008). Therefore, in designing UV systems, the collective impact of different design choices will dictate the effectiveness of operator supervision.

Through human-in-the-loop experimentation, designers would need to evaluate the effectiveness of alternate design choices. However, with a large solution space in terms of design options, comprehensive experimentation can be overly lengthy and expensive. This cost can arise from both costs of implementing any design changes as well as costs of running the experiments themselves (personnel required, setup costs, etc.). Moreover, these costs can become prohibitive if the aim is not to test a certain design change, but to search for a design setting that satisfies a certain output condition (i.e. an optimization process). In addition, human-in-the-loop experimentation is difficult for futuristic systems for which no actual implementation exists. In such cases, experimentation is substituted with approximations from similar systems or must wait for a prototype to first be built.

1.2 Research Statement

1.2.1 Approach

Given the cost and time limitations of human-in-the-loop experimentation for guiding design choices in futuristic heterogeneous unmanned systems, this thesis will investigate the possibility of a discrete-event-simulation model to provide rapid prototyping evaluation capabilities. The following research questions will be answered:

- What attributes need to be captured and how can they be represented when modeling a heterogeneous UV system?
- What type of accuracy/robustness can be expected from a discrete-event-simulation model?
- How can a discrete-event-simulation model aid in the design and assessment of heterogeneous UV teams and related technologies?

1.2.2 Value Proposition

A comprehensive discrete-event-simulation model of human-UV system interaction can assist in the design of heterogeneous systems by identifying those designs that are worth investigating further in the costly evaluation phase versus those that should be eliminated because they do not satisfy system requirements (Figure 1.6).

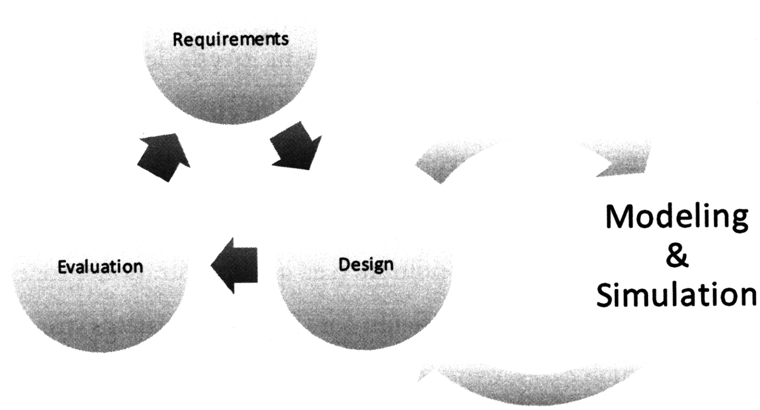


Figure 1.6 Supporting the design phase through a modeling and simulation effort

A model with descriptive and predictive abilities can minimize the need for experimentation with representative users, since it can both help diagnose the cause of failures and inefficiencies, and indicate how potential design modifications will affect the performance of the system. Given the demand for UV systems, which is stimulated by increasing market size and growth trends, and the military's ambitions in terms of futuristic concepts of operations, a model with the above attributes can be very beneficial in streamlining the requirements, design and evaluation cycle.

1.3 Thesis Outline

The remainder of the thesis is organized as follows.

In Chapter 2, titled "Background: Modeling Unmanned Vehicle Systems," the different attributes of a representative human-UV supervisory control system including multiple heterogeneous UVs are elicited. The research that has addressed each of the attributes is presented and the implications of transitioning to a single-operator-multiple-heterogeneous-UVs paradigm are discussed. Previously introduced models of human-UV interaction that describe the interaction through capturing the confluence of the different attributes are also presented. The chapter concludes by presenting the goal of this thesis, which is to introduce a discrete event simulation model that addresses the gaps left by previous work in modeling supervisory control of heterogeneous UVs.

In Chapter 3, titled "Discrete Event Simulation Model for Supervisory Control of Heterogeneous Unmanned Vehicles," a multiple unmanned vehicle discrete event simulation (MUV-DES) model is introduced. This model takes as inputs the human-UV attributes identified in Chapter 2 and serves as a model of human-UV interaction during mission execution. The model utilizes queuing theory to represent the human as a server attending to vehicle-generated exogenous and endogenous tasks, which are defined by their arrival and service processes. The chapter also highlights the method by which the human-UV attributes can be mapped to the model constructs. Finally, the chapter concludes by discussing the model benefits by highlighting the MUV-DES model's ability to overcome the previous research gaps identified in Chapter 2.

In Chapter 4, titled "Model Validation," validation efforts are presented whose aim was to build confidence in the model's accuracy, as well as the model's robustness at modeling supervisory control of heterogeneous UV teams. First, a description is presented of an online experimental test bed, RESCHU, which was designed to support online experimentation in supervisory control of multiple heterogeneous UVs. An experiment conducted with RESCHU that was used to gather large quantities of data for the MUV-DES model validation process is then described. Model accuracy and robustness of the MUV-DES model are then addressed by presenting the findings from the following validation techniques: replication validation, predictive validation, extreme condition testing, internal model stochasticity testing, sensitivity analyses, historical data validation, and construct validation.

Chapter 5, titled "Model Synthesis," demonstrates how the proposed MUV-DES model aids in designing and assessing heterogeneous UV teams and related technologies. Three applications of heterogeneous UV systems are presented where exploration of a large design space is required. In each of the cases, the ability of the MUV-DES model to streamline the design process is elicited. As part of exemplifying the model's ability to aid UV system designers in evaluating the impact of potential design modifications, the MUV-DES model is utilized to make design recommendations for the online experimental test bed introduced in Chapter 4. These design recommendations were implemented and an additional experiment was conducted in order to validate the predicted impact of the design recommendations.

Finally, Chapter 6, titled "Conclusions", consists of a discussion of the conclusions of this research, a discussion of the model's applicability to other supervisory control domains, potential avenues for future work, and the contributions that the research in this thesis has made to the domain of supervisory control.

The research methodology presented in Figure 1.7 serves as a graphical representation of the thesis outline.

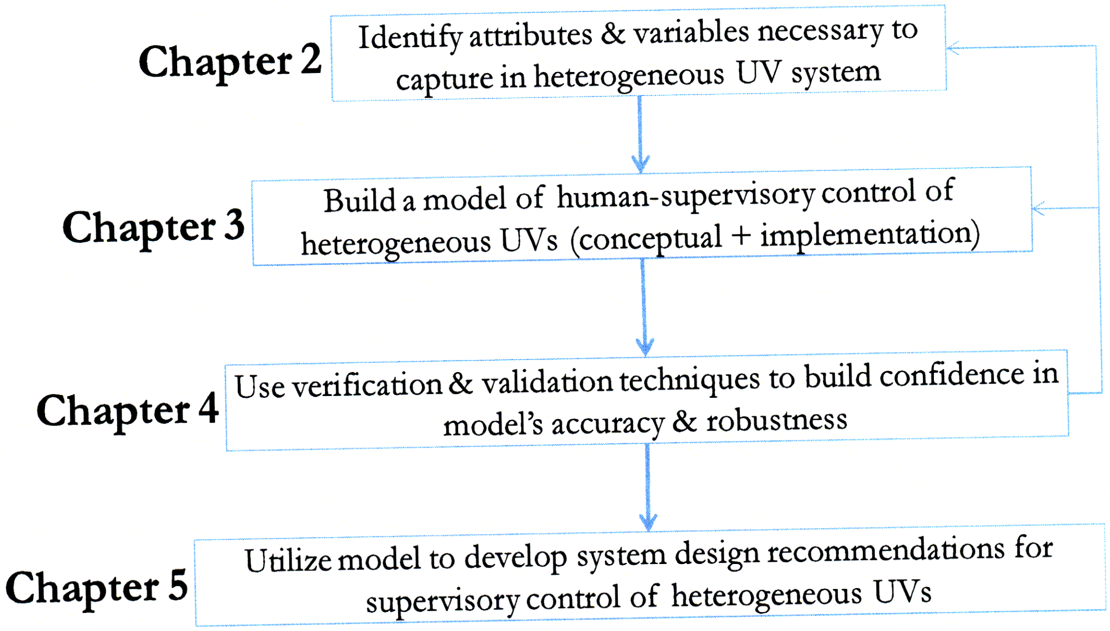


Figure 1.7 Research Methodology/Thesis Outline

2 BACKGROUND: MODELING UNMANNED VEHICLE SYSTEMS

An increasing body of literature has addressed different aspects of human-UV interaction. For example, while research in adjustable autonomy (Goodrich, Olsen et al., 2001; Miller & Parasuraman, 2007; Parasuraman, Barnes, & Cosenzo, 2007) has investigated the aspect of role allocation between the human operator and vehicles, research in vehicle-task assignment algorithms (Alighanbari & How, 2006; Valenti, Bethke et al., 2007) has investigated the aspect of inter-vehicle task allocation in a multi-UV team. This chapter begins by identifying a list of human-UV interaction attributes that are important to consider in the supervisory control of multiple heterogeneous UVs. The research that has addressed these attributes is elicited and the implications of transitioning to a single-operator-multiple-heterogeneous-UVs paradigm are discussed. Previously introduced models of human-UV interaction that describe human-UV interaction through capturing the confluence of the different attributes are also presented. The inadequacy of these models to capture the design implications of interest in supervisory control of heterogeneous UV teams is then identified. The chapter is then concluded by discussing the research methodology proposed to fill this gap.

2.1 Human-UV Interaction Attributes

Several researchers have identified attributes that describe different aspects of human-UV interaction (Crandall & Cummings, 2007b; Cummings & Mitchell, 2008; Donmez, Pina, & Cummings, 2008; Olsen & Goodrich, 2003; Steinfeld, Fong et al., 2006). Based on this previous work, it is proposed that for each of the human and vehicle elements, the team structure, the behavior of an individual entity in the team and the allocation of tasks to team members must be considered.

In this chapter, vehicle team structure, role allocation and level of vehicle autonomy, and vehicle task allocation are identified as describing the behavior of the vehicle element in human-UV interaction. These three attributes define human-UV interaction through shaping the tasks that operators must accomplish. Since this thesis considers a paradigm of a single operator supervising multiple UVs, the structure of the human team, as well as the task allocation between human team members, is not considered. Consideration of multi-operator teams is left for future work. Therefore, for the single human element, only the nature of operator interaction is identified as describing the behavior of the human element in human-UV interaction. In summary, this thesis focuses on the following attributes:

- Vehicle team structure
- Role allocation and level of vehicle autonomy
- Vehicle task allocation
- Nature of operator interaction

These four attributes collectively define single-human-multi-UV interaction for supervisory control of multiple heterogeneous UVs, and will therefore be relevant to the designers of such systems. For each attribute describing human-UV interaction, the previous research highlighted the importance of identifying variables that serve as key performance parameters (Crandall & Cummings, 2007b; Cummings & Mitchell, 2008; Donmez, Pina, & Cummings, 2008; Olsen & Goodrich, 2003; Steinfeld, Fong et al., 2006). In the following sub-sections, for each of the four attributes, previous research that has introduced variables that serve as performance parameters will be discussed. While this discussion will focus on each of the attributes independently, section 2.2 will discuss the importance of the interdependence between the attributes.

2.1.1 Vehicle Team Structure

An important attribute in supervisory control of multiple UVs is the vehicle team structure. The first critical variable that defines vehicle team structure is the size of the team. An increasing body of literature has examined the capacity of single operators to supervise multiple UVs (Cummings & Guerlain, 2007; Cummings & Mitchell, 2008; Cummings, Nehme et al., 2007; Dixon, Wickens, & Chang, 2005; Olsen & Wood, 2004; Ruff, Narayanan, & Draper, 2002).

Initial work investigated the possibility of reducing operator workload as a means to merge the dual operator functions of navigation and payload management. For example, Dixon et al. (2005) showed that auditory and automation aids could be used to support single operator control of a single UAV (both navigation and payload). On the other hand, Ruff et al. (2002) showed experimentally that in the case of unmanned aerial vehicle control, workload does not always decrease as human tasks are automated. This is because increasing levels of automation could result in the human being removed from the decision-making process, which could result in lower situational awareness and human-trust potentially decreased. More recent work has addressed operator capacity in terms of multi-UV control. For example, in terms of actually predicting how many identical UAVs a single operator can control, Cummings and Guerlain (2007) showed that operators could experimentally control up to twelve Tactical Tomahawk missiles given significant missile autonomy.

In general, these studies showed that a single operator could theoretically control multiple unmanned vehicles, but that appropriate automated offloading aids would need to be provided. However, these studies also point out that, due to human limitations, workload does not always decrease as human tasks are automated. Therefore, these studies showed that the team size aspect of human-UV interaction is not only dependent on the level of autonomy attribute, but also on the nature of operator interaction.

One issue with these predictions is that they are experimentally-based, which limits their generalizability. In addition, this previous research has mainly focused on the supervision of a homogeneous set of UVs. However, as UV system mission goals become increasingly demanding, the composition of UV teams is likely be driven by a requirement for interoperability among UVs of varying attributes (such as operational domain, air vs. land). Also, in addition to heterogeneity across vehicle types, even a single UV can have multiple payloads, which will ultimately lead to heterogeneity in operator tasks (for example, a UV can have both and electro-optical camera and a laser designator).

This thesis is concerned with multi-vehicle teams where the team structure is not restricted to homogeneous composition, which is what previous studies examined, but can encompass any level of heterogeneity. Specifically, the aim is to explore the impact of heterogeneity across and within vehicles on the operator, as this will ultimately decide operator limitations.

2.1.2 Role Allocation and Level of Autonomy

Role allocation is concerned with the role of the operator and automation in terms of authority in decision making, as well as for solution generation. The level of automation (LOA) concept (Fitts, 1951; Kaber & Endlsey, 2004; Parasuraman, Sheridan, & Wickens, 2000; Sheridan & Verplank, 1978) addresses role allocation through the introduction of taxonomies that classify human-vehicle interaction schemes. The LOA scales generally represent a continuum that varies between full operator responsibility and full vehicle autonomy.

In a UV system, the role allocation schemes that are feasible are largely dependent on the ability of the vehicle to act independently from the operator. For example, vehicles that require remote manual control can only function under a level of automation that assigns full responsibility and decision making to the operator. In contrast, vehicles with auto-pilot capabilities can be assigned levels of automation that have a more balanced allocation of roles such that some of the decision making and responsibility is assigned to the vehicle. Therefore, the vehicle's achievable level of autonomy will often dictate feasible role allocation schemes. One operational definition of the vehicle's level of autonomy is the concept of neglect tolerance (Crandall, Goodrich et al., 2005; Olsen & Goodrich, 2003). Neglect tolerance is defined as the random process that describes the change in vehicle performance over time as the vehicle is neglected by the operator. Additionally, neglect time (NT) is defined as the mean length of time for which a vehicle can be neglected before its performance falls below some acceptable threshold.

Further research has explored concepts such as adjustable autonomy (Goodrich, Olsen et al., 2001; Miller & Parasuraman, 2007), which have considered more dynamic methods for the division of labor between the human and UV. Adjustable autonomy entails a paradigm where the UV can dynamically vary the degree to which it acts autonomously, allowing it to exploit human abilities in order to improve its performance (Scerri, Pynadath, & Tambe, 2001). As such, role allocation between human and machine could vary depending on the system state, such that the most appropriate level of automation can vary given different situations or functions. For example, a UV that under normal conditions awaits operator approval for new trajectories could resort to management-by-exception under high task load conditions (such that the automation takes a more active role in synthesizing/executing decisions unless vetoed by the human operator). This would ensure that the operator's high task load will not impact the vehicle's ability to travel to targets. Although adjustable autonomy could, in certain cases, be a solution that overcomes human limitations, it could also result in automation actions that are not transparent, potentially leading to mode confusion and lack of operator trust (Parasuraman, Sheridan, & Wickens, 2000). A similar but distinct concept to adjustable autonomy is that of mixed initiative interaction (Bradshaw, Feltovich et al., 2004; Hearst, 1999; Horvitz, 1999). Mixed initiative interaction refers to a paradigm of human-UV interaction with a flexible interaction strategy in which each agent (human or vehicle) contributes what is best suited at the most appropriate time.

As the operator's role in futuristic UV systems changes from manual control to one of mission management, and as the level of autonomy infused in UVs increases, the role allocation between human and machine is likely to be impacted such that much of the decision making and solution generation is assumed by the UVs. However, the work on adjustable autonomy and mixed initiative interaction has reinforced the importance of the human in the system, even at high levels of autonomy (Bradshaw, Feltovich et al., 2004). This is because tasks that require human judgment and reasoning (such as analyzing sensor data, responding to emergent tasks, etc.) will require a human operator in order to be performed effectively and efficiently.

2.1.3 Vehicle Task Allocation

In addition to role allocation, task allocation across and between the vehicles will also be important to consider. The multi-vehicle task allocation problem can also be phrased as a question: given some vehicles and some tasks, which vehicle(s) should execute which tasks(s) (Gerkey & Mataric, 2004)? Although task allocation is relatively simple in the homogeneous case (one type of vehicle and one type of task), it is possible to have a multitude of different pairings of vehicles and tasks in the heterogeneous case. When heterogeneity occurs in vehicle capability and mission tasks, there are different combinations of vehicle-task pairs that become possible. Various architectures have been proposed for vehicle task allocation, generally in the categories of centralized and decentralized allocation schemes.

In a centralized allocation scheme, a central planner system identifies and allocates tasks required to achieve the underlying mission objectives to the different vehicles in a team. On the other hand, decentralized task allocation entails vehicles planning individually while taking into account the tasks that other vehicles in the team are completing. Decentralized task allocation can be used in order to reduce the reliance on a central re-planner, thereby increasing the rate at which vehicles can react to emergent threat and/or targets of opportunity, as well as the robustness of the overall system (Alighanbari & How, 2006; Brunet, Choi et al., 2008; Pavone, Bisnik et al., 2007). However, such a task allocation scheme presents several challenges, mainly the communication overhead required to exchange information about a vehicle's current state, as well as future plans. Communication becomes increasingly challenging as the level of coordination required between vehicles increases.

Coordination can either be intentional, where different vehicles are allocated a task requiring them to coordinate through inter-vehicle communication, or emergent, where the cooperation is an opportunistic result of interacting in the same environment (Gerkey & Mataric, 2004; Parker, 1994). Thus, another variable of concern is the level of coordination between vehicles. Cooperation has been defined as a level of coordination that entails coordinated task assignment with the additional knowledge of the future implications of a vehicle's actions on the performance of other vehicles (Alighanbari & How, 2005). The importance of cooperation between unmanned vehicles was highlighted in a study by Chandler (2004), where cooperative control design approaches for UAV tactical missions were analyzed. The results showed that insufficient shared information between unmanned vehicles can destroy team coherence and the ability of the vehicles to achieve team objectives (Chandler, 2004). One issue with the initial definition of neglect tolerance is its inability to explicitly capture the effects of vehicle cooperation; the definition suggests that neglecting a vehicle has an effect on its own performance and not on that of other, possibly, cooperating vehicles. More recent work by

Wang et al. (2008) has extended the concept of neglect tolerance by introducing inter-vehicle dependence to support cases of cooperating vehicles.

Vehicle task allocation is important with respect to the paradigm of control this thesis is interested in, supervisory control of multiple heterogeneous UVs, as alternate vehicle task allocation schemes can result in different demands imposed on the operator. For example, a team of UVs where cooperation is allowed could result in less demand from the operator due to vehicle team self-reliance. Conversely, cooperation that requires the operator to play a role in the coordination can result in additional demands on the operator. Not accounting for the effect of task handling can result in models that underestimate or overestimate operator task-load, and therefore operator performance.

2.1.4 Nature of Operator Interaction

Although the previous three attributes are important in dictating the level of interaction required from the operator, modeling human-UV interaction also requires modeling the nature of operator interaction, which plays a critical role in defining how the operator handles his/her tasks.

2.1.4.1 Serial vs. Parallel Interaction

Previous research (Carbonell, Ward, & Senders, 1968; Senders, 1964) has modeled the operator in supervisory control settings by breaking down operator-allocated functions into low-level sub-tasks. There are several methods for deriving these sub-tasks, such as using symbolic methods for cognitive modeling among others (Card, Moran, & Newell, 1983; Laird, Newell, & Rosenbloom, 1987). There are two main schools of thought: one that has modeled the human as a serial controller (Carbonell, Ward, & Senders, 1968; Schmidt, 1978; Senders, 1964) and another that modeled the human as a parallel processor capable of attending to tasks simultaneously, dealing with information on more than one channel (Wickens & Hollands, 2000). These two schools of thought are not necessarily conflicting, in that the appropriateness of the human model depends on the level of processing under consideration. While the serial processing model has been applied to higher level tasking, the parallel process has been shown to be applicable for the most part for lower-level perceptual-type processing such as that involved in driving-type tasks (Liu, Feyen, & Tsimhoni, 2006).

When considering supervisory control tasks for complex systems such as those in UV systems, humans are generally required to handle high level tasking that involves application of human judgment and reasoning. As such, humans will act as serial processors in that they can only solve a single complex task at a time (Broadbent, 1958; Welford, 1952). Serial modeling approaches have successfully modeled the human as a single server in a queuing network attending to tasks sequentially. For example, Schmidt (1978) has suggested a single-server queuing system as appropriate for modeling an air traffic controller in charge of both conflict assessment and resolution, as well as general routing type tasks. Similarly, Carbonell et al. (1968) utilized a single-server system to study a pilot's monitoring of his/her instrument panel. Further work has attempted to combine both the serial and parallel approaches by modeling the human as composed of multiple servers in a queuing network, one server for each parallel channel (Liu, 1997; Liu, Feyen, & Tsimhoni, 2006). The model was successfully used to capture perceptual-type processing in a driving-type task with performance data of the model being similar to human subjects performing the same task.

In addition, Senders (1964) argues that the operator is not just a single-channel controller, but one that is dealing with sampled data as opposed to being a continuous controller. This is consistent with supervisory control of UVs, which requires a human operator handling intermittent events as they are generated, providing limited attention to each event. Such a paradigm of limited attention is captured in the concept of interaction efficiency, which was proposed by Crandall et al. (2005) and defined to be the random process that describes the change in vehicle performance over time as the operator interacts with the vehicle. Interaction time (IT), a derivative measure, was defined as the mean length of time for which the operator must interact with the vehicle to ensure it is still working towards mission accomplishment.

2.1.4.2 Attention Allocation Strategies

In general, due to serial operator behavior associated with high level tasks, operators will have to ration their attention between the various tasks, providing limited attention to each one.

The method by which operator attention is allocated will have an impact on the effectiveness of operator interaction. In terms of operator modeling, two attention allocation strategies have been identified from previous work, which are critical in the supervisory control of multiple unmanned vehicles: switching strategy and complexity mitigation strategy.

An important issue in supervisory control of multiple entities is the order by which entities are selected for service. The switching strategy is the procedure used for selecting the next task to service in a serial operator model. The order by which the tasks are serviced affects the total time that tasks spend waiting for service (i.e. a first come first serve switching policy is not necessarily the optimal strategy). Moreover, when a human operator switches between two different tasks, this is accompanied by a mental model switch that comes at a time cost, or switch cost. The switch cost is not limited to switching between cognitively complex tasks, but exists even when humans switch between cognitively simple ones (Rogers & Monsell, 1995). For example, both Goodrich et al. (2005) and Squire et al. (2006) demonstrated that the existence of context switching costs in multi-vehicle control is unavoidable, and that the amount of time required to switch between vehicles can be substantial. This further emphasizes the importance of considering the operator switching strategy in futuristic systems, as switching between different combinations of heterogeneous vehicles is likely to lead to alternate switch costs.

Research by Tulga and Sheridan (1980) examined the problem of multi-task attention allocation, specifically the impact of different switching strategies when the problem is one of allocating attention to multiple simultaneous task demands which appear randomly, last for various periods, and offer varying rewards for service. They developed a model of the human decision maker in this setting and compared it to human-in-the-loop experimental results. It was observed that there were two competing strategies: a) servicing important tasks first, where important tasks are those that provide a high reward when serviced, and b) servicing urgent tasks first, where urgent tasks are those that have very little slack in terms of when they can be handled (i.e. they have fast approaching deadlines). Tulga and Sheridan (1980) showed that for systems where the human operator is experiencing high task load, servicing important tasks first while ignoring the closest-to-deadline factor results in the best performance.

Similarly, Neth et al. (2006) explored human multi-tasking behavior in a synthetic task environment that consisted of multiple independent, heterogeneous tasks competing for the operator's attention. The tasks consisted of multiple progress bars that were independent of one another and heterogeneous in the rewards they offered in return for human attention. Performance was measured with respect to a scoring function that awarded points as the operator engaged the tasks which required attention as time progressed. Human performance profiles were compared to those of several artificial agents with the result that human operators performed worse than the majority of the agents (with respect to the scoring function). Findings showed that operators learn quickly, but that their performance tends to asymptote at a subpar level when compared to that achieved by the best performing artificial agents. Neth et al. (2006) explained that a contributing factor to the sub-optimal human performance, when compared to the best performing artificial agents, was the non-optimal ordering by which the heterogeneous tasks were serviced.

However, the implications that optimal strategies can be deduced and should be utilized in operator training or enforced through decision support are overly simplistic. Although Neth et al. (2006) present their synthetic task environment as capturing aspects of supervisory control, monitoring, and complex system management, dissecting their system into the attributes of human-autonomy interaction reveals that the interactions are much simpler than those expected in a human-UV system. Specifically, the role allocation attribute assigns little decision making and responsibility to the operator (the operator's sole task is to click on a button underneath a progress bar), and the task-allocation attribute is too simplistic for modeling multi-UV control in that the various tasks (progress bars) are independent from one another. The difference is that in a human-UV system, the different attributes are much more inter-related thereby increasing the system complexity. Recent work by Crandall and Cummings (2008) highlighted the ability of operators to identify and apply switching strategies that outperform "optimal" strategies derived using a model of the human-UV system. In addition, Crandall and Cummings (2008) point out that even if system designers could derive switching strategies deemed better performing, which is difficult especially in uncertain command and control environments, these strategies will not be effective unless they are accepted by human operators. Post-experimental subjective questionnaires presented to subjects revealed that subjects preferred more flexible decision support that did not enforce a specific switching strategy, but allowed operators to exercise human judgment and reasoning.

Given this previous research, switching strategies are likely to be an important element of attention allocation in the supervisory control of multiple heterogeneous unmanned vehicles. For this reason, designers of heterogeneous UV systems should be concerned with designing systems that are robust to different switching strategies. However, with the increase in number of and type of vehicles for heterogeneous UV teams, it is unclear how different switching strategies will impact operator performance in such settings. Moreover, the performance of different switching strategies is likely to depend on other attributes as was highlighted above.

Another operator strategy is the mitigation of problem complexity by defining abstractions that reduce the complexity of the problem at hand. Research on air-traffic controllers, who play a supervisory role similar to that of UV operators, has shown that they use structure-based abstractions that reduce the complexity of the three-dimensional (3D) problem (Histon, Hansman et al., 2002). Structure-based abstractions are defined as abstractions that reduce the apparent complexity of the input space, thereby simplifying the decision making process. Several structure-based abstractions were identified that help an air traffic controller with projecting future locations of aircraft, such as: (a) following standard flows that are based on explicit structural elements in the airspace, (b) forming groupings of aircraft with common properties, and (c) concentrating on critical points such as the intersection of aircraft flows. Histon et al. (2002) explained that each of the structure-based abstractions reduces cognitive complexity by simplifying the task of projecting future traffic situations.

Similarly, the use of a grouping strategy by operators to counter increased workload has also been observed in the unmanned vehicle world. In examining how people manage vehicles under different autonomy levels and different coordination capabilities, Goodrich et al. (2007) found that operators tend to group vehicles into informal teams as a way of coping with increased workload. Miller et al. (2005) introduced a human-automation integration architecture, called Playbook©, that provides a means of human-automation communication about plans, goals, methods and resource usage. Playbook© enables human operators to flexibly interact with vehicles similar to the way coaches interact with well-trained humans using a sports team's playbook. In this approach, vehicles are managed as a team and not as individuals, as would be the case in a sequential management style. The Playbook© architecture translates these plays into a decision support tool that supports complexity mitigation by encouraging operators to deal with vehicles at the group level.

Given this previous research, grouping of vehicles is likely to be an important element of complexity mitigation in the supervisory control of multiple heterogeneous unmanned vehicles. However, with the increase in number of and meaning of various groupings for heterogeneous UV teams (such as by vehicle, by task type, by geographical proximity, etc.), it is unclear how such structure-based abstractions will manifest themselves in these settings. Moreover, the applicability of complexity mitigation strategies is likely to depend on vehicle-based attributes such as the role allocation and level of autonomy. For example, in order for operators to adopt vehicle-grouping strategies, human-automation communication will have to be constrained to mixed-initiative concepts that support dynamic allocation of functions between humans and automation (Linegang, Haimson et al., 2003).

2.1.4.3 Operator Situational Awareness

While attention allocation focused on clearly observable operator interactions, an important human behavior precursor not accounted when considering attention allocation is operator situational awareness (SA). SA is defined as the combination of perception of elements in the environment, the comprehension of their meaning, and the projection of their status in the future (Endsley, 1995). Whereas operator attention allocation strategies capture operator choices in terms of attention rationing, operator SA captures the ability of operators to make effective choices as a function of information processing efficiency. As such, SA can significantly influence human behavior and hence, human-UV system performance (Donmez, Pina, & Cummings, 2008). Because SA is dynamic, it can influence operator responses over time and as a result can dynamically impact supervisory control performance. In the single-operator-multiple-UV setting, Cummings & Mitchell (2008) demonstrated that models that do not explicitly account for losses in operator SA will either over or under-predict system performance metrics, and thus be inherently unreliable.

Whereas alternate operator attention allocation strategies result in vehicle downtime and added interaction times due to context switching and decision making, operator SA impacts the times that vehicles have to wait before being noticed by the operator. Cummings & Mitchell (2008) proposed that in addition to a vehicle having to wait while the operator is busy attending to other vehicles, vehicles could also experience wait times due to operator loss of situational awareness (WTSA). WTSAs occur when operators are not aware of vehicles requiring their attention. Cummings & Mitchell (2008) showed that WTSA can account for the largest part of vehicle wait time, and significantly reduce the overall number of vehicles that a single operator could control.

In summary, operator situational awareness has been shown to have a significant impact on the nature of unmanned vehicle operator attention allocation and overall system performance. In addition to attention allocation, it is therefore important to consider SA in a model of supervisory control of heterogeneous unmanned vehicles.

Given this previous research, it can be concluded that a) the human operator should be modeled as a single server responding to discrete samples of data generated by the vehicles being supervised, b) IT is the key performance variable that links the human operator to the unmanned vehicles system, c) attention allocation strategies will need to be taken into account in order to capture the full impact of heterogeneity, and d) operator situational awareness could be a key limiting factor in the supervisory control of multiple unmanned vehicles.

The attributes of human-UV interaction and their associated variables that have been identified in the previous discussions are summarized in Table 2.1. Identifying variables that define different aspects of human-UV interaction is critical in understanding the different design choices that systems designers will face, as well as to develop a supervisory control model of multiple heterogeneous UVs. The variables in Table 2.1 are not independent. For example, the level of automation at which vehicles operate will impact the size of the vehicle teams that can be handled by the operator. Similarly, depending on the level of coordination between vehicle tasks, different levels of automation could be required to support the coordination. Most importantly, team structure, level of autonomy, and task allocation variables will be constrained by operator interaction times and attention allocation strategies.

By capturing the interaction between the different variables, a model of human-UV interaction can be used to understand the impact of alternate design choices on system effectiveness.

Table 2.1 Human-UV system attributes and associated design variables

Attribute	Variables
Vehicle Team Structure	- Vehicle team size - Vehicle team heterogeneity
Role Allocation / Level of Autonomy	- Level of automation - Neglect times
Vehicle Task Allocation	- Level of decentralization - Level of coordination
Nature of Operator Interaction	- Interaction times - Situational awareness - Attention allocation strategies

2.2 Models of Human-UV Interaction

The development of models to predict optimal UV team size has been addressed in a number of previous research efforts. Using the ideas of NT and IT, a temporal-based model has been proposed for modeling human-UV interaction (Olsen & Goodrich, 2003; Olsen & Wood, 2004). The timeline presented in Figure 2.1a can be seen as composed of segments, each of length NT+IT. In a single robot example, the operator interacts with the robot for length of time IT, and then ignores it for length of time NT during each segment. As shown in Figure 2.1b, supervising multiple robots is then captured by adding additional segments of operator interaction while the first vehicle is being neglected (the NT time partition is replaced by ITs for the additional robots).

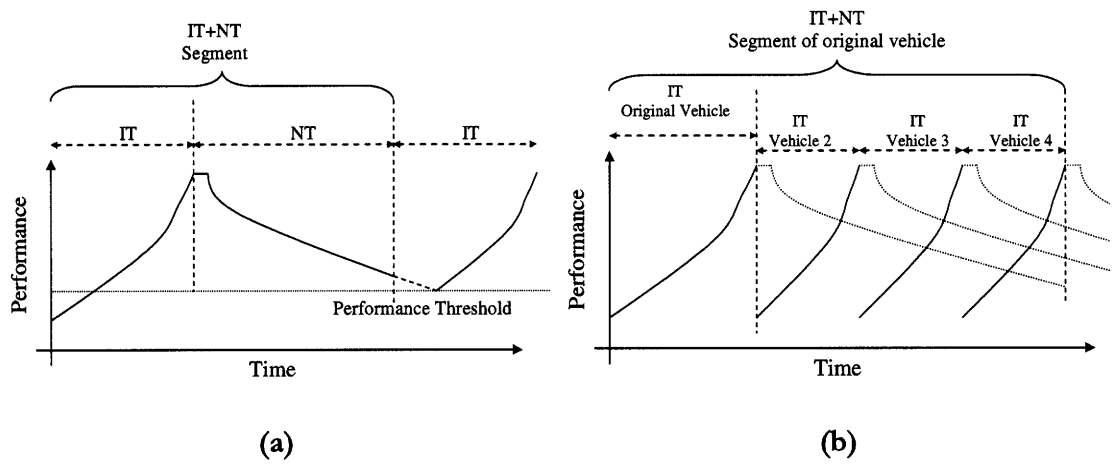


Figure 2.1 The relationship of NT and IT for multiple vehicles (Cummings, Nehme et al., 2007)

The notion of Fan-out (FO) as a function of NT and IT as seen in Figure 2.1b describes the upper bound on the number of independent, homogeneous robots that a single person can manage (Olsen & Goodrich, 2003; Olsen & Wood, 2004):

$$FO = \frac{NT + IT}{IT} = \frac{NT}{IT} + 1 \qquad (1)$$

NT and IT are used in Equation 1 to calculate the total number of robots a single human operator can control such that he/she can make use of the neglect time of the one robot and convert it into ITs for additional robots. Other work has modified the concept of Fan-out to extend it to heterogeneous robots by taking into account vehicles with different NTs and ITs (Crandall, Goodrich et al., 2005; Goodrich, Quigley, & Cosenzo, 2005).

However, as appealing as it is due to its simplicity, in terms of human-automation interaction, the Fan-out estimations in Equation 1 make several assumptions that cause model overestimation of the number of vehicles that can be controlled (Cummings & Mitchell, 2008; Cummings, Nehme et al., 2007):

- Requests for interaction from vehicles arrive serially and are handled instantly, so that no queues develop while robots are waiting on the operator

- The operator is efficient and does not lie idle while vehicles need attention

- The operator appropriately allocates his/her attention to the correct vehicle in need

- Reliance on averages (NT and IT)

If these assumptions do not hold, then an additional critical variable must be considered when modeling human control of multiple vehicles, which is the concept of Wait Time (WT) (Cummings, Nehme et al., 2007). In supervisory control, any sequence of tasks requiring complex cognition will form a queue and consequently, wait times will build. Wait times occur when a vehicle is operating in a degraded state and requires human intervention in order to achieve an acceptable level of performance. In the context of a system of multiple vehicles, wait times are significant in that as they increase, the actual number of vehicles that can be effectively controlled decreases, with potential negative consequences on overall mission success. Cummings and Mitchell (2008) proposed a modification to Equation 1 to include the concept of wait times as shown in Equations 2 and 3.

$$WT = \sum_{i=1}^{X} WTI_i + \sum_{j=1}^{Y} WTQ_j + \sum_{k=1}^{Z} WTSA_k \tag{2}$$

$$FO = \frac{NT}{IT + \sum_{j=1}^{Y} WTQ_j + \sum_{k=1}^{Z} WTSA_k} + 1 \tag{3}$$

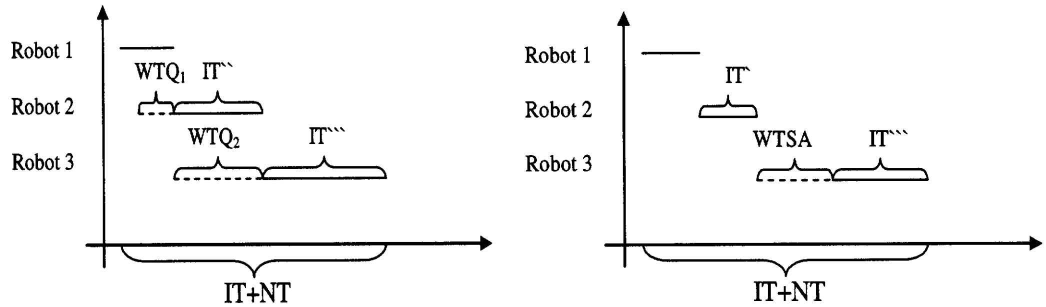
Equation 2 categorizes total system wait time as the sum of:

- The interaction wait times, which are the portions of IT that occur while a vehicle is operating in a degraded state (WTI)

- Wait times that result from queues due to near-simultaneous arrival of problems (WTQ)

- Wait times due to operator loss of situation awareness (WTSA)

In Equation 2, X denotes the number of times an operator interacts with a vehicle while the vehicle is in a degraded state, Y indicates the number of interaction queues that build, and Z indicates the number of time periods in which a loss of situation awareness causes a wait time (Cummings, Nehme et al., 2007). Figures 2.2a and 2.2b further illustrate the relationship of wait times to interaction and neglect times.



(a) (b)

Figure 2.2 (a) Queuing wait times versus (b) situational awareness wait times (Cummings, Nehme et al., 2007)

Although queuing wait times are a function of serial operator behavior when supervising complex tasks, wait times due to interaction and situational awareness are a function of two important human limitations. Wait times due to interaction capture the first limitation which includes the time that an unmanned vehicle idly waits while a human operator attempts to reacquaint himself with the vehicle's current task incurring switching times, as well as any times incurred in the operator's decision-making process. This is consistent with the claim by Goodrich et al. (2005) that longer switching times can dramatically decrease the upper bound on the number of manageable robots. Loss of situational awareness can also significantly decrease the upper bound due to WTSA as was discussed in Section 2.1.4.3.

While more pessimistic than the original Fan-out Equation (1), the revised Fan-out Equation (3) can really only be helpful for gross "ballpark" predictions of operator capacity. This methodology could provide system engineers with a system feasibility metric for early manning estimations (Cummings, Nehme et al., 2007). However, when considering heterogeneous unmanned vehicle systems, there are several gaps left by previous work that causes these models to fall short in addressing designer needs, which will be discussed next.

2.2.1 Lack of Cost Trade Space

First, the Fan-out Equation, revised or otherwise, theoretically predicts the maximum number of vehicles an operator can effectively control, but what is effective is often a dynamic constraint (Cummings, Nehme et al., 2007). Effectiveness is likely to depend on mission-based constraints such as maximizing the number of targets visited, operator performance-based constraints such as minimizing operator workload, or vehicle performance-based constraints such as minimizing average vehicle wait times. The problem with the Fan-out equations is that they do not take into account such explicit performance constraints, and therefore cannot represent any kind of cost trade space (serving instead as upper-bound estimators).

2.2.2 Non-Holistic Models

Second, these previous models have each addressed one or more aspects of human-UV interaction but do not collectively address the different variables; the Fan-out equations are not holistic in their modeling of human-UV interaction attributes. For example, Cummings and Mitchell (2008) addressed human limitations, something that was lacking in the earlier work on Fan-out (Olsen & Goodrich, 2003), but neither explicitly addresses the heterogeneity that could be present in UV teams or the interaction of the variables presented in Table 2.1. This creates a situation where different models are appropriate for different designer intentions. A model that will support designers of future UV systems must be holistic, allowing the designers to vary one or more attributes simultaneously.

2.2.3 Partial Addressing of Heterogeneity

Third, although these previous models have recognized the possibility of heterogeneity, they addressed the issue solely by accounting for the different demands that heterogeneous vehicles/tasks could require of the operator through different NTs and ITs. Since the vehicles and associated tasks are disparate in the heterogeneous case, there is a large diversity in possible attention allocation schemes and the method by which operators allocate their attention across the different tasks and vehicles is likely to affect system performance. Not accounting for the variability that heterogeneous teams introduce could render any design recommendations from extant models inaccurate. In fact, Mau and Dolan (2006) investigated the effects of several alternative scheduling algorithms with the objective of minimizing the total time vehicles/tasks spend waiting for operator attention (called downtime). Utilizing the concept of NT to define the rate of task generation and IT to represent the service time for those tasks, they showed that carefully crafted priority-based scheduling algorithms can result in considerably better performance in terms of downtime than a simpler first-in-first-out (FIFO) algorithm, which was assumed by the Fan-out based models.

2.2.4 Cause-and-effect Human Models

Finally, these previous models capture operator interaction through simple cause and effect. For example, the operator is assumed to react immediately to a task requiring operator attention without taking into account any delays due to loss of operator situational awareness. In the development of complex systems such as those with operators supervising UVs, models of the operator are critical in order to define system boundaries (such as how many UVs one operator can be expected to control), as well as system requirements (such as how much human interaction a system should enable). As was highlighted in Section 2.1, heterogeneity in vehicles/tasks is likely to place stronger emphasis on the importance of operator cognitive processes, including the ability to maintain situational awareness. Since this is not included in the extant models, this could cause these models to under or overestimate the efficiency of the overall interaction.

2.3 Thesis Research Area

In order to support heterogeneous UV system design, a simulation-based model is presented that addresses the gaps identified in previous work. The modeling technique utilized is that of discrete event simulation (DES). Due to the time-critical, complex event-driven nature of human supervisory control, DES, which models a system as it evolves over time by representation of events (Law & Kelton, 2000), can be used to model supervisory control systems. Previous examples of related models include the use of queuing-based DES models to describe the method by which pilots allocate and make sense of alerts and cues from cockpit instrumentation (Carbonell, 1966), and when air traffic controllers manage multiple aircraft (Schmidt, 1978).

In addition to the event driven nature of DES, which is particularly appropriate for modeling supervisory control in command and control settings, there are several other properties of DES that allow a DES model to capture the dynamic behavior of human-UV interaction.

First, the queuing-based nature of DES captures the operator's serial behavior in attending to complex supervisory control tasks. Also, the temporal-based nature of DES allows for the capturing of temporal-based metrics such as vehicle wait times and operator utilization, both of which have been identified as critical in determining supervisory control performance limitations. In addition, capturing of dynamic operator cognitive limitations through simulation lends itself to reproduction of system-based performance metrics.

In summary, a DES approach allows for the dynamic modeling that captures the intricacies of operator-vehicle interaction during mission execution. Chapter 3 proposes a DES model for grouping the different attributes defined in this chapter into a holistic model that allows designers of futuristic heterogeneous UV systems to understand their collective impact. Chapter 3 also discusses the proposed model's ability to overcome the gaps of previous work by: (1) supporting the representation of a cost-trade space; (2) capturing the collective impact of changes in the design variables on the efficiency of operator supervisory control; (3) addressing the implications of heterogeneity; (4) modeling operator limitations through accounting of wait times due to loss of SA. Chapter 4 presents the results of a multi-stage validation process which was aimed at building confidence in the model's accuracy and robustness. The validation process was supported by an online research environment that was specifically developed to capture attributes of futuristic unmanned vehicle systems and through which a large subject base can be accessed. Finally, in Chapter 5, the model's utility is evaluated by addressing the following question: "How can such a model aid in the design and assessment of heterogeneous UV teams and related technologies?"

3 DISCRETE EVENT SIMULATION MODEL FOR SUPERVISORY CONTROL OF HETEROGENEOUS UNMANNED VEHICLES

In this chapter, a queuing-based multiple unmanned vehicle discrete event simulation (MUV-DES) model is presented that can help designers of heterogeneous unmanned vehicle (UV) systems understand the impact of design decisions on the effectiveness of operator supervision. An overview is first presented that discusses the model assumptions, as well as the attributes (from Chapter 2) captured by the model. The model is then described through the introduction of the DES-based constructs that enable the model to capture human supervision of multiple heterogeneous unmanned vehicles. This is followed by identifying the design variables which form inputs to the model and showing how they map to the DES constructs. Additionally, the output metrics that can be measured by the MUV-DES model for the purposes of evaluating designs are discussed. Finally, the chapter concludes by discussing the benefits of the proposed model through highlighting the model's ability to address the different research gaps identified in Chapter 2.

3.1 Overview

There are several assumptions that this model makes. The type of UV system that this model addresses is one where a single human operator is responsible for supervising a team of homogeneous or heterogeneous unmanned vehicles. The human operator could be situated in a ground-based, sea-based, or airborne control station through which he/she supervises the team of UVs. The operator is assumed to act in a supervisory control manner, although the exact role-allocation and the level of vehicle autonomy are system variables. Due to the supervisory control nature of the human-UV role allocation, it is assumed that the operator interacts with the vehicle at discrete points in time (i.e. there is no manual control or teleoperation of the vehicles that would require continuous attention).

The attributes that the model captures, which consist of the attributes from the previous chapter (Table 2.1), are shown in Figure 3.1 grouped by those related to the vehicle team (team structure, role allocation and level of autonomy, and vehicle task allocation), and those related to the human operator (nature of operator interaction). In addition, another attribute captured by the MUV-DES model which can affect the system is the environment unpredictability. Before explaining how the design variables associated with these attributes drive the model, the DES model constructs are explained.

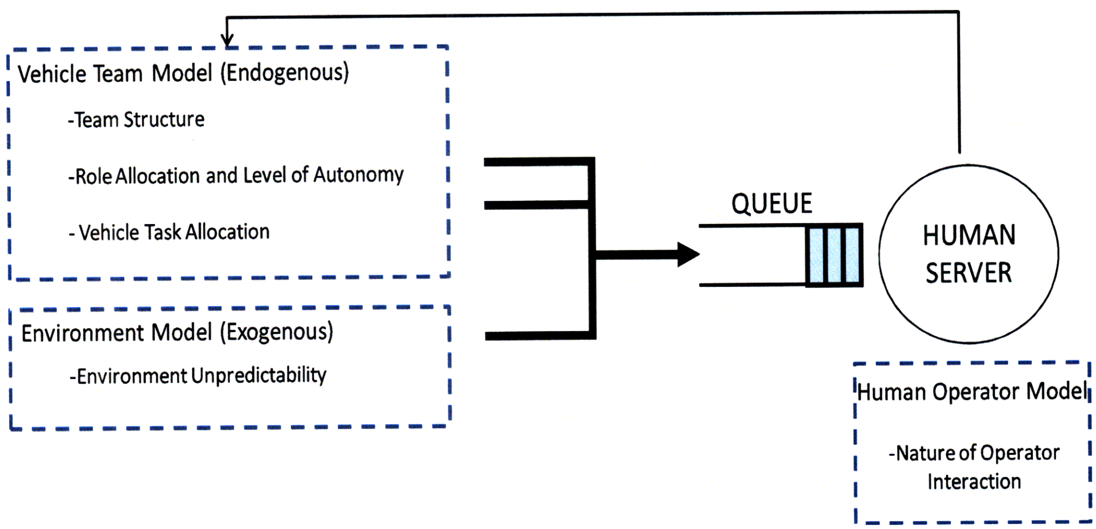


Figure 3.1 A high level representation of the multiple heterogeneous UV supervisory control discrete event simulation model

3.2 Model Constructs

In order to build a model that addresses the supervisory control tasks in UV systems, the goal of the modeling effort was to focus on representing average operator interaction that is context specific for human-UV systems. There are four elements that are critical to any DES model; the events in the system, the arrival processes for these events, the service processes for these events, and the queuing policy. These are discussed in the next sections.

3.2.1 Events

There are three general categories of events in this multi-UV, single operator system. The different event categories[1] are: a) vehicle generated events, b) operator-induced events, and c) environmental events (these different event categories are represented in Figure 3.2).

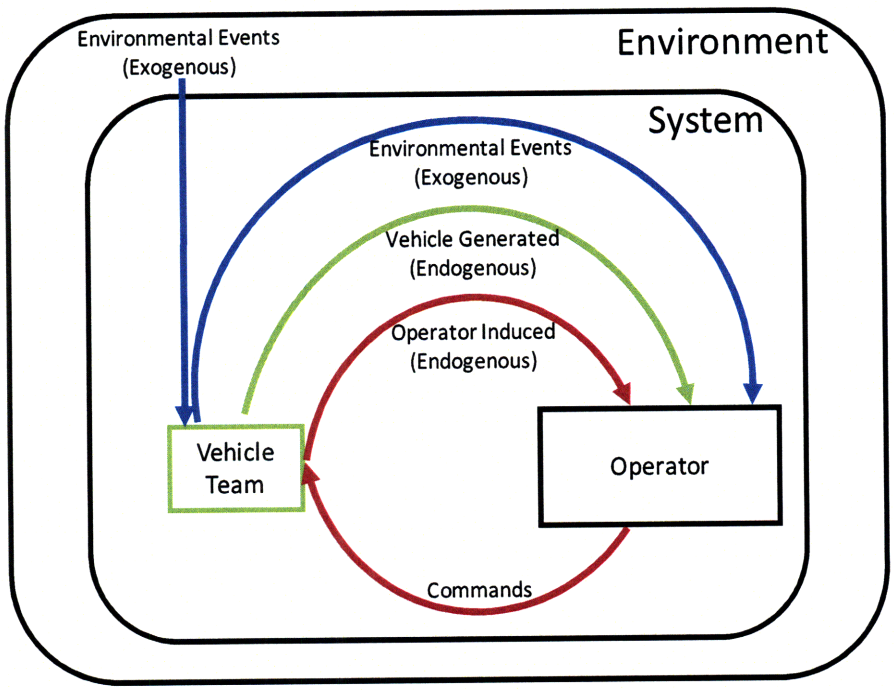


Figure 3.2 Event types

3.2.1.1 Vehicle-Generated Events

Vehicle-generated events are endogenous events that arise expectedly due to the nature of the mission and vehicle capability. An example of an endogenous vehicle-generated event is a ground vehicle that captures imagery which results in an event that requires operator identification of the captured image. This is an expected task, and the human operator needs to use judgment and reasoning to verify the contents of the image. Such a situation is expected in such an unmanned vehicle system as the system designers would be aware of the inability of

[1] Note that the term "event category" will be used to classify the event as being vehicle-generated, operator-induced, or environmentally-generated, the term "event type" will be used to identify a specific task in the human-UV system that is being modeled by events in the MUV-DES model (such as an analyze image event type), and the term "event" will be used to refer to specific instances of an event type, such as an analyze image event that is generated at the beginning of the mission.

the automation to analyze imagery and the need for the operator to apply human judgment and reasoning.

3.2.1.2 Operator-Induced Events

Operator-induced events, which are also endogenous to the human-UV system, address the ability of the human operator at any point of initiating a re-plan for a vehicle, even if the vehicle has not generated a task that requires human judgment and reasoning. This can occur when the operator modifies an existing plan (re-planning) with the expectation that the intervention will lead to improved performance. For example, in the case of a vehicle that is travelling to a target, the operator could re-plan an automation-generated path in order to better meet a time-on-target restriction. Although both vehicle-generated and operator-induced events are endogenous to the system, the former are pushed on to the operator by the system, whereas the latter are generated by the operator.

3.2.1.3 Environmental Events

Environmental events are events that are exogenous to the system, arising unexpectedly due to environment unpredictability. Exogenous events create the need for operator interaction, such as an emergent threat area or a meteorological condition, which require re-planning vehicle trajectories. These are examples of emergent situations that system designers could not account for a-priori, but are expected given the nature of the UV missions.

3.2.2 Arrival Processes

Associated with each event type in the system is an arrival process. Two types of arrival processes are critical in modeling supervisory control of multiple heterogeneous UVs: independent and dependent. Figure 3.3 shows a MUV-DES model with a single human server (representing the human operator), one independent arrival process and two variations of dependent arrival processes. Models of actual systems can contain any number of event types each represented by an appropriate arrival process.

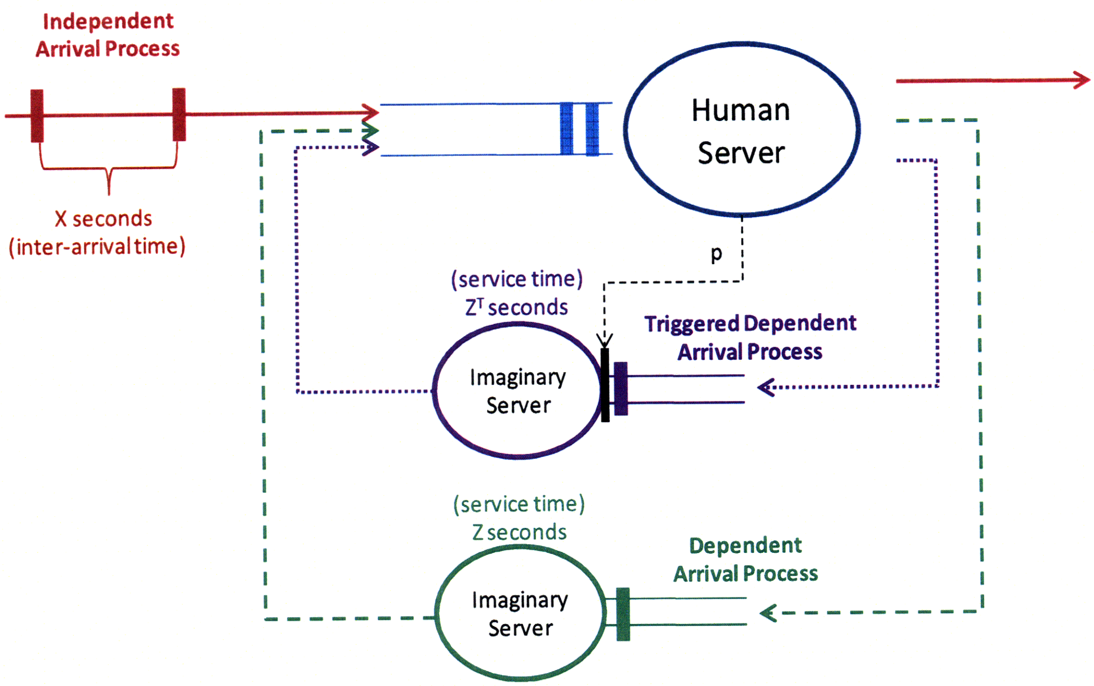


Figure 3.3 Three different arrival processes arriving to the MUV-DES server

3.2.2.1 Independent Arrivals

Exogenous events, which stem from sources external to the mission (weather, target movements, etc.), are typically generated in a manner independent from operator handling of such events. For example, an emergent threat area, which can create a conflict in need of operator intervention, is generated independently of whether or not the previously occurring emergent threat area conflict was handled by the operator. Note that the dependency or lack thereof is with respect to when previously-generated events are serviced and not when they were generated. Therefore, independent arrival processes as defined here can still have arrivals that are dependent on when previous events arrived to the system. Since the arrival of an event is independent from the service of previously-generated events, a basic queuing system arrival process (Bose, 2002) is appropriate for modeling independent arrival processes.

For such an arrival process, the defining attribute is a random variable, X, representing the inter-arrival time between events (the sub-system with solid lines in Figure 3.3). A probability density function, $f_X(x)$, can be used to define the distribution of inter-arrival times. In a UV system with m exogenous event types, the arrival processes can be modeled such that for each exogenous event type i, where $1 \leq i \leq m$, the probability density function $f_{X_i}(x)$ of the continuous random variable X_i is defined, where X_i represents the inter-arrival time between events of type i.

The resultant arrival rate for independent event type i is:

$$\lambda_i^{independent} = \frac{1}{E[X_i]}, \qquad where\ E[X_i] \neq 0 \qquad (1)$$

3.2.2.2 Dependent Arrivals

Since endogenous events associated with a specific vehicle are generally preconditioned on some other event being attended to first, the arrivals are generally dependent on the servicing of previously generated events. The dependency can be on (a) the last generated event of the same type being serviced first (in the case where only one event can exist at any point in time), or (b) an event of some other type being serviced first (in the case where the occurrence of an event is triggered by some other event being serviced first). An example of (a) is the need for the operator to analyze captured images. Once such an event is generated, the operator must attend to the event first before a second "analyze image" event can be generated by the same vehicle. In this case, the arrival of a specific event type is correlated with the time at which the last arriving event was serviced. An example of (b) is the arrival process of events that require a UV to be re-assigned to a new goal. Assuming in such a case that goal re-assignment occurs following the completion of target acquisition at the current goal, the arrival process for the re-assignment event is dependent on the target acquisition event being serviced first. In this case, an event representing goal re-assignment is generated once an event representing target acquisition has been serviced.

In order to model dependent arrival processes, the model presented in this chapter uses a closed queuing network paradigm to model arrivals for the event type in question (Lazowska, Zahorjan et al., 1984). In a closed queuing network paradigm, the population of events is finite. At any point in time, a single event is either waiting in one of the servers' queues or is being serviced. Once an event is serviced by one of the servers, it proceeds to the next server in the network.

To model (a), the case where arrivals are dependent on events of the same type being serviced first, a queuing network with a population of one event and two servers is used (the sub-system with dashed lines in Figure 3.3). One server represents the actual human server in the MUV-DES model, while the other server is an imaginary server whose role is to capture the process that takes place between the end of an endogenous event's service and the next time the endogenous event occurs. For example, once the dependent "analyze image" event is serviced, the vehicle will need to travel to the next goal before another "analyze image" event will be generated.

Modeling the arrival process using a queuing network with two servers will result in a single event instance in the sub-system which arrives in the operator's queue once operator interaction is needed. While the only event in the sub-system is waiting or undergoing service, no other event can be generated (since the sub-system's event population is one). Once the event is serviced, the vehicle will no longer need attention for the event in question, and this is captured through the event proceeding to the imaginary server. The service time for the imaginary server is a random variable, Z, which represents the time between the completion of a service for this dependent event and the next event of the same type arriving to the operator's queue. Since Z's represent the time a vehicle can operate without human intervention for a specific event type, they effectively represent degrees of autonomy. This is similar to the concept of neglect time except that in this case, the neglect time is for a specific event and not for the whole vehicle (i.e. other events associated with the same vehicle can still be generated while a specific event type is being neglected).

To model (b), the case where arrivals are dependent on an event of a different type being serviced first (the triggering event), the concept of blocking is used in addition to the previous concepts. Blocking is used in queuing networks to model the temporary stop of flow of events through a queue, nominally due to another queue having reached its capacity limitation (Balsamo, Persone, & Onvural, 2001; Onvural, 1990; Perros, 1984). In the MUV-DES model, the concept of blocking takes place prior to a triggering event being serviced in order to stop the flow of the dependent events through the queue of the imaginary server (the sub-system with dotted lines in Figure 3.3). In order to model the triggered-dependent process, the triggering event type must first be defined. Having identified the triggering event type, the arrival process is modeled such that servicing of a triggering event causes the removal of the block and therefore allows the dependent event to arrive to the operator's queue some time later. Since the triggering of a dependent event arrival is stochastic, either occurring or not, a Bernoulli distribution with success probability p can be associated with the triggering process. In this second type of dependent arrivals, the variable of concern is the time between the completion of a service for the triggering event and the arrival of the dependent event to the operator's queue. This is captured through a random variable Z^T, the service time for the imaginary server.

In both of the above dependent cases, the arrival processes can be described by the probability density functions $f_Z(z)$ and $f_{Z^T}(z^T)$, depending on the type of dependent arrival process. In a UV system with n vehicles, the dependent arrival processes can be modeled such that for each vehicle i, where $1 \leq i \leq n$, and for each endogenous event j, where $1 \leq j \leq g(i)$, where $g(i)$ is the total number of dependent event types associated with vehicle i, the probability density function $f_{Z_{ij}}(z)$ is defined, where Z_{ij} is the time between an event of type j being serviced and the arrival of the next event of type j. When the dependent event is of the triggered type, the probability density function that needs to be defined is $f_{Z^T{}_{ij}}(z^T)$ where ${Z^T}_{ij}$ is the time between the servicing of the triggering event type, k, and the arrival of the event of type j to the operator's queue.

When the time between the service of a dependent event and the arrival of the next event for a specific event type depends on the state of the system, a family of random variables $\{Z(s), s \in \sum\}$ indexed by the parameter s where s is the state of the system varying over a set $\sum$ representing all possible states, can be used to describe the event type. A probability density function is then defined for each random variable in the set. The method by which the vehicle state can change is described in Section 3.2.3.2, which discusses the service impact concept.

Modeling certain systems could require other variations of arrival processes, such as a process where events arrive independently from each other, but where there is a limit to the maximum number of events that could be in the system at any point in time. Techniques such as balking and reneging can be used to capture variations in arrival processes (Rao, 1967). In the case of balking, the arriving event refuses to enter the queue for some predefined reason, while in the case of reneging, the event leaves the queue after having entered it. Due to the fact that balking and reneging require a decision on the part of the event, they were not considered in the MUV-DES model.

3.2.2.3 Modified Arrival Processes

Having defined the different event arrival processes required to capture the event categories previously identified, this section proposes a method by which lack of operator situational awareness (SA) can be captured by the MUV-DES model. This is done by modifying the previously defined arrival processes to account for any wait times incurred due to loss of operator SA.

As discussed in Chapter 2, SA can significantly influence human behavior and hence, human-UV system performance. One negative side effect of low SA is the creation of additional vehicle wait times due to loss of situational awareness (WTSA). A lack of SA on the part of the operator results in increased time for the operator to notice the needs of the system, which then results in increased wait times for the system (Cummings, Nehme et al., 2007). Cummings & Mitchell (2008) showed that WTSA can account for the largest part of vehicle wait time, and significantly reduces the overall number of vehicles that a single operator could control. It was therefore desired to capture the effects of WTSA in this DES model.

In order to capture WTSA, this model presents a hypothetical relationship between operator utilization and WTSA that is based on the following two relationships. The first relationship is one based on the classic Yerkes-Dodson inverted U-shaped function. The original work by Yerkes and Dodson (1908) related stimulus strength and rapidity of habit-formation (or learning). However, Hebb (1955) proposed a similar relationship for the effects of arousal on human performance. This relationship states that all else equal, both low and high operator arousal are associated with lower performance, with the former being due to human complacency, while the latter being a result of the human's inability to handle excess workload. It is this relationship that the model builds on, such that the arousal level is represented by increasing operator utilization. Operator utilization, ρ, is defined as the percent busy time, or the ratio of the total time the operator is engaged in tasks, i.e. servicing events, to the total time. The second relationship is between SA and performance. Previous work has shown that operator SA can be measured as a function of mission performance (Pritchett & Hansman, 2000). Assuming the following abbreviations, o for performance, and s for SA, the following can be used to describe the above two relationships:

$$\rho \rightarrow o \quad \text{and} \quad o \rightarrow s$$

the relationship proposed for relating operator utilization to SA is based on the law of transitivity:

$$\gamma: \rho \rightarrow s$$

This model therefore builds on an assumption that SA is related to operator utilization through a functional form, γ. The proposed relationship is a parabolic function between WTSA ($\gamma(\rho)$) and operator utilization (ρ) that is concave upwards (Figure 3.4), such that WTSA is high at both low and high operator utilization. When operators are under high levels of utilization, it is assumed that they can be too busy to accumulate the information that is required to build SA. At the same time, when operators are under-utilized, it is presumed that due to a low level of arousal and complacency, they could overlook information from the environment, which would also lead to low SA. Both of these lead to higher WTSA at the extremities of the utilization scale. The parabolic relationship is inspired by the Yerkes-Dodson relationship, which is itself parabolic.

The size of the penalties at different utilization levels is dependent on the exact shape of the γ curve, which is defined by four different parameters. The first of these parameters, S_M, is the minimum penalty due to loss of situational awareness. The second parameter is the point about which the curve is centered (C_P) which represents the utilization level where the SA penalty is at a minimum. The third parameter, I_P, is the width of the interval around C_P which also has minimum SA penalties (if I_P is zero, then the penalty is at a minimum only at C_P utilization and the γ curve is a parabolic curve). Finally, the fourth parameter, S_F, is a scaling parameter that affects the magnitude of the penalties that can be incurred. In the case of a symmetric curve, γ takes on the value S_F at the 0% and 100% utilization levels. Empirical evidence supporting the shape of the curve is presented in Section 4.8.2.

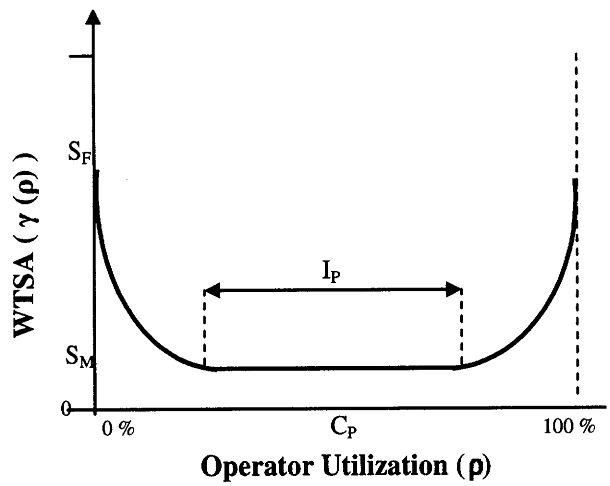


Figure 3.4 γ curve relating WTSA to ρ

Using the proposed functional form between WTSA and operator utilization, the arrival process of dependent events is modified to account for loss of operator situation awareness. The modified arrival process is described by a probabilistic distribution over a random variable (Z' or $Z^{T'}$) that is a function of two main components: a) the random variable associated with the dependent arrival process and, b) a penalty due to loss of operator situational awareness. This is shown for both types of dependent arrival processes in Equations 2 and 3.

$$z' = z + \gamma(\rho) \qquad (2)$$

$$z^{T'} = z^T + \gamma(\rho) \qquad (3)$$

The first term in Equations 2, z, is a realization of the random variable Z defined previously, that describes the time between one service and the next arrival. Similarly, the first term in Equations 3, z^T, is a realization of the random variable Z^T defined previously, that describes the time between the service of the triggering event and the arrival of the dependent event in question. Because the generation of a task does not necessarily imply that the operator notices the generated task, the second term in Equations 2 and 3, $\gamma(\rho)$, represents a penalty due to operator loss of SA. $\gamma(\rho)$ takes on some minimum value when the operator has complete SA and higher values when the operator has degraded SA. As was defined earlier, ρ is the ratio of the total time the operator is engaged in tasks, i.e. servicing events, to the total time elapsed. Because it is required to calculate the penalty γ and therefore ρ for each dependent arrival throughout the mission, the time period over which operator utilization is calculated can span at most the total elapsed time. The exact size of the time window used for calculating utilization, as well as the exact functional form γ that relates WTSA to ρ are design parameters which are a function of interface properties and operator training. In addition, to defining probability density functions associated with independent and dependent arrival processes, the function $\gamma(\rho)$ should be defined over all ρ, where $0 \leq \rho \leq 100$.

3.2.3 Service Processes

3.2.3.1 Service Times

In addition to an arrival process being associated with each event type (endogenous and exogenous), a service process is associated as well. The length of time it takes the operator to deal with an event, interaction time (IT) or Y', is a random variable which can be modeled using a service time distribution. The random variable is a function of two main components; a) the random variable describing the times for which operators need to service tasks, and b) the random variable describing the wait times due to interaction (WTI) (Equation 4).

$$y' = y + wti \qquad (4)$$

The first term in Equation 4, y, is the realization of the random variable Y that describes the length of time for which the operator must interact with the arriving event. The probability density function describing Y, $f_Y(y)$, captures the variability of performance between different operators, as well as the variability in the performance of a single operator. The second term in Equation 4, *wti*, is the realization of the random variable *WTI* that represents the wait times due to interaction that arise when servicing the arriving event. Wait times due to interaction can be due to any decision making, context switching times incurred prior to servicing the arriving event, or the operator having to wait on the system for more information. *WTI*s can vary per event type, due to the dependency of decision making times and switching times on the particular event type. The effect of these *WTI*s is to increase Y'.

The service processes for endogenous event types can be modeled such that for each vehicle i, where $1 \leq i \leq n$, and for each event j, where $1 \leq j \leq g(i)$, where $g(i)$ is the total number of dependent event types associated with vehicle i, the probability density functions $f_{Y_{ij}}(y)$ and $f_{WTI_{ij}}(wti)$ are defined, where Y_{ij} and WTI_{ij} are random variables representing the service time and the wait time due to interaction associated with endogenous event type j. The service processes for exogenous event types can be defined similarly such that for each exogenous event type i, where $1 \leq i \leq m$, the probability density functions $f_{Y_i}(y)$ and $f_{WTI_i}(wti)$ are defined, where Y_i and WTI_i are random variables representing the service time and the wait time due to interaction associated with exogenous event type i.

3.2.3.2 Service Impact

In addition to a service process being defined by service times, it is also important to consider the impact of the serviced event. An important feature that this model relies on allows the impact of an event service to be taken into account such that the system state is updated. The concept of service impact can be used for several purposes, the most important of which is interaction between different event types. Event interaction can take place by having one event type change a system state variable, while having the arrival process of a second event type depend on that same state variable. Event interaction is important, especially for endogenous event types that could have dependent arrival processes where the associated random variable is a function of the system state (Section 3.2.2.2). Two types of event interaction have been identified as important in modeling supervisory control of multiple unmanned vehicles.

First, an event can modify the arrival process of a second event type permanently. For example, servicing an event that represents a vehicle path intersecting a threat area in an untimely manner can lead to the vehicle involved to be damaged. The effect of this could be to slow down the vehicle which would modify the arrival process of subsequent events. Second, an event can modify the arrival process of a second type temporarily; i.e. the arrival process of just the next arriving event. This could occur, for example, when an operator decides to re-plan the vehicle destination in order to reach a target in shorter time. The effect of this would be to speed up the arrival of just the next scheduled event associated with the vehicle's arrival to targets. Other variations of interaction between event types can be modeled in this manner.

3.2.4 Queuing Policy

Finally, the last pillar of a queuing-based DES model is the queuing policy. The queuing policy determines the order by which multiple events that are waiting in the queue are serviced. The MUV-DES model uses variation of queuing policies to capture differences in operator switching strategies.

Examples of switching strategies that can be modeled in MUV-DES include the first-in-first-out (FIFO) queuing scheme as well as the highest attribute first (HAF) strategy (Pinedo, 2002). The HAF strategy is similar to a preemptive priority scheme in that high priority events are serviced first except that there is no pre-emption. Therefore, if an event is generated with a priority higher than any of the events already in the system, it will be moved to the front of the queue, but will not preempt a lower priority vehicle that is already being serviced. Hybrid queuing policies observed experimentally can also be modeled. The queuing policy will then dictate the method by which the queue is handled.

3.2.5 Summary

In summary, event types in the system model need to be identified and then mapped to one of the appropriate event categories. Depending on the arrival process associated with the event category, different parameters need to be estimated. Table 3.1 summarizes these parameters. The total number of streams needed to represent a heterogeneous UV system with n vehicles and e exogenous event types is given by:

$$Total\ \#\ of\ streams\ needed = \sum_{i=1}^{n} g(i) + e \qquad (5)$$

where $g(i)$ is the number of dependent event types associated with vehicle i. The DES pseudo code can be found in Appendix A.

Table 3.1 Relevant parameters for arrival and service processes

Process	Parameter of concern	Parameter notation	Distribution
Independent Arrival	Inter-arrival time between events	X	$f_X(x)$
Dependent Arrival	Time between service of last generated event and arrival of next event or Time between service of triggering event and arrival of event	Z or Z^T	$f_Z(z)$ or $f_{Z^T}(z^T)$
Service	Time required for servicing event	Y	$f_Y(y)$

3.3 Mapping Design Variables to Model Constructs

The next sub-sections will describe how the attributes that define the MUV-DES model (Figure 3.1), through their associated design variables from Chapter 2 (Table 2.1), map to the different model constructs defined in Section 3.2. The mappings are summarized in Table 3.2 and the design variables are shown as inputs to the MUV-DES model in Figure 3.5.

3.3.1 Vehicle Team Structure

The team structure attribute is represented in the model through the team size and team heterogeneity variables. The team size variable is captured in the number of vehicles modeled through endogenous event types. By modeling vehicles with different event types, each with specific arrival and service processes, the MUV-DES model captures the level of heterogeneity variable. A heterogeneous UV team is one where in a team of size n, there exists at least one vehicle, a, and an event type, j, where $f_{Z_{aj}}(z) \neq f_{Z_{bj}}(z)$ or $f_{Y_{aj}}(y) \neq f_{Y_{bj}}(y)$ where $1 < a$, $b < n$ and $a \neq b$. Depending on the underlying research questions, heterogeneity can be classified as existing strictly in service processes, in arrival processes, in event types associated with the vehicles, or any combination of these.

Table 3.2 Mapping design variables to MUV-DES model constructs

Attribute	Variable	MUV-DES Model Construct
Vehicle Team Structure	Vehicle team size	Number of vehicles for which event types are modeled, n
	Vehicle team heterogeneity	Where in a team of size n, there exists at least one vehicle, a, and an event type, j, where $f_{Z_{aj}}(z) \neq f_{Z_{bj}}(z)$ or $f_{Y_{aj}}(y) \neq f_{Y_{bj}}(y)$ where 1 < a, b < n and a $\neq$ b
Role Allocation and Level of Autonomy	Level of automation	Endogenous event types
	Neglect times	Probability distribution for a specific event type, $f_Z(z)$ or $f_{Z^T}(z^T)$
Environment Unpredictability	Uncontrollable events	Event types and associated probability distribution functions, $f_X(x)$
Vehicle task allocation	Coordination	Dependency between event arrivals through event interaction
Nature of operator interaction	Interaction times	Service distributions, $f_Y(y)$
	Attention Allocation (Switching strategy)	Queuing policy
	Attention Allocation (Level of management strategy)	Captured by varying the probability density functions, $f_Z(z)$ or $f_{Z^T}(z^T)$, or the triggering Bernoulli probability, p, associated with operator-induced events
	Situational awareness	WTSA/UT curve, $\gamma(\rho)$, defined over all ρ, where $0 \leq \rho \leq 100$

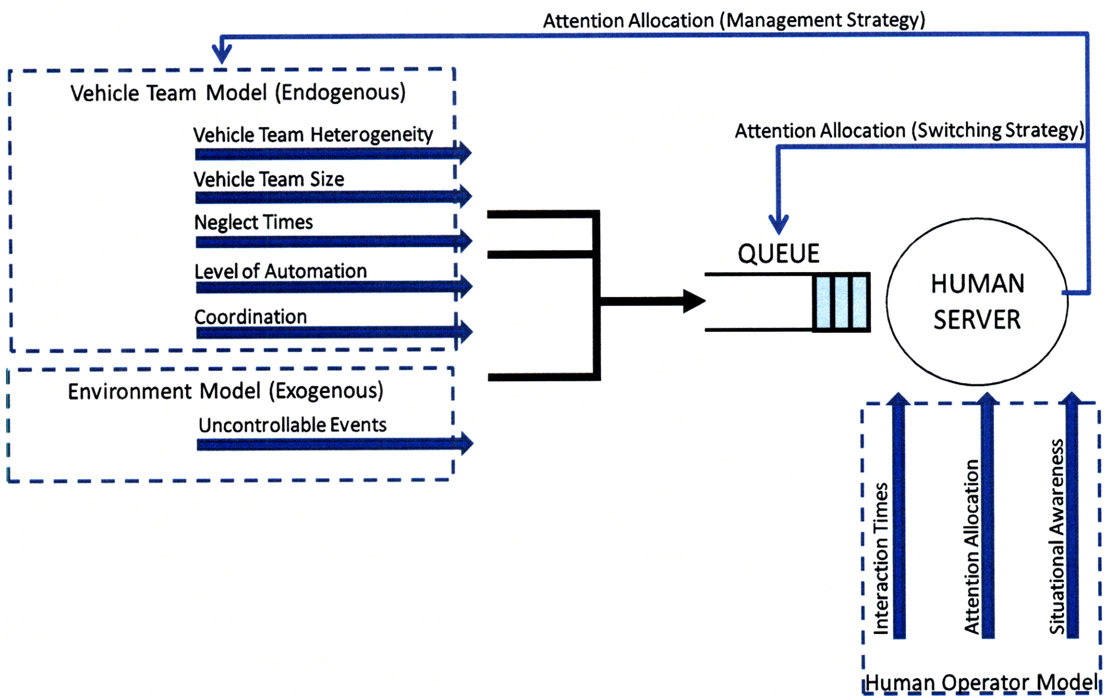


Figure 3.5 A high level representation of the input variables to the MUV-DES model

3.3.2 Role Allocation and Level of Autonomy

The role allocation and level of autonomy attribute is represented in the model through the level of automation and neglect times variables. The level of automation variable is captured through the types of endogenous events represented in the system. Since any event in the model requires human input when it arrives to the human server, defining the events that will be captured in the model is essentially capturing the role of the human operator and hence the role of the vehicles. Neglect times are captured in this model through the probability density functions, $f_Z(z)$ or $f_{Z^T}(z^T)$, associated with each endogenous event type. The type, mean, and variance of the density functions can be varied in order to capture alternate neglect time profiles.

3.3.3 Vehicle Task Allocation

The vehicle task allocation attribute is represented in the model through the level of coordination variable. Because the random variables representing the time between event servicing and the next arrival of dependent events can depend on the vehicle state, different levels of event interaction can be captured by varying the amount and type of dependency. The ability of the model to capture event interaction allows the model to capture coordination between different event types. Although the level of coordination between different vehicles has not been completely addressed in the design of the MUV-DES model, future work is likely to capitalize on the ability of the model to capture event interaction for that purpose.

3.3.4 Nature of Operator Interaction

The nature of operator interaction attribute is represented in the model through interaction times, attention allocation strategies, and situational awareness variables. Interaction time variables are captured in this model through the probability density functions, $f_Y(y)$, associated with event servicing. The model captures two attention allocation strategies: switching strategy and level of management strategy. In order to model the switching strategy of the operator, the type of queue can be varied to represent different switching strategies. For example, as was explained in Section 3.2.4, priority-based switching strategies can be captured through the queuing policy.

The operator level of management strategy, which is the rate at which operator-induced events arrive to the system, is a function of the amount of operator re-planning. The management strategy represents a more global scheme by which the operator distributes his/her attention across the different vehicle events. Since this model supports endogenous events that are both vehicle-generated and operator-induced, the rate at which operator-induced events arrive to the system depends on the operator's desire to interact with the vehicles beyond unavoidable vehicle-generated events. This can be captured by varying the probability density functions, $f_Z(z)$ or $f_{Z^T}(z^T)$, or the triggering Bernoulli probability, p, associated with operator-induced events.

Finally, the situational awareness variable is captured in the model through the $\gamma(\rho)$ function which defines wait times due to loss of SA. While the service and modified arrival process models account for variation in human performance through their stochastic representation of operator performance, it can be assumed that other variables are likely to influence individual operator performance such as training and fatigue. Since the purpose of this modeling attempt was to determine the impact of system variables (i.e., number and type of vehicles), the investigation of the impact of individual operator attributes is left as the subject of future work.

3.3.5 Environment Unpredictability

Finally, the environmental unpredictability attribute is represented in the MUV-DES model through a representation of uncontrollable events that can affect the system. This can be captured by the MUV-DES model through an identification of exogenous event types and by associating each event type with an input stream and a probability density function, $f_X(x)$, describing the arrival process. The type, mean, and variance of the density functions can be varied in order to capture the unpredictability of the environment and can be used to test the robustness of the measured metrics to unexpected events, thereby defining performance boundaries as opposed to exact performance predictions.

3.4 Model Outputs

In addition to supporting the designer of UV systems in varying design variables, the MUV-DES model allows for the measurement of several output variables. These variables include those that arise naturally in DES-based models, as well as specific user-defined metrics that the MUV-DES model allows a designer to capture.

3.4.1 DES-based Metrics

Because of their event-based nature, DES models and their embedded queuing models lend themselves to temporal-based metrics like server utilization and event wait times. Such metrics are critical in the system considered where the operator is a limited resource, with limited attention resources that must be divided across multiple vehicles. Queuing wait times can be used to capture the lengths of time for which vehicle events will have to wait before receiving attention. This is important, as operating vehicles is expensive and wait times associated with vehicles having to wait for operator attention represent inefficiencies that need to be minimized. Operator utilization can be used to capture the effects of alternate design variable settings on operator workload. Due to the fact that operator performance is a function of their utilization, being able to design systems while understanding the impact on operator utilization is critical. Generalizable metrics such as average vehicle wait times and operator utilization can allow for benchmarking and comparisons to be made across applications (Pina, Cummings et al., 2008).

3.4.2 Additional Mission-specific Metrics

In addition to the basic metrics inherently captured by queuing/DES models, the ability of the MUV-DES model to capture the impact of servicing events supports the measurement of other mission specific metrics. Designers could study mission-specific metrics such as the number of a specific task completed, or the rate at which tasks are completed. In addition, more complicated service impacts can take into account the amount of time that an event had to wait before receiving attention, and performance metrics can be affected accordingly. For example, punctually servicing an event that entails a visual task might result in a performance metric being positively incremented. Instead, servicing the visual task with a slight delay might result in smaller reward in terms of performance. Mission specific metrics are important as they can reduce excessive experimental costs by allowing the MUV-DES model to be used with similar output fidelity (Pina, Cummings et al., 2008).

3.5 Model Benefits

An important goal of designing this model was to fill the research gaps identified in Chapter 2. The following sections discuss the benefits of the model by addressing how each of those research gaps is addressed.

3.5.1 Holistic

The MUV-DES model takes as inputs the different human-UV interaction attributes identified in Chapter 2. In doing so, the model is able to capture the holistic effect of the different attributes. The model estimates, through simulation of mission execution, the impact of the design variables on mission performance (including operator and vehicle specific metrics). Therefore, the MUV-DES model allows the designer to understand the impact of changing one or more design variables.

3.5.2 Capturing the Impact of Heterogeneity

Depending on the underlying system being modeled, heterogeneity can exist strictly in the types of vehicle-generated tasks that require operator attention, the rates at which the tasks are generated, the type of operator attention required for attending to the different tasks, or any combination of these. The MUV-DES model can capture these different forms of heterogeneity through different event types, the probability density functions associated with event arrival processes, and/or the probability density functions associated with event service processes. In addition to supporting different demands from heterogeneous vehicles/tasks, the model supports designers in understanding the impact of heterogeneity in terms of the different attention allocation strategies that become feasible. Designers can vary these strategies as was described in Section 3.3.4 to understand the impact of heterogeneity on operator, vehicle and mission performance.

3.5.3 Detailed Operator Model

In addition to capturing the serial operator nature in complex problem resolution, the model has embedded a construct that captures the impact of loss of situation awareness through WTSAs. While a critical aspect of human performance, SA is a challenge to measure, much less model. The methodology used in this MUV-DES model presents a way to incorporate performance-based SA measures in discrete event simulation models in order to make generally effective system observations and predictions.

3.5.4 Cost Trade Space

As was defined in Chapter 2, it is important for the model to capture the fact that effectiveness is often dynamic, depending on mission specifications. The MUV-DES model supports a cost-trade space analysis because it allows the designer to measure multiple operator, vehicle, and mission effectiveness outputs (Section 3.4) while varying one or more design variables (Section 3.3). This can be used to rapidly evaluate the impact of potential design modifications, understand the potential capabilities and limitations of future technology, and replicate current observed behavior in order to help diagnose the cause of failures and inefficiencies. These abilities are discussed in Chapter 5.

However, in order for the model to be reliable for realizing the above benefits, it is required to first build confidence in the model's accuracy and robustness. This is accomplished in Chapter 4.

4 MODEL VALIDATION

In this chapter, the results of a multi-stage validation of the MUV-DES model introduced in the previous chapter are presented. Validation of simulation-based models is important, as derivations from inaccurate models can lead to flawed design recommendations. Several techniques have been proposed for verifying and validating simulation-based models (Balci, 1997; Kleijnen, 1995; Law & Kelton, 2000; Martis, 2006; Sargent, 2005). The literature suggests that model validation should be undertaken with respect to the purpose for which the model was built. As it is often costly and time consuming to determine that a model is absolutely valid, tests and evaluations should be conducted until sufficient confidence is obtained in the model (Sargent, 2005).

Data is a critical element for model validation as it is needed for both building the conceptual model, and for conducting operational validation of the model (Sargent, 2005). However, obtaining appropriate, accurate and sufficient data is costly as well as time consuming, and is therefore often the reason why model building fails (Sargent, 2005). In order to address this possible obstacle, an online experimental test bed was developed in order to support collection of large quantities of data for the MUV-DES model validation in a reasonable length of time.

Following the presentation of the experimental test bed, a discussion of the model validation results is presented. The validation process used in this chapter is a multi-stage validation approach, which has been shown to be comprehensive in its ability to validate multiple facets of a model with respect to the purpose for which the model was built (Sargent, 2005). Overall, the validation process entailed a testing of the model assumptions and a comparison between the input-output relationships of the model and those of an experimental data set, resulting in: a) confidence being achieved in the model's accuracy and b) determination of the model's robustness under different input settings.

The following confidence-building validation techniques were applied and will be discussed in detail in this chapter; a) replication validation, b) predictive validation, c) extreme condition testing, d) internal model stochasticity, e) sensitivity analysis, f) historical data validation, and g) construct validation.

4.1 Data Collection

4.1.1 Experimental Apparatus

In this section, an online experimental test bed is introduced. This test bed allows for virtual heterogeneous unmanned vehicle team supervision and was used to generate data in order to build confidence in the model.

4.1.1.1 Overview

The Research Environment for Supervisory Control of Heterogeneous Unmanned Vehicles (RESCHU) is an online experimental test bed that allows operators to control a team of UVs, composed of unmanned air and underwater vehicles (UAVs and UUVs). All vehicles are engaged in surveillance tasks, with the ultimate mission of locating specific objects of interest in urban coastal and inland settings. While this test bed supports a single UUV type, it contains two UAV types, one that provides high level sensor coverage (akin to a Global Hawk UAV), while the other provides more low-level target surveillance and video gathering (similar to a Predator UAV). The experimental test bed is designed to allow a single operator to supervise one or more of the vehicle types. Because previous research has shown that the simultaneous supervision and payload management (e.g., managing cameras for target identification) of multiple unmanned vehicles, requires automation of navigation tasks for the different vehicles (Cummings, Bruni et al., 2007), this was a basic assumption in the design of the test bed.

4.1.1.2 Interface

The RESCHU interface consists of five major sections (Figure 4.1). The map, shown in Figure 4.2, displays the locations of vehicles (UUVs, medium altitude long endurance (MALE) UAVs, and high altitude long endurance (HALE) UAVs), threat areas, and areas of interests

(AOIs). Vehicle control is carried out on the map, such as changing vehicle paths, adding a waypoint (a destination along the path), or assigning an AOI to a vehicle through a point-and-click interface.

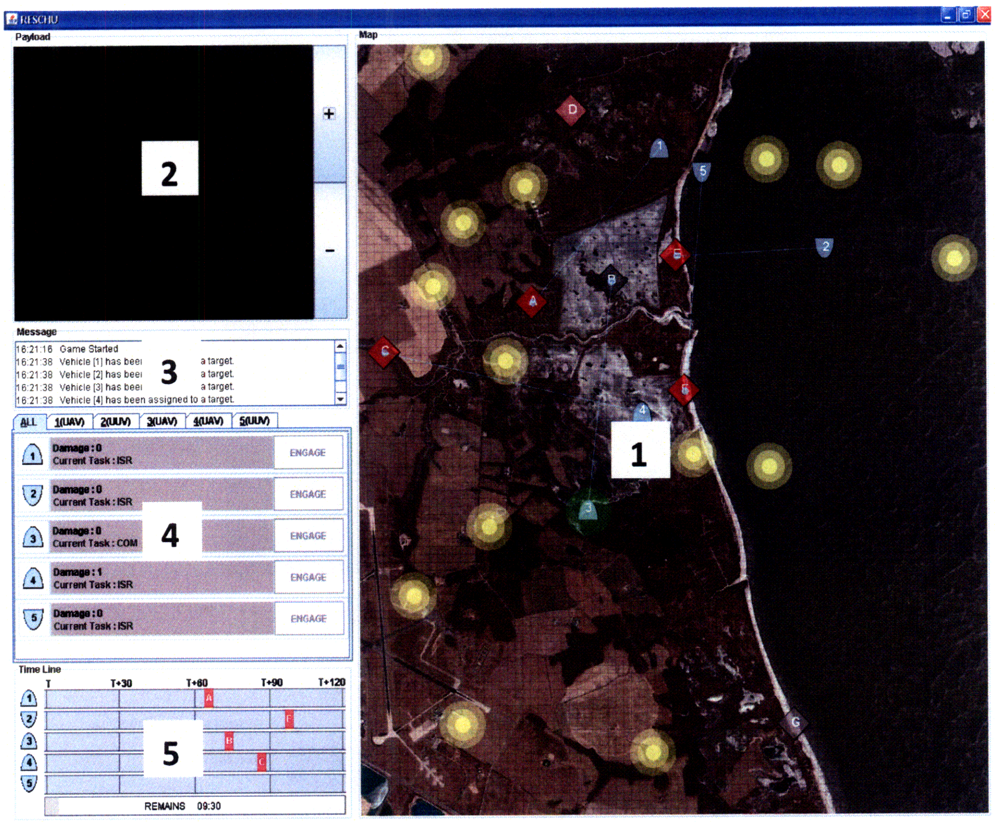


Figure 4.1 RESCHU interface (1: map, 2: camera window, 3: message box, 4: control panel, 5: timeline)

The main events in the mission (i.e., vehicles reaching goals, or automatically being assigned to new targets) are displayed in the message box, along with a time stamp (Figure 4.3a). When the vehicles reach an AOI, a simulated video feed is displayed in the camera window (Figure 4.3a). The operator then has to visually identify a target in this simulated video feed. Example targets and objects of interest include cars, swimming pools, helipads, etc. The control panel provides vehicle health information, as well as information on the vehicle's mission (Figure 4.3b). The timeline displays the estimated time of arrival to waypoints and AOIs (Figure 4.3b). Beneath the timeline is a mission progress bar that shows the amount of time remaining in the total simulation (Figure 4.3b).

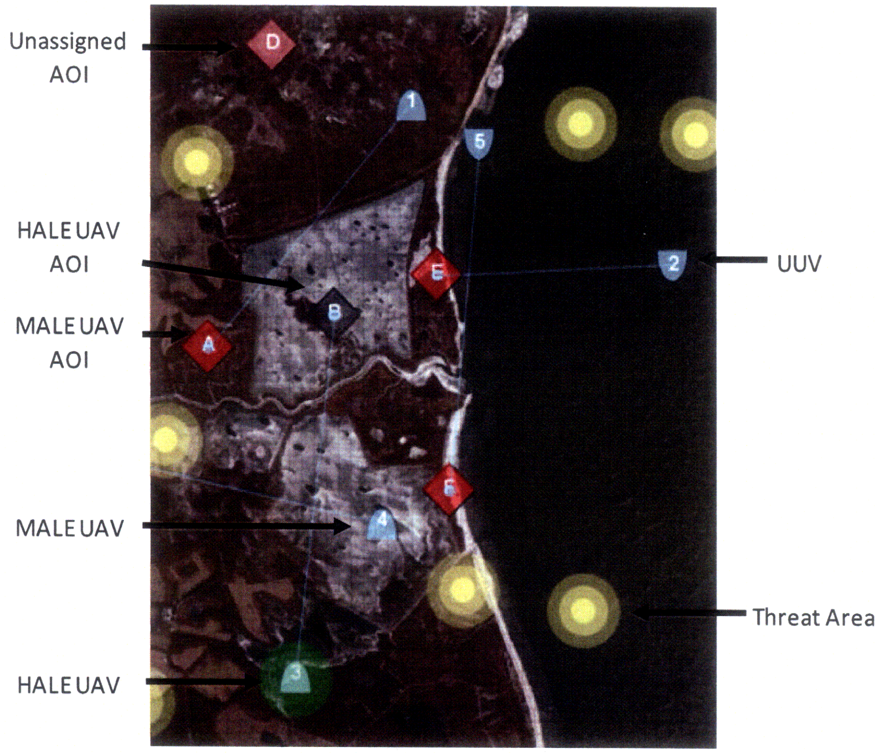
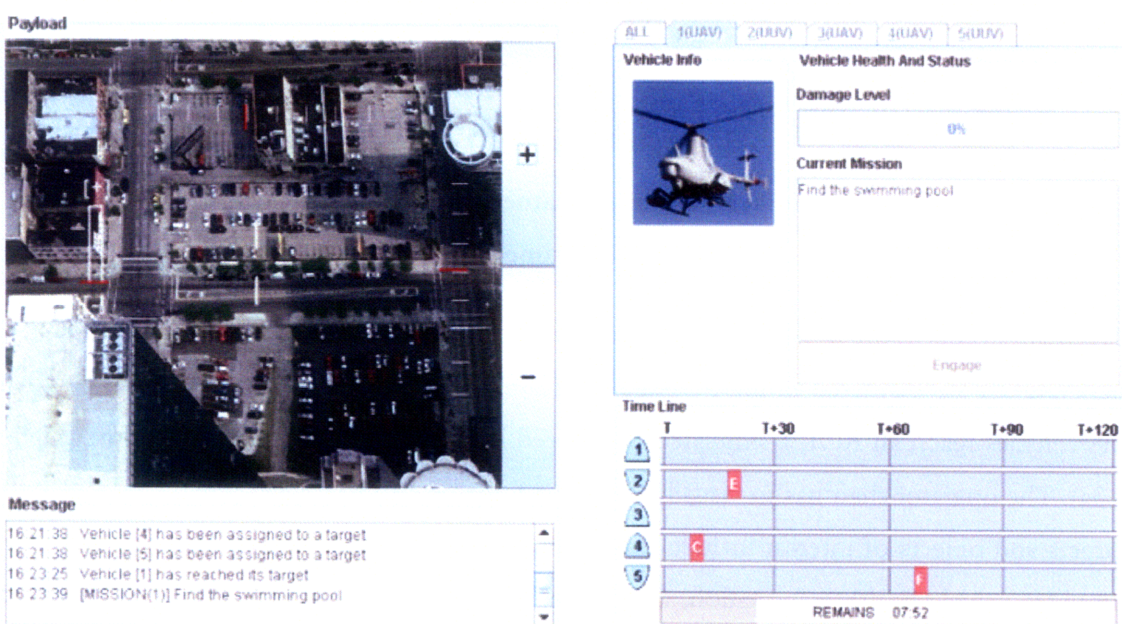


Figure 4.2 The map section of RESCHU



(a) (b)

Figure 4.3 (a) Activated camera view and message box (b) Control panel and timeline

4.1.1.3 Vehicle Types and Functions

As discussed previously, a team of vehicles can be composed of one or more of three types of vehicles: a HALE UAV, a MALE UAV, and a UUV. Both the MALE UAVs and the UUVs travel to areas of interest (red diamond shaped symbols in Figure 4.2) with a pre-determined target that needs to be visually acquired by the operator. UUVs are slower than UAVs. A HALE UAV travels to AOIs that do not yet have a target specified (grey diamond shaped symbols in Figure 4.2), and carries a sensor that supports gathering additional intelligence on AOIs. Engaging a HALE that arrives at a grey AOI reveals the object-finding mission associated with that AOI (i.e., the color of the AOI changes to red). These newly discovered targets should then be acquired by a MALE UAV or a UUV.

4.1.1.4 Operator Tasks

When the vehicles complete their assigned tasking, an automated-path planner assigns the HALE UAV to an AOI that needs intelligence, and the MALE UAVs and UUVs to AOIs with pre-determined targets. The automated-path planner can result in sub-optimal AOI assignments, a property often present in actual systems. The sub-optimal assignments create an additional optional task for the operator of re-assigning a vehicle, for example, the operator may wish to assign a vehicle to a closer AOI. In addition, the operator is able to avoid threat areas by changing a vehicle's goal or adding a waypoint to the path of the vehicle in order for the vehicle to travel around the threat area.

When a vehicle arrives to an AOI, a visual flashing alert is issued to indicate that the operator can engage the payload (the camera), in order to complete the surveillance task. For a HALE UAV, clicking the engage button results in the object-finding mission associated with the grey AOI being revealed (i.e., the AOI then changes color to red). For a MALE UAV or a UUV, engaging the payload causes the camera window to display the simulated live video feed (Figure 4.3a). The operator then has to complete a search task by panning and zooming the camera until the specified object is located. Once the operator submits an object-identification through a right-click interaction (Figure 4.4), the message box notifies the operator on the accuracy of response (this is used to simulate the feedback that real operators get from their commanders or teammates as a consequence of their actions), and the vehicle is automatically re-assigned to a new AOI.

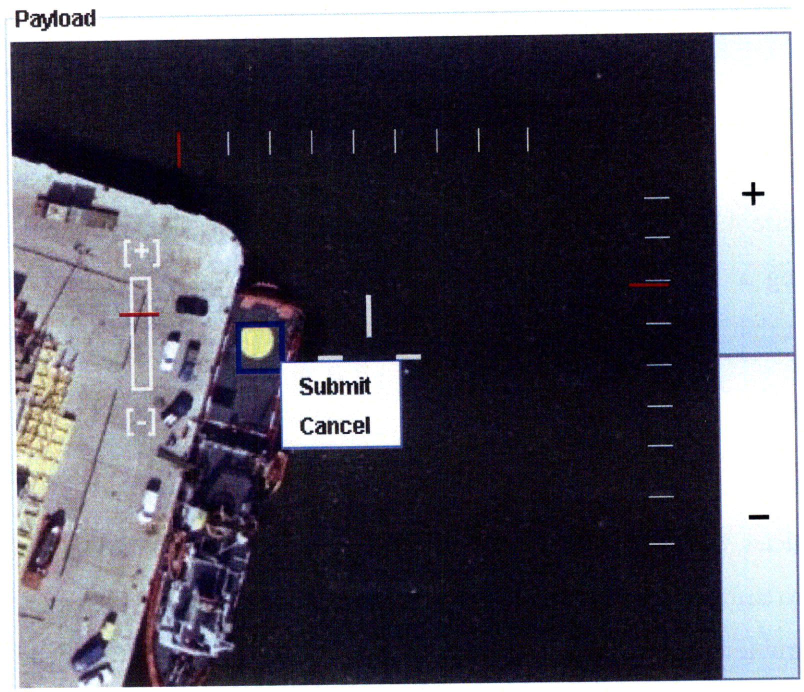


Figure 4.4 Submiting an identified object

4.1.2 Participants and Experimental Procedure

Several experiments were conducted using the RESCHU test bed in order to provide data for model validation. The first such experiment, presented in this chapter, was designed to generate a large data set suitable for model validation. A second experiment, described in Chapter 5, was conducted to validate the effectiveness of design recommendations made by the MUV-DES model.

The first experiment had vehicle team heterogeneity level (as defined in Section 3.3.1) as a between-subject condition: no (n=26), medium (n=25), and high (n=23), representing increasing levels of heterogeneity in terms of vehicle mix per team. There were several reasons for focusing on team heterogeneity as an independent variable. First, by changing team heterogeneity, the spreads in interaction times and neglect times are varied as well. In addition, varied vehicle team size and neglect times are considered in Section 4.7 as part of the historical validation process. Finally, although operator strategies and operator SA can be influenced through interface design, previous work (Chapter 2) has shown that they cannot be directly controlled. It was therefore decided to refrain from varying these two variables.

The objective of conducting the first experiment was multi-faceted. First, it was desired to have performance data associated with different levels of heterogeneity, so as to build confidence in the model's accuracy at replicating human-UV interaction under different conditions. Second, by having vehicle team heterogeneity as an independent variable, the model's ability to replicate statistically significant effects on operator, vehicle and/or mission performance could be evaluated. Finally, having a data set associated with the different levels of heterogeneity allowed for predictive validation by selecting a single data set associated with one of the conditions and predicting the results observed for a second condition.

The no-heterogeneity condition was composed of five MALE UAVs. The medium-heterogeneity condition had three MALE UAVs and two UUVs. Because the UUVs were slower than UUVs, they produced events less frequently. The high-heterogeneity condition required managing two MALE UAVs, two UUVs, and one HALE UAV. HALE UAVs were restricted to grey AOIs, which appeared at a ratio of two-to-five, as compared to red AOIs, which the UUVs and MALE UAVs could visit without assistance from the HALE. Thus, the arrival rates of events for HALE UAVs were different than for both the MALE UAVs and UUVs. Moreover, service times were different between the UV types; the HALE UAVs required just milliseconds of service time (operators clicking the engage button), whereas the MALE UAVs and UUVs required longer service times during which operators had to locate the object of interest. Because the UUVs were slower than UAVs and the HALE UAVs did not have an associated visual task, the no-heterogeneity condition composed of five MALE UAVs was the highest tempo scenario, followed by the medium and then the high-heterogeneity conditions. Thus, the increase in heterogeneity across the three scenarios was inversely related to the rate the operators should interact with the UVs.

The experiment was designed for web-based delivery, with an interactive tutorial and an open-ended practice session. The suitability of using web-based experimentation using this interface is discussed in Appendix B. Participants spent on average 10 minutes performing the practice session. The website was password protected and participation was via invitation. All data were recorded to an online database. Demographic information was collected via a questionnaire presented before the tutorial (Appendix C). Participants were instructed to maximize their overall performance score by 1) avoiding threat areas that dynamically changed, 2) completing as many of the search tasks correctly, 3) taking advantage of re-planning when possible to minimize vehicle travel times between AOIs, 4) ensuring a vehicle was always assigned to an AOI whenever possible.

After participants felt comfortable with the task and the interface, they could end the practice session and start the ten minute experimental session. After test session completion, subjects were required to fill a post-experimental questionnaire before terminating the experiment (Appendix D). After completing the experiment, the participants could see their score, which corresponded to the total number of targets correctly identified.

Seventy-four participants, six females and sixty-eight males, between the ages of 18-50 completed the first study (the mode age group was ages 18-25). Ten participants had previous experience with unmanned vehicles. The participant who scored the highest in the experiment received a $200 gift certificate.

4.1.3 Human-in-the-Loop Experimental Results

The variables of interest for evaluating initial model predictions were the mission score, average search task wait time, and operator utilization. Mission performance was assessed via the mission score, which was calculated as the proportion of the total number of targets correctly identified normalized by the total number possible for the scenario in question. Average search task wait times assessed system performance efficiency, since it demonstrated the effects of operator inefficiencies on the system delay. Operator utilization, or percent busy time, was calculated as the proportion of time the operator actively interacted with the display (e.g., adding a way point, engaging in a visual task, etc.) during the course of the experiment. Utilization therefore excluded any monitoring time expended by operators.

The results, which are summarized in this section, are presented in more detail in Appendixes E and F. A preliminary analysis using a Pearson correlation test demonstrated significant correlations between the three variables of interest, thus a Multivariate Analysis of Variance (MANOVA) was performed to control for the inflation of Type I error. Significant findings were followed with univariate analyses (box plots are shown in Figure 4.5).

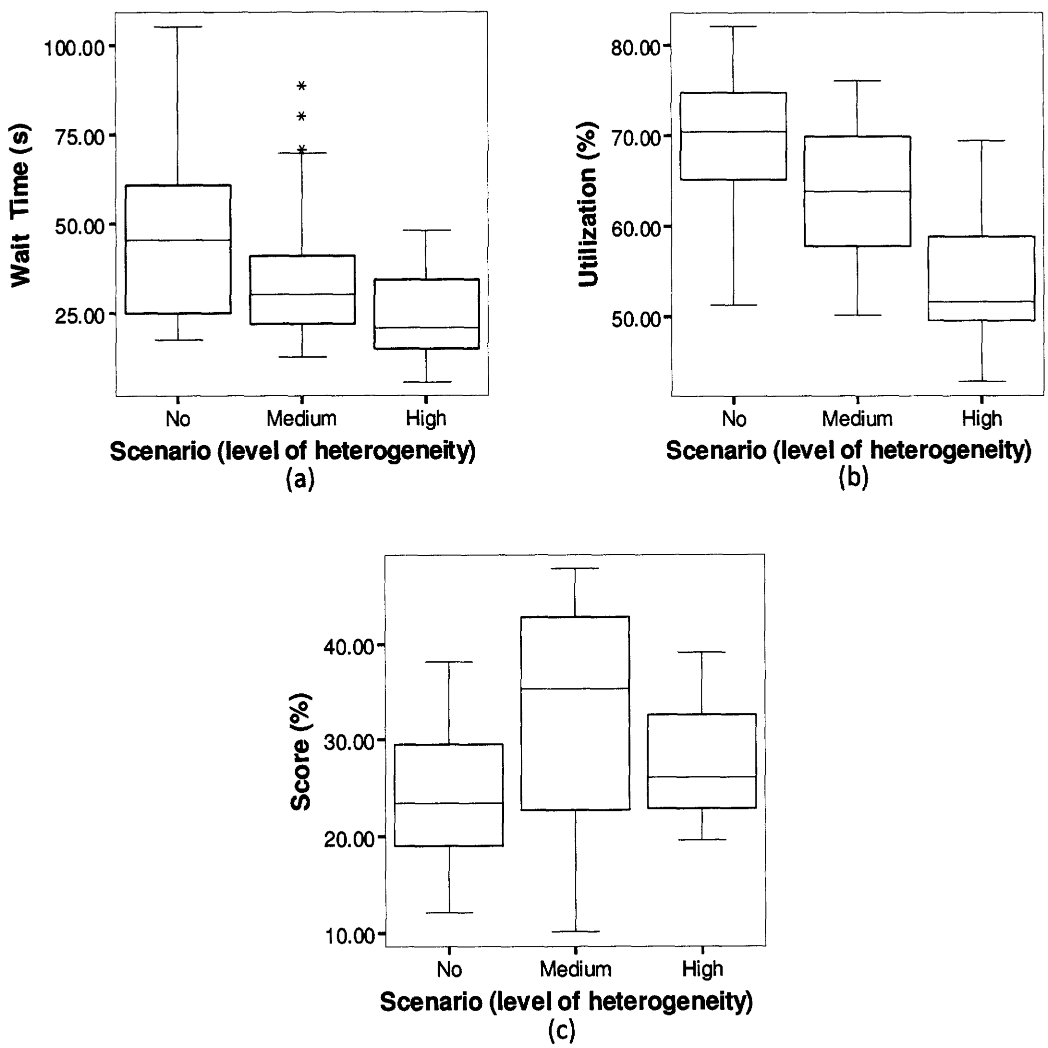


Figure 4.5 Box plots for (a) average wait time, (b) operator utilization, and (c) score results for the RESCHU experimental study

The results revealed that, as expected, utilization was the highest for the no-heterogeneity level, followed by the medium-heterogeneity and then the high-heterogeneity levels. As the level of vehicle team heterogeneity increased, operators interacted less frequently with the vehicles due to longer neglect times and shorter interaction periods. UUVs, which spend a considerable amount of time underwater, result in less interaction with the human operator than UAVs that are in constant communication with the control station. Similarly, the HALE UAVs require shorter interaction than the MALE UAVs. Therefore the heterogeneity in vehicle capabilities and tasks drove the task load imposed on the human operator, which was manifested in operator utilization. As far as the wait time metric, the no-heterogeneity team structure generated longer search task wait times as compared to the high-heterogeneity team structure. This was expected due to the shorter interaction times associated with the high-heterogeneity team in this experiment. Because the operators had to interact for shorter periods with the HALE UAVs, the high-heterogeneity service queues were shorter, thus generating shorter wait times.

Finally, the mission score results show that the no-heterogeneity configuration resulted in a significantly lower score than the medium-heterogeneity team structure. This is due to the significantly higher operator utilization in the no-heterogeneity configuration which negatively impacted operator performance. In addition, there was a non-significant difference in mission score between the high-heterogeneity and medium-heterogeneity conditions. Although the high-heterogeneity condition resulted in significantly lower operator utilization, the results show that this condition resulted in lower scores (non-significant) than for the medium-heterogeneity team structure. One possible explanation is that this is a result of heterogeneity in service processes which is existent in the high-heterogeneity condition but not in the medium-heterogeneity condition. Chapter 5 later shows that even as operator task load is held constant, heterogeneity in service processes can cause larger vehicle wait times. The implications of this are that both operator utilization and vehicle wait times impact the mission score, which is a system performance metric.

4.2 Replication Validation

In order to determine the model's ability to describe the observed data, the data sets from the experiment for each team condition were utilized in order to generate the inputs to the MUV-DES model. The model was then used to replicate the observed human-in-the-loop experimental results for each of the conditions.

Vehicle-generated, operator-induced, and exogenous event distributions in the case study were identified. For the RESCHU interface, four event types per vehicle type involving user interaction existed: 1) a vehicle arriving to an AOI and requiring the operator to undertake the search task (a vehicle-generated endogenous event), 2) an opportunity for re-planning the vehicle's path to a closer AOI (an operator-induced endogenous event), 3) an idle vehicle that requires assignment to an AOI (a vehicle-generated endogenous event), and 4) the intersection of a vehicle's path with a threat area (an exogenous environmentally generated event). The fitted distribution types and their parameters for the different event arrivals and service processes using all the data from the experiment were generated using the EasyFit© software and are summarized in Appendix G.

Using the distributions of arrival rates and service times generated from the user study, 5,000 trials were conducted with the MUV-DES model. Mission score, average wait times for the search task event and operator utilization were replicated for each condition of the study in order to compare the results with the human-in-the-loop experiment. The observed average search task wait time, operator utilization, and mission score from the user study are compared with the model's estimates in Figures 4.6 - 4.8, respectively. The model's results for average score, average search task wait times and operator utilization are all within the 95% confidence intervals of the experimental means. Comparisons of these results with those observed in the human-in-the-loop experiments show that the model adequately captures the previously described system dynamics for the different levels of heterogeneity.

In terms of operator utilization, the model is able to capture the impact of increasing heterogeneity, which entailed operators interacting less frequently with the vehicles due to longer neglect times and shorter interaction periods. In terms of vehicle wait times, the model was able to replicate the difference in wait times between the no and high-heterogeneity teams, which was due to the shorter interaction times in the latter case. Finally, in the case of the mission score, replication of the observed scores from the user study increases confidence in the model's ability to capture the effects on system performance of operator, as well as vehicle team performance.

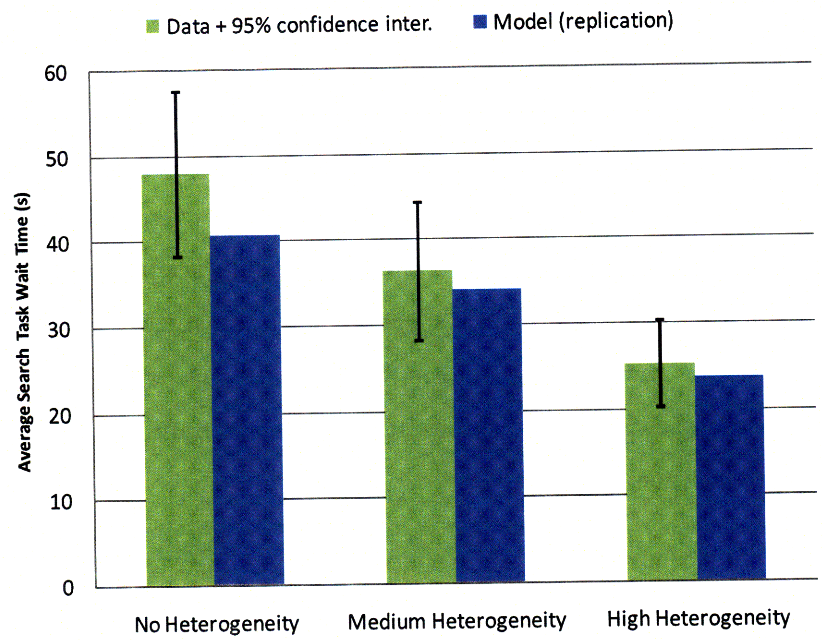


Figure 4.6 Results for average search task wait time compared with model replications

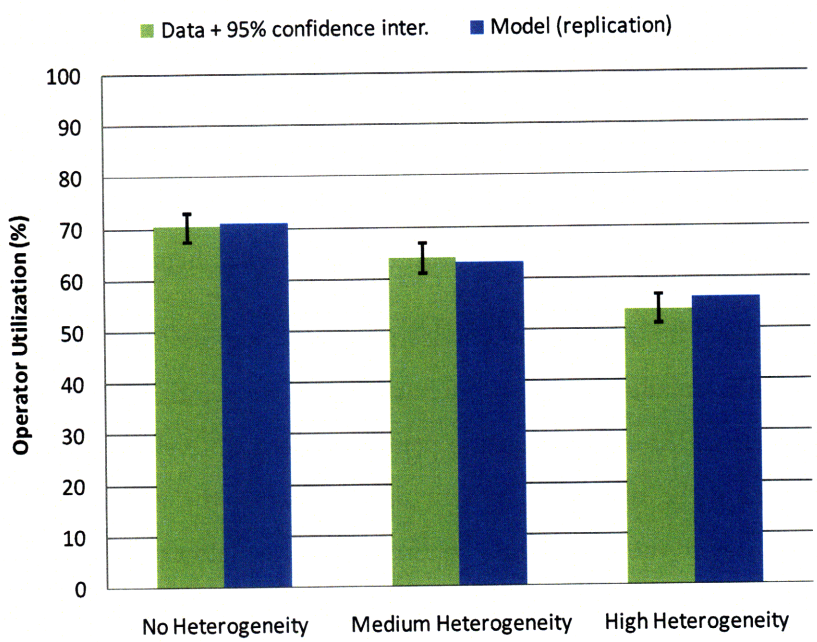


Figure 4.7 Results for operator utilization compared with model replications

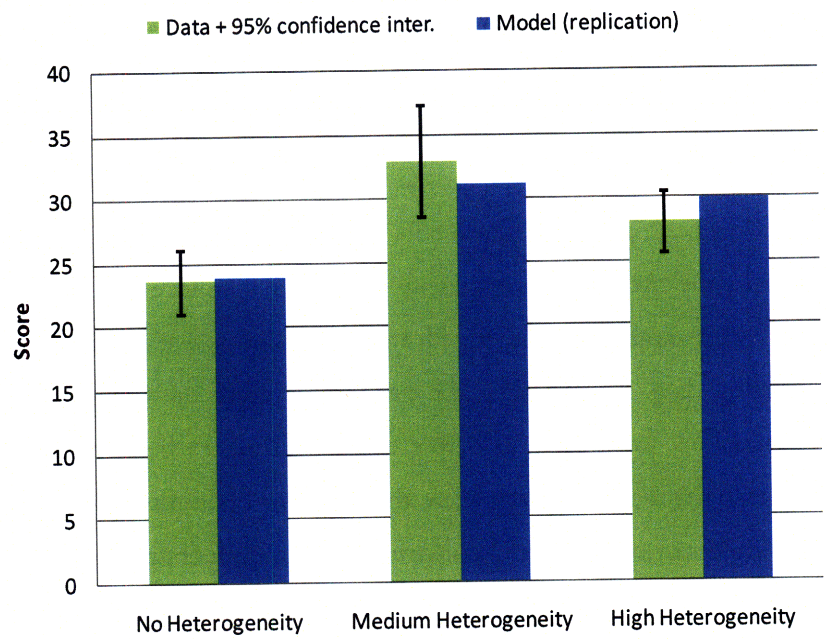


Figure 4.8 Results for the score variable compared with model replications

4.3 Predictive Validation

The next validation technique applied to the model was that of validating the model's ability to predict the effects of changes in the vehicle heterogeneity structure on the dependent variables. This is important, as one of the goals for building this model is to provide predictive abilities in order to aid designers in understanding how potential design modifications will affect the performance of the system. In order to do this, the medium-heterogeneity team was used as a base case to predict the effects of changing the team structure to either the no-heterogeneity or high-heterogeneity teams. The medium-heterogeneity condition was chosen because it represented the median condition in terms of operator task load.

To predict the impact of changes in the team structure from that of the medium-heterogeneity condition to each of the other two heterogeneity conditions, it was important to determine which human-UV interaction attribute(s) would be affected and in what way. These changes would then be reflected in the MUV-DES model constructs by editing the existing distributions (i.e., those belonging to the medium-heterogeneity condition).

In going from the medium-heterogeneity team to the no-heterogeneity team, the change was composed of two UUVs replaced by two MALE UAVs, thus the appropriate arrival and service processes were substituted. In going from the medium-heterogeneity team to the high-heterogeneity team, the change was composed of one of the MALE UAVs replaced by a HALE UAV. However, simple distribution substitution was not possible because there were no HALE UAVs in the medium-heterogeneity team and therefore the arrival processes of vehicle-generated events could not be derived. It was decided to use Monte Carlo simulations to derive the missing data. By encoding the geographic areas on which AOIs could exist, Monte Carlo simulations were used to derive average distances between randomly located AOIs. The samples from the simulations were then used to build distributions that were then used together with the rest of the model constructed for the medium-heterogeneity case. The observed average wait time, operator utilization, and system performance variables from the user study are compared with the model's predictions in Figures 4.9 – 4.11, respectively.

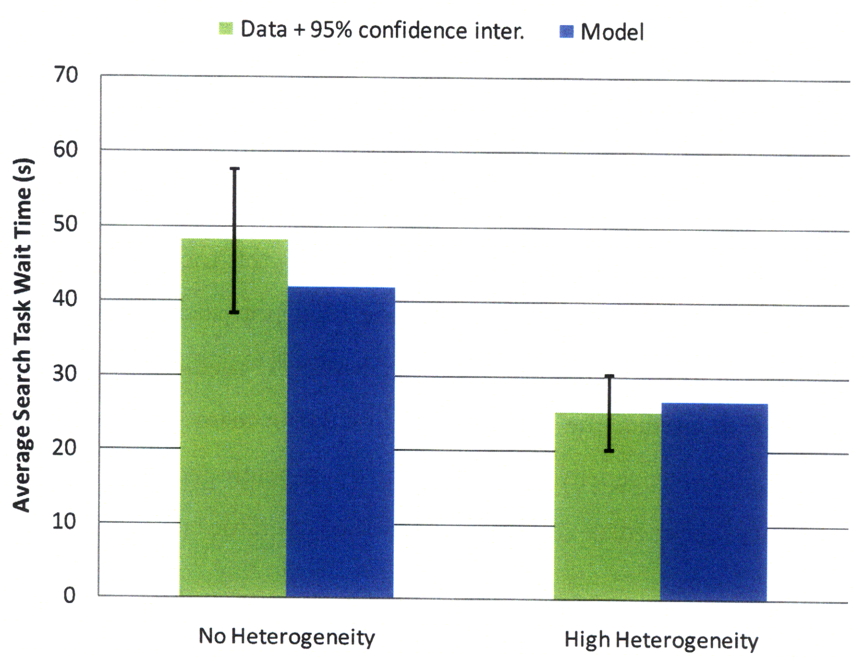


Figure 4.9 Results for average search task wait time compared with model predictions

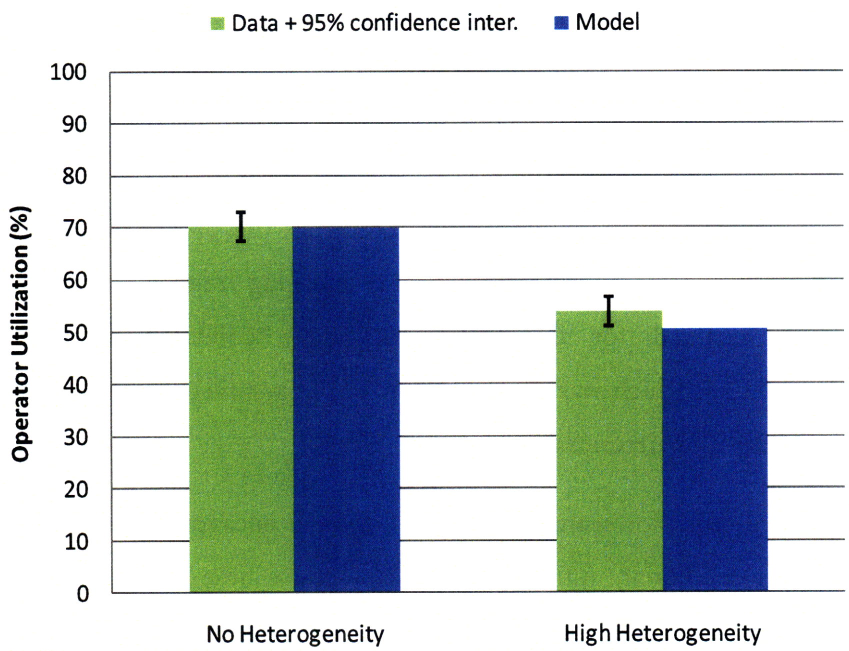


Figure 4.10 Results for operator utilization compared with model predictions

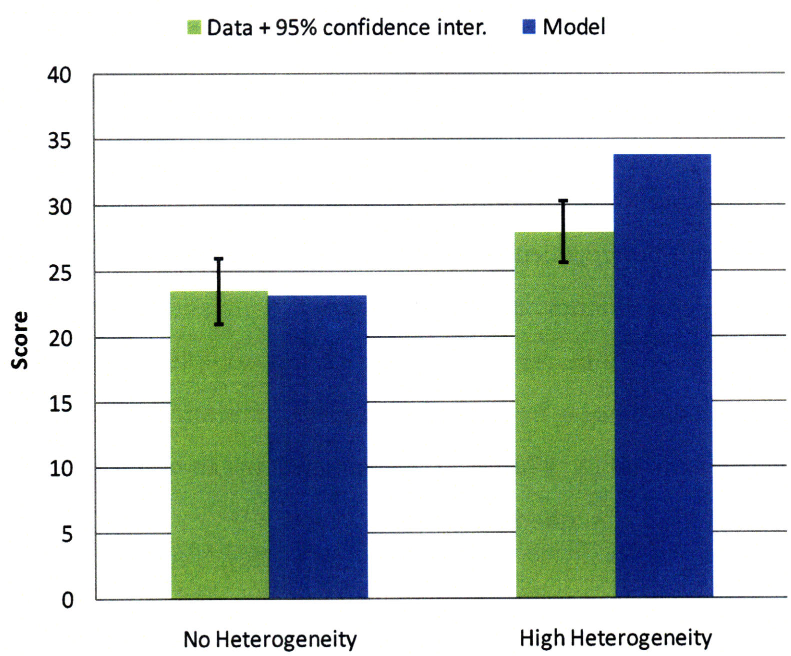


Figure 4.11 Results for mission score compared with model predictions

Although the model's results are within the 95% confidence intervals for the no-heterogeneity team, they tend to be less accurate when predicting for the high-heterogeneity condition, specifically for mission score and operator utilization. Comparing the distributions derived using the experimental data set from the high-heterogeneity condition against those used in making the predictions (i.e., those derived by modifying the processes from the medium-heterogeneity condition) revealed that the operator switching strategy as well as the operator interaction times differed for the high-heterogeneity condition from the medium-heterogeneity condition, which was used to estimate the parameters for making predictions. The discrepancies are each described below.

The average service times for search task events differed between the medium to the high-heterogeneity team conditions. Although the mean service time for search task events was 24.4 seconds in the medium-heterogeneity case, it was 27.6 seconds for the high-heterogeneity condition, despite the lower task load. One explanation is that the lower tempo in the high-heterogeneity case (which led to 53.7% operator utilization versus 63.9% in the medium-heterogeneity case) caused operators to feel more comfortable with spending longer times on the search task (the average search task service time for the no-heterogeneity condition was 24.3 seconds, similar to that in the medium-heterogeneity condition). However, no causal relationship can be established between operator utilization and service times without a controlled study designed for that purpose. The impact of underestimating service times when predicting for the high-heterogeneity condition was underestimation of operator utilization due to the shorter service times and overestimation of mission score due to the greater number of objects that could be serviced (since each task took less due to the shorter service time). However, the inaccuracies in predicting operator utilization and mission score are not reflected in vehicle wait times. This can be explained by an inaccurate estimation of the switching strategy which is described next.

In the medium-heterogeneity condition, there were un-allocated red AOIs to which subjects could re-plan the MALE UAVs or UUVs. Therefore a possible switching strategy that subjects could utilize was one where operator-induced goal re-plan events were prioritized over search task events. Because such a switching strategy results in search task events having to wait while the operator re-plans a vehicle's goal, the resulting search task event wait times are greater than under a first-in-first-out strategy or a priority scheme were search task events are prioritized over goal re-plans. In the high-heterogeneity condition, the existence of grey AOIs decreased the number of red AOIs that subjects could re-plan UUVs or MALE UAVs to (5 red AOIs and 2 grey AOIs instead of 7 red AOIs as was the case for the medium-heterogeneity condition). Therefore, there were fewer opportunities for re-planning in this condition, which was confirmed by the large number of subjects that exhibited a priority scheme were search task events were prioritized over AOI re-plans. This presented a second discrepancy since this was not accounted for when utilizing the data set associated with the medium-heterogeneity condition to make predictions for the high-heterogeneity condition. The result was an overestimation of search task event wait times, which compensated for the lower wait times that should have resulted as a consequence of the underestimation of service times. Overall, the underestimation and overestimation canceled each other out, which resulted in an accurate prediction for search task event wait times.

4.4 Extreme Condition Testing

Extreme condition testing involves checking that the model outputs are plausible for extreme design inputs. This is important as a model should be valid under the complete input domain space, and not just median input settings. Success at extreme condition testing allows confidence to be extrapolated beyond initial test points. Of the inputs to the model, the service times and inter-arrival times between events were the inputs that could take on extremely large or small values. In the case of event service times, extremely small values result in events being serviced almost instantaneously whereas large values result in the server holding up all other events while a single event is being serviced. In the case of event inter-arrival times, large values result in events being rarely produced and therefore, an almost idle operator, whereas really small values result in many events arriving. It is this last case that was identified as important to consider, because a design variable of prime concern in multiple heterogeneous UV systems is the number of vehicles in the team. Increasing the number of vehicles results in additional events arriving, similar to reducing the inter-arrival time for a single event type. Since there is a theoretical limit to the maximum number of vehicles one person can effectively control, even for a "perfect" operator (which could be represented by the earlier notion of Fan-out), it was desired to ensure that as the number of vehicles increased, the model continues to behave appropriately by predicting operator utilization that approaches 100%.

Operator utilization, ρ, was defined in Chapter 3 as the ratio of the total time the operator is engaged in tasks, i.e. servicing events, to the total time elapsed. Since WTIs are part of the total service time, they are accounted for in the definition of operator utilization. However, because the sequence of operator actions in real life does not necessarily match that of the theoretical representation, the identification of WTI is not always straightforward. For example, in the RESCHU test bed, WTI was not easily identifiable for the goal re-plan event type. In the case of re-planning to a new goal, WTI, which was mainly composed of the time it took the operator to decide on a new goal, could be assumed to have taken place sometime after the vehicle's automatic assignment to a new goal and prior to the operator commencing the vehicle goal change. The exact start and end times of the WTI period were impossible to extract when the operator attended to other events after the vehicle's automatic assignment but prior to commencing a goal re-plan. Therefore, WTI could not always be measured for each goal re-plan event occurrence and could not be included when calculating operator utilization from the data set.

Although WTI was not accounted for when estimating operator utilization from the experimental data set, its existence affected the dynamics of operator interaction. It was therefore important to account for the goal re-plan WTI in the MUV-DES model. This was possible by estimating WTI from those re-plan events in the experimental data set where the time between the vehicle's automatic assignment and the start of the goal re-plan did not include other operator interaction. These times were used to generate a probabilistic distribution representing WTI for the goal re-plan event to be used in the MUV-DES model. Therefore, the goal re-plan event WTI was accounted for in the MUV-DES model and could be measured as part of operator utilization.

Figure 4.12 shows a graph of operator utilization (with WTI accounted for and not) versus the number of identical vehicles in the team being supervised (estimated using the MUV-DES model). As can be seen, when WTI was accounted for, the model approached the 100% mark as the number of vehicles was increased. The point of this test is to show how the model acts under an extreme condition of interest, and not to make any conclusions as to the maximum number of vehicles that can be controlled in general. This depends on the inter-arrival times of events, the service times, and the threshold for operator utilization that the mission specifications set. Thus, the maximum number of approximately 9 vehicles is for the RESCHU simulation only.

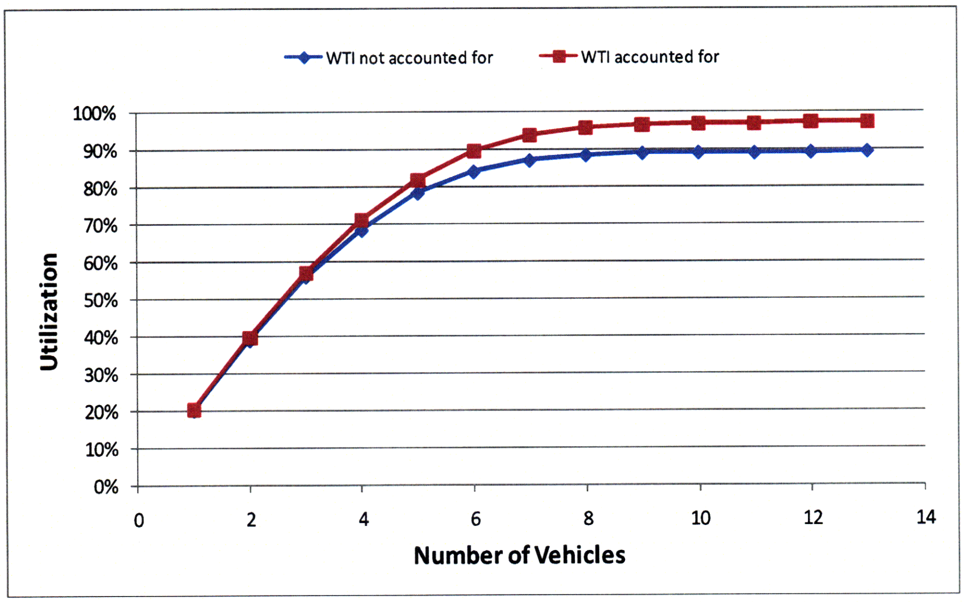


Figure 4.12 Operator utilization vs. number of vehicles in the team (MUV-DES model)

4.5 Internal Model Stochasticity

Another element of concern for validation is the variability exhibited by the MUV-DES model. Extremely large variability in the model's outputs can cause the model's outputs to be questionable. If extremely large variability is a property of the system being analyzed, then this may question the appropriateness of modeling the system using a DES in the first place (Sargent, 2005). Figure 4.13 displays the standard deviation derived from the experimental data set as well as that produced by the MUV-DES model for the three metrics of interest: average wait times for the search task event, operator utilization, and mission score (the abbreviations 1, 2, and 3 are used instead of no-heterogeneity, medium-heterogeneity and high-heterogeneity respectively). The graph shows that the standard deviation exhibited by the model closely matches that from the data for the utilization variable, and is smaller in the case of the score and wait time variables.

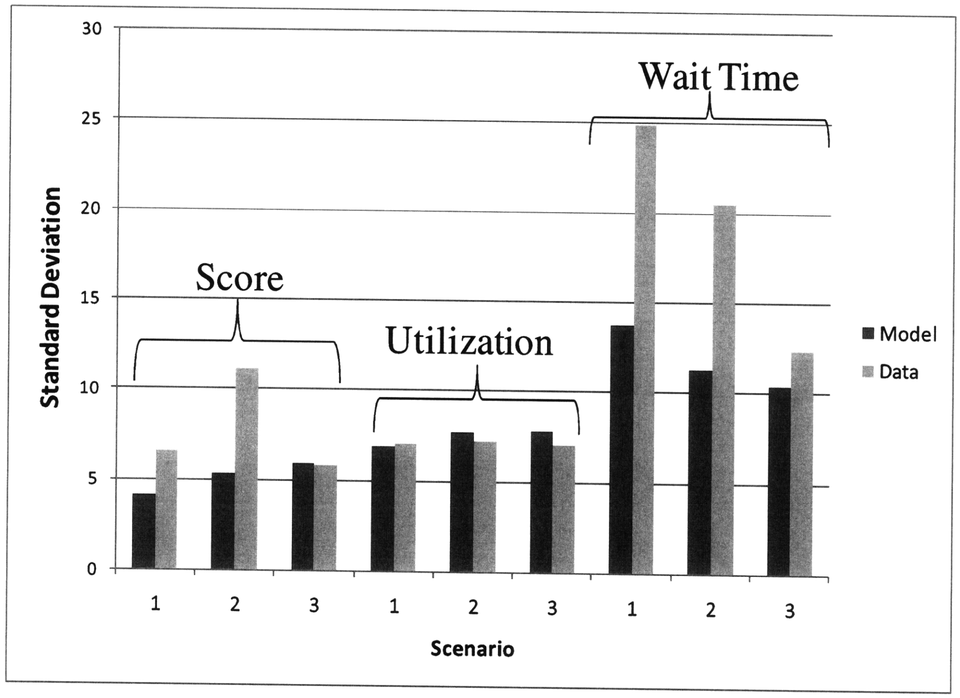


Figure 4.13 Comparing standard deviations between the MUV-DES model vs. experimental data set

The reason for the model under-estimating the standard deviation in the case of wait times can be attributed to two different properties of the MUV-DES model. First, the fixed switching strategy (or fixed queuing policy), which is an a priori input, is equivalent to assuming that all operators exhibit the same switching strategy. This is artificial however, as different users are likely to exhibit varying strategies which result in varied wait times. Since different switching strategies lead to different queuing wait times, this impacts the resulting standard deviation in wait times. Second, the WTSA/UT curve, which is used to dynamically derive operator reaction times to events in the system, is based on a mean reaction time at each utilization level. This also results in less-varied WTSAs, which form another component of the total wait time output. As far as underestimating the mission score, this is due to the fact that mission performance is dependent on both vehicle and operator performance, as was highlighted in Section 4.2. Since vehicle wait times were less varied, the number of objects accurately identified also exhibited less variation.

Therefore, designers that are interested in replicating or predicting the variability in vehicle or system performance should consider using probabilistic distributions for both the switching strategy and the WTSA associated with each utilization level. In the case of switching strategies, designers could consider a more complicated model that allows for a distribution of switching strategies to be modeled as opposed to a single, fixed switching strategy. Recent work by Crandall et al. (2008) has proposed modeling switching strategies using a probability distribution. The challenge in applying such a technique however, is that it is not always clear what strategy operators are in fact pursuing, and therefore extracting samples from which to build the distribution is difficult. In addition, assuming that a distribution could be derived, the process of selecting from this distribution when deciding what strategy should be put into effect is challenging and adds considerable complexity to the model. In the case of the WTSA/UT curve, future designs could associate a random variable with each utilization level as opposed to an average WTSA as was the case in Chapter 3. Since such a model would add significant overhead in terms of estimating the inputs to the model, this should only be done when absolute variability is a prime concern for the underlying research being conducted.

4.6 Sensitivity Analysis

Another important test when validating any computational model is a sensitivity analysis. In a computational model, it is often desired for the outputs to not be overly sensitive to errors in input estimates because estimates of the design variable inputs that feed the model are likely to be imperfect in most cases. In addition, those variables for which the model outputs are highly sensitive need to be identified in order to ensure that they are made sufficiently accurate prior to using the model.

In order to conduct a sensitivity analysis, it was decided to focus on the means of the service and arrival distributions for the search task. The reason for selecting the search task is because this event type is the one with the largest impact on vehicle, operator and system performance. This is due to the fact that the service time for the search task event type (~25s) is large when compared to goal re-plan or avoiding threat area event types (~3s), and because the number of search task events serviced is directly proportional to the mission score. Errors were introduced by shifting the mean of the service and arrival probability density functions, therefore disturbing the estimates from those estimated from the experimental data set. The random variables associated with the arrival and service processes were modified as follows:

$$z' = (1 + e) * z + \gamma(\rho)$$

$$y' = (1 + e) * y + wti$$

where the disturbance factor e was varied between 5% and 20% in 5% increments. z is a realization of the random variable that describes the time between the last-generated event being serviced and the next arrival, γ is the functional form relating WTSA to operator utilization (ρ), y is the realization of the random variable that describes the service time, and wti is the realization of the random variable that describes the wait time due to interaction. The output errors were then estimated by comparing the original replication outputs of the model vs. the outputs after introducing the errors to the input distributions (Figures 4.14-4.15).

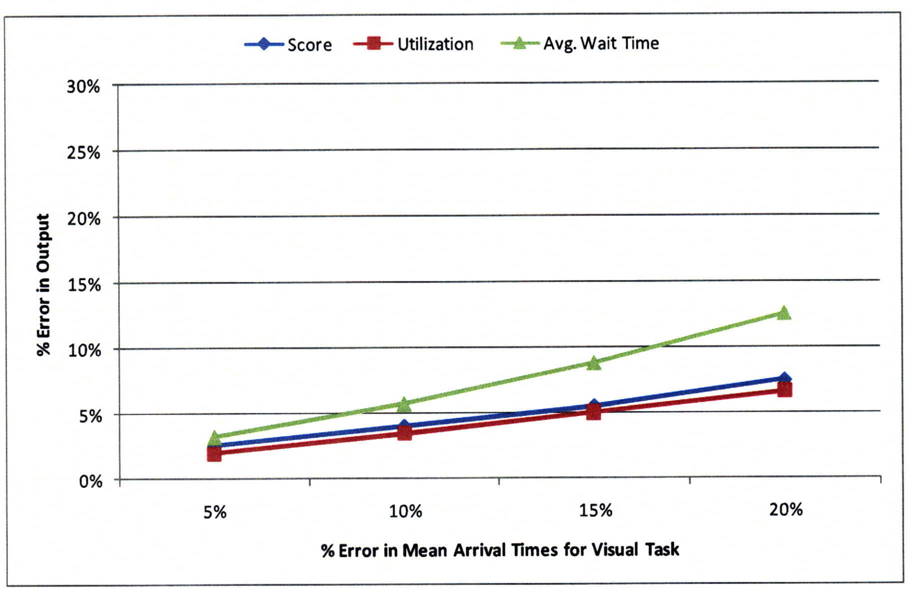


Figure 4.14 Impact of errors in estimates of mean visual task arrival times

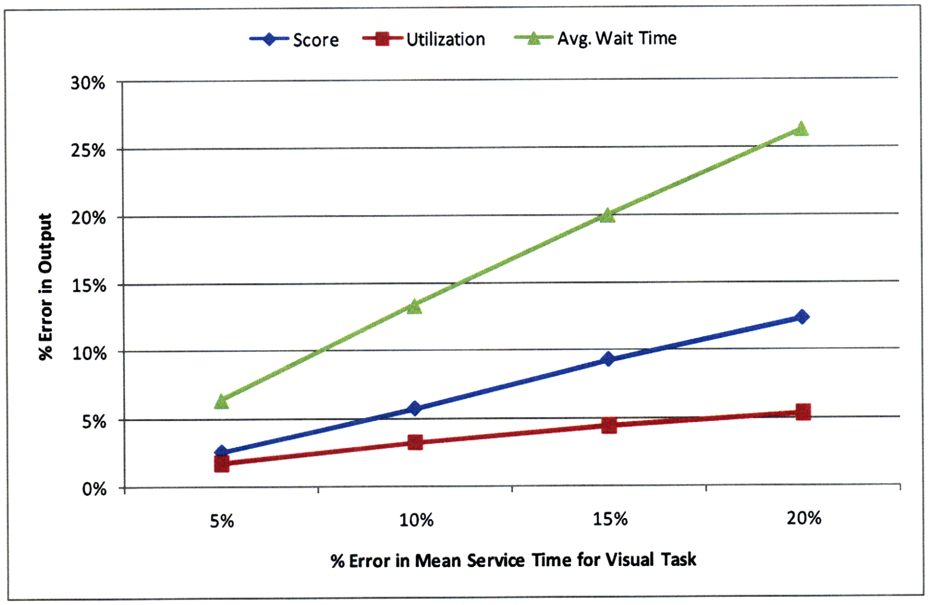


Figure 4.15 Impact of errors in estimates of mean visual task service times

Several observations can be made from Figures 4.14-4.15. First, errors in the service distribution (Figure 4.15) cause larger errors in the output variables than those produced by errors in the arrival distribution (Figure 4.14). Second, the wait time output variable is the most sensitive to errors in the input variables, especially in the case of errors in the service distribution. The larger impact of errors in service times versus errors in inter-arrival times can be explained by the direct dependence of a single event's wait time on the amount of time that it has to wait before receiving attention. Whereas inter-arrival times impact this through changing the number of events likely to be in the queue ahead of the event in question, service times have a larger effect as they increase the time it takes to process each of those events. Finally, because the score metric is a function of wait times, it is also impacted more severely by errors in service times.

Because the wait time variable is an important metric that designers would wish to measure or control (Section 2.2), service times should be measured accurately prior to using the model, especially if the goal is to generate accurate absolute estimates of wait times. Since the wait times measured using the MUV-DES model were within the 95% confidence interval of those extracted from the RESCHU experimental data set, it can be concluded that this was the case for the service distributions estimated in this thesis.

4.7 Historical Data Validation

In this section, additional confidence was built in the MUV-DES model by using historical human-in-the-loop performance data that existed pervious to model construction. The significance of this is the external validity that results from accurately replicating independent data, specifically data that has been used in a number of previous human-in-the-loop experiments to validate other discrete event simulation predictions (Crandall & Cummings, 2007a; Crandall & Cummings, 2007c). This particular data set belonged to an experimental study that investigated operator performance issues in the control of multiple simulated homogeneous UVs in the context of a search-and-rescue mission. The design was relevant to the validation purposes of this chapter as it a) presented a historical data set that was generated independently from the research associated with the MUV-DES model development, and b) the data presented results from a controlled experiment that could be used to validate the model's ability to replicate the effect of changing team size as well as level of autonomy.

4.7.1 Experimental Procedure and Results

A brief description of the experimental setup is given here. A more detailed description can be found in Appendix H. The experimental design was a 2x4 factor study, and the purpose was to investigate the effect of increasing team size and alternate decision support schemes on operator performance and utilization in a search and rescue setting with unmanned ground vehicles. The human-UV interface was the two-screen display shown in Figure 4.16. The mission of the human-UV team was to remove as many objects as possible from a maze in an 8-minute time period (the left screen contained the map of the maze along with the positions of the UVs and objects with known locations). To simulate the real world task of having to identify objects, users were asked to identify a city on a map of the mainland United States using *Google Earth*-style software (the right screen was used for this purpose).

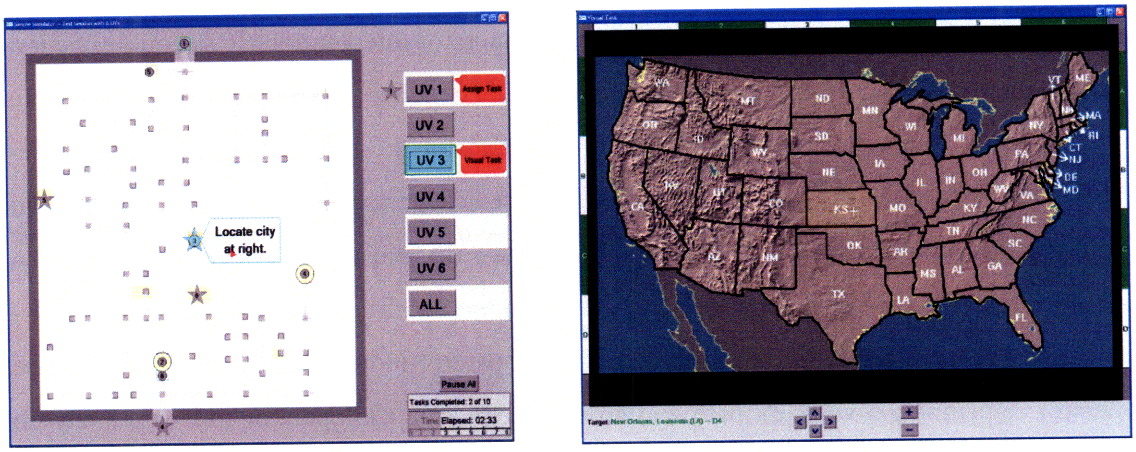


Figure 4.16 Two-screen interface by which an operator directed the UVs

The decision support condition was a between-subjects factor with two levels: no decision support (NDS) and full decision support (FDS). UV team size was a within-subjects factor; each participant performed the search-and-rescue mission for team sizes of two, four, six, and eight UVs. The dependent variables included the number of objects collected, which was indicative of mission performance, and operator utilization (calculated as the percent busy time), which was indicative of operator performance.

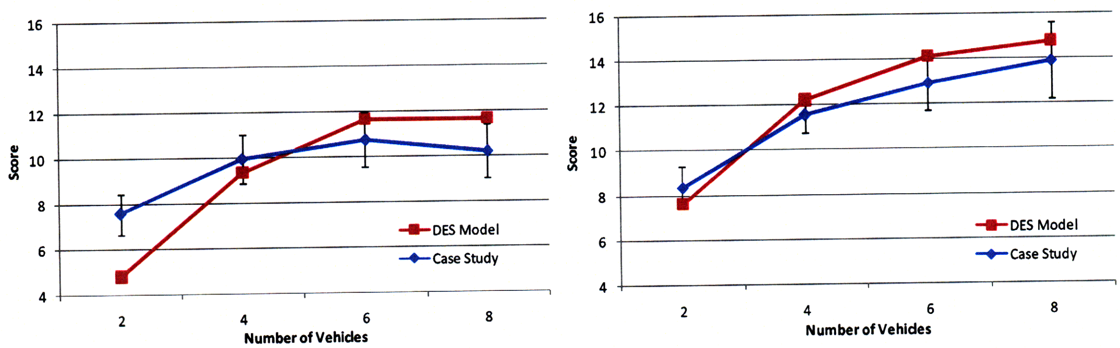
The results presented in Appendix H, showed that changing the team size had a significant effect in terms of mission score and operator utilization up to six vehicles. Beyond that, the addition two more vehicle to the team did not result in significant score or utilization increases. In addition, decision support had a significant effect for both score and operator utilization, causing an increase in the former and a drop in the latter. The next step was to evaluate whether the MUV-DES model could generate performance results similar to those observed in this prior study.

4.7.2 MUV-DES Results

In order to compare the MUV-DES model results to the historical human-in-the-loop experimental results, vehicle-generated and operator-induced distributions in the case study were identified (there were no exogenous events in this study). Vehicle-generated events included both locating a city on the map and goal assignment in the NDS condition. In the FDS condition, only locating a city on the map was modeled as a vehicle-generated event since this was the only event type that necessitated operator intervention (UVs could function without goal-assignments from the user in the FDS condition). Thus, five data sets were measured from the experimental data: 1) a distribution for the inter-arrival times of vehicle-generated events after the UV was serviced for a vehicle-generated event (time between an event requiring the operator to locate a city on the map and the next event requiring the operator to assign the vehicle to a goal), 2) a distribution for the inter-arrival times of vehicle-generated events after the UV was serviced for an operator-induced event (time between the operator re-planning a vehicle's path and an event requiring the operator to locate a city on the map), 3) a distribution for service times of vehicle-generated events (time it takes the operator locate a city on the map), 4) a distribution for the inter-arrival times of operator-induced events (time between the operator re-planning one vehicle's path and the next time the operator re-plans a vehicle's path), and 5) a distribution for the service times of operator-induced events (time it takes the operator to re-plan a vehicle's path). The fitted distribution types and their parameters for different events arrivals and services identified using EasyFit© software are presented in Appendix H.

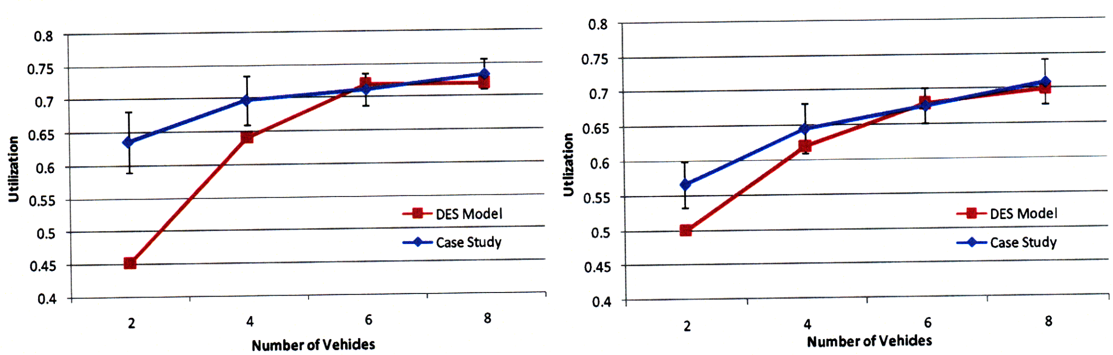
The complete model of the human-UV team also requires a performance model. In the user study, the team scored points when an object was removed from the maze. In the NDS condition, this required two vehicle-generated events to occur (goal-assignment and locating a city). Thus, the MUV-DES model awarded a point for the servicing of every two vehicle-generated events. In the FDS condition, only one vehicle-generated event (locating a city) needed to be performed. Thus, in this condition, the MUV-DES Model awarded a point for every serviced vehicle-generated event.

Using the distributions of arrival rates and service times generated from the data in the user study, 10,000 trials were conducted with the DES, in order to compare the results with the human-in-the-loop experiment. The observed system performance and operator utilization from the user study are compared with the model's estimates in Figures 4.17 and 4.18, respectively. For the FDS condition, the model's results for system performance score are all within the 95% confidence intervals. Likewise, the FDS utilization results are all within the 95% confidence intervals except in the 2-UV case. In this case, the model underestimates operator utilization by approximately one standard deviation (6.7% from the mean operator utilization). In the NDS condition, the model's results of system performance are within the 95% confidence intervals for the 4- and 6-UV conditions, but were low (1.6 standard deviations) in the 2-UV condition and slightly high in the 8-UV condition (0.6 standard deviations). Additionally, results of operator utilization are within the 95% confidence intervals for 6- and 8-UV teams, but not the 2- and 4-UV teams, (2 and 0.8 standard deviations respectively).



(a) (b)

Figure 4.17 Mean performance scores from MUV-DES model and historical data set for a) NDS and b) FDS conditions



(a) (b)

Figure 4.18 Mean utilizations from MUV-DES model and historical data set for a) NDS and b) FDS conditions

Two observations about the accuracy of the model can be made from these results. First, model results are more accurate for larger teams than small teams. This trend appears to be caused, at least to some degree, by overly high penalties associated with low utilization in the SA model. As a result, the human is modeled as not servicing UVs as often as they did in the 2-UV condition (operators were generating more tasks for themselves than necessary, which the model did not capture). This leads to the under estimation of both operator utilization and system performance in this condition. The second trend observed in Figures 4.17 and 4.18 is that the model's estimates are better in the FDS condition than in the NDS condition. Again, it appears that this trend is due to difficulties in modeling human behavior when automation is not present, as human behavior is more difficult to model than is automation's. Thus, systems that rely more on human behavior (i.e., the NDS condition) are more difficult to accurately model than systems that rely more on automated behavior (i.e., the FDS condition).

In summary, the model was used to replicate the system performance and operator utilization of multiple simulated homogeneous UV teams. Comparisons of these results with those observed in the human-in-the-loop experiments show that the model adequately captures these system dynamics for the homogeneous case. Despite variations in the accuracy of the model's results, the model captured the general trends in system performance and operator utilization as the size of the team and the level of decision support changed. This is important since it means that the model gives adequate descriptions of the behavior of different system architectures in a cost effective manner.

4.8 SA Construct Validation

In addition to gaining confidence in the model's effectiveness as a whole, it was desired to build confidence in the WTSA/UT input. Validation is first conducted by showing the model improvement when including the SA model for replicating the outputs observed in the historical data set. This data set, described in Section 4.7, includes operator utilization and mission performance data for different team sizes and levels of autonomy. The more extensive RESCHU data set is then used to build confidence in the exact shape of the curve, which has been proposed to be parabolic and concave upwards. This data set, described in

Section 4.1, includes vehicle wait times, operator utilization and mission performance data for different levels of vehicle team heterogeneity.

4.8.1 Model Improvement (Using Historical Data Set)

Using the random distributions generated for the historical data set discussed previously, 10,000 trials were conducted using the MUV-DES model whose SA model was varied between three alternate configurations.

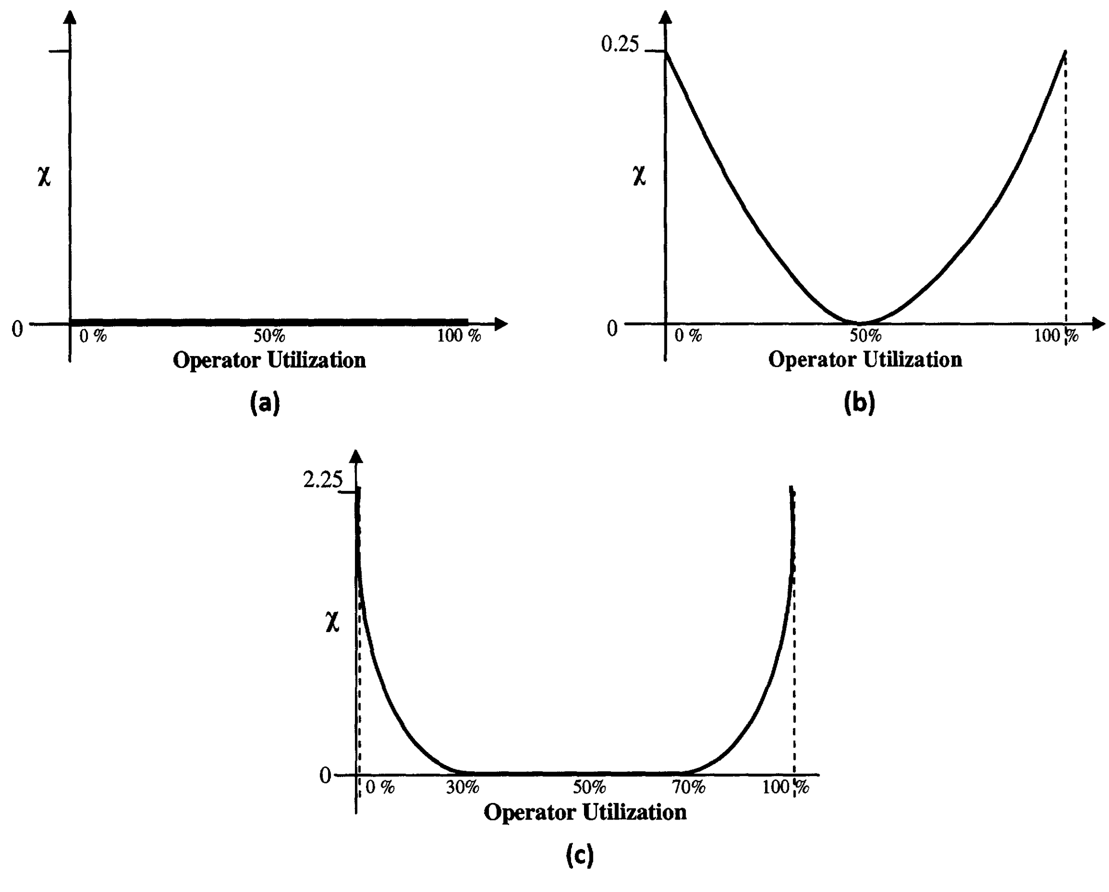


Figure 4.19 Situational awareness curve variations

The first SA model was one where the penalty was zero under all levels of utilization (I_P=100) (Figure 4.19a). This is equivalent to ignoring SA in the MUV-DES model. The second model was a parabolic curve that was concave upwards where C_F was set to 50%, I_P set to 0, and S_F set to 0.25 (Figure 4.19b). This is equivalent to using a simple parabolic curve with no magnitude scaling, as S_F is equal to 0.25 when scaling is ignored. Finally, the third model was one where C_F was set to 50%, I_P set to 40%, and S_F to 2.25 (Figure 4.19c). The magnitude scale factor for this parabolic curve was chosen by selecting the one that minimized the mean squared error for both performance score and operator utilization. This is equivalent to using a curve where the penalty for loss of SA is zero when operator utilization is between 30% and 70%, and with a magnitude scaling that minimizes the mean squared errors.

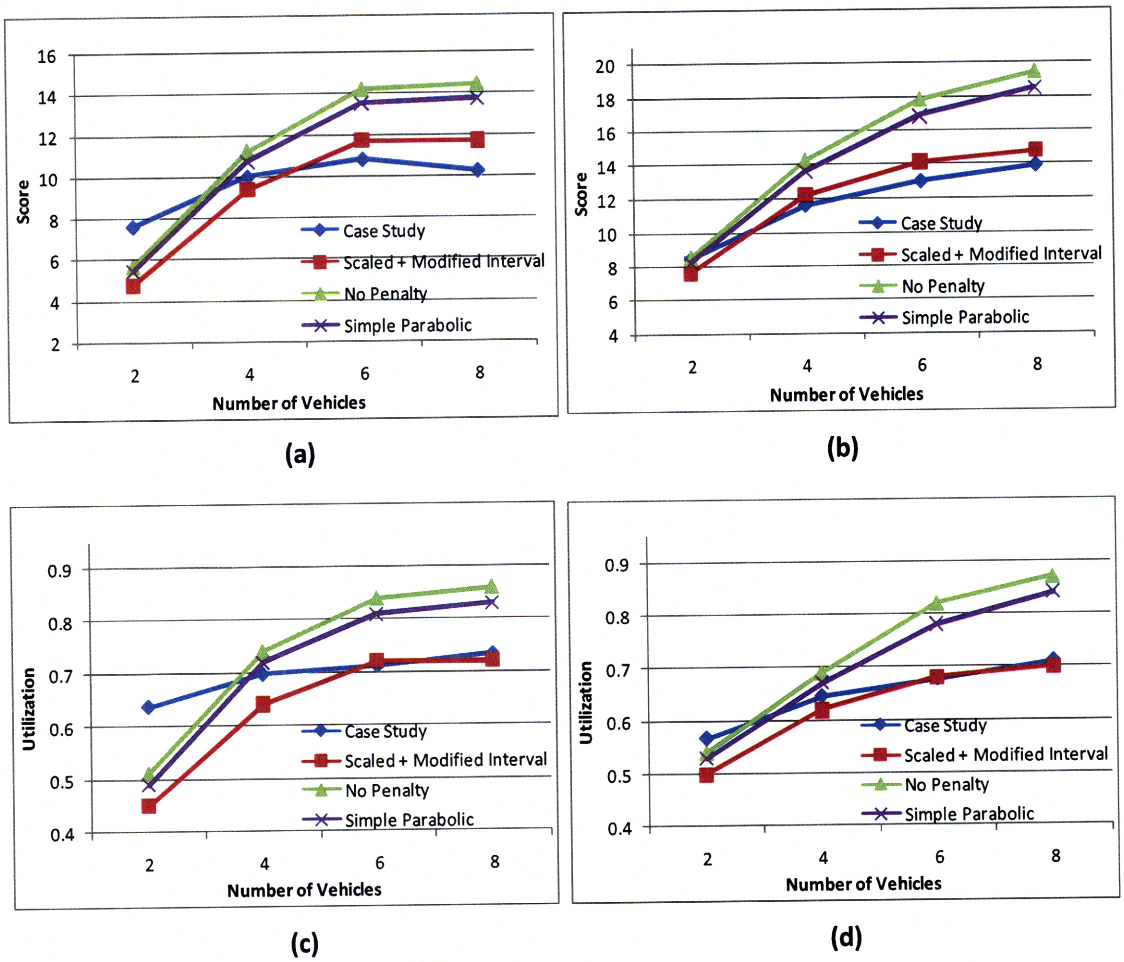


Figure 4.20 Effects of situational awareness model on (a) mean scores (NDS), (b) mean scores (FDS), (c) operator utilization (NDS), and (d) operator utilization (FDS)

The actual scores in the user study for both the no-decision support and full-decision support cases are compared with results from the discrete event simulator for each of the three different SA curves in Figures 4.20a-b. The performance score mean squared error for each of the three SA curves was calculated (for both decision support cases). The SA model with no penalty performed the worst followed by the model with the simple parabolic curve. The model with the scaling and its interval of zero penalty modified performed the best.

It can therefore be concluded that the inclusion of the SA component in the model had an overall positive effect in terms of performance prediction accuracies, specifically when the model was scaled appropriately. When the effect of low SA (i.e., high or low utilizations) was completely ignored, the model tended to over-predict performance for the 4, 6 and 8 vehicle conditions and under predict for the 2 vehicle condition. Including the SA effect, which increased the inter-arrival time between vehicle-generated events because of operator-induced wait times, led to decreased performance scores (and thus improved predictions), except in the 2 vehicle condition where the model was already under-predicting performance. The results also demonstrate that improvement in the predictions when including the SA component is dependent on using a curve with appropriate shape parameters. Using a simple curve with no shape or magnitude modifications leads to a small improvement in the predictions while modifying the shape and magnitude parameters is required to make the penalties significant and the predictions more accurate.

The inclusion of the SA component in the model also had an overall positive effect in terms of utilization predictions as can be seen in Figures 4.20c-d. When the effect of low SA (due to high or low utilizations) was completely ignored, the model tended to over-predict operator utilization for the 4, 6 and 8 vehicle conditions and under-predict for the 2 vehicle condition. Including the SA effect led to an overall improvement in utilization predictions because the SA model tends to have a regulating effect on the number of tasks processed by the operator. When the operator is processing too many tasks, utilization is high which in turn leads to low SA (as computed by the model). The low utilization translates into increased penalties which in turn lead to fewer tasks processed by the operator. When the SA model is excluded from the MUV-DES model, the lack of the regulation effect leads to utilization over-estimates.

While these results with homogeneous data are encouraging, they cannot be extended to the heterogeneous case without further validation. The next step was to extend the results observed using the historical data set by presenting validation results carried out using data from the heterogeneous UV simulator (RESCHU).

4.8.2 Validating shape of WTSA/UT curve using the RESCHU data set

In the previous section, the historical data set did not allow for the WTSA/UT curve to be constructed from the experimental data, as there were no reaction time samples that could be attributed to loss of operator SA (since the experiment was designed for other purposes). The analysis was therefore conducted using the hypothetical parabolic WTSA/UT curve that was introduced in Chapter 3. On the other hand, the RESCHU experiment was designed so that reaction times to threat area intersections with vehicle paths could be used as measure of WTSA. The RESCHU data set is therefore used to construct the WTSA/UT curve and further validate the proposed parabolic shape.

Using the RESCHU data set, WTSA was measured as the time from an emergent threat area intersection with a vehicle's path to the time when the participant responded to this intersection. The response to emergent threat areas was chosen to measure WTSA since avoiding threat areas was the highest priority task for the participants, and it required decisive and identifiable actions, and thus was a performance-based SA measure (Pritchett & Hansman, 2000). In addition, using the response to threat intersections to derive WTSA samples was generalizable across the three heterogeneity conditions tested in the experiment. While there were other possible sources of WTSA such as delays in engaging the search task, only the failure to proactively reroute UVs around the threat areas was used because WTSA could be measured separately from other vehicle wait times. For example, the delays in engaging the search task could be a result of lack of SA (i.e., WTSA), but also a result of operator purposefully choosing to do something else such as re-planning paths for higher priority vehicles (therefore introducing wait times due to queuing).

The next step was to associate an average WTSA for different utilization values. For each subject, four utilization values were calculated for 2.5 minutes time windows of the 10 minute experiment. For each time window, average WTSA was calculated per subject using the reaction time samples that fell within that window. The WTSA/utilization samples were then associated with a specific 10% utilization bin. Because there were not enough data points at lower and higher utilization levels, WTSA estimates are shown only for those bins where enough data was available.

The average WTSA for different values of utilization bins are presented in Figures 4.21, 4.22 and 4.23 for the increasing levels of heterogeneity. Due to missing data, the number and spread of utilization bins for each condition differed. In the case of the no-heterogeneity condition, only 4 bins had enough samples, all at higher utilization values. This is due to the high operational tempo in that condition. For the other two conditions, a larger number of bins could be presented due to the more modest tempo.

Repeated measures Analysis of Variance (ANOVA) revealed that there was a significant difference between different utilization intervals for the no-heterogeneity condition ($F(3,110)=2.54$, $p=0.0598$), the medium-heterogeneity condition ($F(5,119)=2.43$, $p=0.0392$), and high-heterogeneity condition ($F(5,121)=8.11$, $p<0.0001$). Pair-wise comparison statistical results are presented in Appendix I.

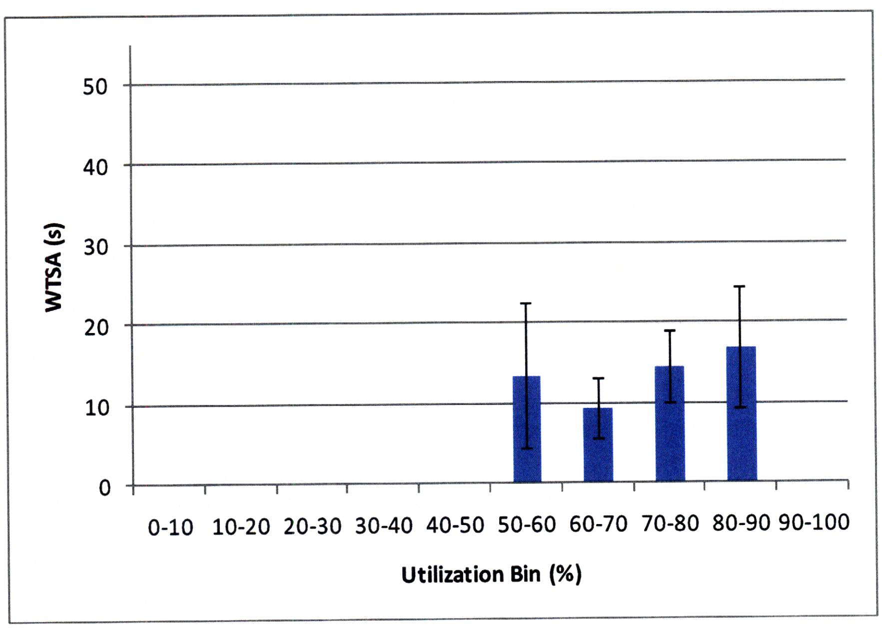


Figure 4.21 WTSA/UT relation with standard error bars for the no-heterogeneity condition

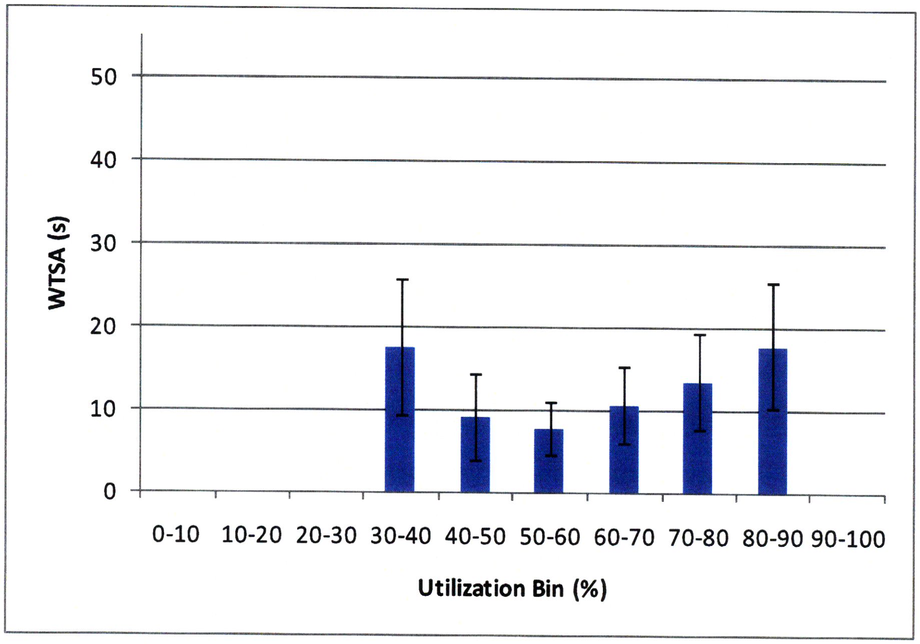


Figure 4.22 WTSA/UT relation with standard error bars for the medium-heterogeneity condition

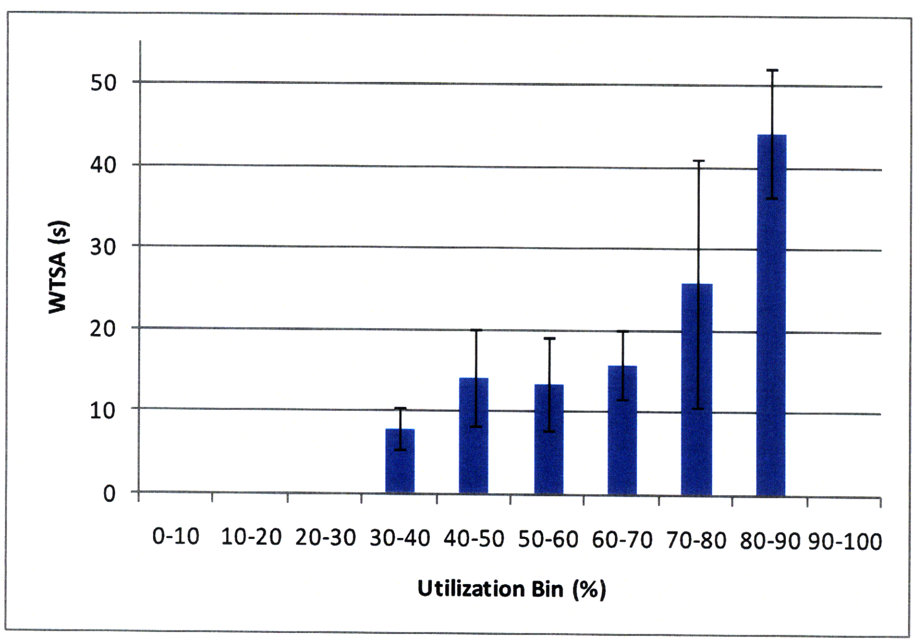


Figure 4.23 WTSA/UT relation with standard error bars for the high-heterogeneity condition

In the no-heterogeneity case, pair-wise comparisons showed that both the 80-90% and 70-80% utilization bins resulted in significantly longer WTSA than the 60-70% utilization bin. In the medium-heterogeneity case, pair-wise comparisons showed that the 80-90% utilization bin resulted in longer WTSA than the 60-70%, 50-60%, and 40-50% utilization bins. In addition, the 30-40% utilization bin also resulted in longer WTSA than the 60-70%, 50-60%, and 40-50% utilization bins. Finally, in the high-heterogeneity case, pair-wise comparisons showed that the 80-90% utilization bin resulted in longer WTSA than the 60-70%, 50-60%, and 40-50% utilization bins. In addition, the 60-70% utilization bin resulted in longer WTSA than the 30-40% utilization bin.

These results demonstrate that WTSA is longer at higher utilization levels than at medium utilization levels, consistent with the predicted curve. On the other hand, the medium and high-heterogeneity conditions contradict in terms of supporting the hypothesized curve at low utilization values. While the medium condition is in agreement with the hypothesized curve in Figure 3.4 in that lower utilization values have higher WTSA than medium utilization values, the high-heterogeneity case resulted in the reverse trend. As most of the data points (for all three conditions) fell towards higher utilization values and the 10-minute experiment was not designed with any significant latencies in required actions, building further confidence in the functional form of the WTSA to utilization relationship at lower utilization values is left for future work. Nonetheless, as hypothesized, the convexity of the WTSA/UT curve and the detrimental impact of higher utilization in terms of operator reaction time are in agreement with the hypothesized relationship.

In order to investigate the detrimental aspect of higher utilization, the experimental data from all three vehicle heterogeneity levels were combined and analyzed to also assess the associations between the utilization intervals and the mission performance efficiency (i.e., score). A mixed linear regression model was developed by backward selection technique to remove insignificant regression terms. After controlling for vehicle heterogeneity level ($F(2,71)=8.37$, $p=0.0005$), time window ($F(3,208)=24.30$, $p<.0001$), and vehicle heterogeneity level/time window interaction ($F(6,208)=2.08$, $p=0.0566$), utilization was identified to be significantly associated with score ($F(5,126)=4.21$, $p=0.0014$). The estimated score means for different utilization intervals are shown in Figure 4.24.

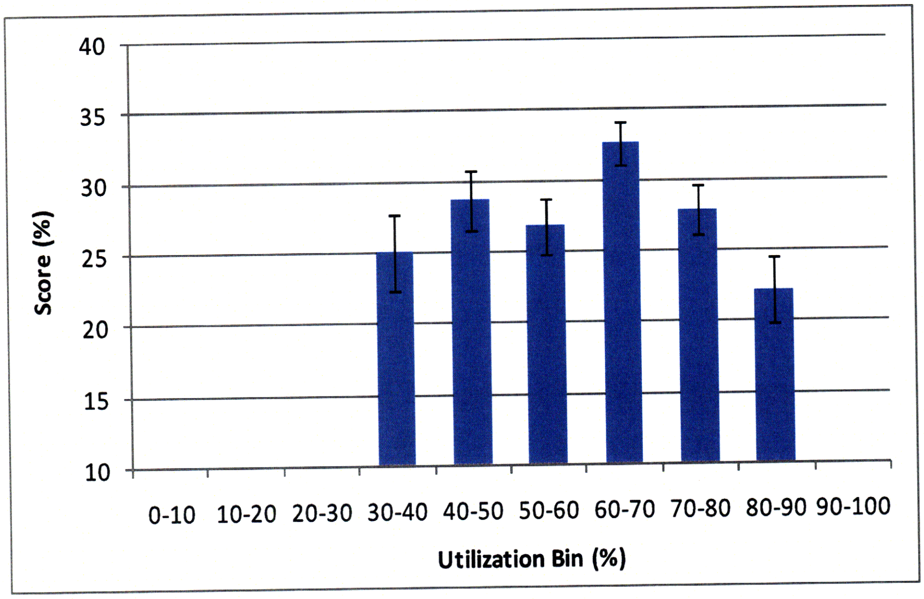


Figure 4.24 Experimental results for score vs. utilization with standard error bars (all conditions)

Pair-wise comparisons revealed that 80-90% utilization corresponded to significantly lower scores than 70-80%, 60-70%, and 40-50% utilization. 60-70% utilization also resulted in higher score than 30-40%, and 70-80% utilization (the rest of the statistical results are presented in Appendix I).

4.9 Summary

The validation process conducted in this chapter resulted in a) confidence being achieved in the MUV-DES model's accuracy, and b) the identification of the robustness of the MUV-DES model to changing inputs, as summarized below:

4.9.1 Model Accuracy

First the MUV-DES model was used to replicate human-UV behavior for a heterogeneous unmanned vehicle system. The results showed that the model is able to capture vehicle, operator, and system performance as the level of heterogeneity is varied. The model was able to capture the impact on operator utilization and vehicle wait times of increasing heterogeneity, which included varied neglect times and interaction periods. Moreover, the model was able to replicate the detrimental impact of heterogeneity in service times on system performance. In addition, as an external validation, the MUV-DES model was also used to replicate system behavior for a historical data set. This validation showed that the model was accurate for larger team sizes and higher autonomy, which represent the types of teams for which the model was constructed. Therefore, the results showed that with sufficient data to estimate the distributions necessary to drive the MUV-DES model, accurate replications of observed human-UV behavior can be replicated for vehicle, operator and mission metrics.

Second, the model's ability to accurately predict the impact of changing vehicle team heterogeneity was validated. The model was able to accurately predict the impact of a team structure design change that led to a decrease in heterogeneity. However, when the model was used to predict the impact of increasing heterogeneity, the increased complexity of the resultant system resulted in the model predictions being less accurate. This was attributed to the human-based variables (switching strategy and service times) changing in response to the vehicle team structure change. It is therefore concluded that in order for the model to result in accurate predictions, the MUV-DES model must account for interaction between design variables.

Finally, confidence was gained in the hypothetical WTSA/UT curve, which is a significant component of the MUV-DES model. The results showed that the model's accuracy improves when including the WTSA/UT curve in the MUV-DES model. This re-emphasizes the experimental observations made by Cummings and Mitchell (2008) about the significance of WTSA in terms of limiting operator performance when controlling multiple UVs. Constructing the curve using WTSA/utilization samples showed that the hypothesized impact of increased utilization in terms of decreased SA and performance was accurate. At lower utilization levels, the results showed that the shape of the curve might depend on other parameters in addition to operator utilization. Finally, the impact of heterogeneity in service times, which was hypothesized to be detrimental in terms of WTQs, seemed to also impact

WTSAs by resulting in a WTSA/UT curve where average WTSA was larger at higher utilization levels.

4.9.2 Model Robustness

In addition to testing the model's accuracy for replication and prediction, the validation techniques applied in this chapter evaluated the model's robustness to changing input conditions.

First, an extreme condition test was applied by driving the number of vehicles to large numbers. Operator utilization was measured using the MUV-DES model by accounting for wait times due to interaction for all event types, something the analysis of the data set was not capable of doing. For larger team sizes, the model behaved as expected by predicting operator utilization that approached 100%.

Second, the internal model stochasticity was tested in order to ensure that the model is consistent in its output behavior from one replication to another. The test showed that the MUV-DES model is consistent and that data variance does not exceed that observed in the experimental data set. The test revealed the model's underestimation of wait time variance, which was explained by noting that the model takes as an input a fixed switching strategy and mean WTSAs for each utilization level.

Finally, a sensitivity analysis was conducted in order to ensure that the model is not overly sensitive to errors in input estimates. The model was generally found to be robust to errors in the input distributions, with the wait time output being the most sensitive, especially in the case of errors in the service time distribution.

Having built confidence in the MUV-DES model's accuracy and robustness, the next chapter, Chapter 5, will explore the model's ability to support designers of heterogeneous unmanned systems.

5 MODEL SYNTHESIS

As was discussed in the introduction, in terms of the requirements, design, and evaluation loop, heterogeneous futuristic systems present a wide variety of choices to designers of unmanned vehicle teams, including the associated interfaces. These different design choices can influence the ability of operators to supervise the teams effectively (discussed in Chapter 1). Therefore, in building effective UV systems, the collective impact of different design choices must be taken into consideration to ensure effectiveness of operator supervision.

Given the limitations of human-in-the-loop experimentation for guiding design choices in futuristic heterogeneous unmanned systems, the MUV-DES model can be used to provide rapid prototyping evaluation capabilities. The model is able to describe the behavior of the system in question, given a specific set of input conditions, as well as predict the effects of changes in one or more system design variables on the behavior of the system. The accuracy of the model in regard to these two benefits was validated in Chapter 4 for the domain of supervisory control of multiple heterogeneous unmanned vehicles.

An accurate model with the above properties can help designers of heterogeneous UV systems in different contexts. Three areas are discussed in this chapter. First, the model can be used to indicate how potential design modifications to an existing system will affect overall performance. Second, designers can use the MUV-DES model when designing futuristic systems to construct hypotheses and understand limitations of future technology in terms of defining system boundaries. Finally, the model can be used for existing systems when the desire is to replicate current observed behavior, in order to help diagnose the cause of failures and inefficiencies. This chapter will provide examples for these three situations.

5.1 Potential Design Modifications for Existing Systems

Designers of existing systems might wish to evaluate the potential of design modifications to improve vehicle, operator, and/or system performance. Designers would likely be concerned with identifying those potential design modifications that are worth investigating further in costly developmental and evaluation phases. As an example, this section considers the interface associated with the RESCHU test bed that was used to validate the model in Chapter 4. In the RESCHU experiment, three types of teams were used as experimental treatments. For the purposes of this illustrative case, it is assumed that the RESCHU interface is an accurate representation of an actual interface in a real system, and that all three team configurations (no-, medium-, and high-heterogeneity) are in operational use. Designers might be interested in evaluating potential design modifications that could lead to improved performance. For the sake of abbreviation, T1 represents the no-heterogeneity team, T2 represents the medium-heterogeneity team and T3 represents the high-heterogeneity team. These abbreviations, as well as the full team descriptors, are used interchangeably in the rest of this section. Two different use cases are presented; the impact of operational design changes, and the impact of developmental design changes.

5.1.1 The Impact of Operational Design Changes

In this section, a simple, single variable design change is investigated by considering the number of vehicles in the team. Such an analysis is important for system designers who might be required to consider the impact of alternate vehicle-resource allocation schemes. The analysis exemplifies the use of the model in capturing the impact of operational design changes.

5.1.1.1 Design Variables and Parameters

Changing the team structure is a problem of resource allocation that does not require any developmental design changes. However, it does assume that pre-mission addition of vehicles to a team is possible through the availability of vehicle assets and system capability to handle the additional vehicle(s), and that the removal of vehicles is also possible through the availability of storage.

Although there are several vehicle types, the number of which could be varied (MALE UAVs, UUVs, and HALEs), for the sake of this analysis, team structure design changes were restricted to the number of MALE UAVs (*number_of_uavs).* Both HALE vehicles and UUVs are generally limited in availability, and play critical roles in a heterogeneous team setting that cannot be replicated by MALE UAVs. For example, removing a HALE or UUV from a team is generally not an option if the mission specifies that such imagery is needed. The number of UUVs and HALEs as well as the rest of the variables from Table 3.2 (including the operator switching strategy, service times, and situational awareness) are parameters in this study. The values assigned to each of the parameters are those derived for the original models for T1, T2 and T3, as described in Chapter 4.

In addition to identifying the variables that should be manipulated for a specific design objective, a range should be defined for each of the design variables in order to proceed with a design space exploration. Depending on the design modifications that are feasible, different ranges are possible. For this illustrative case, it will be assumed that a total of 7 MALE UAVs are available for each team type.

5.1.1.2 Design Space Exploration using Simulation

In order to proceed with a design space exploration, objective functions need to be defined. For all three teams, two main objective functions are a) maximizing mission performance, and b) minimizing operational costs.

The first objective function, maximizing mission performance, is accomplished by maximizing the total number of objects correctly identified. Because the RESCHU scenarios are surveillance-type missions, the goal is to detect and identify as many objects as possible. In analyzing the RESCHU data set in Chapter 4, the mission performance metric used was that of the total number of objects correctly identified normalized by the total number possible for the team configuration under consideration. This scoring function was selected in order to allow performance comparisons to be made across different mission types. Since in Section 5.1, the desire is not to compare performance across the different mission types but to improve the performance for each individually, the normalization factor is not considered.

The second objective function, maximizing vehicle performance, is achieved by minimizing vehicle wait times. In addition, given the results from previous studies that show significant performance degradation when operators work beyond 70% utilization (Cummings &

Guerlain, 2007; Schmidt, 1978), a constraint was identified which required operator utilization not to exceed 70%. Finally, a design vector whose manipulation is desirable in light of the design objectives is presented. These are summarized below:

$$\min \boldsymbol{J}(\boldsymbol{x})$$

$$\text{s.t.}\, \boldsymbol{g}(\boldsymbol{x}) \leq 0$$

$$where\, \boldsymbol{J}(\boldsymbol{x}) = \begin{pmatrix} J_1(\boldsymbol{x}) \\ J_2(\boldsymbol{x}) \end{pmatrix} = \begin{pmatrix} avg\_search\_task\_wait\_time \\ -\, number\_of\_objects\_correctly\_identified \end{pmatrix},$$

$$and\ \boldsymbol{g}(\boldsymbol{x}) = (g_1(\boldsymbol{x})) = (operator\_utilization\ - 70\%),$$

$$\boldsymbol{x} = \ (x_1) = (number\_of\_uavs)$$

The next step is to determine the effectiveness of manipulating the *number_of_uavs* variable in achieving the objective functions and satisfying the constraints. Each of the seven experimental design vectors, which consist of different levels for the *number_of_uavs* variable, were input into the MUV-DES model. The different design variable settings were accounted for by modifying the team structure variable (i.e. changing the number of MALE UAVs). Five thousand trials were conducted using the MUV-DES model in order to calculate *avg_search_task_wait_time* and *number_of_objects_correctly_identified*. The objectives calculated from the experiments for T1 are shown in Figure 5.1 for all the feasible design vectors (all vectors except for 6 and 7 vehicles were feasible in that they satisfied the utilization constraint). The results for T2 and T3 are shown in Appendix J.

Manipulating the *number_of_uavs* variable results in the formation of a tradeoff curve between the two objective functions. For a team size of one vehicle, *avg_search_task_wait_time* and *number_of_objects_correctly_identified* are both at a minimum. As the number of vehicles is increased, both outputs increased in value. Therefore, the two outputs are competing objective functions with respect to changing *number_of_uavs*. It can also be noted from Figure 5.1 that as *number_of_uavs* is varied, the rate of the change in *number_of_objects_correctly_identified* is different than the rate of change in *avg_search_task_wait_time*. With smaller team sizes, the effect of changing *number_of_uavs* has a much larger effect on *number_of_objects_correctly_identified* than *avg_search_task_wait_time.* However at about 3 vehicles, the trend is reversed. For example, in going from 6 to 7 vehicles, *avg_search_task_wait_time* increased by 32% while *number_of_objects_correctly_identified* increased by only 2.5%.

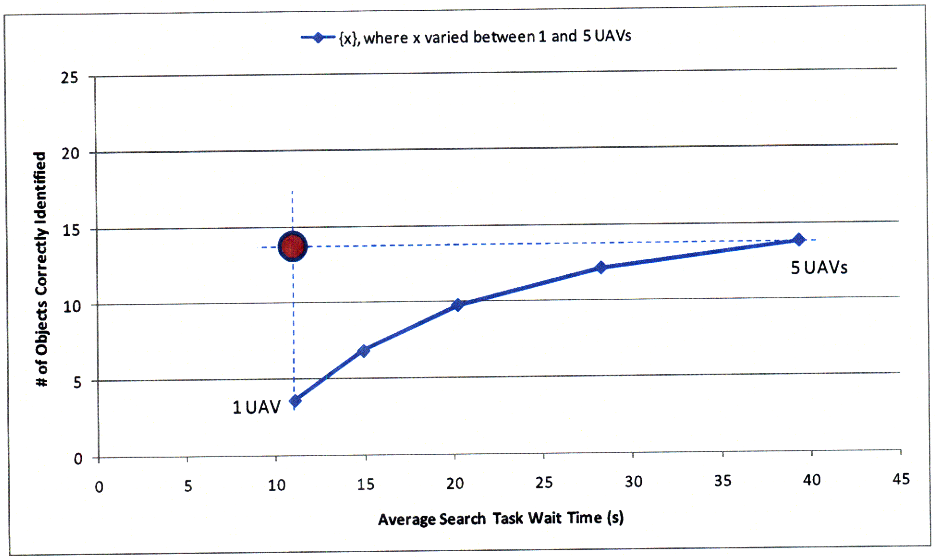


Figure 5.1 Tradeoff curve formed as the number of UAVs was varied for the no-heterogeneity condition

5.1.1.3 Design Recommendations

The next step would be to select a design vector that results in an optimization of the objective functions. Because this is a multi-objective problem, estimation of the Pareto front using a process such as the Weighted Sum method (Zadeh, 1963), is required. With the Weighted Sum method, a new objective is formed that is composed of the weighted sum of each objective function as shown in Equation 1. The scaling factors sf_1 and sf_2 were chosen by selecting the largest possible *number_of_objects_correctly_identified* and largest *avg_search_task_wait_time.*

$$\min J_o = \lambda_1 * \frac{-ObjectsIdentified}{sf_1} + \lambda_2 * \frac{WaitTime}{sf_2}, s.t. \sum_{i=1}^{2} \lambda_i = 1 \qquad (1)$$

In this case, where only the *number_of_uavs* is varied, all the design points lie on the Pareto front. Since a Pareto front is already identified, the next step is to select a point along the Pareto front that results in the best trade-off between the two objectives by identifying λ_1 and

λ_2. It will be assumed that both objectives are important to the designer with a slight emphasis attributed to the *number_of_objects_correctly_identified*. Therefore, $\lambda_1 = 0.6$ and $\lambda_2 = 0.4$ were defined as appropriate for this optimization.

Tables 5.1, 5.2, and 5.3 display for each feasible design vector for each team (T1- T3), the corresponding *number_of_objects_correctly_identified*, *avg_search_task_wait_time*, and J_o when $\lambda_1 = 0.6$ and $\lambda_2 = 0.4$. The optimal design choice in each case is highlighted in grey. In the case of T1, the optimal design setting is 4 MALE UAVs. Since 5 MALE UAVs was the original setting for T1, the design recommendation is to decrease the number of MALE UAVs by one. In the case of T2, the optimal design setting is 3 MALE UAVs. Since 3 MALE UAVs was the original setting for T2, the recommendation is to refrain from making any changes to the team structure. Finally, in the case of T3, the optimal number of MALE UAVs is also 3. Since 2 MALE UAVs was the original setting for T3, the final recommendation is to add an additional MALE UAV to T3.

Table 5.1 Output values corresponding to each feasible design (T1)

number_of_uavs	*number_of_objects _correctly_identified*	*avg_search_task _wait_time*	J_o
1	4	11.0	-0.04379
2	7	14.9	-0.14446
3	10	20.2	-0.21928
4	**12**	**28.3**	**-0.23965**
5	14	39.4	-0.2

Table 5.2 Output values corresponding to each feasible design (T2)

number_of_uavs	*number_of_objects _correctly_identified*	*avg_search_task _wait_time*	J_o
1	8	17.4	-0.18621
2	10	23.9	-0.24603
3	**12**	**33.7**	**-0.25157**
4	14	47.1	-0.2

Table 5.3 Output values corresponding to each feasible design (T3)

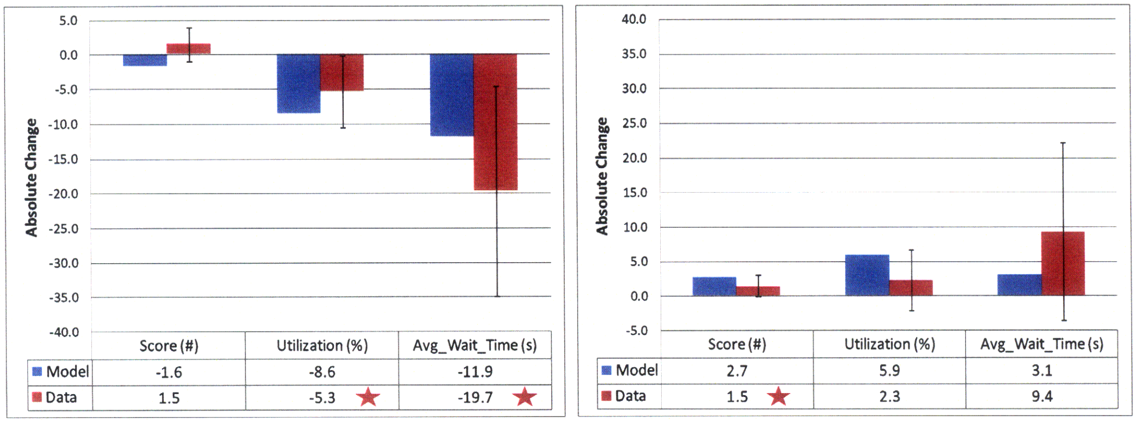
number_of_uavs	*number_of_objects _correctly_identified*	*avg_search_task _wait_time*	J_o
1	7	16.4	-0.18918
2	9	23.6	-0.24353
3	**11**	**32.6**	**-0.24615**
4	12	43.9	-0.2

5.1.1.4 Validating Design Recommendations

It was next decided to experimentally validate the impact of the team structure design recommendations suggested for T1 and T3 as an additional validation check. In order to determine the accuracy of the DES model's predictions, both the magnitude and the direction of the predicted changes were compared with experimental data.

The design recommendations for team structure were implemented in RESCHU by removing one MALE UAV from T1 and adding one MALE UAV to T3. In addition, the number of AOIs was increased by one for the scenario associated with T3 in order to keep the inter-arrival times constant between AOIs, after accounting for the additional vehicle in the team. Following these changes, a human-in-the-loop study was conducted. Fourteen subjects participated in an experiment involving the control of T1 and seventeen subjects participated in controlling T3. The experimental setup was the same as that in Chapter 4.

An analysis of variance (ANOVA) suggested that in the case of T1, the change in team structure resulted in a non-significant increase in the number of objects correctly identified of 1.5 objects ($F(1,38)=1.57$, $p=0.218$), a significant utilization reduction of 5.3% ($F(1,38)=4.275$, $p=0.046$), and a significant wait time reduction of 19.7 seconds ($F(1,38)=6.679$, $p=0.014$). In the case of T3, the ANOVA revealed a marginally significant improvement of 1.5 objects correctly identified ($F(1,38)=4$, $p=0.053$), a non-significant utilization increase of 2.35% ($F(1,38)=1.181$, $p=0.284$), and a non-significant wait time increase of 9.4 seconds ($F(1,38)=2.187$, $p=0.147$). Descriptive statistics are presented in Appendix F. The MUV-DES model predictions as well as the experimental results for the impact of the team structure design change are presented in Figure 5.2.



(a) (b)

Figure 5.2 Predictions using the MUV-DES model and experimental results for a) T1 and b) T3 team structure changes

In the case of T1, the significant changes in operator utilization (-5.3%), and wait times (-19.7s) are effectively captured using the model predictions (Figure 5.2a). The slight over-estimation of the magnitude of the operator utilization decrement can be attributed to the model not accounting for the lower inter-arrival times between search task events due to the removal of one UAV from the team. The lower inter-arrival times caused higher-than-expected operator utilization (and therefore a smaller magnitude change) and higher–than-expected number of objects collected in the experimental results. Nonetheless, the predicted changes for operator utilization and search task wait times are within the 95% confidence intervals of the mean differences.

In the case of T3, the model led to a design recommendation with the intention of increasing the number of objects correctly identified without affecting the rest of the outputs in a considerable manner. This was reflected in the marginally significant increase in the number of objects correctly identified (score) and the non-significant changes in operator utilization and average search task wait times (Figure 5.2b). The non-significant changes were captured using the MUV-DES model in the small change predictions made which are also within the 95% confidence intervals of the mean differences.

In summary, the experimental results confirmed that the MUV-DES model was accurate in predicting benefits as a result of decreasing T1's vehicle team size by 1 vehicle and increasing that of T3 by 1 vehicle.

5.1.2 Developmental Design Recommendations

In this section, a more complex multi-variate design change that evaluates the impact of changing interface elements on the ground control station (GCS) is presented. In addition to investigating changes in the service times and switching strategies, which are directly affected by the GCS interface design, operator reaction time to exogenous events is considered. For RESCHU, subject feedback to question 6 in the post-experimental questionnaire (which asked subjects if the system should have any additional capabilities) revealed that subjects desired a more salient notification of vehicles arriving to AOIs. Providing some kind of aural alert in response to this need would be the appropriate design intervention, which would likely result in reduced reaction times to these exogenous events. The following analysis exemplifies the use of the MUV-DES model in capturing the impact of developmental design changes such as the introduction of an aural alert.

5.1.2.1 Design Variables and Parameters

In this section, the operator switching strategy, the service times for the search task event, and reaction times to the search task event will be considered as design variables in addition to the *number_of_uavs*.

- *switching_strategy*: The operator switching strategy, which can be manipulated by providing decision support that suggests to the operator a certain ordering for servicing the different event types.

- *search_task_service_time*: The service times for the search task, which can be reduced by using intelligence data and/or automated decision support to guide the operator in locating the objects in the camera window.

- *reaction_time*: The reaction time for search task events (i.e. vehicles arriving to AOIs), which can be reduced through the introduction of alerts that improve operator situational awareness (SA), therefore reducing reaction times.

In summary, the design variables consist of the number of UAVs in the team and the three interface-related variables. The rest of the variables in Table 3.2, including the level of autonomy (neglect times), are assumed constant for the sake of this analysis, thus part of the parameter vector. Although the level of autonomy is not investigated in this analysis, this model, when used in a futuristic what-if scenario, could be used to derive autonomy requirements (this will be exemplified in Section 5.2). The design variable vector is therefore as follows:

$$\boldsymbol{x} = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} = \begin{pmatrix} number\_of\_uavs \\ switching\_strategy \\ search\_task\_service\_time \\ reaction\_time \end{pmatrix}$$

For this illustrative case, it is assumed that threat area events are still of the highest priority, and that *switching_strategy* can vary between a prioritization of goal re-plans over search task events and vice versa (i.e., the queuing policy variable from Table 3.2). As for *search_task_service_time*, it is assumed that this can vary between the base case, which is that observed for the existing RESCHU system, and an average service time that is, for this example, 5 seconds shorter. Such a reduction in search task service times could be realized by a new technology, such as an automation-aided target recognition system. Finally, it will be assumed that *reaction_time* can vary between the base case, which is that observed for the existing RESCHU system, and an average reaction time that is 5 seconds shorter as a result of the added aural alert discussed previously. As in the previous section, the *number_of_uavs* will still be assumed to vary between 1 and 7 vehicles. The design variables are summarized in Table 5.4, with associated units and ranges

.

Table 5.4 Design variables, units, and ranges for developmental design space exploration

Design Variable	Unit	Range
number_of_uavs	vehicles	1-7
switching_strategy	priority order	{re-plan over visual, visual over re-plan}
search_task_service_time	seconds	{base case, base case – 5s}
reaction_time	seconds	{base case, base case – 5s}

5.1.2.2 Design Space Exploration using Simulation

The next step is to determine the effectiveness of manipulating the different variables in achieving the objective functions and satisfying the constraints. It is desirable to investigate the collective impact of multiple design changes together as a set, as the manipulation of a single variable might not be sufficient for satisfying all the objective functions. The same objective functions as those identified in Section 5.1.1.2 are utilized in this section.

In order to explore the domain space of feasible designs, a full factorial study was used to test the four design variables at each of the possible values with fifty-six experiments, an approach that would be prohibitively costly in an actual system. For all three team configurations, each of the fifty-six experimental design vectors was input into the MUV-DES model. In each experiment, five thousand trials were conducted using the MUV-DES model in order to calculate *avg_search_task_wait_time* and *number_of_objects_correctly_identified*.

The objective functions (*number_of_objects_correctly_identified* and *avg_search_task_wait_time*) calculated from the experiments with T1 are shown in Figure 5.3. Since simulations using T2 and T3 resulted in similar output behavior, and because the aim of this section is to exemplify model application, the rest of this section will focus on the results observed for just T1 (the figures for T2 and T3 are shown in Appendix J).

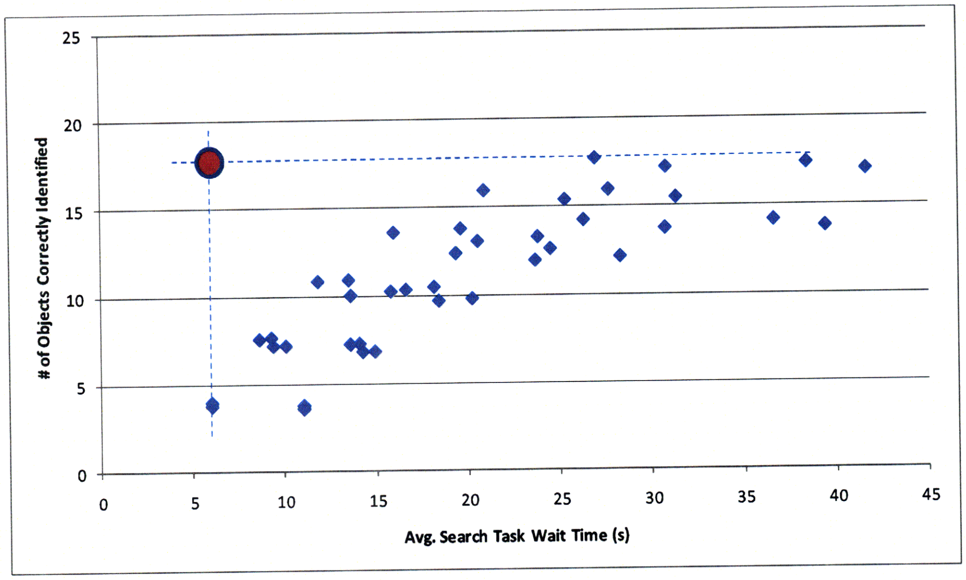


Figure 5.3 Design space exploration for the no-heterogeneity condition (outputs correponding to feasible designs shown)

Since the utopia point is at the top left hand corner of Figure 5.3, there is a tradeoff relationship between the two objective functions. In order to better understand the cause of this relationship, the impact of varying the different design variables needs to be dissected. It is therefore decided next to look at the effects of each variable individually, in order to shed more light on which design decisions should be pursued.

5.1.2.3 Analysis of Main Effects

Figure 5.4 shows a select set of data points from Figure 5.3 with data points grouped into four curves. Each curve corresponds to output values where the only variable in the design vector varied is *number_of_uavs* (i.e. the rest of the design vector consisting of the three interface-related variables is held constant). The base curve consists of points where the *switching_strategy* variable is assigned the replan-over-visual level, and the *search_task_service_time* and *reaction_time* variables are assigned the base-case levels from Table 5.4; other curves will be compared to this base curve. Each of the other three curves varies just one of the interface-related variables from the vector associated with the base curve to their 2nd level value from Table 5.4.

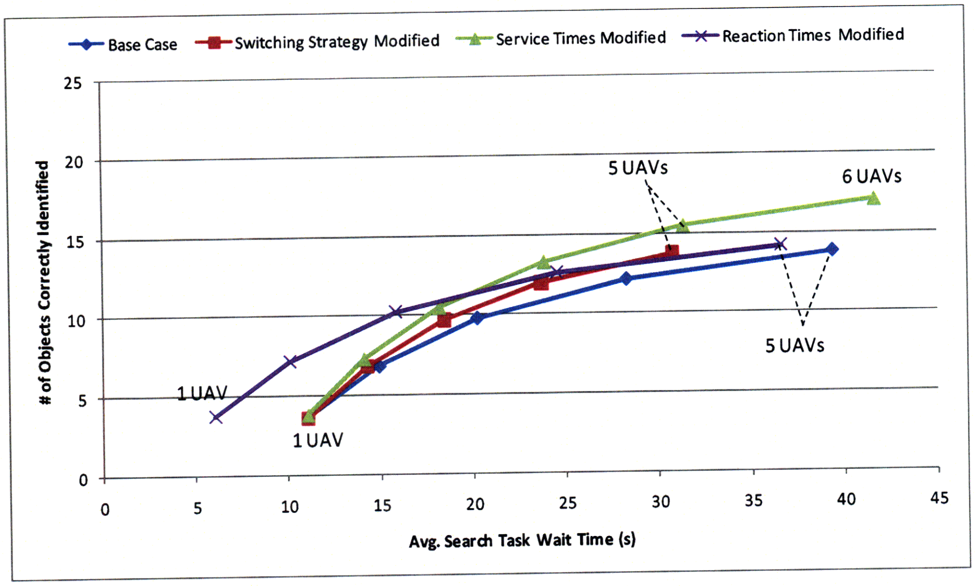


Figure 5.4 Tradeoff curve shift for independent variation of the switching strategy, interaction times, and penalties due to loss of SA

Figure 5.4 demonstrates that manipulating the interface-related variables leads to a shift in the tradeoff curve formed by varying the *number_of_uavs*. Changing the switching strategy, reducing service times, and reducing reaction times all lead to a shift in the tradeoff curve, such that the curve gets closer to the utopia point. However, the exact magnitude of the shift with respect to each of the axes differs for each of the different design changes as summarized in Table 5.5.

Table 5.5 Impact of modifying the three interface-related variables for developmental design space exploration

	Impact on *number_of_ objects_correctly_identified*	**Impact on *avg_search_ task_wait_time***
switching_strategy	Minimal impact	Reduced at larger team sizes
search_task_service_time	Increased at larger team sizes	Reduced at larger team sizes
reaction_time	Minimal impact	Reduced at small team sizes

Comparing the switching strategy modification curve to the base curve reveals that the effect was mainly a reduction in the *avg_search_task_wait_time* objective, while the impact on the *number_of_objects_correctly_identified* is minimal (as evidenced by the purely leftward shift of the curve). Similarly, comparing the service time modification curve to the base curve also reveals a reduction in *avg_search_task_wait_time*. However, when reducing service times, added benefits include an increase in the *number_of_objects_correctly_identified* (as evidenced by the upward shift of the curve, in addition to left shift), and an additional feasible solution for a team size of 6 vehicles (due to a reduction in operator utilization). In addition, in both the switching strategy modification and the reduction in service times cases, the magnitude of the impact on *avg_search_task_wait_time* and *number_of_objects_correctly_identified* depends on the size of the team. The impact is greater when the team size is larger.

Finally, comparing the curve associated with a reaction time reduction to the base curve reveals a reduction in *avg_search_task_wait_time* for smaller team sizes. The reason is that reducing reaction times to search task events impacts wait times due to loss of SA (WTSA), which is only one dimension of *avg_search_task_wait_time* (the other dimension was WTQ). When controlling a small number of vehicles, WTSA is a dominant component of *avg_search_task_wait_time*, as shown in Figure 5.5. The reason is that at small vehicle team sizes, WTSA increases due to the low operator utilization, while WTQ is small due to the low task load. As the team size increases, operator utilization becomes more moderate, thus reducing WTSA but at the same time increasing WTQ. Finally, as the team size continues to grow, both WTSA and WTQ grow in magnitude.

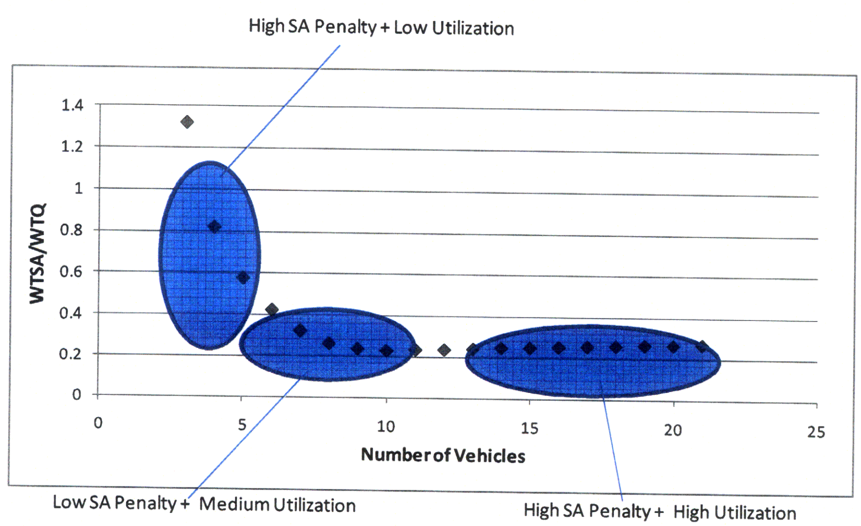


Figure 5.5 Ratio of WTSA to WTQ as the number of vehicles is increased

5.1.2.4 Design Recommendations

Having investigated the impacts and the limits of design changes to the interface-related variables using the previous design space exploration (Table 5.5), the next step is to select a design vector that results in an optimization of the objective functions. Because changing the switching strategy, reducing service times, and reducing reaction times all lead to improvements in both objective functions, an additional cost-based objective function is needed in order to produce a tradeoff relationship.

For example, the risk of each of the design changes in terms of implementation can be considered. One risk in implementing a switching strategy design change is that affecting operator actions is not guaranteed, especially under off-nominal conditions such as high tempo situations or during interruptions. In addition, the performance of a specific switching strategy is not guaranteed under changing environmental conditions. The risk in terms of an automation-aided target recognition algorithm that reduces service times lies in the successful design of the algorithm, which is not trivial. Moreover, although such a design modification is effective due to its ability to improve both objective functions simultaneously, significant reductions in service times are often difficult to achieve because there is often a minimum amount of time needed by the operator to interact with tasks in order to apply human judgment and reasoning. Finally, the risk in reducing reaction times through the introduction of additional alerts that highlight vehicle arrival to targets lies in whether the reaction time reductions will be realized.

By associating numbers to the previously identified risks, a cost-based objective function could be defined. This function could be used to derive a Pareto front using an optimization-based analytic method. Alternatively, designers could take a cost-benefit analysis or probabilistic risk assessment approach. A cost-benefit analysis is especially important when resources (both temporal and monetary) for making design changes are limited, so such an approach can aid in identifying which changes should be pursued and which should be eliminated from consideration. On the other hand, probabilistic risk assessment characterizes risks based on the magnitude of possible adverse consequences and the likelihood of occurrence. Such assessment could be beneficial, especially when designers wish to evaluate the risk of failure in terms of possible vehicle damage or loss.

In summary, the number of design variables and the number of levels considered for each variable was small enough that a full factorial design space exploration could be conducted. This resulted in the identification of a Pareto front. A larger number of variables or levels would require a more limited design space exploration through the use of orthogonal arrays such that only a selection of the design space is explored. Depending on the complexity of the problem at hand, this would be followed by the use of either a gradient-based technique or a heuristic-based technique, such as Genetic Algorithms, Simulated Annealing, or Tabu Search, in order to derive the Pareto fronts from which optimal design points could be selected. An example using Simulated Annealing together with a human-vehicle model can be found in Cummings, Nehme et al. (2007).

5.2 Supporting Requirements Generation for Futuristic Systems

5.2.1 Motivation

In the previous section, the problem of evaluating the impact of potential design modifications on the performance of a system in operation was addressed. When designing futuristic heterogeneous unmanned vehicles systems, designers will be presented with a design space which can be large, especially with advances in technology. Designers might be interested in determining the design limitations imposed by heterogeneity in vehicle capabilities and tasks. Identifying these limitations is critical when generating requirements for futuristic systems in order to define system boundaries. However, due to the size of the design space and the lack of existing system implementations, human-in-the-loop evaluation is extremely difficult.

The MUV-DES model can be used for this purpose by supporting the evaluation of hypothetical scenarios, team architectures, and different interface capabilities. Using the MUV-DES model to identify the limitations imposed by heterogeneity on vehicle team structure designs is exemplified in the next section.

5.2.2 Supporting Requirements Generation Example

An interesting result observed in Chapter 4 and shown more concretely in Section 5.1 is the detrimental impact of heterogeneity in service times on vehicle performance. In Chapter 4, although operator utilization in T3 was lower than T2, T3 resulted in a lower mission score. It was hypothesized in Chapter 4 that this observation was due to the heterogeneity in service times present in T3, which caused large average search task wait times. In Section 5.1.5, the negative impact of service time heterogeneity also existed when comparing the modified T1 team, which is homogeneous, to the modified T3 team, which exhibits heterogeneity in service times. An analysis of variance (ANOVA) suggested that the modified T3 team had a statistically lower operator utilization ($F(1,29)=9.934$, $p=0.004$), and also had a non-significantly higher average wait time ($F(1,29)=0.550$, $p=0.464$). Thus, although operator performance was better with the modified T3 team (through lower utilization), vehicle performance (through avg. wait times) was worse. Again, this can be attributed to the heterogeneity in service times, which is only present in the case of the modified T3 team.

Designing heterogeneous teams may be required due to mission objectives, but given the previously stated results concerning heterogeneity, designers might be concerned with the limitations imposed by heterogeneity on team structure design. The MUV-DES model can be used to evaluate the impact on vehicle wait times of varying degrees of service process heterogeneity, therefore identifying these limitations.

In Table 3.2, team heterogeneity was defined as a situation where in a team of size n, there exists at least one vehicle, a, and an event type, j, where $f_{Z_{aj}}(z) \neq f_{Z_{bj}}(z)$ or $f_{Y_{aj}}(y) \neq f_{Y_{bj}}(y)$ where $1 < a, b < n$ and $a \neq b$. Using the MUV-DES model, a hypothetical homogeneous team of three MALE vehicles was constructed such that initially, all the vehicles had the same arrival and service distributions. The service time heterogeneity of the team was increased by varying just the mean service times for the search task event across the three vehicles, keeping the arrival distributions constant. It order to control for operator task load, the average of the mean service times for all three vehicles was kept constant. This was accomplished by keeping one vehicle's mean service time at the base value, and then increasing that of another vehicle while decreasing by an equal amount that of a third vehicle, therefore keeping the average service time of search task events constant. The random variable modifications can be translated as follows:

$$\text{Vehicle 1: } y' = y + wti$$

$$\text{Vehicle 2: } y' = (1 + e) * y + wti$$

$$\text{Vehicle 3: } y' = (1 - e) * y + wti$$

where e represents the degree of heterogeneity, which is a measure of the spread in the expected values of the service time probability density functions associated with the different vehicles in the system, and was varied between 0% and 100% in 10% increments. y is the realization of the random variable that describes the service time, and *wti* is the realization of the random variable that describes the wait times due to interaction. For each value of e, 5000 trials were conducted using the MUV-DES model in order to determine the mean and variance of the average search task event wait time. The average search task event wait time was considered since this is the dependent variable that was observed to be negatively impacted by service time heterogeneity in the problems that motivated this analysis. Due to the fact that the mean service time of events was maintained constant, operator utilization did not change with increasing heterogeneity. The results are depicted in Figures 5.6 and 5.7.

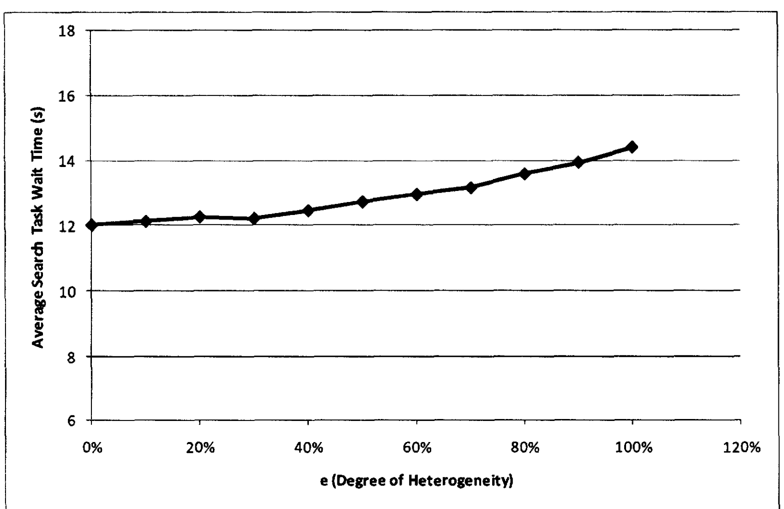


Figure 5.6 Impact of service time heterogeneity on the mean avg. search task wait time

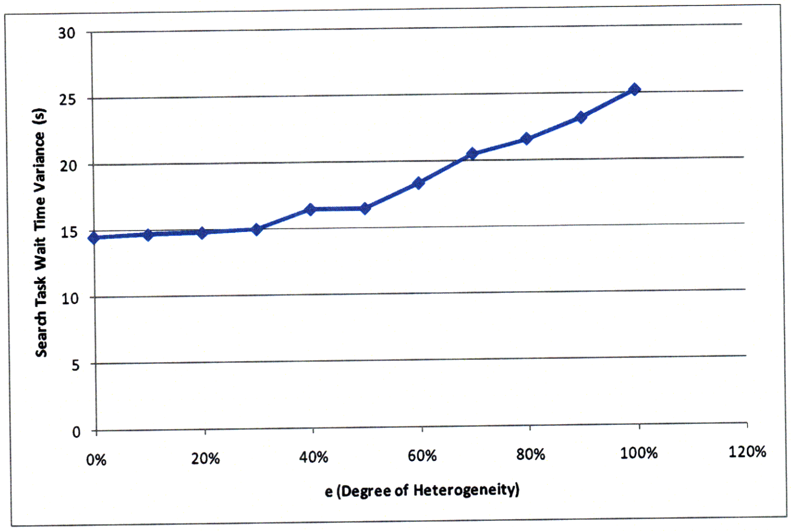


Figure 5.7 Impact of service time heterogeneity on the variance of the avg. search task wait time

Figure 5.6 indicates that mean average search task wait time increases with increasing degree of service time heterogeneity. In addition, the variance exhibited by the average search task wait time also increases with increasing degrees of heterogeneity in service times (Figure 5.7). In both cases, a degree of heterogeneity greater than 40% resulted in the avg. wait time mean and variance departing from the original levels. Designers can use such an analysis in order to identify system boundaries. A 40% degree of heterogeneity is equivalent to a mean service time of 17s, 28s, and 39s for each of the three vehicles. The interpretation is that when the ratio of the vehicle with the longest service time to that of the vehicle with the shortest service time is greater than 2, overall mission performance is negatively impacted as a result of the heterogeneity (independently of operator task load). In light of the discovery of the above limitations, requirements can be generated in order to keep the system within the identified boundary, and when that is not possible, to generate requirements that mitigate the negative impacts of surpassing the boundary.

When requirements limiting the degree of service time heterogeneity are not feasible, such as in missions where vehicles with varied payloads are required, requirements that mitigate the negative impact of service time heterogeneity can be generated. For example, designers can generate requirements that limit operator task load in order to ensure that operator utilization is maintained towards the 60-70% bin, where wait times due to loss of SA are the lowest. Moreover, wait time variance can be mitigated by generating requirements that limit interface flexibility in terms of supporting varied operator attention allocation strategies such that the variability induced by operators is minimized.

5.3 Failure Diagnosis for Existing Systems

5.3.1 Motivation

Existing systems can often be plagued by inefficiencies and sub-optimal performance. There are several techniques for diagnosis of such systems. Some forecast the possibility of failure, and others are reactive, meaning they are employed after a problem has occurred. For example, the technique of root cause analysis attempts to identify the root causes of problems, as opposed to addressing the observable symptoms. Common to most root cause analysis techniques are the following steps: definition of the problem, gathering of data and evidence, identifying causal relationships, and identifying which causes, if removed or changed, will prevent recurrence (Wilson, Dell, & Anderson, 1993). It is this last step where the MUV-DES model can provide significant benefits. Attempting to identify which causes, if removed or changed, can prevent recurrence is costly, lengthy and often dangerous to do with an actual system.

First, the MUV-DES model can be used to replicate current system behavior in order to identify the sources of inefficiency. Because the MUV-DES model is highly structured, it can be used to shed light on the causes behind observed inefficiencies. For example, the ability of the model to differentiate between wait times due to loss of SA (WTSA), wait times due to queuing (WTQ), and wait times due to interaction (WTI) might encourage designers to use the model in order to understand the sources of wait times that are negatively impacting vehicle performance. Second, the MUV-DES model can be used to identify which causes, if changed or removed, can prevent recurrence by predicting the impact of such design changes.

The analysis presented in the subsequent section exemplifies using the MUV-DES model in order to identify the source of the large wait times plaguing the T1 team configuration represented in the RESCHU system.

5.3.2 Diagnostic Example

An interesting phenomenon observed in Chapter 4 is the high average search task event wait time associated with team configuration T1. Given these results, designers might be interested in diagnosing the system in order to understand the source(s) of the inefficiency. The following diagnosis will use a simple procedure of peeling away layers of symptoms in order to identify the root causes of the problem (Figure 5.8).

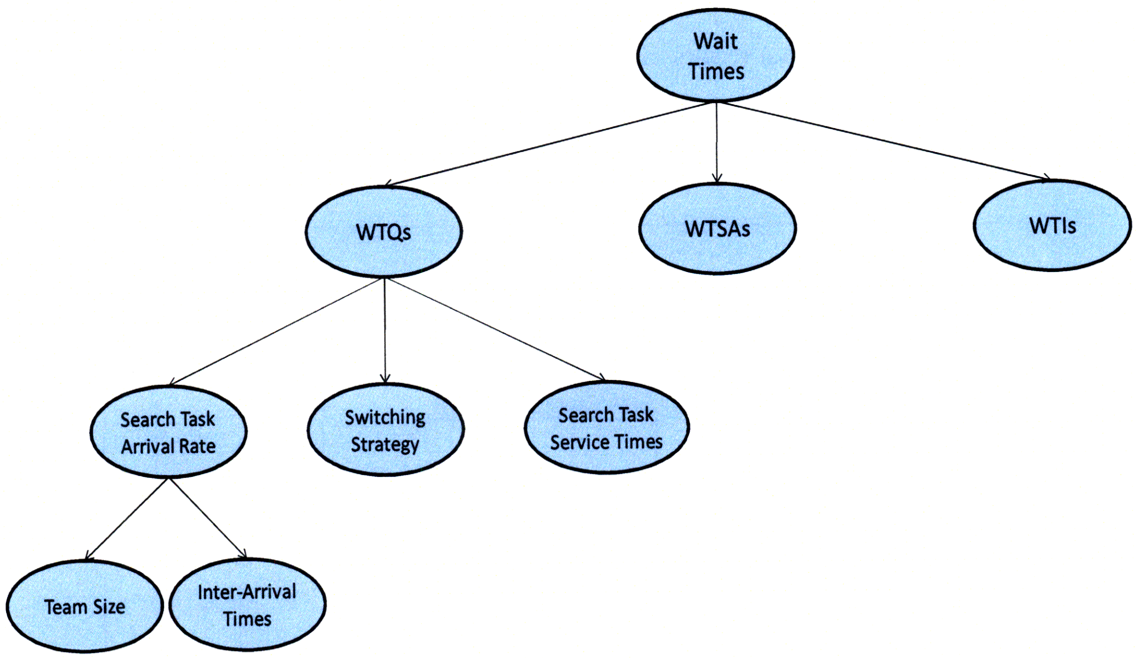


Figure 5.8 Diagnosis tree

The first step is to identify the sources of wait times. As was described in Chapter 2, wait times are composed of WTQs, WTSAs, and WTIs. Because WTIs are included in the search task service time, the focus will be on WTSAs and WTQs. Using the MUV-DES model to run 5000 simulations results in a total average search task wait time of 39.5s, with 28% composed of WTSA and 72% composed of WTQ. Since WTQs form the larger proportion, the rest of the analysis will continue by focusing on the WTQ portion of average search task wait times.

The second step is to identify the sources of WTQs. WTQs are affected by the operator switching strategy, the search task arrival rate, and the search task service times (the threat area arrival rate and service times were not considered since their service times are much smaller than those of the search task event type; approximately 92% smaller).

Manipulating the switching strategy between one where search task events are prioritized and the original modeled strategy, which prioritizes goal re-plan events, results in a 32% decrease in WTQ (from an original value of 28s). Reducing the service times by 50% results in WTQ decreasing by 71%. Since the search task arrival rate is affected by both the inter-arrival times of search task events and the team size, reducing the arrival rate of search task events (by doubling the inter-arrival time) results in WTQ decreasing by 47%. Finally, reducing the team size by one vehicle results in WTQ decreasing by 40%. The results are summarized in Figure 5.9.

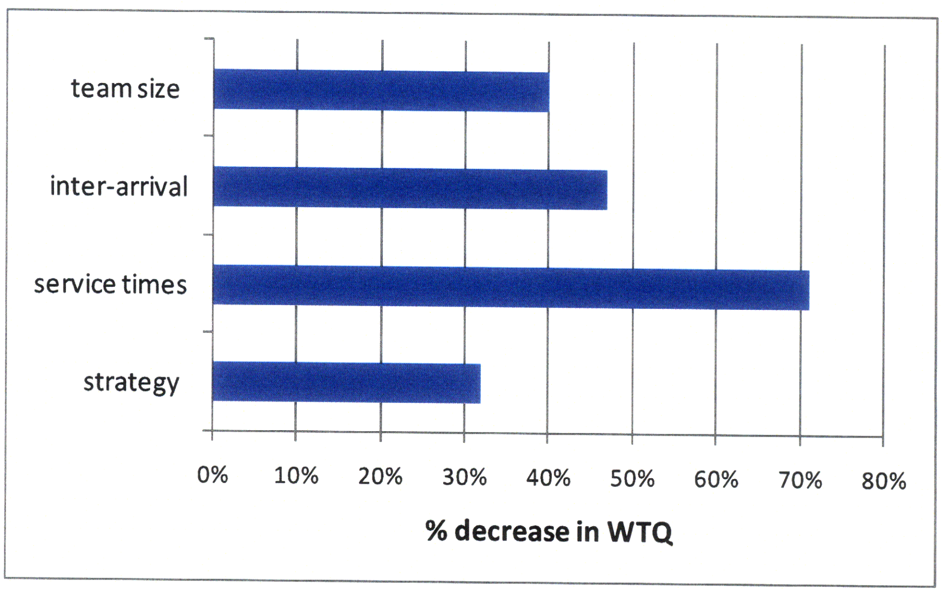


Figure 5.9 Impact of variable changes on queuing wait times

Following this analysis, the impact of each of the root causes on the wait time variable can be compared. In this case, a conclusion can be made that the service times of the search task event is the major cause behind the large wait times observed for T1 (note that additional analysis could entail normalized sensitivities). Understanding the sources of the inefficiencies can allow designers to identify potential design modifications that address sources of detrimental performance. For example, a potential design modification that can be considered is the use of intelligence data and/or automated decision support to guide the operator in the search task event with the aim of reducing the resulting service times. The next step would be to conduct a more detailed analysis, similar to that presented in Section 5.1 that evaluated the impact of potential design modifications with respect to the wait time objective function.

Similar to the diagnosis of wait times presented in this section, the MUV-DES model could be used to generate other relationships in order to understand the reasons behind detrimental operator or system performance. In addition, the MUV-DES model can be used with more complicated diagnosis techniques that take into account the interaction between different causes. Although the use of the MUV-DES model for diagnosis is initially reactive, expertise and continued use of the model can cause it to become more pro-active, and thus able to anticipate failures.

5.4 Summary

The analyses in this chapter illustrated the potential use of the MUV-DES model to guide designers of current and futuristic unmanned systems.

The first application discussed was one where designers were concerned with evaluating the impact of potential design modifications on system performance. Once objective functions, constraints, and design variables were identified, the MUV-DES model was used in the context of a design space exploration in order to identify the performance associated with different design vectors. The analysis was able to shed light on interaction effects between design variables, as well as main effects for each variable. As an additional check, model recommendations based on team size analysis were shown to be accurate using an additional experiment.

The second application discussed was one where designers were concerned with exploring the design space associated with systems of the future in order to determine the design limitations imposed by heterogeneity in vehicle capabilities and tasks. By analyzing the impact of varying degrees of service time heterogeneity, a system boundary in terms of an acceptable degree of heterogeneity was identified. Requirements were then proposed that could keep the system within the identified boundary, as well as those that mitigate the negative impacts of surpassing the boundary.

Finally, the last application presented was one where designers desired to diagnose causes behind system inefficiencies. The MUV-DES model was used in this case to identify which causes, if changed or removed, eliminate the system inefficiencies. Understanding the sources of the inefficiencies can allow designers to identify potential design modifications, both operational and developmental, as was the case in section 5.1. Although such diagnosis is initially reactive, continued use and gained expertise can result in it becoming more proactive, and therefore able to anticipate failures.

6 CONCLUSIONS

Given that advanced technology is expected to shift the burden of low-level tasking from human operators to automated agents in network-centric operations, an inversion of the current many-to-one ratio of operators to unmanned vehicles is likely to be a focal point of future research and development. Since network-centric operations require interoperability among UVs of varying attributes, heterogeneity in vehicle capability and tasks is likely to exist in these future systems. This will lead to a large solution space in terms of design options, causing validation of designs through human-in-the-loop experimentation to become overly lengthy and expensive. In order to address this, a simulation-based model, the MUV-DES model, was developed to support rapid prototyping of UV systems. This chapter summarizes the important results in the MUV-DES model's conception, development, and validation. In addition, the chapter discusses the generalizability of the MUV-DES model, highlights possible avenues for future work, and discusses the contributions of this research.

6.1 Modeling Supervisory Control of Multiple Heterogeneous UVs

6.1.1 MUV-DES Model

Through review of the existing literature, four main attributes of human-UV interaction in the supervisory control of multiple heterogeneous unmanned vehicles (UVs) were identified: vehicle team structure, role allocation and vehicle autonomy, vehicle task allocation, and nature of operator interaction. For each of these attributes, key design variables were identified that should be considered in a model of supervisory control of multiple heterogeneous UVs. A multi-UV discrete event simulation (MUV-DES) model was then developed that incorporated the human-UV interaction attributes and variables. The MUV-DES model utilizes queuing constructs (events, arrival process, service processes, and queuing policies) to model a heterogeneous UV system.

6.1.1.1 Model Inputs and Outputs

By mapping human-UV interaction variables to the DES and related queuing constructs, the MUV-DES model captures the operator, vehicle, and environmental variables as model inputs. In terms of vehicle modeling, role allocation/level of autonomy are captured through vehicle-generated events with which the humans must interact and the rates at which these events are generated. In addition, vehicle task allocation is captured through inter-event interaction. The nature of operator interaction is captured through the serial nature of the single server queuing model and two attention allocation strategies: a switching strategy, which is captured through the queuing policy, and a level of management strategy, which is captured through the rate of operator-induced events. Finally, environmental unpredictability is represented through exogenous events that represent the uncontrolled variables that can affect the system.

In addition to capturing the serial nature of operators dealing with complex tasks, the MUV-DES model has embedded a fundamental construct that captures wait times due to loss of SA. The MUV-DES model does not model individual SA per se, but rather an aggregate effect of inefficient information processing, which likely has many sources that exist at a level difficult to capture in a DES model. While the MUV-DES model captures inefficient attention allocation and management strategies, there are likely many more sources of cognitive inefficiency that are manifested through system delays.

The MUV-DES model allows for grouping of the different human-UV interaction attributes into a holistic model that provides designers of futuristic systems the ability to investigate multiple variables and understand their collective impact. Because of their event-based nature, DES models and their embedded queuing models lend themselves to temporal-based metrics like operator (server) utilization and event wait times. Such metrics are critical in a system where the operator is a limited resource, with finite attention resources that must be divided across multiple vehicles. In addition to these basic metrics, the ability of the MUV-DES model to capture the impact of serviced events on the system state supports the measurement of other mission specific metrics, such as mission success. Mission-specific metrics are important as they can reduce excessive experimental and analysis costs.

6.1.1.2 Model Benefits

Together, the proposed inputs and outputs allow the MUV-DES model to overcome existing gaps in terms of addressing supervisory control of heterogeneous UVs that were introduced in Chapter 2, which include: a) the MUV-DES model is holistic because it takes as inputs the confluence of human-UV interaction attributes and measures their collective impact, b) the MUV-DES model captures both heterogeneity stemming from heterogeneous vehicles/tasks through varied neglect times and interaction times, and the impact of such heterogeneity on the nature of operator interaction in terms of varied operator attention allocation strategies, c) the methodology used in this MUV-DES model incorporates the dynamic impact of operator utilization on SA in discrete event simulation models in order to make effective system observations and predictions, and d) the MUV-DES model supports trade space analysis because it allows the designer to measure multiple operator, vehicle, and mission effectiveness outputs while varying one or more design variables.

By addressing the existing modeling gaps, the MUV-DES model can be used to simulate missions requiring supervisory control of multiple heterogeneous UVs. The ability to measure dynamic mission effectiveness allows the MUV-DES model to be used to satisfy several research objectives including: evaluating the impact of potential design modifications, identifying the limitations of heterogeneity in order to determine system boundaries and influence requirements generation, and replicating observed behavior in existing systems, in order to help diagnose the causes of failures and inefficiencies. These three applications are summarized in Section 6.2.

6.1.2 Model Confidence

In Chapter 4, the results from applying confidence-building techniques to the MUV-DES model were presented. Since accurate and sufficient data for model validation is typically costly as well as time consuming to obtain, an Internet-based Research Environment for Supervisory Control of Heterogeneous Unmanned-vehicles (RESCHU) was developed in order to support collection of large quantities of data in a reasonable length of time. An experiment conducted using RESCHU with team heterogeneity as an independent variable and operator, vehicle, and system performance as dependent variables (as expressed by operator utilization, average search task event wait times, and mission score), was used to build confidence in the model. The conclusions from the validation process can be broken down into those that pertain to model accuracy and those that are associated with the robustness and limitations of the model.

6.1.2.1 Model Accuracy

First, the MUV-DES model was used to replicate vehicle, operator, and system performance exhibited in a human-in-the-loop experiment as the level of heterogeneity was varied. Comparing the results observed in the human-in-the-loop experiments with those generated by the MUV-DES model showed that accurate measures of observed human-UV behavior can be replicated for vehicle, operator, and mission metrics (model results within 95% confidence intervals of experimental data means). In addition, as an external validation, the model was also used to replicate system behavior for a historical independent search and rescue UV supervisory control experimental data set. The MUV-DES model replicated the impact of increasing team size and improved decision support on system and operator performance, and was accurate for conditions that involved larger team sizes and higher autonomy, which is generally the paradigm of control considered in this thesis. Given the search-and-rescue context of the independent test bed and the intelligence-gathering objectives of the RESCHU experimental test bed, the model was able to be used successfully across two completely different settings. This is important since it means that the model gives adequate descriptions of the behavior of different system architectures in a cost effective manner.

Second, using a data set associated with a RESCHU single experimental condition, confidence was built in the model's ability to predict the effects of changes in the team structure on vehicle, operator, and mission metrics. The model was able to accurately predict the impact of a team structure design change that led to a decrease in heterogeneity. However, when the model was used to predict the impact of increasing heterogeneity, predictions were less accurate (with respect to 95% confidence intervals). Further analysis revealed that vehicle team structure changes, especially when they lead to additional complexity (which is often the case when heterogeneity increases), can lead to the operator varying his/her behavior (including the switching strategy and service times). Therefore, the effectiveness of the model at predicting the effect of changes in team composition was shown to be dependent on accurately capturing the interaction between the change in team structure and operator behavior.

Finally, confidence was gained in the SA sub-model that related WTSA to utilization. First, the effectiveness of the WTSA/UT curve in terms of increasing model accuracy was demonstrated. The MUV-DES model with the inclusion of the WTSA/UT curve provided enhanced accuracy for describing experimental observations, as well as for making performance predictions. Second, results showed that the hypothesized impact of increased utilization in terms of decreased SA and performance was accurate. However, at lower utilization levels, the results showed that the shape of the curve might depend on other parameters in addition to operator utilization. While a critical aspect of human performance, SA is a challenge to measure, much less model. The WTSA/UT curve is a useful abstraction that does not model individual SA per se, but rather an aggregate effect of inefficient information processing, which likely has many sources that exist at a level difficult to capture in a DES model.

6.1.2.2 Model Robustness

Several of the validation techniques involved testing the suitability of the model under different input settings. Such tests are critical in identifying those conditions under which the model is suitable, in addition to exposing the limitations of the model's applicability.

First, using extreme condition testing, the impact of increasing vehicle team size on operator utilization was investigated. As the number of vehicles was increased, the model behaved as expected by predicting operator utilization that approached 100%. Unlike experiments with human subjects where operator utilization can only be estimated, the methodical nature of the MUV-DES model allowed a more representative estimate of operator utilization to be captured in order to identify when operator saturation takes place. This was possible by capturing both service times as well as wait times due to interaction.

Second, the variance in the model outputs, which resulted from the stochasticity of the model parameters, was compared to that exhibited in the experimental data set. The test showed that the MUV-DES model is consistent across replications and that the variance in outputs does not exceed that observed in the experimental data set. However, the model resulted in a smaller standard deviation for the wait time and mission score dependent variables than observed in the experimental data. This was attributed to two model assumptions. First, the fixed switching strategy, an a priori input, assumed that all operators exhibit the same switching strategy (a problem noted in the previous section). Second, the WTSA/UT curve, which is used to dynamically derive operator reaction times based on operator utilization, assumes average reaction times. Therefore, designers interested in replicating or predicting the variability in vehicle or system performance should consider using probabilistic distributions for both the switching strategy and the wait times due to loss of SA associated with utilization levels. However, when the underlying research objective is focused more on comparing alternate system designs as opposed to obtaining absolute predictions for vehicle or mission performance variability, the previous approximations are suitable.

Finally, a sensitivity analysis was conducted in order to ensure that the model outputs were not excessively sensitive to errors in the input variables. This is important as errors will always exist in parameter estimation and a model that is not robust to errors can result in inaccurate outputs. In general, the model was found to be robust to errors in the input distributions. Nonetheless, the wait time output was found to be the most sensitive, especially in the case of errors in the service time distribution (25% error in wait time predictions in response to a 20% error in mean service time estimates). Service time distributions should therefore be sufficiently accurate prior to using the model.

In summary, the model was able to describe the behavior of the system in question, given a specific set of initial conditions, as well as predict the effects of changes in one or more system design variables on the behavior of the system. Confidence was therefore gained in the accuracy of the model at providing these two benefits for the domain of supervisory control of multiple heterogeneous UVs.

6.1.3 Model Limitations

Also resulting from the validation process was the identification of modeling limitations. The latter can be classified into two main categories: a) limitations in capturing skill-based tasks, and b) limitations in capturing low task load situations.

First, the MUV-DES model is not suitable for modeling UV systems where the human operator is carrying out skill-based tasks. This is a result of the assumption in the MUV-DES model's construction that the operator is serially attending to complex tasks and that IT is the key variable linking the operator to the UV system. Modeling skill-based tasks could require a parallel processing model such that operators could be generating solutions in parallel with the execution of solutions that have already been identified. Also, operator actions would need to be modeled at lower levels using techniques such as symbolic methods for cognitive modeling.

Second, the MUV-DES model is not appropriate for modeling systems that entail a low average operator task load. This is due to the fact that in situations of low task load, the MUV-DES model is unable to capture the unpredictable operator actions that could result from human creativity. Moreover, it is likely that in low workload situations, the usefulness of the model in terms of supporting human-UV system designers is diminished as information on wait times and operator utilization becomes more variable.

Although the previous two limitations curb the applicability of the MUV-DES model, the future paradigms that drove the development of this model do not share the system properties associated with the identified limitations. First, the systems of interest include vehicles that are highly autonomous (this is a premise on which the need for a modeling tool was built). Such systems with highly autonomous vehicles will generally entail operators applying human judgment and reasoning to knowledge-based tasks. Second, due to manning concerns, future systems of multi-UV control will likely entail efficient designs that present a reasonable task load to the operator. Systems with low average operator task load are therefore less likely to be of interest.

6.2 Model Application

The analyses in Chapter 5 exemplified the potential use of the MUV-DES model to generate design recommendations for the RESCHU system using three alternate underlying research objectives: a) evaluation of the impact of potential design modifications, b) research-oriented requirement generation for futuristic designs, and c) diagnosis-oriented design recommendations. A summary of the important results from these analyses are detailed below.

6.2.1 Evaluation of Potential Design Modifications

The first application considered was one where the underlying research objective was to evaluate the impact of potential design modifications in achieving desired objective functions. When evaluating the impact of potential design modifications, the impact of heterogeneity increases the size of the design space. Optimization-based techniques, which require an underlying simulation model, can be used in order to derive Pareto fronts from which designers can select a desired optimal solution. Using the MUV-DES in such a manner can help identify the changes that are worth investigating further in costly developmental and evaluation phases. By identifying objective functions, constraints, parameters, and design variables, the MUV-DES model was used to pursue a design space exploration to identify the performance associated with different design vectors.

First, in order to exemplify the use of the model in capturing the impact of operational design changes, the number of MALE UAVs in a team was considered as a design variable. As the number of vehicles in the team was varied, a tradeoff relationship was formed between the system performance and vehicle performance objective functions. Model recommendations based on team size analysis were shown to be accurate using an additional experiment, further validating the model (using 95% confidence intervals). The second operational design analysis, which exemplified the use of the model in capturing the impact of developmental design changes, considered a more complex multi-variate design change that focused on the impact of changing interface elements (operator switching strategy, service times and reaction times). As opposed to the team size variable, modifying design variables associated with the interface (by reducing service times, reducing reaction times, and improving the operator switching strategy) led to both mission and vehicle performance shifting towards the utopia point (i.e. in a direction perpendicular to moving along the Pareto front). It was concluded that an additional risk-based metric should be considered. By associating numbers to the risk involved in the implementation of design changes, a cost-based objective function could be defined which could be used to derive a Pareto front using an optimization-based analytic method. Alternatively, designers could also take a cost-benefit analysis or probabilistic risk assessment approach. Since probabilistic risk assessment characterizes risks based on the magnitude of possible adverse consequences and the likelihood of occurrence, such assessment could be especially beneficial when designers wish to evaluate the risk of failure in terms of possible vehicle damage or loss.

6.2.2 Research-Oriented Design Recommendations

The second MUV-DES application considered was one where the underlying research objective was to explore an even less-constrained design space in order to determine the potential capabilities and limitations of future technology. The example considered in this thesis was inspired by observations made during model validation (Chapter 4) that revealed a possible link between service time heterogeneity and under-performance. The MUV-DES model was used to investigate the impact of different degrees of service time heterogeneity on vehicle performance and, as a result, a system boundary in terms of an acceptable degree of service time heterogeneity was identified. Specifically, a degree of heterogeneity greater than 40% was identified as detrimental, increasing both the mean and variability in vehicle wait times. In light of the discovery of this limitation, requirements can be generated in order to keep the system within the identified boundary, and when that is not possible, to generate requirements that mitigate the negative impacts of surpassing the boundary. The MUV-DES model has already been used to evaluate the impact of team structure heterogeneity on the effectiveness of human supervision of a small team of unmanned underwater and aerial vehicles (Kilgore, Harper et al., 2007).

6.2.3 Diagnosis-Based Design Recommendations

Finally, the last application presented was one where designers wish to diagnose causes behind system inefficiencies. Because heterogeneity adds complexity to the causal relationships between design settings and observed performance, the MUV-DES model can be used in order to identify the causal relationships through diagnosis-based approaches. The example considered in this thesis was inspired by observations made while validating the RESCHU system (Chapter 4) that entailed low vehicle performance (driven by high wait times). The MUV-DES model was used in this case to diagnose the causes behind the detrimental vehicle performance using a root cause analysis-type approach.

The MUV-DES model differentiated between wait times due to queuing versus those due to loss of SA (28% composed of WTSA and 72% composed of WTQ). Such a diagnosis using the experimental data would have been extremely costly due to the interleaving of wait times due to queuing, those due to interaction, and those due to loss of SA. The MUV-DES model was able to show that the high service time for the search task event was the main root cause of high wait times, because it increased wait times due to queuing; reducing the service times by 50% results in WTQ decreasing by 71%. Understanding such sources of inefficiencies allows for the identification of potential design modifications that address the discovered sources of detrimental performance. For example, a potential design modification that can be considered is the use of intelligence data and/or automated decision support to guide the operator in the search task, with the aim of reducing the resulting event service times. Nonetheless, proposed design modifications should be supported with additional analysis, similar to that presented in Section 5.1.

6.2.4 Meta-Analysis for Heterogeneity

As part of identifying the attributes and associated variables necessary to capture in a model of a heterogeneous UV system, the importance of considering the heterogeneity in team structure was identified. The multiple dimensions of heterogeneity were found to introduce a number of problems in applying previous models of homogeneous UVs to the heterogeneous case. Specifically, the heterogeneity in vehicles/tasks placed a stronger emphasis on the importance of operator cognitive processes such as the ability to maintain situational awareness, an element not included in the extant homogenous models. In addition, since the vehicles and associated tasks are disparate in the heterogeneous case, the larger diversity in possible attention allocation schemes was identified as important to consider. The impact of increased variability in operator strategies and wait times due to loss of SA were investigated as part of model validation (Section 4.5) and identified to be possible sources of increased variability in mission performance. Together, these results highlight the potential for heterogeneity in vehicle team structure to lead to decreased system robustness.

Analyzing human-in-the-loop experimental results (Section 4.1.3) revealed a possible association between heterogeneity in service times and reduced mission performance. This was later confirmed through a more detailed analysis (Section 5.2) that investigated the relationship between increased service time heterogeneity (with the vehicle team size kept constant) and average vehicle wait times. A degree of service time heterogeneity greater than 40% was identified as the threshold beyond which both mean wait times and their variance increased significantly. The interpretation is that when the ratio of the vehicle with the longest service time to that of the vehicle with the shortest service time is greater than 2, mission performance is negatively impacted as a result of the heterogeneity (independent of operator task load).

During the validation process (Section 4.3), it was also observed that there existed a possible interaction between vehicle team structure heterogeneity and operator-based variables (interaction times, attention allocation strategies, and situational awareness). Results suggested that changes in vehicle team heterogeneity that led to increased supervision complexity could lead to operators varying their interaction times and/or attention allocation strategies. Such interactions would have to be taken into account in order to properly model the impact of changing vehicle team structure heterogeneity.

Finally, constructing WTSA/UT curves (Section 4.8.2) for the three teams of varying levels of heterogeneity revealed a possible interaction between the level of heterogeneity and the scaling of the curve (which is representative of the size of WTSAs). Specifically, the WTSA/UT curve was scaled higher for the condition with service time heterogeneity. Therefore, in addition to service time heterogeneity being a possible source of higher WTQs (as was shown in Section 5.2), service time heterogeneity seemed to impact WTSAs by resulting in a WTSA/UT curve where average WTSA was larger at higher utilization levels.

These observations considered the impact of heterogeneity on the robustness and performance of the system, as well as its interaction with other design variables, which is represented in a causal loop in Figure 6.1. Specifically, the impact of arrival process heterogeneity was distinguished from that of service process heterogeneity, and contributed to a small increase in wait times due to queuing, which in turn resulted in a small degradation in mission performance (Figure 6.2). On the other hand, service process heterogeneity had a much larger negative impact on mission performance (Figure 6.3). In addition to having a larger direct impact on wait times due to queuing, service process heterogeneity increased wait times due to loss of SA. Since wait times due to loss of SA can impact operator attention allocation strategies, this could lead to further impact on wait times due to queuing. Since larger WTQs, as well as larger WTSAs both negatively impact mission performance, the impact of service process heterogeneity was more severe.

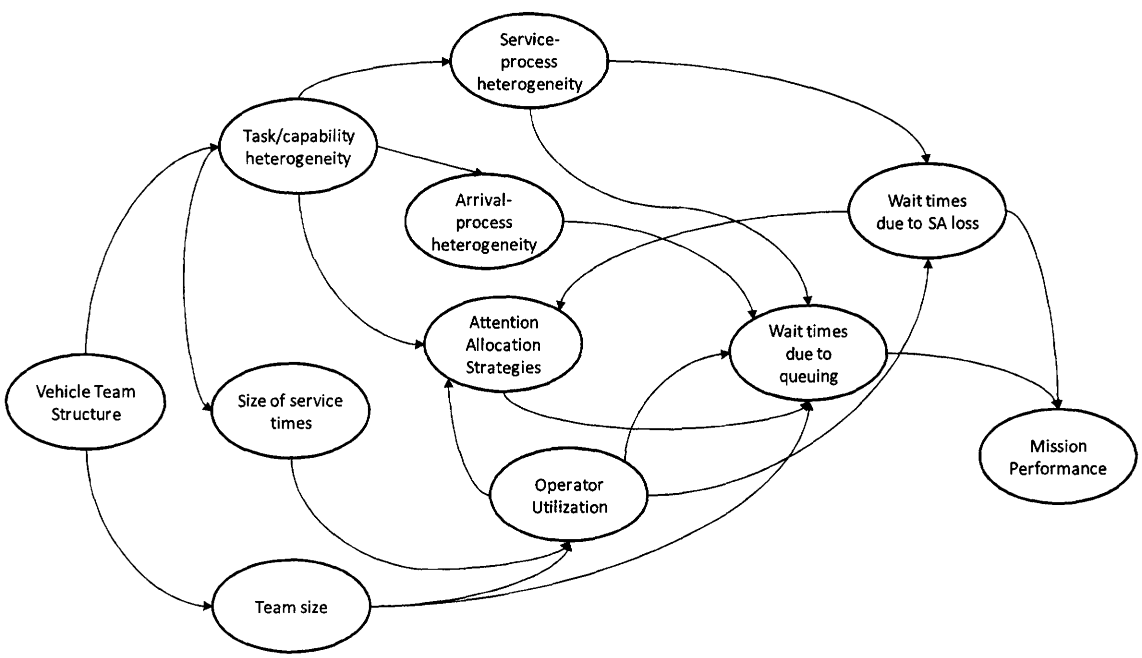


Figure 6.1 Causal loop with a focus on the impact of vehicle team structure

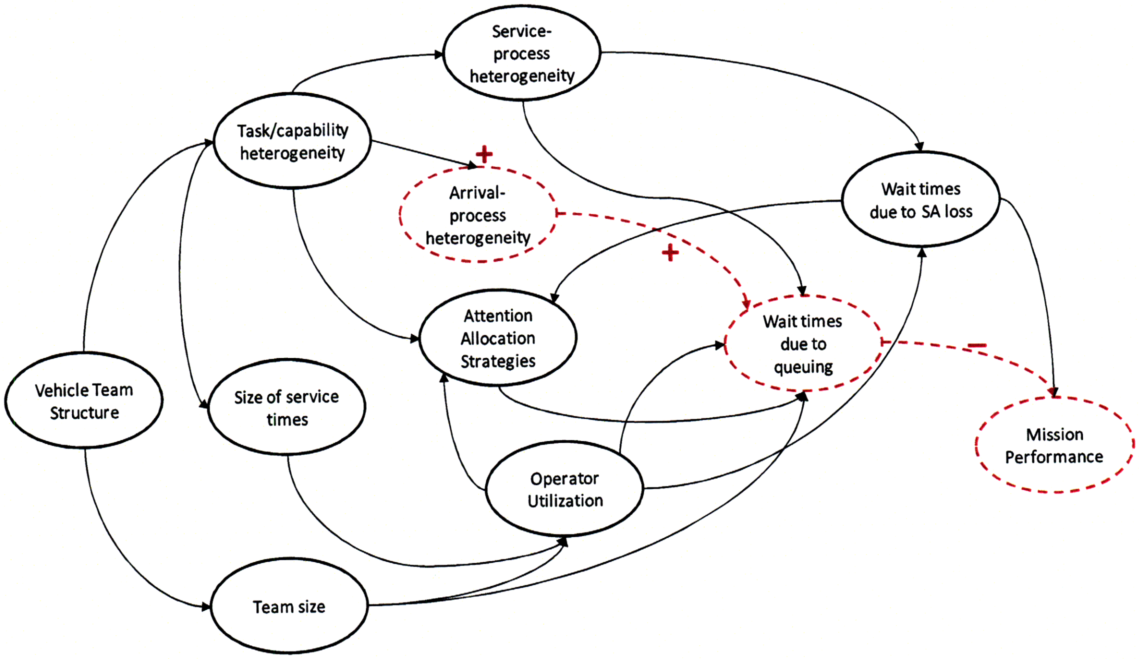


Figure 6.2 Causal loop with a focus on the impact of arrival process heterogeneity

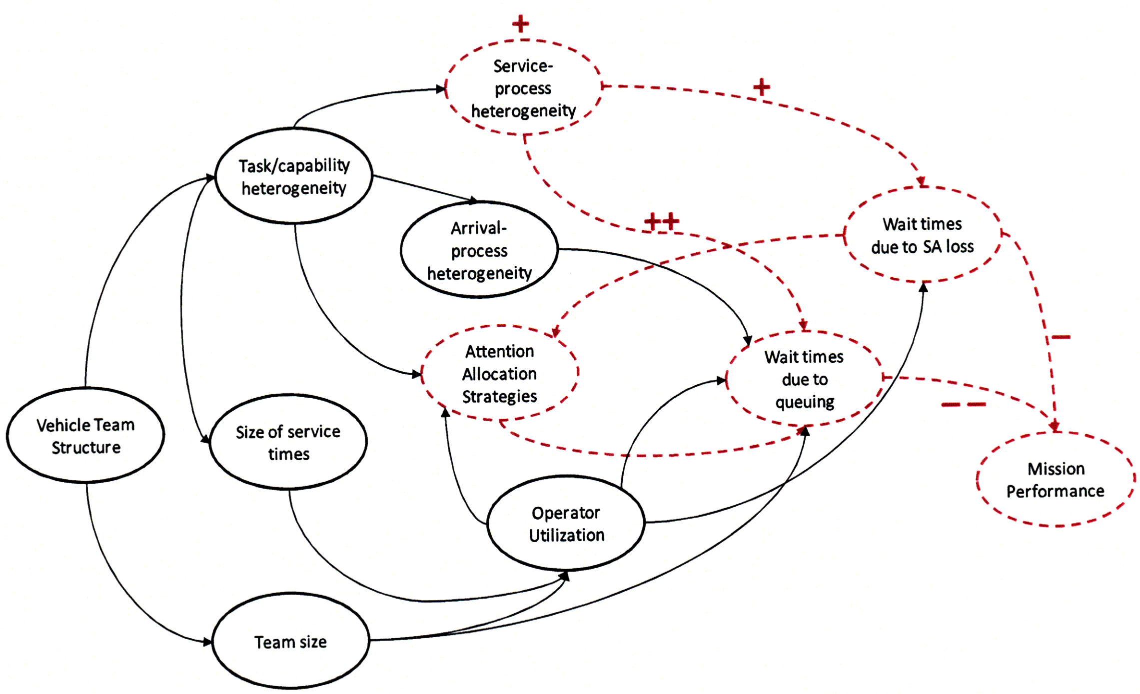


Figure 6.3 Causal loop with a focus on the impact of service process heterogeneity

In addition to these observations, however, the main objective of this research was to develop a tool by which researchers could investigate the impact of heterogeneity in vehicle team structure on the efficiency of human-UV interaction. It is thus expected that additional conclusions could be made that are relevant to the particular systems being modeled.

6.2.5 Other Applications

Although the research in this thesis was motivated by demands in the unmanned vehicle industry, the applications of the work are applicable to the general domain of supervisory control of multiple heterogeneous entities with embedded autonomy. The supervisory control problems that the MUV-DES model is best suited for are those that require knowledge-based behaviors and divided attention (Rasmussen, 1983), such as air traffic control. In such systems, the design of operator interfaces and the level of heterogeneity across platforms can influence the ability of an operator to effectively manage multiple semi-autonomous entities. For example, consider a process control plant, which entails a plant operator monitoring both the automation's control of the process, as well as for significant events. Plant operators will often be responsible for monitoring multiple displays including embedded notification-based alarm systems, which could be heterogeneous in terms of the number of events they generate, as well as the length of time required for resolution. Because such monitoring and event resolution requires knowledge-based behavior and divided attention, the MUV-DES model can be used when designing both displays, as well as the level of heterogeneity across sub-systems.

6.3 Future Work

Three main areas for future work are identified: interaction effects between design variables, the multi-operator case, and embedding additional constructs.

First, as part of building confidence in the model's predictive ability, it was observed that when predicting the impact of team structure changes, lack of consideration of interaction between human-UV system variables (i.e. the variables that form inputs to the model) can result in lower predictive accuracy. Due to the importance of interactions between changes in variables (specifically team-based and operator-based variables) and their impact on the model's prediction accuracy, future work could extend the MUV-DES model to capture such interactions. However, because such interaction effects can be costly to capture, the value added in terms of the underlying research objective should be taken into account.

Second, future paradigms of multiple unmanned vehicle supervision are likely to involve teams of operators as opposed to an individual operator, which was assumed in this thesis. Extending the MUV-DES model to the multiple operator case can aid in futuristic system design. This can be done by extending the MUV-DES model to the multiple server case, accounting for different operator task allocation schemes, and including the effects of team interaction. In addition, the online test bed (RESCHU) can be extended to the multi-player case in order to support validation of such a team-based model.

Finally, this research has implications towards developing more realistic models of human supervisory control and human-system performance. In terms of capturing operator cognitive limitations, the research in this thesis focused on situational awareness. However, there are other operator characteristics which can significantly influence UV-team performance, such as operator trust (Lee & See, 2004) and operator fatigue. Future research should address these issues by identifying and incorporating additional meaningful sub-models within the overall DES model. Incorporating additional structures should be justified by considering the added value of such changes and the contribution of the changes to the overall objective of the underlying research.

6.4 Contributions

The aim of the work presented in this thesis was to answer the research questions identified in Chapter 1. In achieving the research goal, several contributions have been made to the domain of supervisory control of multiple autonomous entities. These contributions include both theoretical contributions in terms of model development, as well as implementation contributions in terms of providing infrastructure for future work in this field. These contributions are discussed below.

By answering the first question, "What attributes need to be captured and how can they be represented when modeling a heterogeneous UV system?", the research resulted in the:

1- Identification of four attributes and associated variables that collectively capture human-UV interaction in the supervisory control of heterogeneous UV teams. The attributes are vehicle team structure, role allocation & level of autonomy, vehicle task allocation, and the nature of operator interaction.

2- Development of a discrete-event-simulation model of human supervisory control of multiple UVs. This was accomplished through the use of discrete event simulation/queuing-based constructs including events, arrival processes, service processes, and queuing policies. To support the modeling efforts, a DES environment that specifically targets modeling human supervisory control of unmanned vehicles was developed. The simulation test bed can support further research in the field of modeling human supervisory control of multiple UVs by allowing researchers to use the environment to capture the effects of model variations. A utility patent on this technology is pending.

3- Quantification of situational awareness effects through the WTSA/utilization relationship. While a critical aspect of human performance, situation awareness has been notoriously difficult to model. The research demonstrated one way its effects can be quantified by utilizing previous relationships (those linking SA to performance and performance to operator utilization) to link wait times due to loss of SA to operator utilization.

By answering the second question, "What type of accuracy/robustness can be expected from a discrete-event-simulation model?", the research resulted in the:

4- Development of an online experimental test-bed, RESCHU, which can be used together with the modeling technique to rapidly collect large quantities of human-in-the-loop data. This online experimental test bed resulted in the collection of data on seventy-five subjects in a period of approximately two and a half weeks, which is a relatively short period of time.

Finally, by answering the third question, "How can a discrete-event-simulation model aid in the design and assessment of heterogeneous UV teams and related technologies?", the ability of the MUV-DES model to streamline the design process was exemplified by applying the model to satisfy three different research objectives. As a result, this thesis begins to fill the gap in the modeling of human-UV interaction in heterogeneous multi-UV control.

In recent years, the use of unmanned vehicles has become increasingly prominent. Unmanned aerial vehicles, ground vehicles, surface vehicles, and undersea vehicles have been used in applications ranging from military operations to border security. In addition, successes of UVs in military applications have encouraged the civil sector to look towards such technology in their future vision. Given the demand for UV systems, which is stimulated by increasing market size and growth trends, the MUV-DES model could play an important role in streamlining the requirements, design and evaluation cycle, therefore supporting the design of efficient and reliable unmanned systems.

APPENDIX A: MUV-DES PSEUDO CODE

This Appendix presents the pseudo code on which the MUV-DES model simulation is based (Table A.1).

Table A.1 Pseudo code for MUV-DES simulation

```
Set simulation_clock = 0,
For all independent and all dependent non-triggered event generators{
        Schedule time for next event generation and insert generator into simulation event list
}

temp = remove event generator with earliest timestamp from event list
Update time
While (time < Total_Simulation_Time){
        Case (temp){
          Dependent:
             event_instance = temp.generate_event()
             insert_into_operator_queue(event_instance)
          Independent:
             event_instance = temp.generate_event()
             insert_into_operator_queue(event_instance)
             Schedule time for next event generation and insert generator into simulation
             event list
          Server:
             Terminate servicing of event j
             Execute all effects associated with event j
             event_instance = pop new event from operator queue
             insert_into_operator_queue(event_instance)
}
---------------------------------------------------------------------------------------------------------------
```

```
Insert_into_operator_queue(event_instance){
    If (server not busy)
        Let server begin servicing event event_instance
        Schedule time for end of service and insert server into simulation event list
    Else
        Insert event_instance into operator queue
}
```

Appendix B: Suitability of Using Web-Based Experimentation Using RESCHU

Overview

Before using RESCHU to conduct experiments for the validation purposes of Chapter 4, an initial experiment was conducted in order to validate the suitability of using the RESCHU interface for web-based experimentation. The premise for this study was that given the interactive tutorial, the low bandwidth required by the simulation, and the restricted web-based delivery method, there would be no statistical significance between remote and local subjects in terms of performance.

Experimental Apparatus

The interface including the experimental setup was the same as that described in Chapter 4.

Participants and Experimental Procedure

The experimental design was a 2x3 factor study with location of experiment and vehicle team heterogeneity level as between-subject factors. The location factor had two levels, laboratory-located experimentation and remote-located experimentation. The vehicle team heterogeneity factor had three levels. The no-heterogeneity condition was composed of five MALE UAVs. The medium-heterogeneity level had three MALE UAVs and two UUVs. The maximum level of heterogeneity, the high-heterogeneity case, required managing two MALE UAVs, two UUVs, and one HALE UAV. The dependent variables were operator utilization, vehicle wait times, and the total number of AOIs engaged.

Participants spent on average 10 minutes performing the practice session. The website was password protected and participation was via invitation. All data were recorded to an online database. Demographic information was collected via a questionnaire presented before the tutorial (Appendix C). Participants were instructed to maximize their overall performance score by 1) avoiding threat areas that dynamically changed, 2) completing as many of the search tasks correctly, 3) taking advantage of re-planning when possible to minimize vehicle travel times between AOIs, 4) ensuring a vehicle was always assigned to an AOI whenever possible.

After participants felt comfortable with the task and the interface, they could end the practice session and start the ten minute experimental session. After test session completion, subjects were required to fill a post-experimental questionnaire before terminating the experiment (Appendix D). After completing the experiment, the participants could see their score, which corresponded to the total number of targets correctly identified.

Thirty participants participated in this study (5 in each condition). The participant who scored the highest in the experiment received a $200 gift certificate.

Experimental Results

The results corresponding to univariate analyses for each of the three dependent variables of concern are presented in Tables B.1-3. Due to the fact that the location factor had a non-significant main effect for all three dependent variables of interest (utilization: $F(1,24)=0.958$, $p=0.338$, avg. wait time: $F(1,24)=11.527$, $p=0.897$, engaged: $F(1,24)=2.541$, $p=0.124$), it was determined that the RESCHU simulator was suitable for web-based experimentation.

Table B.1 Tests of between-subjects effects for operator utilization

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	917.737(a)	5	183.547	2.337	.073
Intercept	108982.251	1	108982.251	1387.605	.000
Scenario	793.624	2	396.812	5.052	.015
Location	**75.208**	**1**	**75.208**	**.958**	**.338**
Scenario * Location	48.906	2	24.453	.311	.735
Error	1884.955	24	78.540		
Total	111784.944	30			
Corrected Total	2802.693	29			

Table B.2 Tests of between-subjects effects for avg. wait time

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	4603.763(a)	5	920.753	1.364	.273
Intercept	83212.185	1	83212.185	123.241	.000
Scenario	1679.752	2	839.876	1.244	.306
Location	**11.527**	**1**	**11.527**	**.017**	**.897**
Scenario * Location	2912.485	2	1456.242	2.157	.138
Error	16204.730	24	675.197		
Total	104020.678	30			
Corrected Total	20808.493	29			

Table B.3 Tests of between-subjects effects for number of targets engaged

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	437.067(a)	5	87.413	6.507	.001
Intercept	6336.533	1	6336.533	471.702	.000
Scenario	354.467	2	177.233	13.194	.000
Location	**34.133**	**1**	**34.133**	**2.541**	**.124**
Scenario * Location	48.467	2	24.233	1.804	.186
Error	322.400	24	13.433		
Total	7096.000	30			
Corrected Total	759.467	29			

APPENDIX C: PRE-EXPERIMENTAL QUESTIONNAIRE

- *Please indicate your sex:*
  - *Male*
  - *Female*

- *Please indicate your age:*
  - *< 18 (if you answer yes here, please quit the experiment and contact the experiment supervisor)*
  - *18 – 25*
  - *25 – 35*
  - *35*

- *Please indicate your occupation (if student, indicate your year and degree)?*

  ____

- *Are you currently or have you ever served in the armed forces of any country?*

  ____

  If yes,
  - *Country:* __________
  - *Service: __Army __Navy __Air Force*
  - *Years of service:*____________

- *Do you have any experience with remotely piloted vehicles (land, sea, sub-sea, air)?*
  - *NO*
  - *YES*

    If yes, please state what vehicle types and number of hours:
    Vehicle types: __Land __Sea __Sub-Sea __Air
    Number of hours? _______

- *Have you participated in a controlled experiment with unmanned vehicle simulators before?*
  - *Yes*
  - *No*

- *Do you currently have or do you have a history of color blindness?*
  - *Yes*
  - *No*

- *How much experience do you have playing video games?*
  - *Never*
  - *< 1 year*
  - *1 – 2 years*
  - *> 2 years*

Appendix D: Post-Experimental Questionnaire

1. How did you feel you performed?

   Very Poor   *Poor*   *Satisfactory*   *Good*   *Excellent*

2. How busy did you feel during the mission?

   Extremely Busy   *Busy*   *Not Busy*   *Idle*

3. On a scale of 1 to 5 (with 5 being the highest demand)
   a. How much mental demand did you perceive? *That is, how much mental and perceptual activity was required (e.g., thinking, deciding, calculating, remembering, looking, searching, etc.)? Was the task easy or demanding, simple or complex, exacting or forgiving?*

      1   2   3   4   5

   b. How much temporal demand did you perceive? *That is, how much time pressure did you feel due to the rate or pace at which the tasks or task elements occurred? Was the pace slow and leisurely or rapid and frantic?*

      1   2   3   4   5

4. On a scale of 1 to 5 (with 5 being the highest awareness), how aware were you of
   a. what was happening in the game

      1   2   3   4   5

   b. where your vehicles were

      1   2   3   4   5

   c. which targets your vehicles were assigned to

      1   2   3   4   5

   d. your vehicles that were waiting for your response

      1   2   3   4   5

   e. changing location of threat areas

      1   2   3   4   5

   f. threat areas on the path of your vehicles

      1   2   3   4   5

   g. what you needed to do next

      1   2   3   4   5

5. Was there some aspect of the interface that you found problematic like UV behaviors, specific buttons, time length, tutorial or general understanding?

6. What other capabilities would you like (if any) the system to have to assist you?

7. Any other comments you might have

APPENDIX E: RESCHU EXPERIMENTAL RESULTS

This appendix presents the statistical analyses associated with the first RESCHU experimental study. A preliminary analysis using a Pearson correlation test demonstrated significant correlations between the three variables of interest: utilization vs. score (ρ=-0.25, p=.03), utilization vs. average search task wait times (ρ=0.50, p<.0001), and score vs. average search task wait times (ρ=-0.58, p<.0001). Because these three measures are somewhat correlated, a Multivariate Analysis of Variance (MANOVA) was performed to control for the inflation of Type I error. The MANOVA results indicated that there were significant effects of heterogeneity level (Wilks' Lambda=0.4, $F(6,138)=13.33$, $p<.0001$). The univariate analysis suggests that the effect of heterogeneity level is attributable to the differences observed in all three variables of interest. There were significant differences between different heterogeneity levels for utilization ($F(2,71)=33.31$, $p<.0001$), score ($F(2,71)=8.13$, $p=.0007$), and total vehicle wait times ($F(2,71)=7.75$, $p=.0009$). Table E.1 presents the pair-wise comparison results.

Table E.1 Pair-wise comparisons

Pair-wise comparison	Δ	df	*p*
Total vehicle wait times			
No vs. medium heterogeneity	11.63 s	71	.1067
No vs. high heterogeneity	**22.74 s**	**71**	**.0006**
Medium vs. high heterogeneity	11.12 s	71	.1454
Utilization			
No vs. medium heterogeneity	**6.29 %**	**71**	**.0062**
No vs. high heterogeneity	**16.45 %**	**71**	**<.0001**
Medium vs. high heterogeneity	**10.16 %**	**71**	**<.0001**
Score			
No vs. medium heterogeneity	**-9.23**	**71**	**.0004**
No vs. high heterogeneity	-4.42	71	.1499
Medium vs. high heterogeneity	4.82	71	.1103

APPENDIX F: DESCRIPTIVE STATISTICS FOR RESCHU DEPENDENT VARIABLES

This appendix contains the descriptive statistics for experiments 1 and 2 conducted using the RESCHU test bed (Tables F.1-2).

Table F.1 Descriptive statistics for RESCHU experiment 1

Condition	Dependent Variable	Mean	Std. dev.	Max.	Min.
No Heterogeneity	Score (%)	23.6	6.5	38.1	12.1
	Utilization (%)	70.2	7.0	82.0	51.3
	Avg. wait time (s)	47.9	24.9	105.2	17.2
Medium Heterogeneity	Score (%)	32.8	11.1	47.8	10.1
	Utilization (%)	63.9	7.2	76.0	50.2
	Avg. wait time (s)	36.3	20.5	88.7	12.4
High Heterogeneity	Score (%)	28.0	5.8	39.1	19.5
	Utilization (%)	53.7	7.0	69.5	42.8
	Avg. wait time (s)	25.2	5.8	48.1	5.6

Table F.2 Descriptive statistics for RESCHU experiment 2

Condition	Dependent Variable	Mean	Std. dev.	Max.	Min.
No Heterogeneity Modified	Score (# objects)	15.1	3.5	20.0	8.0
	Utilization (%)	64.9	9.0	78.8	47.7
	Avg. wait time (s)	28.2	18.7	67.9	10.9
High Heterogeneity Modified	Score (# objects)	10.1	3.0	18.0	6.0
	Utilization (%)	56.1	6.4	70.3	46.5
	Avg. wait time (s)	34.5	26.8	108.6	8.6

APPENDIX G: GENERATED DISTRIBUTIONS FOR MUV-DES MODEL OF RESCHU

Tables G.1-G.3 present the fitted distribution types and their parameters for the different event arrivals and service processes using all the data from the first RESCHU experiment. These were generated using the EasyFit© Software package.

Table G.1 Event arrival and service distributions for the medium heterogeneity team

Event type	Event generator	Distribution	Parameters
MALE UAV			
Type1	Search task arrival	Gamma	$\alpha = 4.04$, $\beta = 26.80$
	Modified search task arrival due to re-plan	Gamma	$\alpha = 3.21$, $\beta = 19.67$
	Search task service	Log-normal	$\mu = 2.96s$, $\sigma = 0.64$
Type2	Re-plan ratio	Bernoulli	$p = 0.51$
	Re-plan service	Exponential	$\lambda = 2.52s$
Type4	Threat area arrival	Exponential	$\lambda = 95.15s$
	Threat area service	Gamma	$\alpha = 1.64$, $\beta = 1.49$
UUV			
Type1	Search task arrival	Normal	$\mu = 174.07s$, $\sigma = 102.33$
	Modified search task arrival due to re-plan	Log-normal	$\mu = 4.61s$, $\sigma = 0.37$
	Search task service	Log-normal	$\mu = 2.78s$, $\sigma = 0.66$
Type2	Re-plan ratio	Bernoulli	$p = 1$
	Re-plan service	Exponential	$\lambda = 2.52s$
Type3	Idle ratio	Bernoulli	$p = 0.42$
	Idle duration	Exponential	$\lambda = 59.1s$
	Idle service	Exponential	$\lambda = 2.52s$
Type4	Threat area arrival	Exponential	$\lambda = 168.63s$
	Threat area service	Gamma	$\alpha = 1.64$, $\beta = 1.49$

Table G.2 Event arrival and service distributions for the high heterogeneity team

Event type	Event generator	Distribution	Parameters
MALE UAV			
Type1	Search task arrival	Gamma	$\alpha = 4.61$, $\beta = 21.97$
	Modified search task arrival due to re-plan	Log-normal	$\mu = 3.86s$, $\sigma = 0.54$
	Search task service	Log-normal	$\mu = 3.14s$, $\sigma = 0.59$
Type2	Re-plan ratio	Bernoulli	$p = 0.58$
	Re-plan service	Normal	$\mu = 3.19s$, $\sigma = 7.32$
Type3	Idle ratio	Bernoulli	$p = 0.10$
	Idle duration	Exponential	$\lambda = 34.88s$
	Idle service	Normal	$\mu = 3.19s$, $\sigma = 7.32$
Type4	Threat area arrival	Exponential	$\lambda = 105.49s$
	Threat area service	Log-normal	$\mu = 0.75s$, $\sigma = 0.56$
UUV			
Type1	Search task arrival	Gamma	$\alpha = 2.89$, $\beta = 73.51$
	Modified search task arrival due to re-plan	Log-normal	$\mu = 4.73s$, $\sigma = 0.58$
	Search task service	Log-normal	$\mu = 2.95s$, $\sigma = 0.69$
Type2	Re-plan ratio	Bernoulli	$p = 1$
	Re-plan service	Normal	$\mu = 3.19s$, $\sigma = 7.32$
Type3	Idle ratio	Bernoulli	$p = 0.69$
	Idle duration	Exponential	$\lambda = 35.91s$
	Idle service	Normal	$\mu = 3.19s$, $\sigma = 7.32$
Type4	Threat area arrival	Exponential	$\lambda = 182.80s$
	Threat area service	Log-normal	$\mu = 0.75s$, $\sigma = 0.56$
HALE			
Type1	Search task arrival	Normal	$\mu = 154.38s$, $\sigma = 56.05$
	Modified search task arrival due to re-plan	Normal	$\mu = 94.42s$, $\sigma = 39.57$
	Search task service	Normal	$\mu = 0.1s$, $\sigma = 0.1$
Type2	Re-plan ratio	Bernoulli	$p = 0.38$
	Re-plan service	Normal	$\mu = 3.19s$, $\sigma = 7.32$
Type4	Threat area arrival	Exponential	$\lambda = 151.00s$
	Threat area service	Log-normal	$\mu = 0.75s$, $\sigma = 0.56$

Table G.3 Event arrival and service distributions for the no heterogeneity vehicle team

Event type	Event generator	Distribution	Parameters
MALE UAV			
Type1	Search task arrival	Gamma	$\alpha = 4.08$, $\beta = 26.27$
	Modified search task arrival due to re-plan	Gamma	$\alpha = 2.80$, $\beta = 18.65$
	Search task service	Log-normal	$\mu = 2.94s$, $\sigma = 0.63$
Type2	Re-plan ratio	Bernoulli	$p = 0.48$
	Re-plan service	Exponential	$\lambda = 3.2s$
Type4	Threat area arrival	Exponential	$\lambda = 72.9s$
	Threat area service	Lognormal	$\alpha = 0.58$, $\beta = 0.56$

APPENDIX H: HISTORICAL DATA SET

This appendix describes the experimental apparatus, experimental procedure, experimental results, and MUV-DES model parameter estimates associated with the historical data set used in the external validation of the MUV-DES model.

Experimental Apparatus

Three aspects of the experimental test bed used in this study are described in this subsection; *mission*, *interface*, and *UV behavior*.

Mission. The mission of the human-UV team was to remove as many objects as possible from a maze in an 8-minute time period. The objects were randomly spread through the maze, which was initially unknown. However, as each UV moved about the maze, it created a map which it shared with the subject and the other UVs in the team. The team could only see the positions of six of the objects initially. In each minute of the session, the locations of two additional objects were shown. Thus, there were 22 possible objects to collect during a session.

An object was removed from the maze (i.e., collected) using a three-step process. First, a UV moved to the location of the object in the maze (i.e., target designation, mission planning, path planning, and UV monitoring). Second, the UV picked up the object (i.e., sensor analysis and scanning). In the real world, performing such an action might require the human operator to assist in identifying the object with video or laser data. To simulate this task, users were asked to identify a city on a map of the mainland United States using *Google Earth*-style software. Third, the UV carried the object out of the maze via one of two exits.

Interface. The human-UV interface was the two-screen display shown in Figure H.1. On the left screen, the map of the maze was displayed, along with the positions of the UVs and objects with known locations. The right screen was used to locate the cities. A participant could only control one UV at a time. When a user desired to control a certain UV, he/she clicked a button on the interface corresponding to that UV. Once the participant selected the UV, he/she could direct the UV by designating a goal location and modifying the UV's intended path to that goal. Designating a goal for the UV was done by dragging the goal icon corresponding to the UV in question to the desired location. Once the UV received a goal command, it generated and displayed the path it intended to follow. The subject was allowed to modify this path using the mouse.

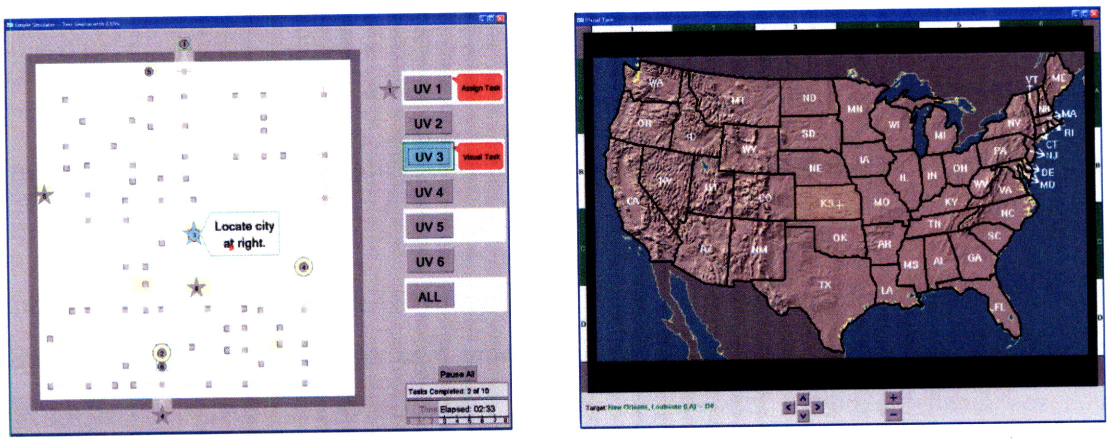


Figure H.1 Two-screen interface by which an operator directed the UVs

UV Behavior. The UVs' map of the maze took the form of an undirected graph. Since the maze was not fully known, a UV had to choose between (a) moving along the shortest path of the known maze to its user-specified goal and (b) exploring the unknown portions of the maze in hopes of finding a shorter path. Each UV used Dijkstra's algorithm on the resulting graph to determine the path it intended to follow.

Two different versions of UV autonomy were employed in the user study. In the first condition, called the *no-decision support* (NDS) condition, each UV's goal destination was determined completely by the human operator. Once the UV arrived at its user-defined goal destination, it did not move again until it received a new command from the user.

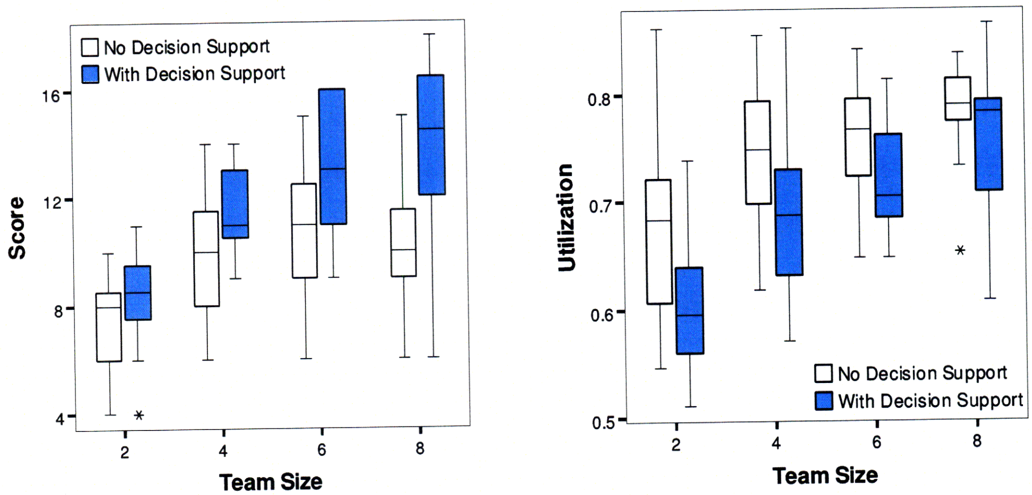
In the second condition, called the *full-decision support* (FDS) condition, each UV automatically selected a new goal when it was left idle. Specifically, a management-by-exception level of automation was used in which a UV left idle at its goal destination, but not on an object in the maze, waited 15 seconds for the user to intervene. If the user did not intervene, the UV automatically moved to the nearest unassigned object (if the UV was searching for an object) or the nearest exit (if the UV was already carrying an object). Additionally, if the user did not intervene, UVs automatically chose to exit the maze via the (estimated) nearest exit in the final 45 seconds of a session. The FDS condition also had one other additional decision support tool to assist the user in locating cities on the map. This decision support tool decreased the search time for a city on the map by about 5 seconds on average.

Participants and Experimental Procedure

The experimental design was a 2x4 factor study, and the purpose was to investigate the effect of increasing team size and alternate decision support schemes on operator performance and utilization in a search and rescue setting with unmanned ground vehicles. The order in which the participants used each team size was counter-balanced throughout the study. Each participant was first randomly assigned to a decision support condition (NDS or FDS), and then was trained on all aspects of the system. Participants then completed three comprehensive practice sessions. Following these practice sessions, each participant performed four test sessions (each with a different team size). Participants were paid $10 per hour; the highest scorer also received a $100 gift certificate. Thirty-two participants between the ages of 18 and 45 participated in the study, 16 in each condition.

Experimental Results

The results from the case study are shown in Figure H.2. A repeated measures ANOVA showed that team size had a significant main effect on the score ($F(3, 90)=41.874$, $p<0.001$). Pair-wise comparisons showed a significant difference in score between all team sizes ($p=0.04$ between the four and eight UV team sizes, $p=0.01$ between the four and six UV team sizes, and $p<0.0001$ for the rest), except between the six and eight UV team sizes, which was not significant. Analysis of decision support type showed a significant main effect on the score variable ($F(1, 30)=9.84$, $p=0.004$).



(a) (b)

Figure H.2 Box-plots for (a) score, and (b) utilization for the case sudy results

The repeated measures ANOVA for utilization showed that team size had a significant main effect ($F(3, 84)=27.97$, $p<0.001$). Pair-wise comparisons showed a significant difference in score between all team sizes ($p=0.005$ between the four and eight team sizes, and $p<0.0001$ for the rest), except between the four and six UV team sizes ($p=0.43$) and between the six and eight UV team sizes ($p=0.12$). The decision support type also showed a significant main effect for utilization ($F(1, 28)=8.45$, $p=0.007$).

MUV-DES Model Parameter Estimation

Table H.1 presents the fitted distribution types and their parameters for different events arrivals and services identified using EasyFit© Software. In most cases, the distribution that best fit the data was the lognormal distribution. A distribution such as the lognormal distribution that is skewed to the left was expected for both service times and arrival rates. This is due to the fact that the data sets showed that in both cases, data points existed where an abnormally long amount of time passes between events or while the operator is servicing a vehicle.

Table H.1 A summary of the distributions derived for both vehicle-generated and operator-induced events

# of Vehicles	Decision Support	Event Category	Arrival Distribution after Re-Planning	Service Time	Arrival Distribution without Re-planning
2	NDS	Vehicle Generated	Lognormal σ=1.1871 μ=3.6762	Lognormal σ=0.83977 μ=1.6274	Lognormal σ=1.0759 μ=3.5615
2	NDS	Operator Induced	Gamma α=0.4444 β=151.3	Exponential λ=0.05971	
2	FDS	Vehicle Generated	Lognormal σ=1.2099 μ=3.550	Lognormal σ=0.76877 μ=1.4283	Exponential λ=0.0397
2	FDS	Operator Induced	Lognormal σ=0.5138 μ=4.307	Lognormal σ=0.4093 μ=3.026	
4	NDS	Vehicle Generated	Lognormal σ=0.9985$\beta\mu$=3.246	Lognormal σ=0.7005 μ=1.312	Lognormal σ=1.234 μ=3.685
4	NDS	Operator Induced	Lognormal σ=1.130 μ=3.351	Gamma σ=0.66218 μ=1.2805	
4	FDS	Vehicle Generated	Lognormal σ=1.2244 μ=3.6333	Lognormal σ=0.76877 μ=1.4283	Lognormal σ=1.201 μ=3.550
4	FDS	Operator Induced	Lognormal σ=0.637 μ=4.4525	Lognormal σ=0.37596 μ=2.8978	
6	NDS	Vehicle Generated	Lognormal σ=1.0456 μ=3.3778	Lognormal σ=0.73413 μ=1.4213	Lognormal σ=1.328 μ=3.6353
6	NDS	Operator Induced	Lognormal σ=1.1743 μ=3.4169	Gamma α=0.94565 β=13.292	
6	FDS	Vehicle Generated	Lognormal σ=1.2356 μ=3.7949	Lognormal σ=0.63637 μ=1.2603	Lognormal σ=1.2272 μ=2.4277
6	FDS	Operator Induced	Lognormal σ=0.53908 μ=4.4863	Lognormal σ=0.37706 μ=2.9057	
8	NDS	Vehicle Generated	Lognormal σ=1.175 μ=3.3932	Lognormal σ=0.62325 μ=1.3214	Gamma α=0.60371 β=126.63
8	NDS	Operator Induced	Gamma α=0.4444 β=151.3	Lognormal σ=1.0852 μ=1.9596	
8	FDS	Vehicle Generated	Lognormal σ=1.2261 μ=3.9041	Lognormal σ=0.59295 μ=1.2155	Lognormal σ=1.3543 μ=2.3909
8	FDS	Operator Induced	Lognormal σ=0.54382 μ=4.5835	Lognormal σ=0.40095 μ=2.9132	

APPENDIX I: CONSTRUCT VALIDATION RESULTS

This appendix contains the results associated with the validation of the shape of the WTSA vs. utilization and the score vs. utilization curves. The number of bins, their associated utilization values, and the pair-wise comparisons between WTSA/score associated with different bins are presented.

WTSA vs. Utilization Results

For the no-heterogeneity condition, there were four final utilization intervals: 50-60% (n=16), 60-70% (n=42), 70-80% (n=50), and 80-90% (n=16). Table I.1 summarizes the pair-wise comparison data.

Table I.1 WTSA data on individual bins for the no-heterogeneity team

Bin	Number of samples (n)	Significantly longer WTSA than
50-60%	16	
60-70%	42	
70-80%	50	60-70%, (t(118)=2.20, p=0.0299).
80-90%	16	60-70%, (t(102) = 2.36, p=0.0202).

For the medium-heterogeneity condition, there were six final utilization intervals: 30-40% (n=10), 40-50% (n=14), 50-60% (n=21), 60-70% (n=39), 70-80% (n=27), and 80-90% (n=15). Table I.2 summarizes the pair-wise comparison data.

Table I.2 WTSA data on individual bins for the medium-heterogeneity team

Bin	Number of samples (n)	Significantly longer WTSA than
30-40%	10	40-50%, (t(120)=1.95, p=0.0536). 50-60%, (*t*(117)=2.40, *p*=0.0182). 60-70%, *t*(117)=2.01, *p*=0.0472).
40-50%	14	
50-60%	21	
60-70%	39	
70-80%	27	
80-90%	15	40-50%, (*t*(117)=2.13, *p*=0.0351). 50-60%, (*t*(116)=2.72, *p*=0.0076). 60-70%, (*t*(117)=2.24, *p*=0.0268).

For the high-heterogeneity condition, there were six final utilization intervals: 30-40% (n=38), 40-50% (n=44), 50-60% (n=30), 60-70% (n=26), 70-80% (n=7), and 80-90% (n=8). Table I.3 summarizes the pair-wise comparison data.

Table I.3 WTSA data on individual bins for the high-heterogeneity team

Bin	Number of samples (n)	Significantly longer WTSA than
30-40%	38	
40-50%	44	
50-60%	30	
60-70%	26	30-40%, (t(91.6)=3.22, p=0.0018).
70-80%	7	30-40%, (t(147)=2.85, p=0.0050). 40-50%, (t(142)=1.94, p=0.0542). 50-60%, (t(143)= 1.95, p=0.0534).
80-90%	8	30-40%, (t(118)=5.63, p<0.0001). 40-50%, (t(146)=4.83, p<0.0001). 50-60%, (t(132)=4.68, p<0.0001). 60-70%, (t(110)=3.36, p=0.0011). 70-80%, (t(132)=2.04, p=0.0437).

Score vs. Utilization Results

There were six final utilization intervals: 30-40% (n=27), 40-50% (n=41), 50-60% (n=45), 60-70% (n=83), 70-80% (n=63), and 80-90% (n=37). Table I.4 summarizes the pair-wise comparison data.

Table I.4 Score data on individual bins for all teams

Bin	Number of samples (n)	Significantly higher score than
30-40%	27	
40-50%	41	80-90% (t(126)=2.04, p=0.0432).
50-60%	45	
60-70%	83	30-40% (t(126)=2.45, p=0.0155) 50-60% (t(126)=2.43, p=0.0163) 70-80% ((t(126)=2.28, p=0.0245) 80-90% (t(126)=4.11, p<0.0001)
70-80%	63	80-90% (t(126)=2.18, p=0.0313)
80-90%	37	

APPENDIX J: ADDITIONAL DESIGN SPACE EXPLORATION RESULTS

This appendix contains additional results associated with the design space exploration of Section 5.1. The objectives calculated from the experiments for T2 and T3 are shown in Figures J.1 and J.2 as the number of UAVs variable is varied. The objectives calculated from the multi-variate design space exploration for T2 and T3, which included changes to the number of UAVs as well as the interface related variables, are shown in Figures J.3 and J.4.

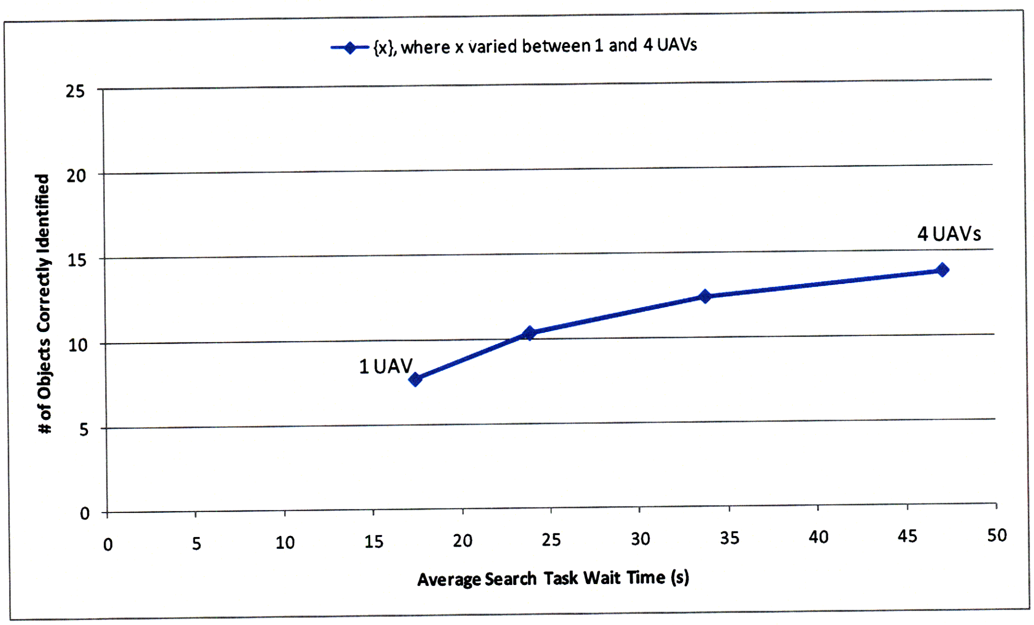


Figure J.1 Tradeoff curve formed as the number of UAVs was varied for the medium-heterogeneity condition

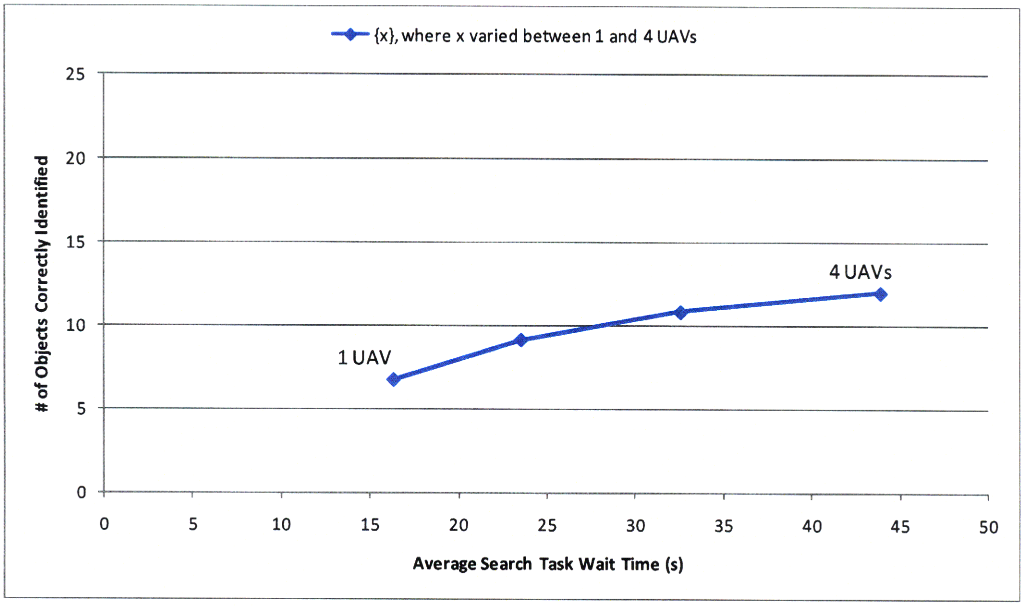


Figure J.2 Tradeoff curve formed as the number of UAVs was varied for the high-heterogeneity condition

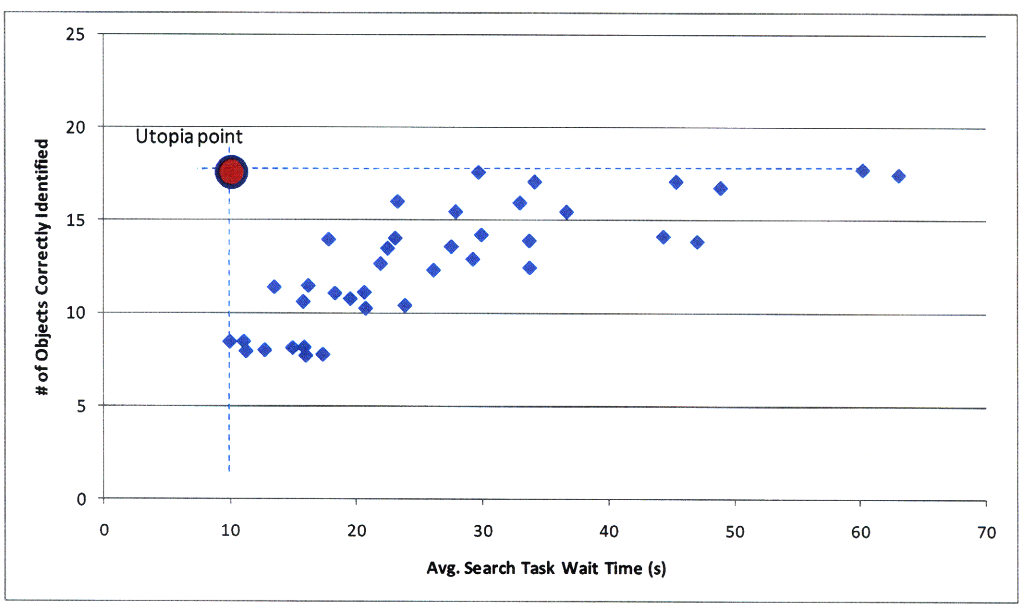


Figure J.3 Design space exploration for the medium-heterogeneity condition (outputs correponding to feasible designs shown)

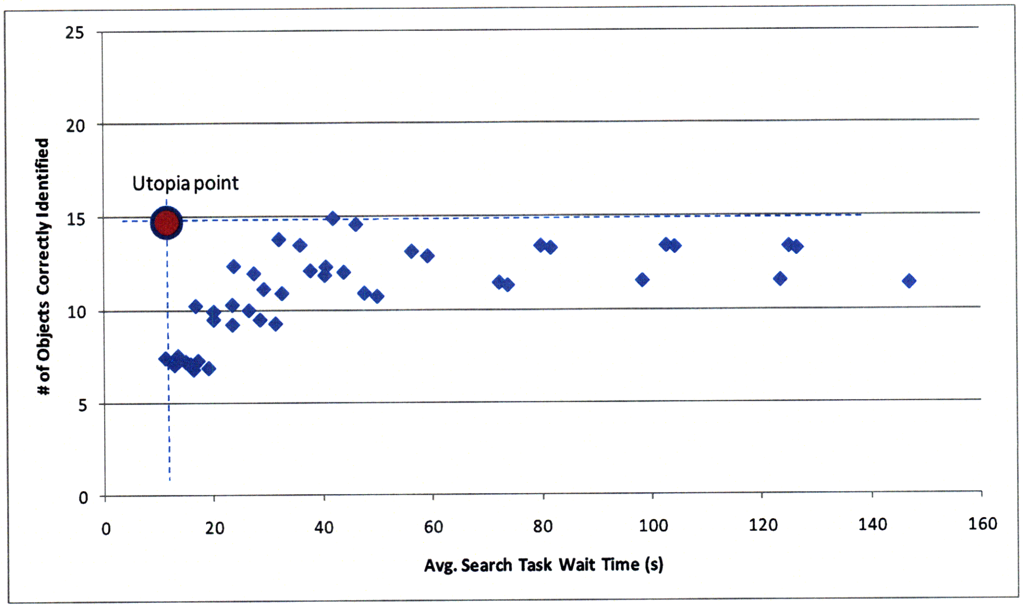


Figure J.4 Design space exploration for the high-heterogeneity condition (outputs correponding to feasible designs shown)

REFERENCES

Alberts, D. S., Garstka, J. J., & Stein, F. P. (1999). Network Centric Warfare: Developing and Leveraging Information Superiority. Washington D.C.: Department of Defense CCRP Publication Series.

Alighanbari, M., & How, J. P. (2005). *Cooperative Task Assignment of Unmanned Aerial Vehicles in Adversarial Environments.* Paper presented at the IEEE American Control Conference, Portland, OR.

Alighanbari, M., & How, J. P. (2006). *Robust Decentralized Task Assignment for Cooperative UAVs.* Paper presented at the AIAA Guidance, Navigation, and Control Conference and Exhibit, Keystone, Colorado.

Balci, O. (1997). *Verification, Validation and Accreditation of Simulation Models.* Paper presented at the 1997 Winter Simulation Conference, Atlanta, Georgia.

Balsamo, S., Persone, V. d. N., & Onvural, R. O. (2001). *Analysis of Queuing Networks with Blocking.* Boston, MA: Kluwer Academic Publishers.

Bharat. (2007). *Emerging UUV & UGV Markets: Forecasts and Business Opportunities to 2020*: Visiongain.

Bolkcom, C. (2005). Homeland Security: Unmanned Aerial Vehicles and Border Surveillance. Washington D.C.: Congressional Research Service.

Bose, S. K. (2002). *An Introduction to Queueing Systems.* New York: Kluwer Academic/Plenum Publishers.

Bradshaw, J. M., Feltovich, P. J., Jung, H., Kulkarni, S., Taysom, W., & Uszok, A. (2004). Dimensions of Adjustable Autonomy and Mixed-Initiative Interaction. In *Agents and Computational Autonomy* (pp. 17-39): Springer Berlin / Heidelberg.

Broadbent, D. E. (1958). *Perception and Communication.* Oxford: Pergamon.

Brunet, L., Choi, H. L., How, J., & Alighanbari, M. (2008). *Consensus-Based Decentralized Auctions for Robust Task Allocation.* Paper presented at the AIAA Guidance, Navigation, and Control Conference, Honolulu, HI.

Cambone, S. A., Krieg, K., Pace, P., & Wells II, L. (2005). Unmanned Aircraft Systems Roadmap. Washington D.C.: Office of the Secretary of Defense.

Carbonell, J. R., Ward, J. L., & Senders, J. W. (1968). A Queuing Model of Visual Sampling Experimental Validation. *IEEE Transactions on Man Machine Systems, 9*(3), 82-87.

Card, S., Moran, T. P., & Newell, A. (1983). *The Psychology of Human-Computer Interaction.* Hillsdale, NJ: Lawrence Erlbaum Associates.

Carrigan, G., Long, D., Cummings, M. L., & Duffner, J. (2008). *Human Factors Analysis of Predator B Crash.* Paper presented at the AUVSI 2008: Unmanned Systems North America, San Diego, CA.

Chandler, P. R. (2004). *Cooperative Control of a Team of UAVs for Tactical Missions.* Paper presented at the AIAA 1st Intelligent Systems Technical Conference, Chicago, Illinois.

Crandall, J. W., & Cummings, M. L. (2007a). *Attention Allocation Efficiency in Human-UV Teams.* Paper presented at the AIAA Infotech@Aerospace Conference, Rohnert Park, CA.

Crandall, J. W., & Cummings, M. L. (2007b). *Developing Performance Metrics for the Supervisory Control of Multiple Robots.* Paper presented at the 2nd ACM/IEEE International Conference on Human-Robot Interaction, Arlington, Virginia.

Crandall, J. W., & Cummings, M. L. (2007c). Identifying predictive metrics for supervisory control of multiple robots. *IEEE Transactions on Robotics - Special Issue on Human-Robot Interaction, 23*(5), 942-951.

Crandall, J. W., & Cummings, M. L. (2008). A Predictive Model for Human-Unmanned Vehicles Systems, (HAL2008-05). MIT Humans and Automation Laboratory, Cambridge, MA.

Crandall, J. W., Goodrich, M. A., Olsen, D. R., & Nielsen, C. W. (2005). Validating Human-Robot Interaction Schemes in Multitasking Environments. *Systems, Man, and Cybernetics, 35*(4), 438-449.

Cummings, M. L., Bruni, S., Mercier, S., & Mitchell, P. J. (2007). Automation Architecture for Single Operator, Multiple UAV Command and Control. *The International C2 Journal, 1*(2), 1-24.

Cummings, M. L., & Guerlain, S. (2007). Developing Operator Capacity Estimates for Supervisory Control of Autonomous Vehicles. *Human Factors, 49*(1), 1-15.

Cummings, M. L., & Mitchell, P. J. (2008). Predicting Controller Capacity in Supervisory Control of Multiple UAVs. *IEEE Transactions on Systems, Man, and Cybernetics, Part A: Systems and Humans, 38*(2), 451-460.

Cummings, M. L., Nehme, C. E., Crandall, J., & Mitchell, P. J. (2007). Predicting Operator Capacity for Supervisory Control of Multiple UAVs. In L. Jain (Ed.), *Innovations in Intelligent UAVs: Theory and Applications*: Springer Berlin / Heidelberg.

Dixon, S., Wickens, C., & Chang, D. (2005). Mission Control of Multiple Unmanned Aerial Vehicles: A Workload Analysis. *Human Factors, 47*(3), 479-487.

DoD. (2007). Unmanned Systems Roadmap (2007-2032): Office of the Secratary of Defense, Washington D.C.

Donmez, B., Pina, P. E., & Cummings, M. L. (2008). *Evaluation Criteria for Human-Automation Performance Metrics.* Paper presented at the Performance Metrics for Intelligent Systems Workshop, Gaithersburg, MD.

Endsley, M. R. (1995). Measurement of situation awareness in dynamic systems. *Human Factors, 37*(1), 65-84.

Feickert, A. (2005). The Army's Future Combat System (FCS): Background Issues for Congress. Washington D.C.: Congressional Research Service.

Fitts, P. (1951). Human Engineering for an Effective Air Navigation and Traffic Control System, Ohio State University Research Foundation Report. Columbus, OH.

Gerkey, B. P., & Mataric, M. J. (2004). A Formal Analysis and Taxonomy of Task Allocation in Multi-Robot Systems. *The International Journal of Robotics Research, 23*(9).

Goodrich, M. A., Johansen, J., McLain, T. W., Anderson, J., Sun, J., & Crandall, J. W. (2007). *Managing Autonomy in Robot Teams: Observations from Four Experiments.* Paper presented at the Human Robot Interaction, Washington D.C.

Goodrich, M. A., Olsen, D. R., Crandall, J. W., & Palmer, T. J. (2001). *Experiments in Adjustable Autonomy.* Paper presented at the IJCA101 Workshop on Autonomy, Delegation, and Control: Interacting with Autonomous Agents, Seattle, WA.

Goodrich, M. A., Quigley, M., & Cosenzo, K. (2005). *Task Switching and Multi-Robot Teams.* Paper presented at the 3rd International Workshop on Multi-Robot Systems, Washington D.C.

Hanson, M. L., Roth, E., Hopkins, C. M., & Mancuso, V. (2004a). *Designing Human System Interfaces for Supervising Multiple UAV Teams.* Paper presented at the AIAA Aerospace Sciences Conference, Reno, Nevada.

Hanson, M. L., Roth, E., Hopkins, C. M., & Mancuso, V. (2004b). *Developing Mixed-Initiative Interaction with Intelligent Systems: Lessons Learned from Supervising Multiple UAVs.* Paper presented at the AIAA 1st Intelligent Systems Technical Conference, Chicago, Illinois.

Harman, W. H. (1989). TCAS: A System for Preventing Midair Collisions. *MIT Lincoln Laboratory Journal, 2*(3), 437-458.

Hearst, M. A. (1999). Mixed-initiative interaction: Trends and controversies. *IEEE Intelligent Systems, 14*(5), 14-23.

Hebb, D. O. (1955). Drives and the CNS. *Psychological Review, 62*(4), 243-254.

Histon, J. M., Hansman, R. J., Aigoin, G., Delahaye, D., & Puechmorel, S. (2002). Introducing Structural Considerations into Complexity Metrics. *ATC Quarterly,* 115-130.

Horvitz, E. (1999). *Principles of Mixed-Initiative User Interfaces.* Paper presented at the SIGCHI Conference on Human factors in Computing Systems, Pittsburgh, PA.

Kaber, D. B., & Endlsey, M. R. (2004). The Effects of Level of Automation and Adaptive Automation on Human Performance, Situation Awareness and Workload in a Dynamic Control Task. *Theoretical Issues in Ergonomics Science, 5*(2), 113-153.

Kilgore, R. M., Harper, K. A., Nehme, C. E., & Cummings, M. L. (2007). *Mission Planning and Monitoring for Heterogeneous Unmanned Vehicle Teams: A Human-Centered Perspective.* Paper presented at the AIAA Infotech@Aerospace Conference, Sonoma, CA.

Kleijnen, J. P. C. (1995). Verification and Validation of Simulation Models. *European Journal of Operational Research, 82*, 145-162.

Laird, J. E., Newell, A., & Rosenbloom, P. S. (1987). Soar: An architecture for general intelligence. *Artificial Intelligence, 33*, 1-64.

Law, A. M., & Kelton, W. D. (2000). *Simulation Modeling and Analysis (3rd ed.).* Boston: McGraw-Hill International Series.

Lazowska, E. D., Zahorjan, J., Graham, G. S., & Sevcik, K. C. (1984). *Quantitative System Performance: Computer System Analysis Using Queueing Network Models.* Upper Saddle River, NJ: Prentice Hall, Inc.

Lee, J. D., & See, K. A. (2004). Trust in automation: designing for appropriate reliance. *Human Factors, 46*(1), 50-80.

Linegang, M., Haimson, C., MacMillan, J., & Freeman, J. (2003). Human control in mixed-initiative systems: lessons from the MICA-SHARC program. *Systems, Man, and Cybernetics, 1*(5), 436-441.

Liu, Y. (1997). Queueing network modeling of human performance of concurrent spatial and verbal tasks. *IEEE Transactions on Systems, Man, and Cybernetics, 27*(2), 195-207.

Liu, Y., Feyen, R., & Tsimhoni, O. (2006). Queuing Network-Model Human Processor (QN-MHP): A Computational Architecture for Multitask Performance in Human-Machine Systems. *ACM Transactions on Computer-Human Interaction, 13*(1), 37-70.

Martis, M. S. (2006). Validation of Simulation Based Models: A Theoretical Outlook. *The Electronic Journal of Business Research Methods, 4*(1), 39-46.

Mau, S., & Dolan, J. M. (2006). *Scheduling to Minimize Downtime in Human-Multirobot Supervisory Control.* Paper presented at the International Workshop on Planning and Scheduling for Space, Baltimore, Maryland.

McCarley, J. S., & Wickens, C. D. (2005). Human Factors Implications of UAVs in the National Airspace. Savoy, IL: University of Illinois.

Miller, C., Funk, H., Wu, P., Goldman, R., Meisner, J., & Chapman, M. (2005). *The Playbook Approach to Adaptive Automation.* Paper presented at the Human Factors and Ergonomics Society's 49th Annual Meeting, Orlando, Florida.

Miller, C., & Parasuraman, R. (2007). Designing for Flexible Interaction Between Humans and Automation: Delegation Interfaces for Supervisory Control. *Human Factors, 49*(1), 57-75.

Mitchell, P. M., & Cummings, M. L. (2005). *Management of Multiple Dynamic Human Supervisory Control Tasks.* Paper presented at the 10th International Command and Control Research and Technology Symposium, MacLean, VA.

Murphy, R. R. (2004). Human-robot Interaction in Rescue Robotics. *IEEE Transactions on Systems, Man, and Cybernetics: Part C - Applications and Reviews, 34*(2), 138-153.

Neth, H., Khemlani, S. S., Oppermann, B., & Gray, W. D. (2006). *Juggling Multiple Tasks: A Rational Analysis of MultiTasking In a Synthetic Task Environment.* Paper presented at the Human Factors and Ergonomics Society 50th Annual Meeting, San Fransisco, CA.

Ollero, A., & Maza, I. (2007). *Multiple Heterogeneous Unmanned Aerial Vehicles.* Berlin Heidelberg: Springer.

Olsen, D. R., & Goodrich, M. A. (2003). *Metrics for Evaluating Human-Robot Interactions.* Paper presented at the Performance Metrics for Intelligent Systems, Gaithersburg, MD.

Olsen, D. R., & Wood, S. B. (2004). *Fan-out: Measuring Human Control of Multiple Robots.* Paper presented at the SIGCHI conference on Human factors in Computing Systems, Vienna, Austria.

Onvural, R. O. (1990). Survey of Closed Queuing Networks with Blocking. *ACM Computing Surveys (CSUR), 22*(2), 83-121.

Parasuraman, R., Barnes, M., & Cosenzo, K. (2007). Adaptive Automation for Human-Robot Teaming in Future Command and Control Systems. *The International C2 Journal, 1*(2), 43-68.

Parasuraman, R., Sheridan, T. B., & Wickens, C. (2000). A Model for Types and Levels of Human Interaction with Automation. *IEEE Transactions on Systems, Man, and Cybernetics, 30*(3), 286-297.

Parker, L. E. (1994). *ALLIANCE: An Architecture for Fault Tolerant, Cooperative Control of Heterogeneous Mobile Robots.* Paper presented at the International Conference on Intelligent Robots and Systems, Munich, Germany.

Pavone, M., Bisnik, N., Frazzoli, E., & Isler, V. (2007). *Decentralized Vehicle Routing in a Stochastic and Dynamic Environment with Customer Impatience.* Paper presented at the Proceedings of the 1st International Conference on Robot Communication and Coordination, Athens, Greece.

Perros, H. G. (1984). Queuing Networks with Blocking: A Bibliography. *ACM Sigmetrics, Performance Evaluation Review, 12*(2), 8-12.

Pina, P. E., Cummings, M. L., Crandall, J. W., & Della Penna, M. (2008). *Identifying Generalizable Metric Classes to Evaluate Human-Robot Teams.* Paper presented at the Metrics for Human-Robot Interaction Workshop at the 3rd Annual Conference on Human-Robot Interaction, Amsterdam, Netherlands.

Pinedo, M. (2002). *Scheduling: Theory, Algorithms, and Systems* (2nd ed.). Englewood Cliffs, NJ: Prentice Hall.

Pritchett, A. R., & Hansman, R. J. (2000). Use of Testable Responses for Performance-based Measurement of Situation Awareness. In M. R. Endsley & D. J. Garland (Eds.), *Situation awareness analysis and measurement.* Mahwah, NJ: Lawrence Erlbaum Associates.

Rao, S. S. (1967). Queuing with Balking and Reneging in M|G|1 Systems. *Metrika, 12*(1), 173-188.

Rasmussen, J. (1983). Skills, Rules, and Knowledge: Signals, Signs, and Symbols, and other Distractions in Human Peformance Models. *IEEE Transactions on Systems, Man, and Cybernetics, 13*(3), 257-266.

Rogers, R. D., & Monsell, S. (1995). Costs of a Predictable Switch between Simple Cognitive Tasks. *Journal of Experimental Psychology: General, 124*(2), 207-231.

Ruff, H. A., Narayanan, S., & Draper, M. H. (2002). Human Interaction with Levels of Automation and Decision-Aid Fidelity in the Supervisory Control of Multiple Simulated Unmanned Air Vehicles. *Presence, 11*(4), 335-351.

Sargent, R. G. (2005). *Verification and Validation of Simulation Models.* Paper presented at the 2005 Winter Simulation Conference, Orlando, Florida.

Scerri, P., Pynadath, D., & Tambe, M. (2001). *Adjustable Autonomy in Real-World Multi-Agent Environments.* Paper presented at the Proceedings of the Fifth International Conference on Autonomous agents, Montreal, Canada.

Schmidt, D. K. (1978). A queuing analysis of the air traffic controller's workload. *IEEE Transactions on Systems, Man, and Cybernetics, 8*(6), 492-498.

Senders, J. W. (1964). The Human Operator as a Monitor and Controller of Multidegree of Freedom Systems. *IEEE Transactions on Human Factors in Electronics, 5*(1), 2-5.

Sheridan, T. B. (1992). *Telerobotics, Automation, and Human Supervisory Control* Cambridge, MA: MIT Press.

Sheridan, T. B., & Verplank, W. (1978). Human and Computer Control of Undersea Teleoperators. MIT Man-Machine Systems Laboratory, Cambridge, MA.

Squire, P., Trafton, G., & Parasuraman, R. (2006). *Human Control of Multiple Unmanned Vehicles: Effects of Interface Type on Execution and Task Switching Times.* Paper presented at the 1st ACM SIGCHI/SIGART conference on Human Robot Interaction, Salt Lake City, Utah.

Steinberg, M. (2006). *Intelligent Autonomy for Unmanned Naval Vehicles.* Paper presented at the SPIE Unmanned Systems Technology Conference, Olrando, FL.

Steinfeld, A., Fong, T., Kaber, D., Lewis, M., Scholtz, J., Schultz, A., et al. (2006). *Common Metrics for Human-Robot Interaction.* Paper presented at the 1st ACM SIGCHI/SIGART Conference on Human Robot Interaction, Salt Lake City, Utah.

Tulga, M. K., & Sheridan, T. B. (1980). Dynamic Decisions and Work Load in Multitask Supervisory Control. *IEEE Transactions on Systems, Man, and Cybernetics, 10*(5), 217-232.

Utt, J., McCalmont, J., & Deschenes, M. (2005). *Development of a Sense and Avoid System.* Paper presented at the Infotech@Aerospace, Arlington, VA.

Valenti, M., Bethke, B., How, J., Farias, D. P. d., & Vian, J. (2007). *Embedding Health Management into Mission Tasking for UAV Teams.* Paper presented at the IEEE American Controls Conference, New York, NY.

Vanden Brook, T. (2007). Drones' Supply Short of Demand [Electronic Version]. *USA Today,*

Wang, J., Wang, H., & Lewis, M. (2008). *Assessing Cooperation in Human Control of Heterogeneous Robots.* Paper presented at the HRI, Amsterdam, Netherlands.

Weil, S. A., Freeman, J., MacMillan, J., Jackson, C., Mauer, E., & Patterson, M. (2006). Designing of a Multi-Vehicle Control System: System Design and User Interaction. In N. J. Cooke (Ed.), *Advances in human performance and cognitive engineering research: Human factors of remotely piloted vehicles.* New York, NY: JAI Press.

Welford, A. T. (1952). The psychological refractory period and the timing of high-speed performance - a review and a theory. *British Journal of Psychology, 43*, 2-19.

Wickens, C. D., & Hollands, J. G. (2000). *Engineering Psychology and Human Performance.* New Jersey: Prentice Hall.

Wilson, P. F., Dell, L. D., & Anderson, G. F. (1993). *Root Cause Analysis: A Tool for Quality Management.* Milwaukee: ASQ Quality Press.

Yerkes, R. M., & Dodson, J. D. (1908). The relation of strength of stimulus to rapidity of habit-formation. *Journal of Comparative Neurology and Psychology, 18*, 459-482.

Zadeh, L. (1963). Optimality and Non-scalar-valued Performance Criteria. *IEEE Transactions on Automatic Control, 8*(1), 59-60.

Zaloga, S. J., Rockwell, D., & Finnegan, P. (2008). World Unmanned Aerial Vehicle System: Market Profile and Forecast 2008 Edition: Teal Group.