OPTIMAL CONTROL POLICIES FOR STOCHASTIC NETWORKS WITH MULTIPLE DECISION MAKERS

By

Howard D. McInvale

Dissertation

Submitted to the Faculty of the

Graduate School of Vanderbilt University

in partial fulfillment of the requirements

for the degree of

DOCTOR OF PHILOSOPHY

in

Interdisciplinary Studies: Civil and Environmental Engineering

August, 2009

Nashville, Tennessee

Approved:

Professor Mark P. McDonald

Professor Sankaran Mahadevan

Professor David M. Dilts

Professor Kenneth R. Pence

Professor Frank L. Parker

INTERDISCIPLINARY STUDIES: CIVIL AND ENVIRONMENTAL ENGINEERING

OPTIMAL CONTROL POLICIES FOR STOCHASTIC NETWORKS WITH MULTIPLE DECISION MAKERS

HOWARD D. McINVALE

Dissertation under the direction of Professors Mark P. McDonald and Sankaran Mahadevan

Decision makers often confront an inability to understand the consequences of interactions within systems of systems (SoS), which can have physical and human components and exhibit hybrid (continuous and discrete) dynamics. The human and physical interactions with the environment and related uncertainties can make optimization and control difficult and result in unintended consequences. As examples, transportation networks may experience lengthy delays or gridlock, and economic stimuli may be ineffective, as a result of suboptimal network control policies. The objective of this dissertation is to motivate, propose and implement a framework that provides decision support in order to manage and operate human-physical networks with hybrid dynamics. The stochastic human-physical analysis framework facilitates the integration of system simulation, uncertainty analysis and optimization under uncertainty for this class of problems.

Specifically, this dissertation: 1) motivates the necessity for a SoS approach to optimizing network control policy; 2) proposes a SoS approach to policy analysis and design under uncertainty; 3) develops an integrated discrete choice and agent-based simulation approach for stochastic human-physical networks with hybrid dynamics; 4) develops and validates computationally inexpensive surrogate models to predict high-fidelity simulation outputs, and uses these models to perform probabilistic reachability analysis and sensitivity analysis; and 5) performs uncertainty propagation and stochastic policy optimization considering both cooperative and non-cooperative decision-makers.

ACKNOWLEDGEMENTS

I want to acknowledge some wonderful people without whom this milestone would not be reached. First, I want to express my sincerest gratitude to my committee for their guidance, support and friendship.

Mark McDonald, Ph.D., Assistant Professor of Civil and Environmental Engineering, Vanderbilt University School of Engineering

Sankaran Mahadevan, Ph.D., Professor of Civil and Environmental Engineering, Vanderbilt University School of Engineering

David Dilts, Ph.D., M.B.A., Professor and Director of Engineering Management, Vanderbilt University School of Engineering; Professor of Operations Management, Owen Graduate School of Management

Kenneth Pence, Ph.D., Associate Professor of the Practice, Engineering Management, Vanderbilt University School of Engineering

Frank Parker, Ph.D., Distinguished Professor of Civil & Environmental Engineering, Vanderbilt University School of Engineering

Dr. Mark McDonald's relevant technical expertise and personal commitment encouraged me. Dr. Sankaran Mahadevan's principled and compassionate intellectual leadership truly inspired me. Dr. David Dilts exemplified uncompromising academic standards and a genuine concern for my well-being. Dr. Ken Pence challenged my assumptions and kept my research directions grounded and purposeful. Dr. Frank Parker imparted valuable seeds of wisdom in me about how to think. I was well-served in my research to have such a balance of perspectives and breadth of experiences advising me. I am humbled by your professional examples, grateful for my time in your tutelage, and honored to call you friends.

I want to convey my heartfelt appreciation to my family. I want to thank my wife, Amy, for her encouragement of, sacrifice for and dedication to me and our family throughout this process. To my children, Jackson, Joseph and Madeline for revealing to me the joy and excitement that accompanies learning new things and reminding me that success is the sum of many small efforts.

TABLE OF CONTENTS

APPENDIX

LIST OF TABLES

LIST OF FIGURES

GLOSSARY OF KEY TERMS

Below is an alphabetical listing of key terms used in this dissertation. The purpose of this glossary is to clarify the definition and/or context of central concepts.

All-At-Once (AAO) Approach- *Chiralaksanakul, A. and Mahadevan, S., 2007.* A multi-disciplinary optimization solution method based on nonlinear programming. Known to be highly centralized, this simultaneous analysis and design algorithm performs the system analysis inside the optimization algorithm and distributes evaluation of governing equations in a way that can result in impressive efficiency.

Decision-Making- *Dilts, D. and Pence, K., 2006.* A cognitive choice resulting from a combination of bounded rationality and political perspective. Typically, individual decisions incorporate "conceptual lenses" where individuals make decision, based on [their environment], using one or a combination of several decision models.

Energy and Information Bonds- *Davenport et al, 2006.* The idea of a "bond" has to do with the link or manner by which interaction occurs. Example descriptions were as follows, "Elements of mechanical systems are energy-bonded; elements from biological systems are information bonded."

First Order Reliability Method (FORM)- *Haldar and Mahadevan, 2000.* The first-order reliability method (FORM) is an analytical method used to determine the probability of a function of continuous random variables assuming less than a certain value The probability of a function being less than or equal to zero (failure of a component limit state) is given as

$$P_{Fail} = P\{g(\mathbf{x}) \le 0\} = \int_{g(\mathbf{x})\le 0} f_{\mathbf{x}}(\mathbf{x})d\mathbf{x}$$

where $f_{\mathbf{x}}(\mathbf{x})$ is the joint probability density of variables $x_1, x_2 \ldots x_n$.

Hybrid Systems- *Lygeros, 2004.* Hybrid systems are dynamical systems that involve the interaction of different types of dynamics, typically discrete and continuous.

Multi-disciplinary Feasible (MDF)- *Cramer, 1994.* A multi-disciplinary optimization solution method based on nonlinear programming in which complete multidisciplinary analysis problem feasibility is maintained at each optimization iteration. The tradeoff in assuring feasibility at each step is increased computational expense.

Multi-disciplinary Optimization (MDO)- *Chiralaksanakul, A. and Mahadevan, S., 2007.* Comprehensive process of finding a set of system design parameters, subject to constraints, such that a prescribed performance measure is optimized. The term "multidisciplinary" refers to the fact that an analysis of the underlying system involves more than one engineering discipline.

Multi-modal- *Jansen, 2004.* Having to do with a set of preferences for flow. In the case of urban transportation, modes of transit could be automobile, subway, bicycle, bus, etc.

Optimization Under Uncertainty (OUU)- *Rockafeller, 2001*. The process of identifying the "best" result in the presence of unknown factors. Problems of optimization under uncertainty are characterized by a necessity of making decisions without knowing what their full effects will be.

Probabilistic Reachability- *Abate, 2007*. A control theory concept aimed at evaluating the likelihood a state of a system will reach a certain target condition during some time horizon, starting from a given set of initial conditions and possibly a control input.

Stochastic Programming- *Holmes, 1994*. A mathematical programming strategy where some of the data incorporated into the objective or constraints is uncertain. Based on scenarios (possible outcomes of the data) for specific and precise joint probability distributions, the outcomes are described in terms of elements of a set, for example, a set of possible demands.

System of Systems (SoS)- *DeLaurentis, 2005*. A collection of trans-domain networks of heterogeneous systems likely to exhibit operational and managerial independence, geographical distribution, and emergent behaviors that would not be apparent if the systems and their interactions are modeled separately.

System of Systems Engineering (SoSE)- *Keating, 2003*. An emerging engineering area aimed at addressing problems associated with the integration of multiple complex systems.

CHAPTER I

INTRODUCTION

> *"Mathematics shows many faces as it works in diverse settings. Statistics measures the quality of information. Optimization finds the best alternative. Probability quantifies and manages uncertainty. Control automates decision making. Modeling and computation build the mathematical abstraction of reality upon which these and many other powerful mathematical tools operate. Mathematics is indeed the foundation of modern decision making."*
>
> -- Paul Davis, Worcester Polytechnic Institute

Interdependent systems with human decision-makers constitute many current systems of systems (SoS) problems. SoS examples are urban transportation networks (Sheffi, 1985), ecomomic infrastructure (Roberts, 2004), military operations and logistics (Sage and Cuppan, 2001), (Pei, 2000), social networks (Lukasik, 1998) and private enterprises (Carlock and Fenton, 2001), just to name a few. Systems of systems often possess multidisciplinary attributes and contain hybrid (discrete and continuous) dynamics. But perhaps the most interesting element of many SoS is the human element. The "humans in the loop" make associated problems inherently difficult to model and analyze. Because humans both affect and react to their environment, the nature of their interdependent relationship is inherently complex. Failure to properly account for the nature of the interactions between these systems can lead to unintended events with unanticipated consequences. A computationally affordable approach to integrate multi-disciplinary system models and uncertainty analysis to provide optimal control policies to SoS decision-makers is needed.

This dissertation motivates, proposes and implements a framework that provides decision support to those who manage and operate human-physical networks with stochastic hybrid dynamics. The stochastic human-physical analysis framework facilitates the integration of system simulation, uncertainty analysis and optimization under uncertainty for this class of problems.

The approach takes advantage of well-established techniques from agent-based modeling, surrogate modeling, probability and statistics and optimization, where possible, but in some cases proposes new techniques for addressing some of the component issues. The necessity for a SoS approach is motivated through a detailed investigation into the transportation systems problem of optimal ramp metering. Then, a basic SoS approach is proposed in the context of an economic stimulus problem. An integrated discrete choice and agent-based simulation approach for stochastic human-physical networks with hybrid dynamics is developed. Computationally inexpensive surrogate models are developed and validated to predict high-fidelity simulation outputs. These models are used to perform probabilistic reachability analysis and sensitivity analysis for network policy. Finally, uncertainty propagation and stochastic policy optimization is performed for cooperative and non-cooperative decision-makers. The result of this work is a methodology that integrates multi-disciplinary models; performs uncertainty analysis; generates output that satisfies reliability and performance requirements; and provides risk-informed decision support to those who operate and manage stochastic human-physical systems with hybrid dynamics.

1.1 Background

1.1.1 History

General Systems Theory has gathered momentum since 1948 through the pioneering work of Wiener, von Neumann, Bertalanffy, Ashby and Forrester, among many others (Francois, 1999). Out of this rich history, a new area of research has grown. Despite falling short of the goal of a grand unified theory of systems, a science of systems has begun to emerge over the last two decades (Troncale, 1985), (Bürger, 1991). More recently, Ossimitz (2003) supports the call for further advances in systems thinking stating,

> "Through analysis of systems (consisting of components which influence each other), interrelated thinking, has become necessary in many areas and should be promoted."

One research area in which this should be promoted is system of systems.

The relatively new field of System of Systems (SoS) research has yet to settle on a common taxonomy or even a widely accepted definition for a SoS. System of Systems is a relatively new term that is being applied to large scale inter-disciplinary problems with multiple, heterogeneous, distributed systems, which may be embedded in networks, at multiple levels. While particular views vary, it is widely agreed that System of Systems is a new and critical discipline for which design and analysis techniques are incomplete (Crossley, 2004). Manthorpe (1996), Kotov (1997), Luskasik (1998), Pei (2000), Sage and Cuppan (2001), and Carlock and Fenton (2001) offer definitions of systems of systems.

One definition common in literature is that SoS is a collection of different elements, which can have physical and human components and exhibit hybrid (continuous and discrete) dynamics, and together produce results not obtainable by the elements alone. The elements can include people, hardware, software, facilities, and policies; that is, all things required to produce systems-level results (Rechtin, 2000). Examples of SoS include environmental systems, economies, supply chains, information systems, biological systems and transportation networks. Evidence of the popularity of SoS research is showcased by numerous recent SoS developments (Valerdi et al, 2007).

- Advent of Institute of Electrical and Electronics Engineers (IEEE) Conference on SoS.
- Inception of the International Journal of SoS Engineering.
- Definition of the SoS signature areas at Purdue University.
- A National Center for Systems of Systems Engineering at Old Dominion University.
- Inclusion of SoS considerations in the Defense Acquisition Guidebook (DAU, 2006).
- Development of systems uniquely labeled System-of-Systems such as Army's Future Combat Systems by Boeing and Science Applications International Corporation (SAIC).
- Creation of the SoS Engineering Center of Excellence by the Office of the Under Secretary of Defense for Acquisition, Technology and Logistics, specifically the Deputy Director of Joint Force Integration.

The need to solve SoS problems is important not only because of the growing complexity of today's systems, but also because SoS problems involve decisions that commit large amounts of money and resources and outcomes that can carry long-term consequences. An example of this kind of consequence is shown in the failure to properly address interactions and complexities of SoS in the national response following Hurricane Katrina in 2005.

Leading researchers in "systems thinking" have published system classifications to describe the relationship among constituent systems (Gharajedaghi, 2005), (Buckley, 1967). Inherent traits (Maier, 1994) are is one way to describe a system of systems. Another method is by identifying the system bonds or interactions defined by the types of flows that exist between the components. Pahl and Beitz (1996) describes the kinds of flows that bond systems (i.e., link their interactions) as information, energy, material, and spatial relations. Gharajedaghi (2005) states, "Mechanical systems are energy-bonded and sociocultural systems are information-bonded." For many SoS, the laws of physics govern the relationship between physical elements, but the human behavior is much more uncertain. I will use this definition for the purposes of this research (DeLaurentis, 2005):

> "A collection of trans-domain networks of heterogeneous systems likely to exhibit operational and managerial independence, geographical distribution, and emergent behaviors that would not be apparent if the systems and their interactions are modeled separately."

Examples of current SoS research challenges were expressed in a 2008 call for proposals by the Department of Energy (DOE), Office of Advanced Scientific Computing Research (ASCR) and the Office of Science (SC). The announcement outlined a need for research including techniques for formulating, analyzing and solving challenging optimization problems in complex multiphysics systems. Additional areas of interest in this category included risk analysis and the quantification and mitigation of uncertainty. Three listed areas of interest motivate this line of research (DOE, 2008). They were:

- Techniques for integrating models with data to support decision-making
- Analysis and algorithms for stochastic optimization
- Related methods for sensitivity analysis, risk analysis, and uncertainty assessment

To meet the need for computational decision support methods that provide accurate, efficient solutions to multi-disciplinary problems, this dissertation draws on concepts from a broad class of academic areas including systems theory, statistical theory, network theory and decision theory. Chapter 2 contains a comprehensive review of the literature pertaining to the topics included in this dissertation. The integration of these topics forged an opportunity to extend the body of knowledge through the research objectives addressed in this dissertation.

1.1.2 Motivation

Most Americans participate in a transportation system of systems everyday when they commute to work and for many of these travelers SoS stands for "Source of Stress." Transportation systems have become ever more linked to broader issues in society and the economy, particularly in America. Americans are the most mobile people on earth, but transportation systems are being pushed to the limits due to population growth, technological change and increased globalization of the economy (TRB, 2005). Transportation systems problems are one of many systems of systems problems for which current solution methods and problem solving approaches are inadequate. Optimally controlling interdependent SoS is especially challenging, given the technical complexity and the difficulty in modeling and predicting human choices. An example of such a problem is the ramp metering to optimally control a transportation network. Of the numerous traffic control strategies employed, perhaps the most prevalent strategy is ramp metering. A ramp metering scheme consists of controlling boundary conditions with metering lights which delay the entrance of cars onto the highway. The intent of such flow control strategy is to improve operating conditions on the highway by restricting the entrance of vehicles. Since traffic engineers control the rate of flow of vehicles onto the highway, it is natural to ask how best to control the rate of flow in order to optimize some system-level performance metric.

In response to the challenges, the vision for traffic management is, "an integrated control of freeway networks involving both ramp metering and route guidance; however, only preliminary measures are currently in use" (Papageorgiou et al, 2003). Papageorgiou et al (2003) identified the following issues in future transportation research specific to road traffic control strategies and their implementation:

- Operational control systems are the exception, rather than the rule
- Employing optimal control algorithms can dramatically improve freeway congestion
- Substantial improvements are achievable via modern traffic control methods and tools
- Improvements are possible at the network-wide level

Methodological developments are required to produce integrated control strategies that are efficient and applicable to large networks. Kotsialos and Papagerorgiou (2001) and Hoogendoorn and Bovy (2001) confirm this assessment of current research related to transportation optimization and control. The goal of this dissertation effort is to further this line of research effort with a contribution to the body of knowledge.

1.2 Problem Statement and Research Overview

Decision makers often confront an inability to understand the consequences of interactions within systems of systems (SoS), which can have physical and human components and exhibit hybrid (continuous and discrete) dynamics. The human and physical interactions with the environment and related uncertainties can make optimization and control difficult and result in unintended consequences. For instance, transportation networks may experience lengthy delays or gridlock and economic stimuli may be ineffective as a result of suboptimal network control policies. The objective of this dissertation is to motivate, propose and implement a framework to provide decision support to those who manage and operate human-physical networks with hybrid dynamics. The stochastic human-physical analysis framework facilitates the integration of system simulation, uncertainty analysis and optimization under uncertainty for this class of problems.

1.3 Research Objectives

The research question addressed in this dissertation is how a system of systems approach can be applied to optimizing control policy in human-physical networks. Such approaches require optimization under uncertainty characterized by the necessity of making decisions without knowing what their full effects will be (Rockafeller, 2001). Given the complexity of SoS operational environments, policy-makers must to choose policies under uncertainty which will bring about system-wide improvement and ensure catastrophic failure is avoided. The objective of this dissertation is to motivate, propose and implement a framework that provides decision support to those who manage and operate these human-physical networks. The proposed research does not directly address SoS architecture or design issues; rather it accepts the architecture and design as given and seeks to improve the overall performance of the existing SoS through identifying optimal operational controls.

To accomplish these goals, this dissertation: 1) motivates the necessity for a SoS approach to optimizing network control policy; 2) proposes a SoS approach to policy analysis and design under uncertainty; 3) develops an integrated discrete choice and agent-based simulation approach for stochastic human-physical networks with hybrid dynamics; 4) develops and validates computationally inexpensive surrogate models to predict high-fidelity simulation outputs, and uses these models to perform probabilistic reachability analysis and sensitivity analysis; 5) performs uncertainty propagation and stochastic policy optimization considering both cooperative and non-cooperative decision-makers. An overview of each of the research objectives follows.

1.3.1 The Necessity of a System of Systems Approach to Optimizing Individual Flow Systems

This objective motivates the necessity for a SoS approach to optimizing network control policy. Using an example ramp metering problem, two optimal ramp metering formulations are examined. The example problem is a physical system model based on actual flow physics (including time and space elements) with user behavior considered assumed and fixed. Shortcomings are identified in single system approaches that use mathematical programming to determine optimal policies for controlling the flow of highway traffic in order to optimize some system-level measure of performance. The sensitivity of ramp metering strategies to choices of performance measures in the objective function is investigated.

Specifically, this objective uses a partial differential equation (PDE) constrained optimization approach to the ramp metering problem and compares results for formulations based on total vehicle miles traveled (VMT) and on total delay. VMT and delay-based formulations are shown to produce optimal strategies that differ in terms of metering restrictiveness and locations. An expanded formulation of the optimal ramp metering problem is proposed which explicitly includes the impact of capacity reduction due to poor queue discipline at diverge bottlenecks into the optimization problem. Using synthetic data, this objective also illustrates how traditional optimal ramp metering formulations may inadvertently cause traffic problems for both the highway and associated surface streets. In light of these results, this objective advocates formulating and solving SoS problems as multi-objective optimization problems which account for the broadest set of transportation system priorities.

The contribution of the work is in showing how optimizers can exploit incomplete mathematical models of the evolution of flow and density on freeway segments in ways which may cause such unintended consequences as a drop in freeway capacity at diverge bottlenecks. This phenomenon is usually not considered, even in the most sophisticated of ramp metering algorithms. These effects are shown to be potentially exacerbated by the choice in objective function. One conclusion is that optimization formulations should take into account the distribution of the origins of drivers who are headed for the diverge bottleneck. Another conclusion is traditional methods can adversely impact connected flow network performance and a broader systems perspective is needed to better address the problem.

In light of these results, the following features are concluded to be desirable in developing optimal network control strategies: (1) when possible, explicitly consider the physics of network and related uncertainties; (2) integrate the coupled nature of human-physical environment where flow exchange is based on user choice; (3) formulate control strategies based on multi-objective optimization which accounts for the broadest set of system priorities. The extension of this research to include these mentioned features are the focus of a later objective.

1.3.2 An Integrated Approach to Policy Analysis and Design Under Uncertainty for SoS

This objective proposes a system of systems approach to policy analysis and design under uncertainty. Human and physical systems often interact with each other and form networks of systems, also referred to as systems of systems (SoS). In developing operational policies for these systems, it is important to model the operation of the system of systems, represent and propagate uncertainties through the operational model, and optimize the system under uncertainty. This objective proposes a framework for modeling and optimizing policy decisions for systems of systems under uncertainty in the context of economic policy planning. The numerical example demonstrates the integrated approach to system simulation, uncertainty analysis, and optimization under uncertainty for a static, multi-sector economic stimulus problem with linear interdependencies. These tools combine to provide decision-makers insights into the impacts of primary and secondary effects resulting from system interdependence. A decoupled approach to optimization under uncertainty is employed using first-order approximations to probability estimates. The example using an economic network illustrates how planners can make more robust decisions under uncertainty using reliability-based optimization methods.

1.3.3 Modeling and Simulating Stochastic Human-Physical Networks with Hybrid Dynamics Using Agent Based Models

This objective develops an integrated discrete choice and agent-based simulation approach for stochastic human-physical networks with hybrid (continuous and discrete) dynamics. Previous models of stochastic human-physical networks often exclude many of the realistic network factors such as hybrid dynamics and human behavior. The human and physical interactions with the environment and related uncertainties can make optimization and control difficult and result in unintended consequences.

Several traditional approaches to address interdependent network problems are based on network equilibration (Nagurney et al, 2002) and (Nagurney and Toyasaki, 2003). However, these techniques lack a computationally efficient way to optimize under uncertainty (Papageorgiou et al, 2003). This objective presents a decision support framework for SoS policy makers to affordably evaluate existing or potential policies for human-physical systems in which network flows and human decisions are coupled. The stochastic human-physical network with hybrid dynamics is modeled and simulated using agent-based stochastic simulation.

The approach developed in this objective facilitates explicit consideration of multi-network physics through agent-based stochastic simulation and incorporates the dual impacts of user decisions on physical system performance and the physical system state on subsequent user choice. For a given network operating environment, modal preferences are estimated for a set of control policies (i.e., modal access costs and network access schedule) and total network demand. Expected network performance and network vulnerabilities can be estimated based on user-informed equilibrium network performance and inherent preferences among the user population.

Resulting trends and sensitivities serve as insights to inform SoS policies aimed to regulate or incentivize preferred user behavior and shape policy decisions such as pricing modal access to promote a desired aggregated flow across the multi-modal network.

1.3.4 Surrogate Modeling, Model Validation and Sensitivity Analysis for Stochastic Human-Physical Networks with Hybrid Dynamics

This objective develops and validates computationally inexpensive surrogate models to predict high-fidelity simulation outputs, and uses these models to perform probabilistic reachability analysis and sensitivity analysis. Gaussian process models and Quadratic Response Surface models are developed to represent the continuous variables (i.e., tolls, fares, signal timing). In order to determine which continuous models are most appropriate for predicting the responses, a quantitative model evaluation method called a Predicted Residual Sum of Squares (PRESS) test is performed (Allen, 1971). To evaluate the predictive strength of the developed models, each of the models is compared to high-fidelity simulation outputs and statistical tests are performed on the fitted models. The quadratic response surface models are shown to be the most appropriate surrogates for the four continuous model outputs. A binary logistic regression model is developed to address the prediction of categorical responses (i.e., pass/fail) and is used to perform probabilistic reachability analysis to assess the probability of the human-physical network with stochastic hybrid dynamics reaching a failed state. Finally, sensitivity analysis is performed to demonstrate the effects of varying control values on output metrics of interest and the SoS objectives.

1.3.5 Uncertainty Propagation and Cooperative/Non-Cooperative Policy Optimization for Stochastic Human-Physical Networks with Hybrid Dynamics

In this objective, system metrics are evaluated under uncertainty and stochastic policy optimization for cooperative and non-cooperative decision-makers is performed. The system objectives of network delay, mass transit ridership, total revenue and network reliability are considered. Surrogate models are used to generate output statistics for control policies evaluated over a stochastic demand using Monte Carlo simulation. Analysis is performed under uncertainty to optimize weighted combinations of policy objectives, so the impacts of stochastic demand and varying objective weights are considered. Finally, policy optimization is performed and compared for two network cases. In the first case, centralized policy optimization is performed for cooperative system leaders that are willing to adhere to control policies set by a central authority. In the second case, a de-centralized optimization is performed for competitive system leaders who seek to myopically optimize the objectives that most benefit their constituent system.

1.4 Research Significance

This research extends the current approaches to optimally controlling stochastic human-physical networks by addressing the hybrid dynamical evolution in a computationally efficient way. The stochastic processes are modeled with a high fidelity simulation and the physical system evolves in response to network controls. Additionally, uncertain initial conditions and model parameters contribute to the system evolution which may or may not reach a failed system state. The best indication of the successfulness of a given control strategy can only be shown in implementation. However, the ability to provide uncertainty-based estimates in a reasonable amount of time is valuable decision support for those who manage and operate stochastic human-physical network with hybrid dynamics.

Based on my review of the literature, it is my opinion that perhaps the greatest technical challenge, in SoSE at the present, is the integration of all the aspects of SoS modeling, analysis, optimization and control. This dissertation presents one approach for addressing this challenge by offering an uncertainty-based, multi-disciplinary optimization approach to integrate multi-disciplinary models; perform uncertainty analysis; generate output that satisfies reliability and optimization analyses; and provide risk informed decision support.

The research described in this dissertation is a contribution to a systems science aimed at meeting the need for computational decision support to those who manage and operate human-physical networks with hybrid dynamics. Such support benefits system leaders by identifying strategies to optimize both collective and competitive objectives. This benefits the public by offering policy solutions to improve the quality of their experience in complex networks, while ensuring the likelihood of catastrophic failure is minimized. It is my goal that this research meaningfully advances the state of the art for finding efficient, reliability-based solutions to some of the interdependent, multi-disciplinary challenges in our world.

CHAPTER II

LITERATURE REVIEW

A review of concepts and literature relevant to this research (organized by topic) is presented here. The chapter is organized by topic for ease of reference. The literature review is organized as follows—

1. System of Systems
2. Multi-Disciplinary Optimization
3. Stochastic Programming
4. Reliability Analysis
5. Systems Modeling
6. Game Theory
7. Uncertainty Analysis
8. Uncertainty-based Design
9. Surrogate Modeling
10. Transportation Theory
11. Logistic Regression
12. Probabilistic Reachability

System of systems (SoS) analysis and design for a given architecture involves five key components: (1) optimization, combinatorial problem solving; (2) dynamics and control; (3) non-deterministic analysis; (4) game theory and economic/competitive behavior; and (5) domain specific modeling (Crossley, 2004). Various methods exist for each of these domains. Perhaps the single greatest challenge in system of systems engineering is integrating the various available tools into a framework for formulating and solving system of system engineering problems. Multi-disciplinary optimization (MDO) is one approach to providing integrated solutions.

MDO is capable of producing solutions which include more accurate analysis because the interactions among disciplines are considered. Reliability analysis techniques are capable of modeling, in a probabilistic framework, many uncertainties inherent to SoS problems. Techniques such as the first-order reliability method (FORM) can provide a failure probability at lower computational expense than required for simulation-based uncertainty analysis methods. Current methods of MDO accommodate system analysis and design in three disciplines: optimization, uncertainty analysis, and domain-specific modeling.

SoS can exhibit evolutionary or emergent behavior and involve multiple decision-makers and humans in the loop. The human-environment interactions occur over space and time. Various modeling strategies exist which are capable of modeling these phenomena at varying levels of sophistication and detail, including input-output analysis, systems dynamics, network models, agent-based simulation. Each of these techniques will likely have their place in the modeling, design, and analysis of systems of systems.

Input-output analysis can be useful in understanding interdependency in a system of systems, but cannot model SoS problems at a microscopic level or handle changing SoS topologies when member systems come online or go offline. Network models describe the topology of a SoS and are a central part of SoS modeling and simulation. Agent-based simulation allows for modeling decision logic of the SoS elements and for capturing emergent behavior in the dynamically changing network topology.

There are clear benefits in using uncertainty-based design methodologies. First, confidence in analysis tools increases as physical, model, and data uncertainties are systematically addressed. Designs are more robust as a wider array of situations is considered. Finally, uncertainty-based design allow the designer to consider extreme circumstances and can allow for planning that accommodates the worst possible conditions.

While the benefits of uncertainty-based design are clear, there are several difficulties which need to be addressed if these benefits are to be realized. Current uncertainty-based methods are more complex and significantly more computationally expensive than deterministic methods, and more efficient methods of performing uncertainty-based design are clearly needed. With the high computational cost of probabilistic design, it is too expensive for use with high fidelity models. Multi-disciplinary analysis only compounds the problems of computational effort, since such problems must generally be solved iteratively.

A comprehensive solution methodology must address modeling, design, analysis and optimization. Because SoS research is less than 20 years old, there are some significant areas that provide great opportunity for study, particularly in the area of MDO when the disciplines are both human and physical. Smith and Mahadevan (2005) developed efficient methods for reliability analysis of multi-disciplinary systems, and has further integrated MDO and RBDO for multi-disciplinary energy-bonded systems. However, the current literature is lacking advance in the following areas: (1) developing MDO methods for SoS with energy and information bonded systems; (2) developing efficient reliability-based methods of solution to MDO problems comprised of energy and information bonded systems, (3) a stochastic programming approach to optimize in a Multi-disciplinary SoS performance, (4) a cohesive method to efficiently integrate surrogate models into the MDO process (DOE, 2008). Fundamental methodological developments in system of systems engineering must be generic enough to be able to accommodate a wide range of diverse models.

Due to the size of SoS problems, surrogate models are critical to reducing the computational expense associated with the modeling, uncertainty analysis, and optimization of systems of systems. Common surrogate techniques for modeling continuous processes include Gaussian Process modeling and quadratic response surface methods. For discrete events, logistic regression has become a standard method of analyzing model relationships (Hosmer and Lemeshow, 2000). Surrogate models facilitate the performance of uncertainty analysis.

Uncertainty analysis is valuable to decision-makers who design, operate and control SoS to determine when and under what conditions the network is likely to fail. An important approach for hybrid systems, known as reachability analysis, is rooted in classical control theory evaluates whether starting from a given set of initial states the system will reach a certain set or not (Abate et al, 2008). For deterministic problems, reachability is a yes/no problem; in stochastic problems, trajectories originating from each initial state have likelihoods of reaching the failed state (Abate, 2007).

Finally, two application domains are explored in this dissertation. The data used in both example problems in this research are synthetic values based on typical figures and are not intended to solve a specific number problem, rather to produce generalized findings. An economic stimulus example is used to introduce the optimization under uncertainty approach. Relevant economic concepts such as Leontief Input-Output modeling are reviewed in this chapter. The second application domain of a transportation network example is used to illustrate the developed stochastic human-physical analysis framework and showcase the surrogate modeling, uncertainty analysis and policy optimization. Therefore, the relevant transportation domain concepts and fundamental theory are also reviewed in this chapter.

2.1 System of Systems (SoS)

2.1.1 General Systems Theory

General Systems Theory has gathered momentum since 1948 through the pioneering work of Wiener, von Neumann, Bertalanffy, Forrester and Ashby, among many others (Francois, 1999). Bartolomei (2007) recognizes key theoretical approaches in systems theory and charts its origin in natural science and its growth through social science, political science, management science and most recently in engineering science. The systems view gives a distinct view of humans and nature (Midgley, 2003). Despite falling short of the goal of a grand unified theory of systems, a skeleton for a science of systems is beginning to emerge (Troncale, 1985).

2.1.2 System of Systems

System of Systems is a relatively new term that is being applied to large scale inter-disciplinary problems with multiple, heterogeneous, distributed systems, which may be embedded in networks, at multiple levels. In addition to the definitions listed in Section 1.1, the Office of the Under Secretary of Defense for Acquisition, Technology, and Logistics, provides a frequently asked questions document about "systems of systems" describing a system of systems as, "a set or arrangement of interdependent systems that are related or connected to provide a given capability" (OUSD-ATL link, 2004). System of Systems Engineering, while still predominantly focused on the defense sector, is beginning to be considered for other challenges; some specific applications that are being explored are in transportation systems, emergency response, and space exploration. Other applications where a system of systems approach can be applied are healthcare, internet design, software integration, homeland security, and other national and global challenges.

Current research efforts in system of systems engineering problems includes, but is not limited to: establishment of an effective frame of reference; crafting of a common lexicon; study of architecting; study of various modeling & simulation techniques, including network theory, systems dynamics, and agent-based simulation; decision making and design under uncertainty and multi-disciplinary analysis and optimization (Crossley, 2004). Central to all of the proposed definitions is the potential for human involvement both as participants and decision-makers. Although the relatively new field has yet to settle on a common taxonomy, the essence of the descriptions in Maier (1998), Crossley (2004) and DeLaurentis (2005) is similar and a common lens through which this research and most SoS research is viewed.

2.1.3 Energy-Bonded and Information-Bonded Systems

Inherent traits are just one way to describe a system of systems. Another method is by identifying the system components across domains and the interaction between these components. Pahl and Beitz (1996) define the types of flows that exist between the systems. Pahl and Beitz describe the kinds of flows that bond systems together as information, energy, material, and spatial relations. Boulding (1956) described open systems as systems in which information, energy, or material was exchanged with an environment. The conceptual idea of describing the system properties that bond it to other systems or to its operating environment is where I derive the use of the terms energy-bonded and information-bonded systems for this research.

Before developing a strategy for using control mechanisms to influence SoS with energy and information bonds, I must first define the terms. Leading researchers in "systems thinking" have published system classifications to describe the relationship among constituent systems. Gharajedaghi (2005) provide the following description:

> While elements of mechanical systems are "energy bonded," those of sociocultural systems are "information-bonded." In energy-bonded systems, laws of physics govern the relationship between elements…but the behavior of active parts (i.e., humans) in information-bonded systems is considered a voluntary association in which bonding is achieved based on perception.

Buckley (1967) explains the dynamics of information bonded systems as a function of the effect of information (or lack thereof), rather than energy transmission. Information-bonded systems can be viewed as a set of elements linked almost entirely by intercommunication of information. It is an organization of meaning emerging from a network of interactions among individuals (Gharajedaghi, 2005). Networks in which both information-bonded systems and energy-bonded systems interact include transportation, biomedical, technology and security networks. The contemporary organizational environment requires successful integration and interoperability of systems of systems (Brooks and Sage, 2005). This research develops a methodology for analyzing and optimizing interdependent human and physical systems in order to provide decision support to policy makers.

2.1.4 System of Systems Engineering (SoSE)

The emerging system of systems context arises when a need or set of needs are met with a mix of multiple systems, each of which are capable of independent operation but must interact with each other in order to full the global mission or missions. The architecture of a system of systems may include existing and yet-to-be-designed aircraft, satellites, ground vehicles, ground equipment, and other independent systems, in addition to their human operators and managers.

The phrase "system of systems" has been in use for several years now, but there is not a single, widely accepted definition of a system of systems. Several researchers have developed their own definitions for a system of systems, and the work by Keating, et al. provides a summary of several of these perspectives of system of systems (Keating et al, 2003). Keating and his co-authors describe their view of systems of systems as meta-systems that are themselves comprised of multiple autonomous embedded complex systems that can be diverse in technology, context, operation, geography and conceptual frame.

A key element of a system of systems is the presence of human decision makers, and this is a critically important modeling element in domains such as homeland security and transportation planning where decision maker behavior drives the behavior of the system. In a system of systems problem, models of human decision making are a vital necessity, whereas in a family of systems problem, they are not. The additional degrees of freedom associated with the ability of each system to operate independently within the family or system of systems add complexity above that encountered in the simpler systems engineering paradigms.

A significant challenge in system of systems design is determining the appropriate mix of independent systems. This is further complicated as yet-to-be-designed systems are considered as potential options for the system of systems. Because the constituent systems are capable of independent operation, the systems could not only cooperate but also compete (Crossley, 2004). The system of systems is a dynamic entity as new systems are added and current systems are replaced or removed. The operation of the system of systems occurs in an uncertain environment (for instance, an Air Traffic Management system of systems must handle weather conditions). Interoperability, which is defined by the DOD as, "the ability of systems . . . to provide data, information, material, and services to, and accept the same from, other systems. . . and to use the data, information, material and services so exchanged to enable them to operate effectively together," (DOD, 2003) also poses a significant challenge.

To summarize, system of systems problems go beyond traditional systems engineering problems because of their size and because of the complexity involved in their architecture, design, and operation. System of systems problems are different from both traditional systems engineering problems and family of systems problems because they have the properties including: Operational Independence; Managerial Independence; Evolutionary Behavior; Emergent Behavior; and Geographical Distribution (Sage and Cuppan, 2001). Systems of systems problems represent major technical challenges for which current solution methods and problem solving approaches are inadequate. This dissertation will develop fundamental methodologies needed to address some of the challenges.

2.1.5 Five Signature Areas for System of Systems Research

Crossley (2005) has identified five areas as necessary for SoS problem formulation and solution: (1) optimization, combinatorial problem solving; (2) dynamics and control; (3) non-deterministic analysis; (4) game theory and economic/competitive behavior; and (5) domain specific modeling.

Optimization is a formal approach which allows decision makers to determine which of the available policies or designs is best. Dynamics and control involve governing the process or performance of the SoS. Non-deterministic analysis is necessary because SoS must perform under conditions set by an uncertain operating environment. Game theoretic models describe the logic of those who make decisions in a SoS. Domain-specific modeling facilitates appropriate representation of the operational environment and relevant interdependencies. The integration of research in these five areas offer a comprehensive set of ideas from which SoS research can continue to grow. This section highlights the most important research challenges from this perspective (Crossley, 2004).

Optimization, Combinatorial Problem Solving

Over the past 20 years, advances in computation have allowed formalized optimization methods to become a part of design efforts for most single complex engineering systems, like aerospace vehicles. These advances have placed system simulation at the heart of the design process. When addressing a single complex system, most design optimization strategies focus on minimizing or maximizing an objective while meeting several constraints. These objectives and constraints typically characterize the performance of the individual system for a typical design mission or missions (Crossley, 2004).

However, optimization strategies alone do not address the impact of a single system's features on the performance of a larger system of systems, nor do they usually address the dynamic, evolving, uncertain environment in which the system of systems must act. Yet there can be no doubt that optimization and traditional operations research techniques a prominent role in SoS research (Crossley, 2004).

A large field of work exists in optimization and operations research that can address organizing a system of systems from existing single systems (Winston, 1991). Resource allocation is currently used in any number of fields of engineering and business to improve the profit, throughput, or other system-level metric. However, these approaches make the assumption that the resources being allocated or assigned have static characteristics that are known.

One important feature of a system-of-systems perspective is a shift in focus regarding the optimization objectives for single systems. For instance, when designing new, single systems, there is often a tendency, owing to the desire to simplify the problem space, to optimize individual systems without considering the broader impacts on the larger system of systems. This generally has the impact of improving that system's performance, but unfortunately a set of locally optimal systems rarely produces the best system of systems.

Additionally, single complex systems designed today are typically designed for a specific, and generally static, mission and operating environment. Because its constituent components are themselves independently operating systems, a system of systems can continually evolve. As new systems are produced, they must be integrated into the system of systems; similarly as current assets reach retirement, they must be removed from the system of systems. These concepts suggest applications of multistage stochastic programming (Sabbagh, 1996); dynamic programming (Bellman, 1957); Markov decision processes (Denardo, 1965); or optimal control (Dreyfus, 1966), in which decisions are made over time. The optimization aspect is also important for the control of a system of systems to ensure optimal performance to complete the assigned tasks and missions. With multiple, independently operating systems in a system of systems, concepts of hybrid, hierarchical, and distributed/decentralized control may provide approaches to ensure that the system of systems maintains an optimal level of performance (Crossley, 2004).

Dynamics and Control

Systems of systems problems represent major technical challenges for which current solution methods and problem solving approaches are inadequate. Particular challenges exist in the area of planning, designing, and operating the SoS. The timescale for each phase is an important factor of the SoS dynamics. Planning may be a 50 year time horizon determining whether a highway network should be constructed. Designing may be a 10 year time horizon involving the actual construction of the highway network. Operating is a short time-scale perspective aimed at controlling the network performance. Research into system of systems optimization and control is limited (Crossley, 2004).

Non-Deterministic Assessment, Decision-Making and Design under Uncertainty

A system of systems, like single, complex systems, will operate in the real world; however, the operating environment is non-deterministic. For instance, an air traffic management system of systems operates in varying kinds of weather conditions, which can be predicted, but not with absolute certainty. Most engineering disciplines are beginning to address this type of uncertainty, and this focus must also be incorporated in system of systems problems (Crossley, 2004).

A system of systems approach further extends the impact of non-deterministic assessment. Much of the motivation behind the move to a capability-based acquisition strategy requiring system of systems solutions is that the capabilities sought by the customer are driven by the desire to have high performance that is robust with respect to varying operating conditions and scenarios. When designing a system of systems, nondeterministic conditions must be addressed, and the system of systems must be architected to be reliable.

Uncertainty analysis in the context of a large scale system generally requires constructing a probability distribution that describes the overall system response give the distributions of the inputs. For a single complex system, the aggregate PDF can often be determined by Monte Carlo Simulation or sometimes by analytical reliability methods (Haldar and Mahadevan, 2000). In a system of systems sense, it is not clear that the system of systems level reliability assessment could or should be conducted in the same manner.

Because a system of systems is comprised of multiple systems capable of independent operation, as one system entity begins to reach a degraded performance or a failure mode, other system entities can alter their independent operations to perform functions that the failed system no longer performs. This is not simply an "m out of n" redundancy issue, because, if properly designed, the system of systems does not necessarily include spare systems whose sole purpose is to replace a failed system (Crossley, 2004).

For instance, if a regional transportation network is viewed as a system of systems, whose constituent systems include busses, trains, and aircraft, and then as a train system becomes unavailable, no spare train may be present in the system of systems. One or more of the busses can be assigned to different or more frequent routes to replace the transportation capacity of the train. When the train becomes unavailable, an aircraft may be assigned to different or more frequent routes. The more desirable option would the one that maintains the highest system-of-system performance. However, deciding upon this best option implies some sort of resource allocation is needed. In order to predict the reliability of the system of systems, the failure of the train system would need to be modeled, and then a series of conditional assessments are needed to determine the remaining reliability of the system of systems. Examples of such assessments include whether one bus is used to replace the train system, if multiple busses are used to replace the train system, if an aircraft is used to replace the train system, etc. The approaches needed to perform this type of assessment currently are not readily apparent, though it appears obvious that model-based simulation of the entire system of systems is necessary to understand the implications of a major disruption to one or more critical systems, and that fault-tolerant and robust designs are needed for systems of systems (Crossley, 2004).

The management of uncertainty for SoS engineering is also a focus of much literature. Techniques in reliability-based design optimization (Zhao and Ono, 1999), (Youn and Choi, 2004), (Chiralaksanakul and Mahadevan, 2004), (Wu et al, 1990), (Wu and Wang, 1998), (Du and Chen, 2000, 2004), (Du and Sudjianto, 2003), (Yang and Gu, 2004), (Parkinson et al, 1993), (Jung and Lee, 2002), (Yu and Ho, 2000), (Gunawan and Azarm, 2005), (Lee and Park, 2002), (Stocki et al, 2001) and stochastic programming are relevant to multi-disciplinary problems.

Economic Theory, Game Theory, & Other Approaches for Modeling Competitive Behavior

A system of systems is comprised of individual systems capable of independent operation, and the individual systems are managed independently. As described in previous sections, using multiple systems in collaboration can provide capabilities well beyond those available from a single system. Further, the ability for each constituent system to operate independently can provide increased robustness for the overall system of systems. However, these aspects provide an additional level of complexity in determining which systems provide which contributions to the overall performance. The best operation for one system may compete with best operation of other systems. Decision making for the system of systems must resolve these issues, and determining an appropriate sharing of capabilities and resources will call upon applications and approaches from game theory and competitive behavior (Mas-Collel et al, 1995). In some circumstances, a system of system may itself compete against other systems of systems using strategies that allow less or non-competitive behavior in one aspect in order to provide overall system-of-systems level performance. This autonomy in decision maker behavior is of critical importance of system-of-system behavior, because in practical system of system problems, the member systems have choice in how they operate. Hence the decision maker behavior is what drives the performance of systems of systems.

In some cases, distributed decision-making can be stated explicitly as a game theory problem, or as a very special case of a multi-disciplinary design optimization problem known as a mathematical program with equilibrium constraints (Nagurney and Dong, 2001). Yet these approaches have significant limitations in that humans are not always economically rational, and cannot possibly have and understand the true state or dynamics of a system of system and are thus boundedly rational (Simon, 1957). In other words, modeling decision maker behavior using optimization methods makes assumptions that are not realistic. System dynamics (Forrester, 1968) and agent-based (Axelrod, 1997) approaches to modeling decision maker behavior are explored. Further review of Game Theory is presented in Section 2.6.

Domain-Specific Modeling

Optimization and control poses some interesting multi-disciplinary issues for system of systems work. For example, the proposed Future Combat System involves manned ground vehicles, unmanned air vehicles, combat robots, soldier robots, communications systems, and the soldier him/herself (USCG, 2004). Each of these systems is a multi-disciplinary system which must be modeled, and there is yet another multi-disciplinary issue in integrating these models. To successfully provide function evaluations for an optimization approach, or other decision making strategy, these components must be appropriately modeled. One important issue is systems be modeled at compatible levels of detail, which will require domain-experts to successfully communicate with other domain-experts. Further, nearly all systems of systems will have humans-in-the-loop, requiring the capability to model and simulate human behavior. Simultaneously modeling all of the constituent systems in a system of systems is difficult. Emerging approaches like grid computing, where a central core or bus computational structure controls distributed, dissimilar simulations and modeling programs, appears to be a promising area for system of systems approaches. Modeling a large number of diverse systems capable of independent operation may see benefits from recent work in agent-based modeling. Some applications of agent-based modeling have included tens of thousands of independent agents and have also incorporated human behavior simulation as part of the modeling strategies (Axelrod, 1997), thereby allowing for the modeling of boundedly rational agents. A further challenge is using these modeling strategies in conjunction with optimal design and optimal control methods.

2.2 Multi-disciplinary Design Optimization (MDO)

Optimization is a tool which can help decision makers reach decisions which are justifiable, by some metric, as the best decision. Multi-disciplinary design optimization (MDO) is an optimization approach which allows decision makers to perform optimization in systems where multiple disciplines and/or system models must be integrated. MDO has been used in a number of fields, including automobile design, naval architecture, electronics, computers, and electricity distribution.

MDO allows designers to incorporate all relevant disciplines simultaneously. The optimum of the simultaneous problem is superior to the design found by optimizing each discipline sequentially, since it can exploit the interactions between the disciplines. However, including all disciplines simultaneously significantly increases the complexity of the problem. Problem formulation is normally the most difficult part of the process. It includes the selection of design variables, constraints, objectives, and models of the disciplines. A further consideration is the strength and breadth of the interdisciplinary coupling in the problem.

Many solution methods work only with single objectives, although MDO problems typically involve several objectives. When using these methods, the designer normally weights the various objectives and sums them to form a single objective. Other methods allow multiobjective optimization, such as the calculation of a Pareto front. The designer must also choose models to relate the constraints and the objectives to the design variables. These models are dependent on the discipline involved. The Multi-disciplinary nature of most design problems complicates model choice and implementation. Often several iterations are necessary between the disciplines in order to find the values of the objectives and constraints, or special formulations of the MDO problem are necessary. Once the design variables, constraints, objectives, and the relationships between them have been chosen, the problem can be expressed in the following form:

$$\min_{\mathbf{x}} f(\mathbf{x})$$

$$s.t.\ \mathbf{g}(\mathbf{x},\mathbf{u},\mathbf{v}) \geq \mathbf{0} \quad \text{where (} \mathbf{g} \text{ is a set of limit states)}$$

$$\mathbf{h}(\mathbf{x},\mathbf{u},\mathbf{v}) = \mathbf{0}$$

where *f* is an objective, **x** is a vector of design variables; **u** and **v** are vectors of state variables for the disciplinary analyses, **g** is a vector of constraints, **h** is a vector of disciplinary analyses. The problem is normally solved using appropriate techniques from the field of optimization. These include gradient-based algorithms, population-based algorithms, or others. Specialized solution techniques such as Multi-disciplinary Feasibility (Kodiyalam, 1998), Interdisciplinary Feasibility (Cramer et al, 1994), Simultaneous Analysis and Design (Alexandrov and Lewis, 2000), and Collaborative Optimization (Braun, 1996) are available for these types of problems. A brief overview of the Multi-disciplinary Feasible method and the All-At-Once (AAO) methods will be given; as they are the two most widely use approaches.

2.2.1 Multi-disciplinary Feasible (MDF) Approach

The most basic of MDO formulations is the MDF approach, also known as 'Nested Analysis And Design' (NAND), 'All-in-One' (AIO), and 'One at a Time'. This formulation is distinct from AAO, presented later in this section. A single system-level optimizer is used, and from the perspective of the optimizer MDF is no different than a 'standard' optimal design problem. A system analyzer coordinates all of the subspace analyzers. The optimizer supplies the system analyzer with a design **x**, and the system analyzer supplies the optimizer with the appropriate response functions, *f*, **g**, and **h**.

Since convergence must be achieved for all analyses after every iteration of the optimizer, some iterative algorithm must be applied at every optimizer iteration to assure disciplinary consistency. Fixed point iteration is a popular solution method for MDF, although other analysis options exist. A formulation strategy is classified as MDF if a complete system analysis is performed for every optimization iteration. The analysis is "nested" within the design (hence the acronym NAND). The optimizer is charged with the responsibility to find the optimal design **d** (the design solution), while the system analyzer is solely responsible to find the set of consistent coupling variables **y** and returns values for the objective function *f*, the limit state functions **g**, and the disciplinary analyses **h**.

The MDF problem statement, shown above, is completely non-hierarchic in nature (no communication restrictions). In a purely computational context, this approach is desirable if the subspaces are weakly coupled (fast analysis convergence), and if the subspace analyses are not computationally expensive. From the point of view of the optimizer, this problem is no different from solving a traditional optimization problem. In an organizational context, MDF allows the continued use of legacy analysis tools without modification. If the organization already performs a complete analysis before making a design decision, MDF is a natural fit.

2.2.2 All-At-Once (AAO) Approach

The All-At-Once Approach (AAO), also referred to as Simultaneous Analysis and Design (SAND) is a highly centralized approach. Instead of utilizing analyzers to complete the analysis for each subspace, algorithms are used that compute only the residuals of the governing equations. The SoS optimizer controls two sets of decision variables: the original design variables **d**, and the state variables **y**. AAO centralizes both design and analysis, but still distributes evaluation of governing equations. This can result in impressive efficiency, but is difficult to map to organizational structures or simulation tools due to its centralization and specialized structure. The formulation of the AAO approach is given below. It includes no auxiliary constraints because the optimizer chooses values for the state variables that assure disciplinary consistency. In this approach, design, system analysis, and subspace analyses are all performed simultaneously.

$$\min_{\mathbf{d},\mathbf{y}} f(\mathbf{d})$$
$$s.t.\ \mathbf{g}(\mathbf{d},\mathbf{y}) \geq \mathbf{0} \quad \text{where (} \mathbf{g} \text{ is a set of limit states)}$$
$$\mathbf{h}(\mathbf{d},\mathbf{y}) = \mathbf{0}$$

2.3 Stochastic Programming

Stochastic programs are mathematical programs where some of the data incorporated into the objective or constraints are uncertain and uncertainty is characterized by probability distributions for the parameters. In practice, the uncertainty information can range in detail from a few scenarios (possible outcomes of the data) to specific and precise joint probability distributions. The outcomes are generally described in terms of elements of a set, for example, the set of possible demands (Holmes, 1994). The field is currently developing rapidly with contributions from many disciplines including operations research, mathematics, and probability and applications in areas such as macroeconomic modeling, freight planning and traffic management (Birge and Louveaux, 1997).

2.3.1 Mathematical Programming

Many decision problems can be modeled using mathematical programs, which seek to maximize or minimize some objective which is a function of the decisions. The possible decisions are constrained by limits in resources, minimum requirements, etc. Decisions are represented by variables. Objectives and constraints are functions of the variables, and problem data. Examples of problem data include unit costs, production rates, sales, or capacities (Bazarra et al, 1990). Let x_i represent production of the i^{th} of n products. The general form of a mathematical program is

$$\begin{aligned} &\min f(x) \\ &s.t.\ \mathbf{g}_1(x) \le \mathbf{0} \\ &\quad \mathbf{g}_m(x) \le \mathbf{0} \\ &\quad \forall\, x_i \textbf{ in } \mathbf{X} \end{aligned}$$

where X is a set of all nonnegative real numbers. The constraints can linear of nonlinear to capture the essence of the model.

2.3.2 Recourse Models

Holmes (1994) gives a good overview of recourse models. When some of the data are random, then solutions and the optimal objective value to the optimization problem are themselves random. A distribution of optimal decisions is generally unrealistic. Ideally, one decision and one optimal objective value are preferred. One logical way to pose the problem is to require one decision now and minimize the expected costs (or utilities) of the consequences of that decision. This paradigm is called the recourse model.

Suppose x is a vector of decisions that one must take, and $y(w)$ is a vector of decisions that represent new actions or consequences of x. Note that a different set of y will be chosen for each possible outcome w. An example two-stage formulation is:

$$\min f_1(x) + E[f_2(y(w), w)]$$

$$s.t.\ \mathbf{g}_1(x) \leq \mathbf{0}, \ldots, \mathbf{g}_m(x) \leq \mathbf{0}, \textbf{ for } \forall\ x_i \textbf{ in X}$$
$$\mathbf{h}_1(x, \mathbf{y(w)}) \leq \mathbf{0}, \ldots, \mathbf{h}_k(x, \mathbf{y(w)}) \leq \mathbf{0}, \textbf{ for } \forall\ \mathbf{y}(w) \textbf{ in Y}$$

The set of constraints h_1 ... h_k describe the links between the first stage decisions x and the second stage decisions $y(w)$. Note that each constraint is required to hold with probability 1, or for each possible w in W. This procedure facilitates making a correction (recourse) to the first stage decision that is the best such correction. Recourse models can be extended to multistage problems, where you make one decision now, wait for some uncertainty to be resolved (realized), and then make another decision based on what happens. The objective is to minimize the expected costs of all decisions taken.

2.3.3 Probabilistically Constrained Models

Holmes (1994) gives a good overview of probability constrained models. In some cases, it may be more appropriate to try to find a decision which ensures that a set of constraints will hold with a certain probability. An example might be a delivery service that experiences random demands, and wishes to find the cheapest way to deliver its packages with a high probability. Again, assume that x is a vector of decisions. The general form of this problem is to the following:

$$\min f(x_i)$$

$$s.t.\quad \mathbf{P}[\mathbf{g}_1(x_i) \le \mathbf{0}, \ldots, \ \mathbf{g}_m(x_i) \le \mathbf{0}] \ge \alpha$$
$$\mathbf{h}_1(x_i) \le \mathbf{0}, \ldots, \mathbf{h}_k(x_i) \le \mathbf{0}$$
$$\textbf{for } \forall\, x_i \textbf{ in } \mathbf{X}$$

2.4 Reliability Analysis

Reliability analysis is a central issue in formulating and solving a system of systems engineering problems because systems of systems must perform reliably under uncertain conditions. However, the computational effort involved in modeling a system of systems may be very expensive. The use of approximate uncertainty analysis techniques which can provide reasonably accurate results with a minimum of computational expense are useful in system of systems engineering and may be an acceptable alternative to expensive simulation-based methods for reliability analysis.

2.4.1 First-Order Reliability Method (FORM)

The first-order reliability method (FORM) is an analytical method used to determine the probability of a function of continuous random variables assuming less than a certain value The probability of a function being less than or equal to zero (failure of a component limit state) is given as

$$P_{Fail} = P\{g(\mathbf{x}) \le 0\} = \int_{g(\mathbf{x}) \le 0} f_{\mathbf{x}}(\mathbf{x}) d\mathbf{x}$$

where $f_{\mathbf{x}}(\mathbf{x})$ is the joint probability density of variables $x_1, x_2 \ldots x_n$.

An analytical evaluation of the above integral is possible in only a few special cases, and hence numerical integration is necessary. FORM has been found to be a computationally efficient and reasonably accurate analytical approximation to the probability integral under consideration for many problems. In FORM, there are three important steps in the calculation of the probability of failure for an individual component failure mode. These are:

1. Transformation of the random variables **x** to the standard normal space **u**.
2. Calculation of the most probable point of failure. This point is the solution to the constrained optimization problem $\mathbf{u}^* = \arg\min(\|\mathbf{u}\| \mid G(\mathbf{u}) = 0)$. Good algorithms for solving this problem have been proposed by (Hasofer and Lind, 1974); (Rackwitz and Fiessler, 1978) and (Der Kiureghian et al, 1994).
3. Calculation of the reliability index β. β is in general equal to $\boldsymbol{\alpha}\mathbf{u}^*$, in which $\boldsymbol{\alpha}$ is the negative normalized gradient row vector of the limit state surface in the **u** space, pointing toward the failure domain. For most practical problems β is greater than zero, in which case β is also equal to $\|\mathbf{u}^*\|$. The probability of failure is approximated as $P_{Fail} = \Phi(-\beta)$.

For more details about the implementation of FORM are found in (Ditlevsen and Madsen, 1996), (Haldar and Mahadevan, 2000), and (Nowak and Collins, 2000).

A very important by-product of FORM is the so-called alpha vector, defined by

$$\boldsymbol{\alpha} = -\frac{\nabla_u G(u)}{\|\nabla_u G(u)\|}$$

The alpha vector is the negative normalized gradient row vector of the limit state function in the transformed space. Also, at optimality in the FORM problem, the alpha vector is collinear with the MPP vector. This vector is important because this vector gives relative importance information about each of the random variables (Ditlevsen and Madsen, 1996), (Haldar and Mahadevan, 2000). This vector can help analysts determine which uncertain parameters are the most important so that information gathering efforts are focused on these variables. Random variables with alpha values of low magnitude can often be modeled as deterministic at the mean.

2.4.2 Inverse FORM

A very important technique in reliability-based design optimization is the inverse FORM method. The inverse FORM problem has a very important role in reliability-based design optimization because it returns a worst-case point at a certain probability level. By obtaining safety at the worst case point for a desired level of reliability, a designer can assure performance of the design at a prescribed level of reliability. The inverse FORM problem solves the optimization

$$\mathbf{u}^* = \arg\min(g(\mathbf{u}) \mid \|\mathbf{u}\| = \beta^t)$$

In this optimization, β^t is a prescribed reliability index. If in the worst case scenario the limit state function of the design is greater than zero, it is obvious that the design has a reliability index greater than the target reliability index.

FORM and inverse FORM share some common concepts, such as the ideas of probability transformation to the standard normal space, the reliability index, and the alpha vector. If the limit state is equal to zero at the solution of the direct and inverse FORM problems, then the direct FORM and inverse FORM have the same solutions. In fact, the inverse FORM solution will always be a FORM solution for a given limit state function plus a constant term. Hence, the alpha vector will be coincident with the MPP vector at optimality for both direct and inverse FORM. Thus, sensitivity results can be obtained from both FORM and inverse FORM. This is an especially useful fact for problems involving policy problems, because inverse FORM can be used to obtain relative sensitivities with respect to random parameters of outcomes such as the number of fatalities or economic losses for which no meaningful, crisp limit state may exist.

2.4.3 System of Systems Reliability

Another important task in uncertainty analysis for system of systems problems is the estimation of system reliability. For purposes of reliability analysis, systems are generally characterized as series, parallel, or general systems. The failure probability for systems in series is calculated

$$P_{Fail,Series} = P\{\bigcup_i g_i(\mathbf{x}) \le 0\}$$

and the failure probability for a system in parallel system is calculated (Haldar and Mahadevan, 2002)

$$P_{Fail,Parallel} = P\{\bigcap_k g_k(\mathbf{x}) \le 0\}$$

Systems can be represented as combinations of series and parallel systems. It is common to evaluate general systems reliabilities using either a cut set formulation or a link set formulation. A cut set is a set of failure events for which occurrence of all the events results in failure of the system. A minimal cut set contains no more than the minimum number of events required for system failure; disjoint cut sets are mutually exclusive sets of events. It is usually easier to define minimal cut sets. A link set is a set of survival events for which occurrence of all the events results in survival of the system. Failure probabilities for a general system can be written as either the union of a set of intersections of events, or by using de Morgan's laws, the intersection of a set of unions of events.

It is often preferable to use the cut set formulation for several reasons. First, it is often easier to identify cut sets than link sets. Secondly, if the cut sets are disjoint, then the failure probability of the general system is the sum of the failure probabilities of the cut sets. Thirdly, failure probabilities for a cut set are easily calculated as a series system failure probability. Finally, bounding formulae applied to a group of minimal cut sets result in conservative estimates of failure probabilities.

As was the case for component reliability, system reliability can be evaluated by simulation methods, but it is computationally expensive to do so. However, system reliability can be approximated with elegant analytical approximations using FORM. Let **B** be the vector of reliability indices for each of the limit states and the elements of the matrix **R** be the dot products of the corresponding $\boldsymbol{\alpha}$ vectors obtained from the FORM analysis for each distress mode. Then for a series system, the system failure probability is given by 1 – Φ(**B**, **R**), where Φ(**B**, **R**) is the standard normal multivariate CDF with correlation matrix **R**. For the bivariate case (Dunnett and Sobel, 1954) shows

$$\Phi(\beta_1, \beta_2, \rho_{1,2}) = \Phi(\beta_1)\Phi(\beta_2) + \int_0^{\rho_{1,2}} \frac{1}{2\pi\sqrt{1-\rho^2}} \exp[-\frac{\beta_1{}^2 + \beta_2{}^2 - 2\rho\beta_1\beta_2}{2(1-\rho^2)}] d\rho$$

If more than two limit states are considered, then one may elect to use bounding formulae such as those in (Ditlevsen, 1979) or evaluate the multinormal CDF using methods such as importance sampling methods in (Mahadevan and Dey, 1998) and (Ambartzumian et al, 1997), multiple linearizations in (Hohenbichler and Rackwitz, 1987), or a moment-based approximation as found in (Pandey, 1998).

2.4.4 Reliability-Based Design Optimization (RBDO)

Formulating and solving problems in system of systems engineering requires integrating system modeling and uncertainty analysis into the decision making process. RBDO is a methodology by which this integration can occur. In an RBDO problem, an objective function is minimized subject to reliability constraints corresponding to various limit states and deterministic constraints. The objective function is a function of the design variables and can include initial costs, failure costs, maintenance costs, structure weight or performance efficiency. The uncertain variables are modeled as random variables. The design variables can be distribution parameters such as means or standard deviations of the random variables or the design variables can be deterministic. RBDO methods fall into three groups depending upon how reliability analysis is incorporated into the optimization process. (Tu et al, 2001) refers to the RBDO methods that use the reliability index directly as Reliability Index Approach (RIA) and those based on quantile functions of the probability distributions as the Performance Measure Approach (PMA).

Nested algorithms, used before the 1990s include a full reliability analysis at every step of the design optimization algorithm. It is well known that nesting these two procedures results in a large number of function evaluations, and studies performed in (Agarwal and Renaud, 2004), (Liang et al, 2004), (Du and Chen, 2004) and (Yang and Gu, 2004) have confirmed that nested methods require many more function evaluations than RBDO methods in which the reliability analysis loop is either decoupled or eliminated via single loop methods.

To reduce the computational expense associated with nested methods, many researchers have developed single-loop approaches to RBDO (Madsen and Hansen, 1992), (Chen et al, 1997), (Wang and Kodiyalam, 2002) and (Agarwal and Renaud, 2004). The methodologies are focused on removing the inner reliability analysis loop by making the optimality conditions of either FORM or inverse FORM constraints in the optimization loop. This general approach using direct FORM is described below.

$$\begin{aligned}
&\min_{\mathbf{d},\mathbf{x}} f(\mathbf{d}) \\
&s.t.\ g(\mathbf{d},\mathbf{x}) = 0 \\
&\frac{\mathbf{u}^*}{\|\mathbf{u}^*\|} = -\frac{\nabla_u g(\mathbf{d},\mathbf{x})}{\|\nabla_u g(\mathbf{d},\mathbf{x})\|} \\
&\|\mathbf{u}^*\| \geq \beta_{required}
\end{aligned}$$

The first and second constraints are simply the necessary conditions for the MPP, the third constraint requires the reliability index for the design to be greater than or equal to the required reliability index. This formulation is the single-loop RIA formulation. If the optimality conditions of the inverse FORM problem are used, a similar formulation can be derived.

$$
\begin{aligned}
&\min_{\mathbf{d},\mathbf{x}} f(\mathbf{d}) \\
&s.t.\ g(\mathbf{d},\mathbf{x}) \geq 0 \\
&\frac{\mathbf{u}^*}{\|\mathbf{u}^*\|} = -\frac{\nabla_u g(\mathbf{d},\mathbf{x})}{\|\nabla_u g(\mathbf{d},\mathbf{x})\|} \\
&\|\mathbf{u}^*\| = \beta_{required}
\end{aligned}
$$

Several researchers have reformulated the nested RBDO problem to decouple the reliability analysis from the design optimization. This has been accomplished in one of two ways, using direct or indirect FORM (i.e., RIA or PMA). Royset et al, (2001) provided decoupled RBDO formulations that use both RIA and PMA approaches. Using RIA, (Torng and Yang, 1993) replaced the probabilistic constraint with its Taylor series expansion to solve:

$$
\begin{aligned}
&\min\ f(\mathbf{d}) \\
&s.t. \\
&\beta(\mathbf{d},\mathbf{x}^*) + \nabla_{\mathbf{d}}\beta(\mathbf{d},\mathbf{x}^*) \geq \beta_{required}
\end{aligned}
$$

Using PMA, (Wu and Wang, 1998), (Wu et al, 2001) and (Du and Chen, 2004) developed decoupled formulations of the RBDO problem as shown.

$$
\begin{aligned}
&\min c(\mathbf{d}) \\
&s.t. g_k(\mathbf{d},\mathbf{x}_k) \geq 0 \qquad \forall k \in K \\
&where \\
&\mathbf{x}_k = \arg\min(g_k(\mathbf{d},\mathbf{x}_k) \mid \|\mathbf{u}_k(\mathbf{x}_k)\| = \beta_{required}) \quad \forall k \in K
\end{aligned}
$$

Zou and Mahadevan (2006) implemented a decoupled formulation in which the probability function is used in place of the reliability index. This formulation has several advantages over previous methods. First, this method can be used to solve problems with systems-level reliability constraints. Secondly, it can handle problems with probability of failure in the objective function. It can also allow for different reliability analysis methods to be used for different limit states.

Most approaches to systems-level reliability analysis are computationally expensive. Furthermore, no proof of convergence exists for either single-loop or decoupled RBDO algorithms (Royset et al, 2001). However, when single-loop methods are successful, they are usually more computationally efficient than decoupled or sequential methods.

2.4.5 RBDO Formulations with Discipline and SoS Reliability Constraints

In order to address the difficulty of including system reliability constraints in RBDO, (McDonald and Mahadevan, 2007) develops an algorithm which can include system reliability constraints in a single-loop formulation for continuous design and random variables. This formulation is similar to previous single-loop methods, but uses the augmented decision space and appropriate multinormal CDF approximations to assure that the system-level probability of failure is less than a specific threshold. In this approach the FORM optimality conditions are satisfied for each component in the system. Although most RBDO methodologies only require the satisfaction of either the direct FORM or the inverse FORM optimality conditions, sometimes both sets of optimality conditions are required.

Component reliability targets are included in the decision space of the optimization problem, along with the design vector and each component limit state MPP. These component reliability targets, along with the correlations, are obtained from the limit state direction cosines $\boldsymbol{\alpha}$, where

$$\boldsymbol{\alpha} = \frac{-\nabla_u G(\mathbf{d},\mathbf{u})}{\left\|\nabla_u G(\mathbf{d},\mathbf{u})\right\|}$$

The component reliability information is then used to calculate the system failure probability, which must be less than the allowable level. This formulation is given.

$$\min_{\mathbf{d},\mathbf{u},\mathbf{B}} f(\mathbf{x})$$

$$s.t.\ g_k(\mathbf{d},\mathbf{u}_k) = 0 \qquad \forall k \in \mathbf{K}$$

$$\frac{\mathbf{u}_k}{\left\|\mathbf{u}_k\right\|} = -\frac{\nabla_u G_k(\mathbf{d},\mathbf{u}_k)}{\left\|\nabla_u G_k(\mathbf{d},\mathbf{u}_k)\right\|} \qquad \forall k \in \mathbf{K}$$

$$\left\|\mathbf{u}_k\right\| = \beta_k \quad \forall k \in \mathbf{K}$$

$$p(\mathbf{B},\mathbf{R}) \leq \hat{p}$$

The objective function lists **B** to denote the entire vector of target reliability indices, whereas β_k refers to the target reliability index for the k^{th} limit state. **R** is the correlation matrix between failure modes, calculated by the equation $R_{ij} = \boldsymbol{\alpha}_i\boldsymbol{\alpha}_j$. It is necessary that **B** be included in the augmented decision space because the optimizer must decide how reliable each component of the system must be. In the constraints, $p(\mathbf{B}, \mathbf{R})$ is the system failure probability (McDonald and Mahadevan, 2007).

McDonald and Mahadevan (2007) show this formulation to be extremely efficient for problems of relatively small size with limit states that do not require Multi-disciplinary analysis. The formulation above is directly extendable to MDO problems solved with the Multi-disciplinary approach.

2.5 System Modeling

System of system engineering requires that the interactions among systems be modeled. This section outlines three popular modeling approaches: Input-Output Analysis, Systems Dynamics, and Agent-Based Modeling.

2.5.1 Input-Output Analysis

The economist Wassily Leontief (1953) proposed a model for economies which are built upon input-output technologies. This model assumes that the production functions for all sectors of the economy are linear and involve the goods produced by other sectors of the economy. The model shows that the vector of final outputs, x, of all sectors required to meet a vector of initial demands, c, can be determined through the matrix equation

$$\mathbf{x} = (\mathbf{I} - \mathbf{a})^{-1}\mathbf{c}$$

where **I** is the identity matrix and a_{ij} is the dollar amount of goods from sector *i* used in production of the goods from sector *j* .

This model has been adapted for other uses besides macroeconomic modeling. Haimes et al (2001) has adapted this model to estimate the consequences to a set of interdependent infrastructures arising from a terrorist attack. In this adaptation, **c** represents a perturbation in the systems' inoperability caused by damage from a terrorist attack, **x** represents the final state of the systems' inoperability, and the a_{ij} relate the extent of inoperability caused to infrastructure *j* to the final inoperability of infrastructure *i.* In the context of this model, inoperability is a continuous variable between 0 and 1 representing the extent of loss of functionality of a system, where a system with inoperability 0 is completely functional and a system with inoperability 1 has failed. The a_{ij} are a measure of how the inoperability of infrastructure *j* is transferred to infrastructure *i.* If a_{ij} = 0, then total failure of infrastructure *j* does not affect infrastructure *i*. If a_{ij} = 1, then if infrastructure *j* fails, then infrastructure *i* will also fail. If a_{ij} = 0.5 then complete failure of infrastructure *j* will cause infrastructure *i* to suffer an inoperability of 0.5. Because inoperability of a system cannot be greater than 1, (Haimes et al, 2001) proposed the solution of the extended Leontief equations, where inoperability is the minimum value of 1 and the result obtained by the Leontief model.

2.5.2 Systems Dynamics

System of systems engineering involves modeling interactions among the member systems. The Leontief input-output model, while simple and easy to construct, cannot account for intricate independencies in a system of systems environment. For this reason, systems thinking and systems dynamics modeling are two useful tools for modeling and studying the emergent behaviors inherent to system of systems engineering problems.

Systems thinking is a mental model that promotes the belief that the component parts of a system will act differently when isolated from their environment or other parts of the system. It includes viewing systems in a holistic manner, rather than through purely reductionist techniques because often in systems the behavior of parts can only be understood in the context of the behavior of the whole. It promotes gaining insights into the whole by understanding the linkages and interactions between the elements that comprise the whole "system," consistent with systems philosophy (Gharajedaghi, 2005).

System thinking recognizes that all human activity systems are open systems; therefore, they are affected by the environment in which they exist. System thinking recognizes that in complex systems events are separated by distance and time; therefore, small catalytic events can cause large changes in the system. System thinking acknowledges that a change in one area of a system can adversely affect another area of the system (Gharajedaghi, 2005).

System dynamics is one approach to modeling the dynamics of complex systems such as population, ecological and economic systems, which usually interact strongly with each other. What makes using System Dynamics different from other approaches to studying complex systems are the use of feedback loops (Forrester, 1961). Stocks and flows are the basic building blocks of a System Dynamics model. They help describe how a system is connected by feedback loops which create the nonlinearity found so frequently in modern day problems. Computer software is used to simulate a system dynamics model of the situation being studied and can allow for a graphical specification of the interdependencies among system variables and can simulate the evolutionary behavior of systems by solving the system of equations.

Model simulations can greatly aid in understanding how the system changes over time, and allow for the understanding of the importance of various uncertainties and the understanding of how best to control the behavior of the system.

2.5.3 Agent-Based Modeling

System of systems, by definition, has multiple decision makers, and often these systems have many humans in the loop. In assessing policies and strategies for systems of systems, it is necessary to model the behaviors of decision making entities. Current research efforts in system of systems engineering problems includes agent-based modeling (Crossley, 2004). Agent-based modeling is described in (Axelrod, 1997) as a specific individual-based method for computer simulation with the intent to construct computational devices (known as agents with some properties) and simulate them in parallel to model the phenomena. The process is one of emergence from the micro-level of the social system to the higher macro-level. The predominant methodological approach to research involving computational modeling characterizes most systems with respect to equilibrium. Agent based modeling, by using simple rules, can result in far more complex and interesting behavior.

Agent based models consist of dynamically interacting rule-based agents. The systems within which they interact can therefore create complexity like that which is present in the real world. According to (Page, 2005), agents are: 1) intelligent and purposeful, but they are not so smart as to reach the cognitive closure implied by game theory; and 2) situated in time and space. They reside in networks and on lattice-like neighborhoods. The situation of the agents and their behavioral rules are encoded in algorithmic form in computer programs. The modeler makes the assumptions most relevant to a given situation and then watches phenomena emerge from the interaction of the agents. Sometimes the result of the agent-based modeling process is equilibrium. Sometimes it is an emergent pattern. Sometimes, however, the result can be unintelligible chaotic behavior.

On some levels, agent based models complement traditional analytic methods such as systems dynamics. However, unlike other methods which focus on the characterization of equilibrium states, agent-based modeling allows analysts to explore the generation of equilibrium states. This type of study explores why certain complex phenomena exist in social systems, and this is one of the most important contributions of agent-based modeling. A more complete explanation of agent-based models is found in (Axlerod, 1997), (Arthur et al, 1997) and (Prietula et al, 1998).

2.6 Game Theory

Game theory was presented briefly in Section 2.1 as part of the five signature areas of system of system research. Game theory studies situations where players choose different actions in an attempt to maximize their returns. This study of the interactions of decision makers is central to formulating and solving system of systems problems. The field came into being with (Morgenstern and von Neuman, 1947). It provides a formal modeling approach to social situations in which decision makers interact with other minds. Game theory extends the simpler optimization approach developed in neoclassical economics.

Equilibration models are rooted in Game Theory (Sheffi, 1985). A game consists of a set of players and a set of rewards for each player for each combination of strategies selected by the players. A game has an equilibrium strategy if and only if there is a strategy in which no single player can be made better off by switching strategies unilaterally (Gibbons, 1992). This principle is very important in game theory, and it has found extensions into other fields. For instance, in the field of urban transportation planning, the user equilibrium traffic flow pattern is the pattern of traffic flow where no traveler is able to reduce their own travel time by unilaterally switching routes in a transportation network (Wardrop, 1952).

Although some game theoretic analyses appear similar to decision theory, game theory studies decisions made in an environment in which decision makers interact. In other words, game theory studies choice of optimal behavior when costs and benefits of each option depend upon the choices of other individuals. This makes game theory a very important tool in making decisions in a system of systems context since systems of systems have managerial independence with operational interdependence (Sage and Cuppan, 2001).

2.7 Uncertainty Analysis

Because uncertainty is present in virtually all SoS, analyzing uncertainty is important for solving SoS problems. The concepts comprising uncertainty analysis will be discussed in the following three categories: representing uncertainty, quantifying uncertainty and propagating uncertainty.

2.7.1 Uncertainty Representation

Uncertainty is traditionally represented probabilistically as the measure of error associated with an aspect of a system. Haldar and Mahadevan (2000) describe uncertainty, specifically in the context of randomness or stochasticity, as the occurrence of multiple outcomes without any pattern. However, some quantities in a system model may not have a probabilistic representation since data may be sparse or may be based on expert opinion. Representations such as fuzzy sets, evidence theory etc. are used, leading to interval analysis of the system model. Transformations have been proposed from non-probabilistic to probabilistic format, through the maximum likelihood approach (Ross et al, 2002). Such transformations have attracted the criticism that information is either added or lost in the process (Cooper et al, 1996).

Most studies only incorporate the first type of uncertainty, namely, physical or inherent variability. Alternatives such as those in Mehta et al (1992) used Bayesian techniques to consider data uncertainty in developing confidence bounds on the output through a nested computation, using either Monte Carlo or analytical methods. These confidence intervals can be combined with probabilistic sensitivity information to increase the system's robustness (Stoebner and Mahadevan, 2000).

Representations of epistemic uncertainty also continue to be explored. Ferson et al (2004) presented challenge problems to motivate research in this direction. One developing technique at Vanderbilt University involves interval representations of uncertainty as a probabilistic approach to represent interval data for input variables in reliability and uncertainty analysis problems, using flexible families of continuous Johnson distributions.

2.7.2 Uncertainty Quantification

Uncertainty quantification is the quantitative characterization and reduction of uncertainty in applications. Three types of uncertainties commonly appear in literature. The first type is uncertainty due to variability of input and/or model parameters and the characterization of the variability is given (i.e., probably density functions). The second type is similar to the first type except that the variability characterization is not fully available. The third type, which is the most challenging, is modeling uncertainty due to an unknown process or lack of knowledge (Tong, 2008).

Both classical and Bayesian statistical analyses are used to first quantify physical variability. Mahadevan et al (2001) and (Mahadevan and Rebba, 2005) suggest that the Bayesian approach is quite valuable when there is only little data. For any random variable that is quantitatively described by a probability density function, there is always uncertainty in the corresponding distribution parameters due to small sample size. As testing and data collection activities are performed, the state of knowledge regarding the uncertainty changes, and a Bayesian updating approach can be implemented.

The quantification of model prediction uncertainty is due to multiple sources: physical variability, inadequate data, measurement errors, and modeling errors. Uncertainty and error quantification is an important challenge in multi-disciplinary analysis and optimization, and has to address uncertainties in physical, data, and modeling uncertainties. Modeling errors may relate to governing equations, boundary and initial condition assumptions, loading description, and approximations or errors in solution algorithms. Model errors will be quantified by comparing model prediction and experimental observation, properly accounting for uncertainties in both. Numerical errors in model predictions are typically quantified first, using sensitivity analysis, uncertainty propagation analysis, discretization error quantification, truncation (residual) error quantification, etc. The variability in experimental measurement can also be quantified. The model form error can be quantified based on all the above errors, following the approach in (Mahadevan and Rebba, 2005).

Finally, the contributions of different sources of uncertainty and error can be integrated to quantify the uncertainty in Multi-disciplinary system response. Obviously, the combination of various sources of error is nonlinear and not straightforward. An effective approach is to use a Bayes network, where individual contributions can be mapped through conditional probability relationships, and the overall effect can be quantified by integration through the Bayes network. This approach is explained more fully in (Mahadevan et al, 2001).

2.7.3 Uncertainty Propagation

Various methods are available to compute the uncertainty in the system response due to uncertainties in the input quantities, but these techniques need to be explored and developed for multi-disciplinary systems. Probabilistic techniques have been pursued extensively, by modeling the inherent variability through random variables. These methods, along with Bayesian techniques, can also consider uncertainty in the statistical distributions that derives from lack of adequate data. Non-probabilistic, interval analysis has been pursued to deal with imprecise information using existing probabilistic approaches and empirical distribution functions. In addition, the errors due to mathematical modeling and numerical approximation and discretization need to be quantified. Thus, a comprehensive strategy is needed for computing the total uncertainty in modeling and simulation, and to determine the contribution of each type of uncertainty to the overall uncertainty. A survey of common uncertainty propagation methods are presented in (DeLaurentis, 2000). Literature from the Sandia National Laboratory also describes the various methods used in (Oberkampf et al, 1998). The discussion in this research is restricted to the probabilistic approaches of Monte Carlo simulation and moment-based methods.

Monte Carlo Simulation

Monte Carlo simulations are the most accurate method of uncertainty propagation. For computationally expensive simulations, a surrogate model can be generated for various confidence levels and then the Monte Carlo simulations can be performed. As a first step in probabilistic modeling, Monte Carlo simulation is commonly used. However, the basic Monte Carlo method is too time-consuming to achieve acceptable accuracy for low probability events. Several efficient sampling schemes and variation reduction techniques such as Latin Hypercube Sampling and adaptive importance sampling have been developed (Mahadevan and Dey, 1997).

Moment-based Methods

An attractive alternative to Monte Carlo simulation is to use analytical approximations that combine probability theory and optimization methods (Rackwitz and Fiessler, 1978), (Cruse et al, 1988). These are based on first-order or second-order approximations of the system performance constraint equations (limit states), and analytical estimation of the reliability through the identification of the most probable combination of the random variables that cause the system to be at the limit state.

The simplest and most commonly used method is the method of moments (Putko et al, 2001). Method of moments is computational much cheaper than the full non-linear Monte Carlo simulations. Traditionally, only the first order moments are available for full non-linear CFD calculations. The accuracy of the method may be improved by using higher order derivatives. Calculation of higher order derivatives is computationally expensive and no known automatic differentiation packages can calculate more than first order derivatives. Use of adjoints has been successfully demonstrated in propagating first order derivatives (Alekseev and Navon, 2003).

Uncertainty propagation is often performed with the aid of surrogate models to create response surfaces. There are several surrogate modeling techniques, such as Gaussian Process modeling in which response variables are modeled as a group of multivariate normal random variables and polynomial chaos first introduced in (Wiener, 1938) as "Homogeneous Chaos." Surrogate modeling and its uses are discussed in more detail in Section 2.9.

2.8 Uncertainty-Based Design

Methods for design optimization under uncertainty can be divided into three classes: sampling methods for expected value optimization problems, robust optimization, and reliability-based design optimization. Sampling methods can be used to solve either robust optimization or reliability-based optimization problems, but it is useful to discuss these methods separately. The sampling methods perform all experiments (whether mathematical simulations or physical tests) simultaneously and then optimize the design based on the results of those experiments. Robust optimization methods use the numerical optimization procedure to specify which simulations are needed and evaluate those simulations one at a time. Reliability-based design optimization (RBDO) methods also use numerical optimization procedures but with the goal of reliability rather than robustness.

There are several existing RBDO methods, each with their own relative strengths and weaknesses. Likewise, there are several options for formulating and solving MDO problems [including Multi-disciplinary Feasibility (Kodiyalam, 1998), Interdisciplinary Feasibility (Cramer et al, 1994), Simultaneous Analysis and Design (Cramer et al, 1994), and Collaborative Optimization (Braun, 1996)].

For nonlinear problems the expected value of the objective function is, in general, not the value of the function evaluated at the expected values of its inputs, so sampling-based approaches are used to compute the expectation. A robust design problem seeks a solution that is relatively insensitive to small changes in the uncertain quantities. A reliability-based design seeks a solution that has a probability of failure that is less than some acceptable value.

Traditional design procedures are based on combinations of factors of safety and knockdown factors. The aerodynamic design procedures used by the industry are exclusively deterministic (Zang et al, 2002). There has been considerable work on "robust controls," but this work has been limited to using interval bounds on the uncertain variables (Ham et al, 2000). Reliability-based design methods have been used within civil engineering for several decades, as noted in (Haldar and Mahadevan, 2000).

To use uncertainty-based design methods, the various uncertainties associated with the design problem must be characterized and managed, and these characterizations must be exploited. Uncertainty in engineering analysis and design arises from several sources (Oberkampf et al, 1999). Some of the "known" sources are: (1) physical uncertainty or inherent variability, (2) informational uncertainty or statistical uncertainty, and (3) modeling error. Uncertainties are typically specified in terms of probability density functions, membership functions, or interval bounds. Better and less resource-intensive methods are needed for both uncertainty propagation and optimization under uncertainty.

There are clear benefits in using uncertainty-based design methodologies. First, confidence in analysis tools will increase as physical, model, and data uncertainties are systematically addressed. Design cost, risk, and cycle time would be reduced as models could be selected with an understanding of the model uncertainty so that computational overkill can be reduced. System performance would increase while assuring reliability of systems because optimization methods would lead to minimum cost designs and optimal flow down of system reliability requirements to the component level. Design will clearly be more robust as a wider array of situations is considered in the design process. Finally, uncertainty-based design can allow the designer to consider extreme circumstances and can allow for planning that accommodates the worst possible conditions.

Regarding MDO, there are several examples of previous research in Multi-disciplinary optimization under uncertainty. Mahadevan and Smith (2006) developed very efficient methods for reliability estimation of Multi-disciplinary systems. Chiralaksanakul and Mahadevan (2004) applied decoupled reliability-based design optimization techniques to several different Multi-disciplinary optimization algorithms. Smith and Mahadevan (2007) implemented several formulations of reliability-based Multi-disciplinary optimization on a few simple problems. Smith and Mahadevan (2007) optimized the integration of component and system design under uncertainty for an aerospace vehicle.

While the benefits of uncertainty-based design are clear, there are difficulties which need to be addressed if these benefits are to be realized. Current uncertainty-based methods are more complex and significantly more computationally expensive than deterministic methods and more efficient methods of performing uncertainty-based design are clearly needed. It is often too computationally expensive to do probabilistic design with high fidelity models. Multi-disciplinary analysis compounds the problems of computational effort, since such problems must generally be solved iteratively. McDonald and Mahadevan (2008) have made progress in this area with a novel single loop method for solving RBDO problems with system and component reliability, as well as, extending it to discrete and continuous design and random variables.

2.9 Surrogate Modeling

The most straightforward approach for uncertainty propagation is known as Monte Carlo simulation. Monte Carlo simulation is a sampling based approach, in which a large number of random realizations of the input parameters are generated, and the simulator is run for each sample. The output samples are then used to make inference about the model output distribution.

The problem with basic Monte Carlo sampling, however, is that it requires a very large number of evaluations of the computer simulation in order to accurately characterize the output distribution, and in practical applications, obtaining this number of evaluations is often not feasible. For this reason, there are several techniques available for reducing the variance in sampling-based estimators. McKay et al (1979) describes latin hypercube sampling as an approach that has the goal of attaining a more even distribution of the sample points in the parameter space. When reliability estimation is of interest, importance sampling (Haldar and Mahadevan, 2000) is popular. Importance sampling uses a sampling density that is concentrated in the failure region, so that samples are evaluated efficiently in the parameter space.

An alternative to efficient sampling techniques is to use an inexpensive approximation to the input/output relationship in lieu of the expensive computer simulation. Such approximations are often called surrogate models, or response surface approximations. A variety of methods are available for developing response surface approximations, including the development of models with reduced degrees of freedom, polynomial regression, multivariate adaptive regression splines (Friedman, 1991) and (Schumaker, 2007), neural networks, non-intrusive polynomial chaos (Isukapalli et al, 1998), and Gaussian process interpolation. The use of Gaussian process interpolation for surrogate modeling has been of particular interest within the scientific community for studies involving both uncertainty quantification and optimization. Examples include (Bichon et al, 2008); (Jones et al, 1998); (Kennedy and O'Hagan, 2001); (Bayarri et al, 2002); (Simpson et al, 2001); (Kaymaz, 2005); (Kennedy et al, 2006); (Oakley and O'Hagan, 2002) and (McFarland et al, 2008).

2.9.1 Surrogate Multi-disciplinary Analysis

Gaussian process models have several features which make them an attractive choice for a surrogate model in the context of multi-disciplinary analysis. The primary feature of interest is the ability of the model to "account for its own uncertainty" (Kennedy and O'Hagan, 2001). That is, each prediction obtained from a Gaussian process model also has an associated variance, or uncertainty. This prediction variance primarily depends on the closeness of the prediction location to the training data, but it is also related to the functional form of the response. Figure 2.1 depicts a Gaussian process model. The uncertainty bounds are related to the closeness to training points and to the curve's shape.

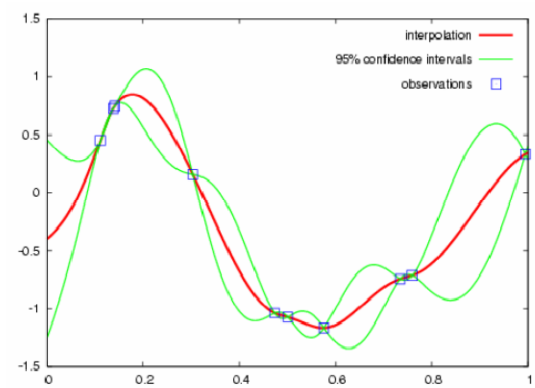


Figure 2.1. Example Gaussian Process Model with Uncertainty Bounds

The basic idea of the Gaussian process model is that the response values are modeled as a group of multivariate normal random variables. A parametric covariance function is then constructed as a function of the inputs. The covariance function is based on the idea that when the inputs are close together, the correlation between the outputs will be high. As a result, the uncertainty associated with the model's predictions is small for input values which are close to the training points, and large for input values which are not close to the training points. In addition, the GP model may incorporate a systematic trend function, such as a linear or quadratic regression of the inputs (in the notation of Gaussian process models, this is called the mean function, while in Kriging it is often called a trend function). The effect of the mean function on predictions which interpolate the training data is small, but when the model is used for extrapolation, the predictions will follow the mean function very closely. Gaussian Process models are used to build surrogate physical disciplinary models over a design and random space. Gaussian process models are also able to characterize its own uncertainty in order to quantify the model error associated with the surrogate model.

2.9.2 Reliability Analysis using Gaussian Process Models

As engineering applications become increasingly complex, they are often characterized by implicit response functions that are both expensive to evaluate and nonlinear in their behavior. Reliability analysis given this type of response is difficult with available methods. Current uncertainty analysis methods focus on the discovery of a single most probable point of failure, and then build a low-order approximation to the limit state at this point. This creates inaccuracies when applied to engineering applications for which the limit state has a higher degree of nonlinearity or is multimodal. Sampling methods, on the other hand, do not rely on an approximation to the shape of the limit state and are therefore generally more accurate when applied to problems with nonlinear limit states. However, sampling methods typically require a large number of response function evaluations, which can make their application infeasible for computationally expensive problems. Bichon et al (2008) presents a promising technique based on Gaussian Process interpolation as one approach for efficient global reliability analysis.

2.10 Transportation Theory

In mathematics and economics, Transportation Theory is the name given to the study of optimal transportation and the allocation of resources. The portion of transportation literature that is most relevant to this research is optimizing and controlling a transportation network.

Highway traffic operations are influenced by the behavior of drivers. A highway can be used by a finite number of vehicles, and the driver perceived safe distances between vehicles determine this limit. For a given speed, as distances become shorter, more vehicles can use the highway. Both the volume of drivers choosing to use the highway (demand) and the maximum volume that can be served (supply) depend on driver behavior. Congestion results from too many people attempting to reach their destinations at the same time using the same highways. The combination of demand, capacity, and certain infrastructure features determines how drivers perceive the traffic conditions. Transportation agencies strive for economical solutions to congestion that satisfy a majority of highway users.

Let us assume that all vehicles move at the same speed S. The time headway between two consecutive vehicles h is the distance x between these two vehicles divided by the speed S: $h = x/S$. The same relationship holds for average values too (i.e., $\bar{h} = \bar{x} / S$). The reverse of average intervehicle time is volume V and the reverse of average intervehicle distance is density D. Thus, the following relationship is obtained, $V = S \cdot D$, implying volume equals speed times density (this relationship is called the fundamental traffic flow relationship). The fundamental diagram is depicted in Figure 2.2, where a flow (q) is a function of density (k) and the slope from the origin is speed (Daganzo, 2008).

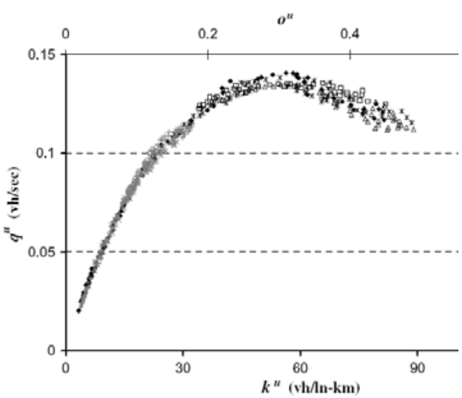


Figure 2.2 Fundamental Diagram of Traffic Flow

SoS problems present major technical challenges for which current solution methods and problem solving approaches are inadequate. Consider, for example, SoS problems related to controlling road traffic on transportation networks. Papageorgiou et al (2003) reported on a team effort to identify directions for future transportation research. Specific to road traffic control strategies, the following points were presented.

- Operational control systems are the exception, rather than the rule
- Employing optimal control algorithms can dramatically improve freeway congestion
- Substantial improvements are achievable via modern traffic control methods and tools
- Improvements are possible at the network-wide level

The vision for this type of traffic management in literature is, “an integrated control of freeway networks involving both ramp metering and route guidance. Very preliminary measures are currently in use, but a lot more developments are required to produce integrated control strategies that are efficient and applicable to large networks or in real-time.” (Papageorgiou et al, 2003).

2.11 Logistic Regression

Logistic regression, also referred to as the logistic model or logit model. It is a statistical model used for predicting the probability of an event occurring by fitting the data to a logistic function (Agresti, 2002). In the early 1940s, the mathematical concept of a probit, short for probability unit, was a prominent statistical scale for normal deviates based on the normal distribution. Berkson (1944) advanced the idea by proposing the use of the logistic function instead of the normal probability function, coining the term logit by analogy to the probit of Bliss (1934). For the inverse of the logistic function, the dependent variable is a logit (the natural log of the event odds), as shown.

$$\log(odds) = \log it(P) = \ln\left(\frac{P}{1-P}\right)$$

and if log(odds) are linearly linked to the independent variable X, then the relation between X and P is non-linear with an S-shaped function as shown in Figure 2.3 (Brannick, 2007).

$$\ln\left(\frac{P}{1-P}\right) = a + bX$$

$$or \quad P = \frac{e^{a+bX}}{1+e^{a+bX}}$$

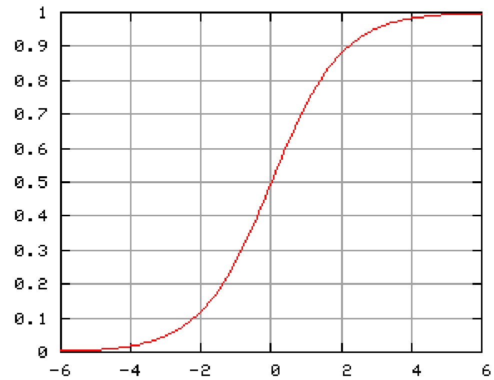


Figure 2.3 Logistic Regression Model

User equilibrium is commonly modeled in accordance with a logistic regression model. In this research domain, an example application would be to model user choice among drivers traveling in a transportation network with multiple modes and routes. Inputs such as perceived travel time for each option would generate corresponding proportions which serve as likelihoods for the user selecting each mode and route.

Over the last decade, logistic regression has become a standard method of analyzing model relationships with discrete responses (Hosmer and Lemeshow, 2000). It is appropriate for data in which there is a binary (success/failure) response variable, such as the discrete response variable in this study problem, network failure. Unlike linear regression, where one estimates the relationship between predictor variables and an outcome variable, logistic regression estimates the conditional probability that a dichotomous outcome occurs. (Hilbe, 2009). The general form for the logistic model is logit(π) = log(odds) = $\alpha + \beta X$, where

$$\pi = P(Y = 1 \mid X) \text{ and is given by } \frac{e^{\alpha+\beta X}}{1+e^{\alpha+\beta X}}$$

Evaluation of this parameterized model yields a probability that the outcome occurs. For example, if the outcome is defined as the network reaching a failed state, then π is one way to represent the failure probability for the network.

2.12 Probabilistic Reachability

Reachability analysis in discrete, continuous or hybrid systems seeks to partition states into two categories: those that are reachable from the initial conditions, and those that are not (Mitchell et al, 2001). The concept of probabilistic reachability centers around determining the probability of reaching a given system state from a given set of initial conditions and subject to a given control. Reachability is an important topic in classical control theory (Abate et al, 2008). For deterministic problems, reachability is a yes/no problem evaluating whether starting from a given set of initial states the system will reach a certain set or not. In stochastic problems, the different trajectories originating from each initial state have likelihoods of reaching the set (Abate, 2007).

System evolution in stochastic human-physical networks is influenced by control policy, so a SoS priority is to choose appropriate controls to minimize the probability that the state of the system will enter the failed state. Given the complex dynamics of practical applications, approximations are needed for reachability computations. Various approximation approaches are proposed in the literature including: over-approximations by ellipsoids (Kurzhanski and Varaiya, 2002), polyhedral (Asarin et al., 2003), and oriented rectangular polytopes (Stursberg and Krogh, 2003), (Yazarel and Pappas, 2004).

Abate (2007) describes the approach presented in (Girard et al, 2006) as a valid approach for hybrid systems. The approach uses an abstraction of the original problem that can be represent and propagate the system dynamics. This general approach is used in this research to quantify the likelihood of a stochastic human-physical network with hybrid dynamics reaching a failed state.

For this problem, the system is represented by a high fidelity computer simulation. Support for reachability analysis in optimally controlling deterministic problems has been pointed out in (Hedlund and Rantzer, 2002) and (Lygeros, 2004). Connections between reachability, and safety for deterministic hybrid systems (mostly applied to air traffic management) has been stressed in (Mitchell et al., 2005) and (Lygeros et al, 1999), and (Tomlin et al, 1998). Reachability for stochastic hybrid systems, such as the class of problem presented in this research, is a recent focus of research. Bujorianu and Lygeros (2003) address theoretical issues regarding the measurability of the reachability events. However, even the most recent approaches consider the problem of reachability analysis for continuous time stochastic hybrid systems without any control input.

CHAPTER III

THE NECESSITY OF A SYSTEM OF SYSTEMS APPROACH TO OPTIMIZING INDIVIDUAL FLOW SYSTEMS

3.1 Introduction

Travelers generally expect a transportation system to be efficient and reliable, yet on many urban freeways, demand often exceeds capacity. In order to address this problem, transportation engineers manage freeway congestion with strategies aimed at controlling the flow of traffic (Daganzo et al, 2002). Of the numerous freeway traffic control strategies employed, the most prevalent strategy is ramp metering. A ramp metering scheme consists of controlling on-ramp fluxes with metering lights which delay the entrance of cars onto the highway. The intent of ramp metering is to improve operating conditions *on the highway* by restricting the entrance of vehicles. Ramp metering decision-makers seek to improve performance by controlling the rate of flow in order to optimize some system-level performance measure. Ramp metering improves freeway traffic flow by distributing traffic over time and space to avoid saturation pressure on bottlenecks (Jin and Zhang, 2001). When demand pressure is low, ramp metering can completely eliminate freeway congestion. However, when demand pressure is high, traffic engineers must prioritize between a “freeway-first” policy with heavily metered ramps and a “balanced” policy considering the interests of traffic on both the freeway and feeder streets. Current ramp metering algorithms usually give priority to freeway traffic (Jin and Zhang, 2001).

Several ramp metering optimization schemes are used in practice, including fixed time strategies found in (Wattleworth, 1965), (Yuan and Kreer, 1971), (Tabac, 1972), (Wang, 1972), (Wang and May, 1973), (Cheng et al, 1974), (Schwartz and Tan, 1977), and (Bayen et al, 2004) and reactive metering strategies such as SWARM (NET, 1996), ALINEA and METALINE (Papageorgiou and Kotsialos, 2000). However, (Muñoz and Daganzo, 2000) and (Cassidy et al,

2002) show that diverge bottlenecks (exit ramps) can create significant delay. Cassidy (2002) shows some of the existing ramp metering algorithms to even increase total delay which can lead to unintended consequences due to underestimation of the effect of poor queue discipline at bottlenecks (Cassidy, 2002).

These queue discipline problems arise because drivers who wish to exit at the diverge bottleneck will not merge until the last instant because the left hand lanes offer better progression. When these drivers cannot merge right, they slow down and impede the progression of the left (and right) hand lanes. Therefore, the percentage of the traffic stream headed for the diverge bottleneck has a large impact on the capacity of the freeway section immediately before the diverge bottleneck. Further, some metering schemes restrict drivers from entering the freeway immediately before the bottleneck, resulting in longer queues at on-ramps near bottlenecks clogging surface streets, which can pose additional challenges in urban areas (i.e., impeding access for emergency vehicles or delaying city busses with many passengers), as well as an increase in the proportion of drivers headed for the bottleneck exit.

The conclusions of (Cassidy, 2002) support the hypothesis that the formulation of a ramp metering optimization problem can influence the resulting metering scheme and, therefore, its subsequent effectiveness as measured by some system-level metric. Two metrics commonly used to evaluate the effectiveness are total system delay (Cassidy, 2002) and total vehicle miles traveled (VMT) (Bayen et al, 2004). As measures of effectiveness, delay is a measure of travel time beyond free flow trip time and preferred to be minimized; VMT is a measure of network flow and preferred to be maximized. Cassidy's results lead to several related questions:

1) What is the most appropriate performance measure to use for optimizing metering strategies?

2) What consequences are likely to arise from using various performance measures?

3) What impacts do diverge bottlenecks have on optimal metering strategies? What can be done about them?

The first task in this objective is to determine the extent to which the choice of system-level measures of performance impacts optimal ramp metering control strategies and to shed further light on the benefits and unintended consequences of "freeway first" control strategies, particularly on freeway sections with diverge bottlenecks. The second task of this objective is to offer an expanded PDE constrained formulation that considers the additional reduction in capacity at the bottleneck caused by poor queue discipline at the bottleneck. An example of poor queue discipline at diverge bottlenecks is drivers entering the passing lane to advance along the queue and exit closer to the off-ramp. This action slows or even clogs the passing lane to through traffic, causing what amounts to further reduction in the capacity at the bottleneck. The impact of poor queue discipline at the bottleneck propagates backward and slows traffic upstream much more than current models consider. The expanded formulation accounts for flow bound for an exit at which a bottleneck would activate and is shown to produce more conservative ramp metering policies that appear to be more realistic.

This objective also investigates the sensitivity of the optimal ramp metering strategy to the choice of the objective function. An alternate formulation is proposed to improve ramp metering results by explicitly including the capacity reduction of the diverge (off-ramp) bottlenecks in the optimization problem. A model formulation is presented that uses two different measures of performance to solve a 26-link highway network topology found in (Bayen et al, 2004) and compare the results. Specifically, this objective examines a ramp metering problem based on the physics of highway traffic flow expressed in the Lighthill-Whitham-Richards (LWR) partial differential equation (PDE) and measured on the basis of total VMT and total delay. The VMT-based model optimizes total number of vehicle miles traveled on a series of freeway sections. The delay-based model optimizes total system delay calculated as the difference between total actual travel time and the corresponding free flow travel time at the speed limit. These models constrain the control strategy such that no backward waves propagate, ensuring gridlock does not occur. This assures the LWR PDE is satisfied for the steady-state steady-flow conditions. This allows us to simplify the problem by considering on-ramp flow fixed over the analysis period and focus the study on the properties of optimal solutions derived by various formulations.

In summary, this objective models the physical flow of a multi-network system and examines the extent to which operational controls can be employed such that appropriate performance measures are optimized and unintended consequences are minimized. Simplified optimal ramp metering formulations are examined based on different objective criteria and accounts for reduced mainstream capacity due to diverging traffic (off-ramps). Specifically, this objective examines optimal ramp metering formulations and offers improvements using mathematical programming approaches to determine optimal policies for controlling the flow of highway traffic in order to optimize a system-level performance measure. This objective investigates the sensitivity of ramp metering strategies to choices of performance measures in the objective function.

This chapter is organized as follows: Section 3.2 describes the PDE based optimization approach to the ramp metering problem; Section 3.3 describes the two candidate model formulations for performing ramp metering optimization; Section 3.4 provides the numerical results for the two formulations; Section 3.5 contains the expanded formulation that is more inclusive of the dynamic nature of the highway physics; Section 3.6 discusses the results of the numerical illustrations; and Section 3.7 provides conclusions.

3.1.1 Background

The first attempts to optimally meter freeways were developed in the 1960s and 1970s. Sheffi (1985) further developed the inputs and link performance models including a discussion of speed-flow-density relationships for general freeway segments. Approaches in the review of relevant literature include ALINEA, METALINE, and PDE-Based control formulations, as well as to recent critiques of these types of strategies. These approaches can lead to suboptimal performance when queue discipline effects are not properly considered. The objective illustrates how common optimization formulations for flow networks are based on mathematical assumptions in modeling which an optimizer will exploit. The resulting design may cause unintended consequences. A simplified formulation of a LWR PDE-based approach to optimal ramp metering illustrates what the consequences of this phenomenon are. ALINEA and METALINE, as well as other approaches based on direct control of the LWR PDE have similar weaknesses as described in (Cassidy, 2002).

One of the major objectives of ramp metering is to influence routing. Commonly used assumptions, such as those in (Payne and Thompson, 1974) are too simplistic, in that they assume that users are either routed directly into the freeway or into an equivalent surface street. This is obviously not realistic, as drivers will use the freeway and will wait in queues on the on ramps and feeder streets in order to use the freeway. Governing queuing theory equations, such as those in (Tarko, 2003) are applied independent of the impact of the control policy on queue discipline. This vulnerability is mitigated in this research by explicitly considering the resulting capacity reduction at diverge bottlenecks.

Another concern in optimizing ramp metering control is modeling reasonable flow phenomena. This research follows the process as included in (Bayen et al, 2004) where the merge phenomenon is considered implicitly in the choice of fundamental diagram. Capacity reduction may be modeled by introducing a bottleneck-bound flow, α_i , which reduces the flux of each increment on link *i* universally by $(1- \alpha_i)$. This objective only considers steady-state, steady-flow ramp metering schemes, thus the calculation of alpha is straightforward. Alpha becomes the flux-weighted average of the fraction of cars at each on ramp with a destination of the bottleneck.

A distance based formulation to maximize vehicle miles traveled (VMT) and a time based formulation to minimize network delay is presented and sensitivities analyzed. By formulating and solving problems with different objectives, two different methods of exploitation are observed. In the VMT-based formulation, upstream progression is maximized by restricting ALL traffic necessary to prevent queues before the bottleneck at the ramp closest to the bottleneck. However, when optimizing with a delay metric, the optimal metering scheme equates the marginal delay on the freeway with the marginal delay on the on ramps.

3.1.2 Optimization Method

The optimization method used to solve this non-linear problem is the generalized reduced gradient method. By simplifying the optimization problem to steady-state, steady flow (SSSF) solutions of the LWR PDE, it is possible to solve it as a static nonlinear programming problem by any standard NLP method. The results prove the exploitation of the LWR equation by the optimizer to alleviate the bottleneck by metering the on ramps closest to the bottleneck to maximize progression upstream, or with equating marginal delays on the freeway and on ramps. The implication of these optimality principles is that when drivers headed beyond the bottleneck are metered in favor of those who will take the exit at the diverge bottleneck and create queue discipline problems, the model based optimal solution is not optimal.

To address this issue, an expanded formulation of the optimal ramp metering problem is developed which explicitly includes the impact of capacity reduction due to poor queue discipline at diverge bottlenecks into the optimization problem. Using synthetic data, it is shown how traditional optimal ramp metering formulations may inadvertently cause traffic problems for both the highway and associated surface streets. It is then shown how accounting for the poor queue discipline that undermine current strategies will lead to control policies based more realistic system physics that will improve network performance. A side by side fundamental diagram and time-space diagram depicts the solution of the LWR equation inclusive of the reduction in capacity at the bottleneck.

3.1.3 Assumptions and Limitations

Several key limitations and assumptions support the work presented in this objective. One important assumption is that it is preferable to operate in the unqueued portion of the fundamental diagram because for equal flows, the speed obtained in the unqueued portion is higher than that obtained in the queued portion. Only when freeway queue storage is desired should a strategy involving queued states of the fundamental diagram be considered. Two simplifications in the research formulations are disallowing shock waves to form on the mainstream and no consideration of on-ramp queue constraints. For the purposes of this portion of research, those who wish to use the on-ramp are assumed to be willing to wait in queue without switching routes.

An assumption for each system is that the "traffic physics" is modeled in accordance with the Lighthill-Whitham-Richards (LWR) partial differential equation (PDE) model (Richards, 1956) and (Lighthill and Whitham, 1956), which describes the evolution of the car density on the highway using a PDE. This PDE relates the time derivative of the car density to the space derivative of the flux function, where the flux function is an empirically determined function which relates the number of cars traveling through a given section of the highway per unit of time to the local car density. From the constraints in the model that the flow for each section is less than the capacity flow, steady-state, steady flow (SSSF) conditions are imposed, yet the LWR PDE is still satisfied.

Another assumption is that the fundamental unit of analysis is the highway and its associated ramps. This assumption allows for clear illustration of the operational implications of a model of traffic flow that does not consider flow phenomena near the bottleneck. The unit of analysis selected is appropriate for the work done, as the focus of the study is solely on the operations of the freeway and on ramps. As many modern ramp metering approaches consider only the operation of the freeway, our unit of analysis is significantly larger than that for many recent papers and algorithms.

Another important assumption is the LWR PDE model accurately captures the conservation of flow over the control volume. The freeway is assumed homogeneous between on and off ramps, with a fundamental diagram that is time and space invariant for each link. The example contains a detailed description of the capacity drop for the diverge bottleneck, as this is a key concern, and assume that the merging capacity drop is taken into account in the fundamental diagram of the freeway segment. This has been shown to be standard practice in such models (Cayford et al, 1997). Additionally, for this objective user decisions are assumed and fixed. A stochastic simulation presented in Chapter 5 will further incorporate more realistic physics and mode/route switching through a user choice model.

3.2 Problem Description

Jin and Zhang (2001) state the purpose of ramp meters is to regulate input demand so that operational balance is achieved in the system. To accomplish this, on-ramp meters (i.e., D^1, D^2 and D^3 in Figure 3.1) restrict entrance from surface street links (i.e., w, x, y in Figure 3.1) onto the freeway. Transportation decision-makers seek to control the flow such that the tradeoff between progression on the freeway, governed by the LWR PDE and queuing on adjacent on-ramps and surface streets optimizes a given system-level measure of performance (i.e., total VMT on the freeway or total system wide delay). Many metering schemes simply transfer the travel cost from the freeway to its on-ramps and surface streets. However, an effective metering scheme should reduce the overall commuter travel cost (Cassidy, 2003).

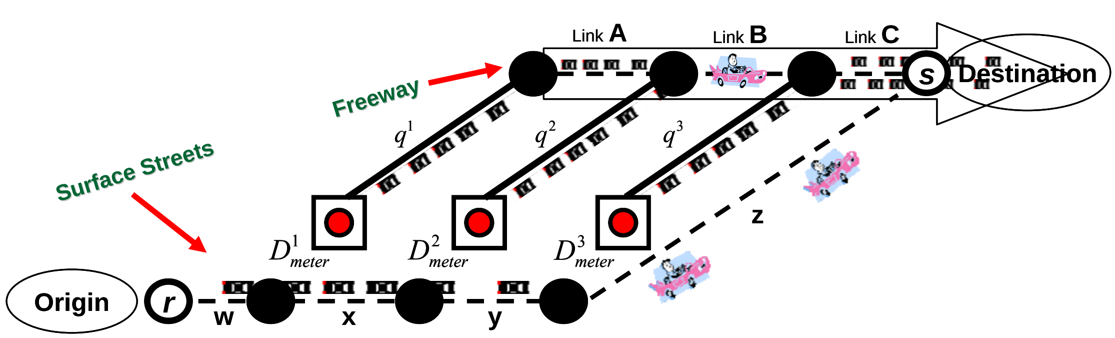


Figure 3.1 Ramp Metering Problem Diagram

3.2.1 Lighthill-Whitham-Richards (LWR) PDE

The physics of traffic flow is modeled in accordance with the Lighthill-Whitham-Richards (LWR) partial differential equation (PDE) which describes the evolution of the car density on the highway using a PDE (Richards, 1956), (Lighthill and Whitham, 1956). The LWR PDE is a macroscopic model based on aggregate variables that summarize information about multiple vehicles to describe the behavior of traffic (Bellemans et al, 2002). This equation is a physical law in traffic theory that relates the time derivative of the car density to the space derivative of the flux function, such that cars do not appear or vanish and flow is conserved across the network (Bellemans et al, 2002).

The example problem for this study was formulated based on the 26-link section of highway containing both on and off ramps as shown in Figure 3.2 (Bayen et al, 2004).

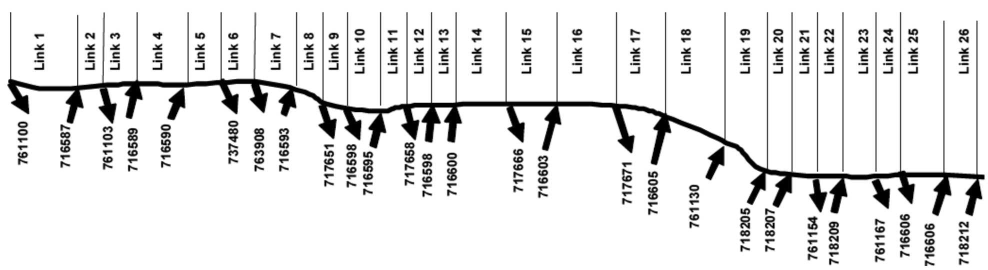


Figure 3.2 Portion of I-210 East, California

For a network of N connected highway segments, indexed by i, at link densities ρ_i (i.e., vehicles per mile), $N_i(\rho_i) = 0$ implies the conservation of vehicles in the traffic context across network N. The length of link i is called L_i and the coordinate on this link is $x_i \in [0, L_i]$. The evolution of car density on the highway is described by the following PDE:

$$N_i(\rho_i) = \frac{\partial \rho_i}{\partial t} + \frac{\partial q_i(\rho_i)}{\partial x_i} = 0$$

in which $q_i(\cdot)$ represents a flux function relating the flux of cars (number of cars through a given section of the highway during a time unit) to the car density at that location. This equation is interpreted as the local rate of change of car density which is equal to the space derivative of the flux of cars (i.e., conservation of mass). The flux $q_i(\cdot)$ can be identified empirically from highway data.

In this example, drivers merge (on ramps) and diverge (off ramps) across a 26-link highway topology. The governing equations for this system are given by:

$$
\begin{aligned}
&N_i(\rho_i) = \frac{\partial \rho_i}{\partial t} + \frac{\partial q_i(\rho_i)}{\partial x_i} = 0 && 1 \le i \le N \\
&\rho_i(x_i, 0) = \rho_i^o(x_i) && 1 \le i \le N \\
&q_i(\rho_i(0,t)) = q_{i-1}(\rho_{i-1}(L_{i-1}, t)) + q_i^{on}(t) && \forall i \in ON \\
&q_i(\rho_i(0,t)) = (1 - \beta_{i-1}(t)) q_{i-1}(\rho_{i-1}(L_{i-1}, t)) && \forall i \in OFF
\end{aligned}
\tag{1}
$$

In the governing equations, $\rho_i^o(x_i)$ denotes the density of cars at time 0. The set of links with merging on-ramps is represented by the *ON* equation. $q_i^{on}(t)$ denotes the inflow of cars into link *i*. *OFF* denotes the set of links with diverging off ramp cares leaving link *i* through an off ramp. Every x_i ranges in [0, L_i], and $t \in [0, T]$. The interpretation of the *ON* equation is the conservation of flow at an on-ramp (the flow into link *i* is the flow from link $i - 1$ plus the additional flow from the ramp). The *OFF* equation expresses the same with off ramps. In the last equation $\beta_i(t)$ represents the proportion of flow leaving link i. The first order approximation that $\beta_i(t) = \beta_i$ does not depend on time is also used.

3.2.2 PDE Constrained Optimization for Control Problems

Constrained optimization seeks to optimize a cost function, subject to constraints inherent to the problem and imposed by the available control. In case one or more of the constraints is the satisfaction of a PDE, the phrase "PDE constrained optimization" is used. The physics of a highway network is described by the LWR PDE and corresponds to what is referred to as the "fundamental diagram." Figure 3.3 depicts a fundamental diagram for a typical traffic flow where a flux q is a function of density ρ and the slope from the origin is speed (Geroliminis and Daganzo, 2007).

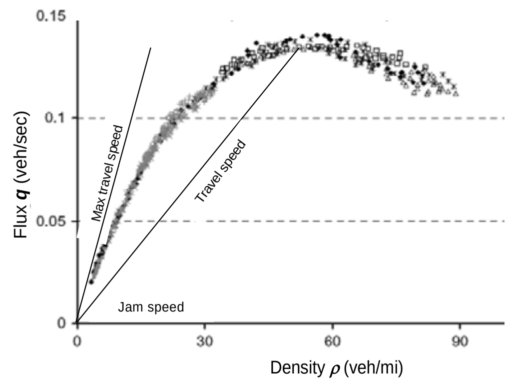


Figure 3.3 Fundamental Diagram of Traffic Flow

Many traffic network problems can be formulated as mathematical programming problems with objectives to minimize the total cost to the network. The optimization program is nonconvex, nonlinear and includes constraints in the form of PDEs (Sheffi, 1985).

min	**Total Cost**
w.r.t.	**Metering Policy** (q^{on})
s.t.	**Link Performance Model** (*LWR PDE*)
	Conservation of Flow

As previously stated, our formulation disallows backward wave propagation by ensuring link flows do not exceed link capacities. The LWR PDE is satisfied for the steady-state steady-flow conditions and under these conditions flow is conserved at steady state. The problem is simplified to consider q^{on} fixed over time by holding metering rates (fluxes) constant over the analysis period. This allows us to focus on the properties of optimal solutions derived by various formulations. The total cost function can be any appropriate metric. A model formulation, using different metrics, is presented in Section 3.3.

3.3 Model Formulation

A formulation for determining metering schedules to optimally control a network of connected highway segments is presented in this section and solved based on two different measures of performance. First, a mathematical program is presented to maximize the total vehicle miles traveled with respect to the ramp meter strategy and subject to satisfying the LWR PDE and ensuring the conservation of flow for each link. Vehicle miles traveled (VMT) is defined as the sum of distances driven by cars on a highway section over a time interval (Chen et al, 2001).

3.3.1 VMT-based Formulation

max	**Vehicle Miles Traveled**
w.r.t.	**Metering Policy** (q^{on})
s.t.	**Link Performance Model** (*LWR PDE*)
	Conservation of Flow

VMT is expressed as a function of $q_i(\rho_i(\cdot))$, as can be seen in the objective function of the following optimization problem:

$$\begin{aligned}
&\max \ \sum_{i=1}^{N} \int_0^{L_i}\int_0^{T} q_i(\rho_i(x_i,t))\,dx_i\,dt \\
&\textbf{w.r.t.} \quad \mathbf{q}, \boldsymbol{\rho} \\
&\textbf{s.t.} \\
&N_i(\rho_i) = \frac{\partial \rho_i}{\partial t} + \frac{\partial q_i(\rho_i)}{\partial x_i} = 0 && 1 \le i \le N \\
&\rho_i(x_i,0) = \rho_i^o(x_i) && 1 \le i \le N \\
&q_i(\rho_i(0,t)) = q_{i-1}(\rho_{i-1}(L_{i-1},t)) + q_i^{on} && \forall i \in ON \\
&q_i(\rho_i(0,t)) = (1-\beta_{i-1}(t))q_{i-1}(\rho_{i-1}(L_{i-1},t)) && \forall i \in OFF \\
&0 \le q_i^{on} \le q_{\max,i}^{on} \\
&q_i \le q_i^{cap}
\end{aligned} \tag{2}$$

The first four constraints in this optimization problem are the governing equations listed in Section 3.2.1 and the last two constraints set a bound on the number of cars that can be let into the highway. On-ramp flux q_i^{on} is the control variable for the problem. Maximum flow on link i is represented by $q_{\max,i}^{on}$; q_i^{cap} represents the capacity of the i^{th} homogeneous section of the freeway.

3.3.2 Delay-based Formulation

min	**Total Delay**
w.r.t.	**Metering Policy** (q^{on})
s.t.	**Link Performance Model** (*LWR PDE*)
	Conservation of Flow

The delay-based formulation is a mathematical program to minimize the total delay to the system with respect to the ramp metering strategy, while subject to satisfying the LWR PDE and ensuring the conservation of flow for each link. The link performance model and conservation of flow are represented the same way as in the VMT formulation. However, the objective function minimizes total delay which is computed as the sum of ramp delay and freeway delay. Ramp delay is the amount of time vehicles spend waiting at an on-ramp. Average ramp delay d for a given link is computed using this time-dependent version of Little's formula from queuing theory in Eq. (3)

$$d_{ramp} = \frac{3600}{C} + 900T\left[x - 1 + \sqrt{(x-1)^2 + \tfrac{4x}{CT}}\right] \quad (3)$$

where the time period T is expressed in hours and capacity C in vehicles/hour with x representing a saturation ratio of volume to capacity. For this problem, T = 1 hr, $C = q_i^{on}$, and $x = \frac{q_{\max}}{q_i}$.

Freeway delay is the difference in free flow speed s_{ff} and actual travel speed s for flow q on link i Average freeway delay for a given link is computed using Eq. (4)

$$d_{freeway} = (q_i)t_{ff}(\tfrac{s_{ff}}{s} - 1) \tag{4}$$

For this problem, the free flow trip time t_{ff} is 60 seconds, implying it takes that long to travel each 1-mile link at the free flow speed s_{ff} of 60 mph. The delay-based approach is the same as the VMT formulation in Eq. (2), except the objective function of the optimization problem is replaced by Eq. (5).

$$\min \sum_{i \in ON} q^i_{\max} d^i_{ramp} + \sum_{i=1}^{N} q_i d_{freeway} \tag{5}$$

3.4 Numerical Results

Each formulation in Section 3.3 was solved from the same initial conditions: free flow speed = 60 mile/hr; jam density = 600 veh/mile; initial flow = 5000 veh/hr; and the percent diverge β_i equal to 25 percent at each off-ramp. Optimal metering was obtained for each on-ramp *i*. The maximum on-ramp flux for each link was given as no more than 1300 vehicles per hour. Values for q_{on} in Table 3.1 and Table 3.2 are the flux for each link at convergence of the optimization problem (vehicles/hour to which link flow is restricted in the optimal policy).

Table 3.1 VMT-based Optimal Metering Policy

Vehicle Miles Traveled Formulation

On-Ramp *i*	q_{on}	q_{max}
2	1300	1300
4	1300	1300
5	1300	1300
8	1300	1300
11	1300	1300
13	1300	1300
14	1300	1300
16	1300	1300
18	1300	1300
19	1300	1300
20	1300	1300
21	**952.9**	1300
23	1300	1300
26	1300	1300
Total VMT (miles)	135,172	
Total Delay (min)	1162	

Table 3.2 Delay-based Optimal Metering Policy

Delay Formulation

On-Ramp *i*	q_{on}	q_{max}
2	1300	1300
4	1300	1300
5	1300	1300
8	1300	1300
11	1300	1300
13	**1283.9**	1300
14	**1285.6**	1300
16	**1260.6**	1300
18	**1210.6**	1300
19	**1215.7**	1300
20	**1223.5**	1300
21	**1236.6**	1300
23	1300	1300
26	1300	1300
Total Delay (min)	1,061	
Total VMT (miles)	134,288	

The VMT formulation optimized total distance and produced an optimal control strategy of aggressively restricting the flow just before the bottleneck. The optimal policy was a total VMT of 135,172 miles with a total delay of 1162 minutes. The delay formulation optimized total time and produced an optimal control strategy that modestly metered many on-ramps upstream from the bottleneck. The optimal policy was a total delay of 1061 minutes with a total VMT of 134,288 miles.

The simulation results produced some interesting insights. First, although both formulations generated an optimal control policy, the metering strategy was sensitive to the performance measure that was optimized. The results illuminated the tradeoff between time and distance as the primary performance measure in the optimization problem. As depicted in Table 3.1, aggressive metering just prior to the bottleneck allowed more drivers to travel unimpeded, provided they desired to travel relatively short distances and exit prior to the bottleneck. However, the long line at the aggressively metered on-ramp has the potential to interfere with adjacent surface streets and contribute to congestion in local urban areas where the capacity to store queued vehicles is much less than on the freeway. Longer delays at meters may incentivize drivers to change routes or even modes of travel. The potential consequences serve as a call for a broader system perspective to better balance the needs of the greater transportation system.

A second insight is that, although the results in Table 3.1 and Table 3.2 represent widely accepted approaches, neither accounts for the upstream impacts due to the reduced capacity at the diverge bottleneck. The extent to which capacity near the bottleneck is over estimated in the formulation determines whether the resulting optimal control solution will guarantee congestion is avoided. An expanded formulation that incorporates the impacts due to reduced capacity is offered in Section 3.5.

3.5 Expanded PDE Constrained Formulation and Results

Despite being among the most prevalent highway traffic control strategies used in practice, even popular ramp metering algorithms do not include some of the effects of the physics of diverge bottlenecks (Cassidy, 2002). Cassidy (2003) suggests much of the literature on how ramp metering might achieve reductions in congestion actually offers techniques which can make conditions worse than if left uncontrolled. The central aspect of the physics of diverge bottlenecks not adequately considered is the additional reduction in capacity at the bottleneck caused by poor queue discipline at potential bottlenecks. As previously stated, drivers entering the passing lane to advance along the queue and exit closer to the off-ramp slows or even clogs the passing lane to through traffic, causing what amounts to further reduction in the capacity at the bottleneck. This section offers an expanded formulation that includes the impact of this reduction in capacity.

The expanded PDE constrained formulation differs from the previous formulation by explicitly considering the flow bound for an exit at which a bottleneck would activate, expressed by α_i. This proportion can be estimated from an origin-destination (O-D) matrix for a given highway based on historical data describing where drivers enter and exit the freeway. The impact can be described as a shrinking of the fundamental diagram mapping density to flow and based on Greehshield's traffic flow relationship in Eq. (6)

$$q_i = (1-\alpha_i)\, V_{ff}\, \rho_i (1 - \tfrac{\rho_i}{\rho_{ff}}) \tag{6}$$

where V_{ff} and ρ_{ff} are the free flow volume and density; q_i and ρ_i are the actual link flow and density (Tarko, 2003).

The (Bayen et al, 2004) VMT-based approach is modified to illustrate the expanded formulation. For the 12 on-ramps prior to the bottleneck at exit 22, α_i values were initialized such that the total expected upstream flow bound for the bottleneck was 25 percent (same as previous formulation). Another difference in the expanded formulation is that β_i is parameterized as $\beta_{i-1}(t, q_{on})$ in the *OFF* constraint. β_i represents the proportion of flow that exits off ramp *i*.

The expanded PDE constrained formulation to minimize total delay is shown below. Ramp and freeway delay are computed as shown in Section 3.3.

$$
\begin{aligned}
&\min \quad \sum_{i \in ON} q^{i}_{\max} d^{i}_{ramp} + \sum_{i=1}^{N} q_i d_{freeway} \\
&\textbf{w.r.t.} \quad \mathbf{q}, \boldsymbol{\rho} \\
&\textbf{s.t.} \\
&N_i(\rho_i) = \frac{\partial \rho_i}{\partial t} + \frac{\partial q_i(\rho_i)}{\partial x_i} = 0 && \forall\, i \in BN\,; 1 \le i \le N \\
&N_i(\rho_i) = \frac{\partial \rho_i}{\partial t} + \frac{\partial q_i(\rho_i, \alpha_i)}{\partial x_i} = 0 && \forall\, i \notin BN\,; 1 \le i \le N \\
&\rho_i(x_i, 0) = \rho_i^{o}(x_i) && 1 \le i \le N \\
&q_i(\rho_i(0,t)) = q_{i-1}(\rho_{i-1}(L_{i-1}, t)) + q_i^{on} && \forall\, i \in ON \\
&q_i(\rho_i(0,t)) = (1 - \beta_{i-1}(t, q_{on}))\, q_{i-1}(\rho_{i-1}(L_{i-1}, t)) && \forall\, i \in OFF \\
&0 \le q_i^{on} \le q^{on}_{\max,\, i} \\
&q_i \le q_i^{cap}
\end{aligned}
\tag{7}
$$

The initial conditions were the same for the alternate formulation as in the formulations in Section 3.3. The additional information regarding the percentage of flow from each upstream link desiring to exit off-ramp 22 was added and the optimization was resolved. Again, each formulation generated an optimal metering strategy to ensure no bottleneck activated. Table 3.3 and Table 3.4 depict the optimal metering policies for the expanded formulation.

Table 3.3 VMT-based Optimal Metering with α Considered

Delay Formulation

	On-Ramp i	α_i	q_{on}	q_{max}
	2	0.08	1300	1300
	4	0.17	1300	1300
	5	0.25	1300	1300
	8	0.33	1258.9	1300
	11	0.42	1052.7	1300
	13	0.50	903.1	1300
	14	0.50	905.4	1300
	16	0.42	838.9	1300
	18	0.33	761.4	1300
	19	0.25	785.9	1300
BN at 22_{exit}	20	0.17	813.6	1300
	21	0.08	845.4	1300
	23		1300	1300
	26		1300	1300

Total Delay (min)	4,565
Total VMT (miles)	108,630

Table 3.4 Delay-based Optimal Metering Policy with α Considered

Vehicle Miles Traveled Formulation

	On-Ramp i	α_i	q_{on}	q_{max}
	2	0.08	1300	1300
	4	0.17	1300	1300
	5	0.25	1300	1300
	8	0.33	1300	1300
	11	0.42	1300	1300
	13	0.50	1300	1300
	14	0.50	1175.4	1300
	16	0.42	1132.2	1300
	18	0.33	0	1300
	19	0.25	112.0	1300
BN at 22_{exit}	20	0.17	1183.9	1300
	21	0.08	1126.1	1300
	23		1300	1300
	26		1300	1300

Total VMT (miles)	109,830
Total Delay (min)	8077

The results from the alternate formulation verified the tendency for VMT-based control policies to meter more aggressively over fewer on-ramps and delay-based control policies to meter more modestly across more on-ramps upstream from the bottleneck. The optimal policies (depicted in Figure 3.4) for both formulations restricted heaviest at on-ramp 18 (where most people are headed to the diverge bottleneck).

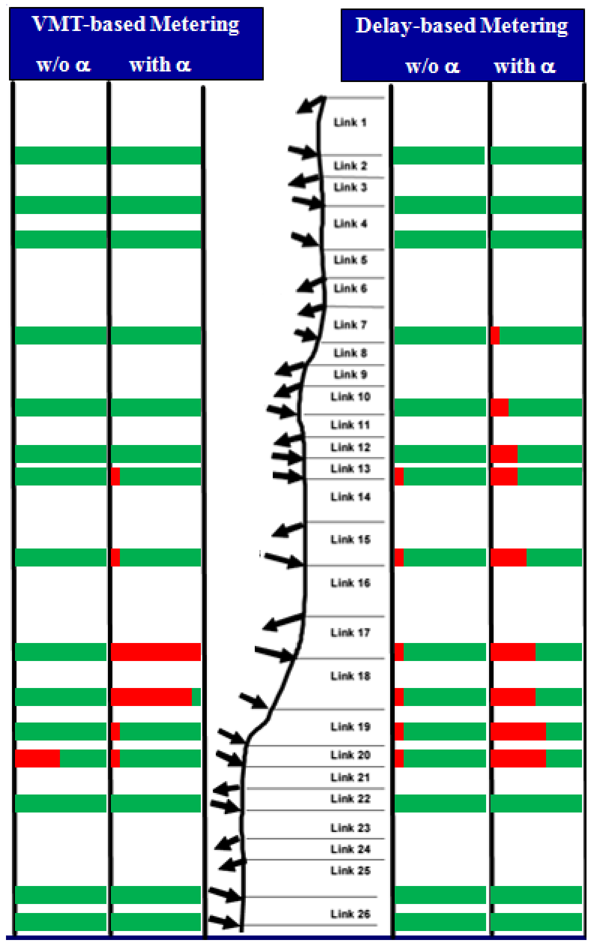


Figure 3.4 Optimal Policies w/ Alternate Problem Formulation

The impact of q^{on}= 0 at on-ramp 18 was no cars entered the freeway at that location for entire analysis period. From a driver's perspective, this is extremely inconvenient. The VMT-based policy allowed maximum entry as far as possible (until on-ramp 14), then metered increasingly heavier until on-ramp 19 where flow increased rapidly to maximize the flow of travelers with destinations beyond the bottleneck. The delay-based formulation began metering at on-ramp 8 and increasingly restricted flow through on-ramp 18, then gradually decreased metering until the bottleneck.

3.6 Analysis

For this objective, a common problem instance was used to illustrate the formulations and results in Sections 3.3, 3.4 and 3.5. However, over the course the study, multiple simulations were performed and sensitivity of the optimal policy to the performance measure was consistently observed. When distance was the priority, aggressive metering just prior to a potential bottleneck was the optimal policy. When time was the priority, a modest, broader optimal policy was the solution. The reason for this observed sensitivity is the scope of the system under consideration. For example, when optimizing VMT, only the distance traveled by vehicles *on the highway* mattered, thus the aggressive, single point metering strategy provided the greatest possible overall forward flow. Time spent waiting to enter the highway was not a factor; the optimizer exploited this by assigning all queued vehicles to the single point location. Conversely, in the delay formulation, the optimizer exploited short trips upstream to promote faster overall flow by restricting flow at more upstream on-ramps.

3.6.1 Discussion of Results

The results in our study showed how optimal control strategies derived from VMT-based and delay-based formulations differ. But which strategy is better for the broader transportation system? For one thing, the VMT-based optimal policy which included completely closing a freeway on-ramp should be scrutinized for its potential to cause tremendous traffic jams around the entrance ramp. Furthermore, with humans-in-the-system, such a drastic policy would likely lead to someone being fired, thus it may not be optimal if it is not realistically feasible. As compared to the commonly used VMT-based formulation solution, the delay-based formulation produced a metering scheme that distributed the metering effects across the highway system and prevented extremely long queues at on-ramps near a bottleneck. Both optimal ramp metering schemes effectively control highway traffic; however, the question of whether the control causes traffic problems for associated surface street systems will depend on the circumstance. VMT-based formulations aggressively constrain flow immediately upstream of potential bottlenecks. This can lead to long lines at on-ramps that interfere with flow on adjacent surface streets. The delay-based formulation spreads a more modest metering policy across more upstream links. In some cases, it may be preferred to store vehicles on the highway rather than in an urban center (even if this means allowing temporary jams to activate). Thus, for many typical highway traffic situations, maintaining steady-state steady-flow conditions through a delay-based control strategy seems more appropriate. Most commuters likely prefer reaching their destination in the least amount of time and value minimizing delay over maximizing VMT. The delay-based formulation is more appropriate when this is the case.

The expanded formulation also exposes a tradeoff between underutilized capacity upstream and delay which was the results of the reduced capacity at the bottleneck. In our initial simulations, the optimization procedure traded between underutilized capacity upstream and decreasing the anticipated percentage of drivers exiting at the bottleneck in order. When a constraint was imposed to maintain alpha equal to twenty-five percent, the excess diverge bottleneck volume was reverted back to the freeway and the optimization required heavier metering to balance the system and maintain steady-state steady flow conditions. This resulted in a non-linear increase in delay, particularly at the on-ramps.

Including the effects of diverge bottlenecks in the problem formulation produced a metering strategy in Section 3.5 that better accounted for the reduced capacity at the bottleneck and its impacts upstream on traffic flow. The resulting control strategy metered more aggressively, which underscored the premise that prior formulations did not include the effects of poor queue discipline on the traffic flow. When the reduced capacity at the bottleneck is not accounted for in the optimization formulation, Figure 3.5 provides a graphical illustration of the impact as determined by solving the LWR PDE.

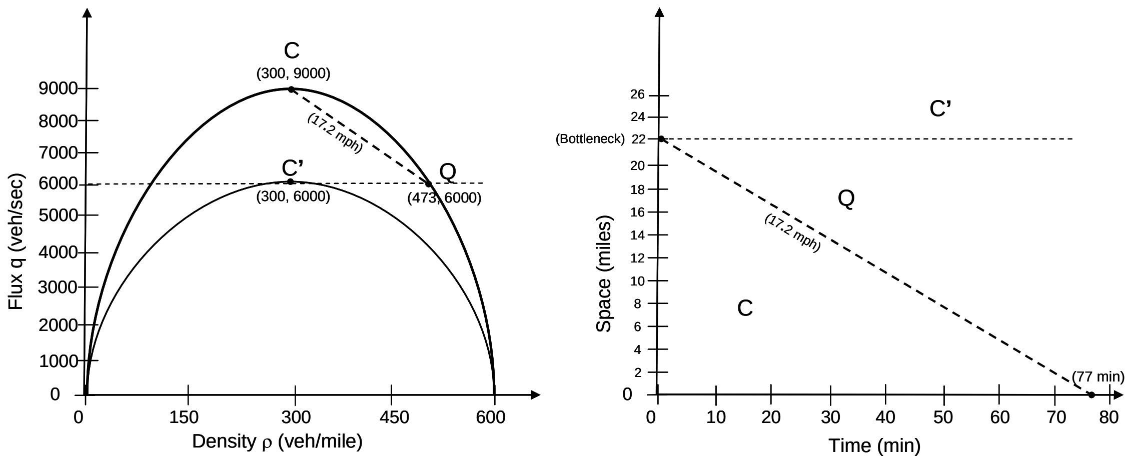


Figure 3.5 Shockwave Propagation Due to Reduced Capacity at Bottleneck

In the absence of queue discipline effects the traffic would move according to the top curve on the fundamental diagram on the left graph in Figure 3.5. However, the reduction in capacity of the segment caused by the poor queue discipline of drivers exiting at the diverge bottleneck results in a situation where, if left unaccounted for in the ramp metering control strategy, would result in upstream traffic approaching the bottleneck in state C with the bottleneck discharging in state C'. Traffic immediately before the bottleneck would then be in state Q and a shockwave would propagate backward from the bottleneck with a speed equal to the slope between points C and Q on the fundamental diagram. In our example, the speed was 17.2 mph and is depicted in Figure 3.5. Facility-specific studies are needed to parameterize the fundamental diagram, but the expanded formulation proposed in this objective accounts for this realistic aspect of highway traffic.

3.6.2 Insights and Observations

Several interesting insights were observed. First, the results show that when drivers who are headed beyond a bottleneck are metered in favor of those who will take the exit at a diverge bottleneck and create queue discipline problems, the model based optimal solution is NOT optimal. Second, ramp metering strategies that do not account for queue discipline can cause unintended consequences to adjacent networks. Third, delay-based optimal metering solutions generated a total distance traveled that was less than the generated total distance traveled in the optimal distance-based metering solution. This is to be expected, since vehicle miles traveled is to be maximized and delay is to be minimized. Furthermore, another insight was that it is not unreasonable to generate an optimal strategy that requires a ramp to completely close. In the VMT case, without considering queue discipline effects, this was the best place for the overall system to apply metering. This makes sense because these vehicles enter the freeway the furthest downstream, and without activating another constraint, such as a minimum allowable influx constraint, there was no reason to meter anywhere else. However, the feasibility and impacts of such an extreme policy were discussed in Section 3.6.

3.7 Conclusion

This objective examined the extent to which ramp metering strategies can control freeway traffic such that appropriate performance measures are optimized and unintended consequences are minimized. The most appropriate metric to use depends on the situation. Many typical city planning factors, such as capacity of surface streets, proximity of on-ramps to critical infrastructure and services and freeway design, should be considered when choosing a metric. Optimal ramp metering strategies were observed to be sensitive to the problem formulation that is solved. VMT and delay based formulations were shown to produce optimal strategies that differed in terms of metering restrictiveness and locations. One particularly important consequence of the VMT-based formulation is that it meters the people accessing the highway immediately upstream from the bottleneck. This is likely a very undesirable situation if the bottleneck is a diverge bottleneck, as this will result in metering people who are not likely to cause queue discipline problems. This will almost certainly reduce bottleneck throughput and increase system-wide delays.

If the percentage of drivers headed for a diverge bottleneck can be estimated, then it is possible to tailor a ramp metering strategy using the methods developed in this objective. An alternate formulation was proposed and shown to provide a strategy that was more inclusive of the physics of diverge bottlenecks, as the fundamental diagram was parameterized by the fraction of users exiting at the bottleneck. Therefore, resulting optimal ramp metering strategies are based on more realistic models of freeway performance and thus have a greater potential for reducing congestion and limiting unintended consequences. Estimation of such information is a notoriously difficult problem, but with the advent of sensor network capabilities provided by devices such as personal GPS navigational systems and cell phones with GPS capabilities, it is likely for more accurate estimation of O-D matrices to be possible in the future.

Ideally, a metering scheme should be specially tailored to the freeway it serves. This may involve solving the ramp metering problem as a multi-objective optimization problem using the same PDE constrained optimization discussed in this objective. Ramp metering solutions designed to effectively control highway traffic may inadvertently cause traffic problems for associated surface streets, especially if queue lengths on and inadequate queue storage capabilities of surface streets are ignored. Therefore, while this study has focused on freeway metering, optimization and control is most effective for the broader transportation system when aspects of and impacts to as many facilities as possible are considered.

The need for a broader systems perspective to better address SoS problems is only one of several general insights into broader SoS problems that come from the research described in this objective. A second insight is that system engineers and planners should take great care in deciding the problem-solving methods to employ and the performance measures with which to evaluate success. State of the art mathematical approaches, such as the ones examined and developed in this objective, still simplify the SoS problem in order to analyze and optimize certain aspects. Finally, a risk of selective or piecemeal solutions to SoS problems is the potential for unaccounted for interdependencies to result in different performance that is missed without an integrated optimization approach.

From the SoS issues identified in this research objective, subsequent tasks should seek to show how optimization and control methods, used as decision support to SoS policy makers, can increase the reliability and performance of an SoS. The lack of an integrated system approach is addressed in the next objective in the context of an economic SoS. A general optimization under uncertainty (OUU) approach is developed as an initial road map from which more detailed features can be added. The economic example in the next objective showcases the appropriateness of the SoS approach in a domain other than transportation. However, in later objectives the transportation example is reintroduced as the OUU approach is extended to incorporate microscopic system performance, hybrid dynamics and user decisions.

CHAPTER IV

AN INTEGRATED APPROACH TO POLICY ANALYSIS AND DESIGN UNDER UNCERTAINTY FOR SYSTEM OF SYSTEMS

4.1 Introduction

System of systems problems appear within and across many domains. In developing operational policies for these systems, it is important to model the behavior of the system of systems, represent and propagate uncertainties through the system model, and optimize the policies under uncertainty. Policy makers are interested in achieving desired outcomes as a result of a given action. For example, creating demand for finished goods from each of a developing nation's sectors will stimulate cross-sector demands and improve the performance of the entire economy. This objective develops an SoS approach to determining optimal control policies for these situations and showcases the appropriateness of such strategies in the context of an economic SoS.

The goal of the economic planner would be to maximize the benefit of the economic investment made in the economy. Another example would be the assignment of tasks to work centers in a work flow network in which each work center is dependent on some or all of the other work centers in the network in order to complete its tasks. The goal of the policy designer would be to assign tasks to work centers such that the overall workload is smallest, the total job is completed most quickly, the job is done at minimum cost, etc. Communications networks can have systems with input-output relationships, and it may be required to maximize metrics of network efficiency with the network operational policies. Queuing networks also exhibit similar properties if there are multiple interdependent servers who can send the items in queue to other servers before or after they have been served, and it may be desirable to minimize the total wait time for a network of servers. Such a network representation is consistent with system of systems research which typically involves the large-scale integration of many independent, self-contained systems. From homeland security and military planning to air traffic control and satellite operations, complex multi-systems are very interdependent.

The synthesis of these very large systems often results in different problems than those presented by the design of a single system and the degree of uncertainty grows with the size of the system (DeLaurentis et al, 2006). SoS attributes such as operational and managerial independence, interdisciplinary domains and emergent behavior, require the relationships between systems to be modeled accurately and comprehensively. Policy design and analysis for systems of systems should also systematically account for the various sources of uncertainty and optimize the system objectives under uncertainty. A challenge for policy design in the context of systems of systems is integrating analyses which are often performed separately. This objective develops a framework for policy design in a system of systems which integrates simulation, uncertainty analysis, and optimization techniques to provide decision support for policy design. The proposed framework represents a network of interdependent systems using input-output models which describe the behavior of the various systems, efficiently analyzes and propagates uncertainties using analytical reliability methods, and optimizes system objectives under these uncertainties. This framework provides insights to alert decision makers of potential direct and indirect impacts from system interactions and make risk-informed policy decisions. The framework described in this objective addresses policy design in systems of systems from the perspective of optimization under uncertainty (OUU). This is important because policies should be effective even under adverse conditions. This framework uses first-order approximations of the probabilistic constraints to develop a decoupled formulation for SoS optimization under uncertainty. Our approach to OUU is depicted in Figure 4.1.

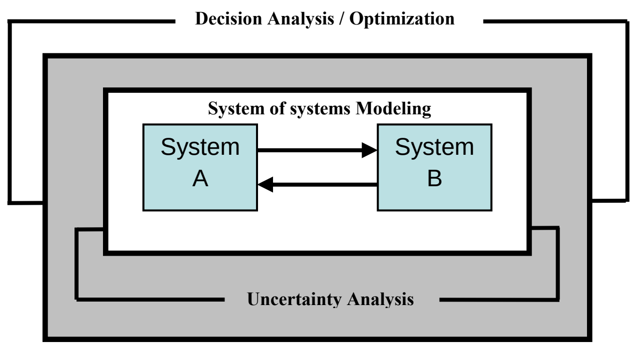


Figure 4.1 Analysis Framework for SoS Decision Support

SoS optimization under uncertainty has three components: (1) a model of SoS behavior and interactions, (2) uncertainty analysis, and (3) policy optimization analysis.

The rest of the chapter is organized as follows. Section 4.2 gives an overview of the elements of the methodology used in this objective, including SoS analysis, uncertainty propagation, and optimization under uncertainty. Section 4.3 develops a novel approach to find robust optimal solutions for SoS policy optimization, proves that under assumptions of convexity in the decision space robust optimal policies do exist, and gives the optimality conditions for robust optimal solutions. Section 4.4 gives a numerical illustration of the proposed methodology. A brief summary of the chapter is in Section 4.5.

4.2 Methodology

Systems of systems may be modeled in different ways, depending on the level of detailed information available. While different mathematical and computational models such as simplified input-output models, systems dynamics, and agent-based simulation are available for describing their behavior, systems of systems are all networks involving their member systems. Networks are an organizational foundation for societies and economies. Nagurney (2003) suggests that network theory, as a methodology, has developed into a powerful and dynamic medium for abstracting complex problems with associated nodes, links and flows. Since systems of systems are abstract networks of autonomous systems, the hypothesis is forwarded that such network analysis techniques are able to model the behavior of systems of systems. This section will show how network theory can be used as the basis for policy optimization for systems of systems. An overview of the Leontief input-output model (Leontief, 1953) is provided, which gives a simplified mathematical tool for the input-output analysis of economic networks. This approach is also adaptable to other types of network systems (Haimes and Jiang, 2001). The Leontief Input-Output model is used to derive the response of an economic system of systems to an external stimulus. Analytical methods are described for uncertainty analysis and methods for optimization under uncertainty. The OUU framework is generic and can be incorporated with other appropriate models of system of systems response. In the following subsections, each of the three elements of the proposed methodology are described, and then their integration is developed.

4.2.1 Leontief Input-Output Model

The Leontief Input-Output Model is a macro-level approach that represents the interactions among various interdependent entities, such as sectors in a national economy, as linear relationships (Leontief, 1953). The input-output model is based on three assumptions: fixed coefficients of production, constant returns to scale, and homogeneity of input resources. As an example, Figure 4.2 is a two-dimensional representation of the Leontief production function, where K is capital input and L is labor input. The L-shaped isoquants illustrate the various combinations of K and L to achieve a given level of system output

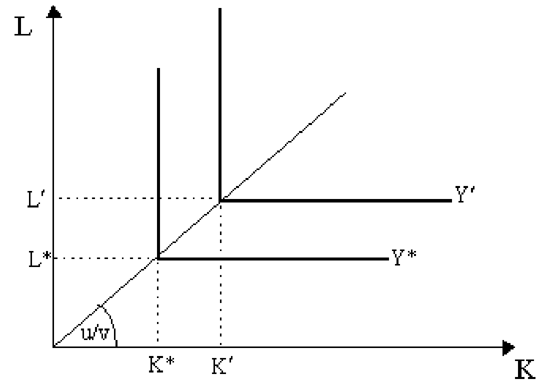


Figure 4.2 Leontief Production Functions

The concepts of general equilibrium and input-output analysis are well-established in literature. Intriligator (1971) provides a more thorough discussion of the Leontief model on which this discussion is largely based. For an economy in which all production functions are of the input-output type, the problem of general equilibrium leads to a linear programming problem and the existence of meaningful solutions can be proved. Such an economy produces n goods and output for the total economy of good j is x_j, for $j = 1,2,\ldots,n$. Some output can be sold, as producer goods, to other firms, where x_{kj} is the amount of the k^{th} good used in the production of the j^{th} good, and the production function for good j is of the input-output type:

$$x_j = min(\frac{x_{1j}}{a_{1j}}, \frac{x_{2j}}{a_{2j}}, \ldots, \frac{x_{nj}}{a_{nj}}) \quad (1)$$

This is the optimal production strategy under the Leontief production function.

The constant parameters a_{kj} are the coefficients of production, nonnegative amounts of the k^{th} commodity required to produce one unit of good j. According to the proportionality equation $x_{kj} = a_{kj} x_j$, for $j = 1,2,\ldots,n$, the inputs of commodities required to produce any commodity are proportional to the output of that commodity, the coefficients of production. The output of any commodity is used either as input for the production of commodities or as final demand. Thus, the balance equation is

$$x_k = \sum_{j=1}^{n} x_{kj} + c_k \tag{2}$$

for $k = 1,2,\ldots,n,$ where x_k is the output of commodity k; the first term on the right is the total use of commodity k in the production of all other commodities; and the second term on the right, c_k, is the final demand for commodity k, including consumption, investment, export and government demand. Combining the balance equation and proportionality equation yields

$$x_k = \sum_{j=1}^{n} a_{kj} x_{kj} + c_k \tag{3}$$

for $k = 1,2,\ldots,n$, where the left side gives output, and the two terms on the right are input and final demand, respectively. The resulting n equations can be written as the single matrix equation, the Leontief equation

$$\mathbf{x} = \mathbf{ax} + \mathbf{c} \tag{4}$$

where where $\mathbf{A} = (\mathbf{I} - \mathbf{a})^{-1}$ and $\mathbf{I}$ is an identity matrix. The Leontief equation can be solved for the output required to produce a given vector of final demand

$$\mathbf{x} = (\mathbf{I} - \mathbf{a})^{-1}\mathbf{c} \tag{5}$$

or

$$\mathbf{x} = \mathbf{Ac} \tag{6}$$

where $\mathbf{A} = (\mathbf{I} - \mathbf{a})^{-1}$. The sum of the elements i^{th} column of the $(\mathbf{I} - \mathbf{a})^{-1}$ is called the sector multiplier since a change in final demand Δc_i requires a change in final demand $\Delta\mathbf{x} = \sum_j \Delta c_i (\mathbf{I} - \mathbf{a})_{ij}^{-1}$. The values of the matrix multiplier are sensitivities, representing sector interdependencies which measure the impact of interactions among sectors in the system of systems. Demand for goods from a sector with a larger multiplier stimulates a larger demand from other sectors.

Other nonlinear models of production and general equilibrium exist in the economic literature (Intriligator, 1971). This objective uses the Leontief input-output model only for the sake of illustration. One may include a more appropriate model with the overall OUU framework if necessary.

4.2.2 Uncertainty Analysis

The model discussed in Section 4.2.1 is used as a baseline model to predict the impact of a major economic investment given values of the uncertain model parameters and the mitigation and prevention strategies implemented. However, since some of the model inputs are random, the model outputs will also be random. Thus methods for propagating uncertainty through the underlying models are needed. Since Monte Carlo Simulation (MCS) is expensive, using MCS with the thousands of iterations required for accurate estimation of output statistics may be impractical, especially in the context of policy optimization. Thus, it may be desired to estimate the statistical moments of a function of random variables using analytical approximation methods which require only a few model evaluations.

One can use second-moment approximations for the output statistics of the mean and variance of system of system level responses as a function of random variables, but second-moment approximations often are inaccurate with nonlinear limit states and non-normal random variables (Ditlevsen and Madsen, 1996), (Haldar and Mahadevan, 2002), (Nowak and Collins, 2000). Given the potential for inaccuracies with second moment approaches, it is desirable to use more accurate methods for estimating the CDF of the system of systems model outputs, especially if extreme events are of particular importance. The First-Order Reliability Method (FORM) is a more accurate analytical method that can be used to determine the CDF of functions of random variables (Ditlevsen and Madsen, 1996), (Haldar and Mahadevan, 2002), (Nowak and Collins, 2000).

The probability of a function of random variables being less than or equal to some value k is given as

$$P\{g(\mathbf{y}) \leq k\} = {}_{g(\mathbf{y})\leq k}\ f_{\mathbf{y}}(\mathbf{y})d\mathbf{y}$$

where $f_{\mathbf{y}}(\mathbf{y})$ is the joint probability density function of the random variables $\mathbf{y}$. An analytical evaluation of the integral is possible only for a few special cases, and hence numerical integration is necessary. FORM has been found to be a computationally efficient and reasonably accurate analytical approximation to the probability integral under consideration for many problems (Ditlevsen and Madsen, 1996), (Haldar and Mahadevan, 2002), (Nowak and Collins, 2000). In FORM, there are three important steps in the calculation of the probability of failure for an individual component failure mode. These are:

• Transformation of the random variables $\mathbf{y}$ to an uncorrelated standard normal space $\mathbf{u}$. Several transformations for correlated and non-normal random variables are available, such as Rosenblatt, Nataf, and Rackwitz-Fiessler (see (Rosenblatt, 1952), (Liu and Der Kiureghian, 1986), and (Rackwitz and Fiessler, 1976)).

• Calculation of the most probable point (MPP) corresponding to the condition $g(\mathbf{y}) = G(\mathbf{u}) = k$. This point is the solution to the constrained optimization problem.

$$\mathbf{u}^* = argmin(\|\mathbf{u}\| \mid G(\mathbf{u}) = k) \tag{7}$$

Good algorithms for solving this problem have been proposed in (Hasover and Lind, 1974), (Rackwitz and Fiessler (1978), and (Zhang and Der Kiureghian (1994).

• The probability of the event $g(\mathbf{y}) < k$ is approximated as $\Phi(-\beta)$, where β is the reliability index. The reliability index is calculated as $\boldsymbol{\alpha}\mathbf{u}^*$ where

$$\boldsymbol{\alpha} = -\frac{\nabla G_u(\mathbf{u})}{\|\nabla G_u(\mathbf{u})\|}^T \tag{8}$$

That is, the alpha vector is the negative normalized gradient row vector of the response function in the transformed space. At optimality in the FORM MPP search, it is important to note that the alpha vector $\boldsymbol{\alpha}$ is collinear with the MPP vector. The alpha vector can help analysts determine which uncertain parameters are the most important so that information gathering efforts are focused on these variables (Haldar and Mahadevan, 2002). Random variables with alpha values of low magnitude can often be modeled as deterministic at the mean.

A further useful technique in reliability-based design optimization is the inverse FORM problem, which returns a worst-case point such that the probability of a more extreme event is $\Phi(-\beta)$. The inverse FORM problem solves the optimization (Du and Chen, 2004).

$$\mathbf{u}^* = argmin\left(G(\mathbf{u}) \mid \|\mathbf{u}\| = \beta^t\right) \tag{9}$$

In this optimization, $\beta^t = \Phi(p_f^t)$, where p_f^t is a target failure probability in design optimization. The inverse FORM formulation allows for the decoupling of the reliability analysis and optimization problems in reliability constrained optimization problems, which are discussed in the next section, and greatly reduces the computational expense involved in solving such problems.

4.2.3 Optimization Under Uncertainty

In the process of policy design for systems of systems, decision makers want to set decision variables at values which will result in the best overall result for the system of systems. However, finding the optimal policy is more difficult in the presence of uncertainty. One existing strategy for the implementation of optimization under uncertainty is stochastic programming (Bertsimas and Tsitsiklis, 1997). In stochastic programming, some of the parameters are random variables and the optimization formulation includes many scenarios, or realizations of the random variables. Feasibility for all possible scenarios can be required, or infeasibility can be penalized in the objective as in Eq. (10):

$$\begin{aligned}
&min\ E_{i\,scenarios}\{f(\mathbf{d},\mathbf{x_i})\} + \sum_{i\,scenarios} q_i^T p_i(g_i(\mathbf{d},\mathbf{x}_i)) \\
&s.t. \\
&\mathbf{h}(\mathbf{d},\mathbf{x}_i) = \mathbf{0} \quad \forall i\,scenarios \\
&\mathbf{g}(\mathbf{d},\mathbf{x}_i) \geq \mathbf{0} \quad \forall i\,scenarios(optional) \\
&\mathbf{d} \in \mathbf{D}
\end{aligned} \tag{10}$$

where $\mathbf{d}$ is a vector of decision variables, $\mathbf{u}_i$ is a set of state variables for each scenario i, and $\mathbf{x}_i$ is a realization of the random variables defining scenario i. Each scenario occurs with probability q_i. In Eq. (10) the penalty function for each scenario p_i is defined to be zero if $\mathbf{d}$ is feasible for scenario i. In addition, the penalty function satisfies a form of monotonicity in that worse violations incur greater penalty. There may be additional side constraints on the decision variables Eq. (10) accounts for uncertainty in optimization problems, but does so at a very large computational cost as the underlying models must be evaluated for every specified scenario.

Often decision makers planning may not be concerned with all possible scenarios, but only with preparing for the scenarios which could realistically occur and result in catastrophic consequences. Since extreme events are often of concern to decision makers, another approach to optimization to optimization under uncertainty is optimization for the worst-case scenario (Rockafellar, 2007). This statement of the OUU problem is given as

$$\begin{aligned} &min_{\mathbf{d}} max_{s \in S}\ f(\mathbf{d}, s) \\ &s.t.\ \mathbf{d} \in \mathbf{D}(t) \quad \forall t \in \mathbf{S} \end{aligned} \tag{11}$$

The policy $\mathbf{d}$ is required to be feasible no matter what parameter value (scenario) occurs; hence, it is required to be in the intersection of all possible feasible sets $\mathbf{D}(t)$ for each of the possible scenarios. The inner maximization yields the worst possible objective value among all scenarios. This approach will be used and only continuous random variables will be considered. If the set $\mathbf{S}$ is continuous, this is a problem of semi-infinite, but there is a need to define the set $\mathbf{S}$ probabilistically. The problem is defined as

$$\begin{aligned} &max_{\mathbf{d},z}\ z \\ &s.t. \\ &P\{f(\mathbf{d}, \mathbf{x}) - z \leq 0\} \leq \Phi(-\beta^t) \end{aligned} \tag{12}$$

and using methods of semi-infinite optimization and reliability-based optimization to solve this problem (Kortanek, 2001), (Jongen et al, 1998). Only the scenarios which occur within a continuous subspace of the transformed normal space are considered (i.e. those scenarios such that $\|\mathbf{u}\| \leq \beta^t$). This statement of the OUU problem allows for the use of techniques developed in the semi-infinite optimization and reliability-based design optimization (RBDO) literature which exploit the computational efficiency of FORM in the evaluation and/or assurance of probabilistic constraints, which is reviewed here for the reader's convenience.

In an RBDO problem, an objective function is minimized subject to reliability constraints corresponding to various limit states and deterministic constraints. The objective function is a function of the design variables and random variables. The design variables can be distribution parameters such as means or standard deviations of the random variables or can be deterministic. Nested algorithms, which were used before the 1990s include a full reliability analysis at every step of the design optimization algorithm. It is obvious that nesting these two procedures results in a large number of function evaluations, and studies performed in (Agarwal and Renaud, 2004), (Du and Chen, 2004), (Liang et al, 2004) and (Yang and Gu, 2004) have confirmed that nested methods require many more function evaluations than RBDO methods in which the reliability analysis and optimization iterations are either decoupled or combined into a single loop. To reduce the computational expense associated with nested methods, many researchers have developed single-loop approaches to RBDO which formulate the RBDO problem as a single optimization problem and solve it using the Karush-Kuhn-Tucker conditions of FORM or inverse FORM as constraints (Madsen and Hansen, 1992), (McDonald and Mahadevan, 2007), (Wang and Kodiyalam, 2002), (Liang et al, 2004). This general approach is described in Eq. (13).

$$\begin{aligned}
&\min_{\mathbf{d},\mathbf{u}} f(\mathbf{d}) \\
&s.t. \\
&g(\mathbf{d},\mathbf{u}) = 0 \\
&\frac{\mathbf{u}}{\|\mathbf{u}\|} = \frac{-\nabla_u g(\mathbf{d},\mathbf{u})}{\|\nabla_u g(\mathbf{d},\mathbf{u})\|} \\
&\|\mathbf{u}\| \geq \beta^t
\end{aligned} \tag{13}$$

The first and second constraints in Eq. (13) are simply the necessary conditions for the MPP, the third constraint requires the reliability index for the design to be greater than or equal to the required reliability index. This formulation is the single-loop formulation using direct FORM Karush-Kuhn-Tucker (KKT) conditions. Wang and Kodiyalam (2002) and (McDonald and Mahadevan, 2007) have developed an alternative approach to single-loop RBDO which uses the gradients of the limit state and the relationship $\mathbf{u}^* = \boldsymbol{\alpha}\beta$ to approximately calculate an MPP at which the design is then optimized, and the gradient evaluations and optimization are repeated until convergence to an appropriate FORM MPP and optimal design is achieved. Unfortunately, this approach to RBDO does not always converge, and other strategies known as “decoupled” or “serial single-loop” methods have been developed with improved convergence properties (Royset et al, 2001).

Several researchers have reformulated the nested RBDO problem to decouple the reliability analysis from the design optimization. This has been accomplished in one of two ways, using direct or inverse FORM. Torng and Yang (1993) and (Zou and Mahadevan, 2006) used FORM and replaced probabilistic constraints with a Taylor series expansion to solve:

$$\begin{aligned}
&min\ f(\mathbf{d}) \\
&s.t. \\
&\beta(\mathbf{d},\mathbf{u}^*) + \nabla_{\mathbf{d}}\beta(\mathbf{d},\mathbf{u}^*) \geq \beta_{required}
\end{aligned} \qquad (14)$$

Wu et al (2001), (Wu and Wang, 1998), (Royset et al, 2001) and (Du and Chen, 2004) used inverse FORM and developed decoupled formulations of the RBDO problem as stated in Eq. (15).

$$\begin{aligned}
&min\ f(\mathbf{d}) \\
&s.t. \\
&g_k(\mathbf{d},\mathbf{u}_k) \geq 0 \quad \forall k \in K \\
&where \\
&\mathbf{u}_k = argmin_{\mathbf{u}_k}(g_k(\mathbf{d},\mathbf{u}_k) \,|\, \|\mathbf{u}_k\| = \beta^k_{required} \quad \forall k \in K
\end{aligned} \qquad (15)$$

For this objective the inverse FORM approach is used to solve the robust optimization problem for a continuous set of possible scenarios. In this problem, a decoupled formulation similar to that used in (Du and Chen, 2004), but modified to assure convergence to a global optimum.

4.2.4 Economic Policy Optimization Under Uncertainty

In performing the analysis of the effects of the economic policy, a mathematical programming with equilibrium constraints approach will be used. However, these constraints will be linearized at the current equlibrium point. Therefore, the benefit of the economic policy would be calculable directly from the Leontief equation

$$\Delta\mathbf{x} = (\mathbf{I} - \mathbf{a})^{-1}\Delta\mathbf{c} = \mathbf{A}\Delta\mathbf{c} \tag{16}$$

However, in making plans to improve the output of the economy, the structure of the economy may not be certain. The objective considered here is that of finding a policy which yields the best results under the worst case. This section describes our approach to the problem of economic policy optimization under uncertainty. Since a policy which will give reliably good results is desired, the policy optimization problem can be stated as

$$\begin{aligned}
&\max_{\Delta\mathbf{c}} z \\
&s.t. \\
&Pr\{\sum_{i\,sectors} \Delta x_i - z \le 0\} \le \Phi(-\beta^t) \\
&\Delta\mathbf{x}_s = \mathbf{A}_s\Delta\mathbf{c} \quad \forall s \le \mathbf{S} \\
&\Delta\mathbf{c}_{min} \le \Delta\mathbf{c} \le \Delta\mathbf{c}_{max} \\
&\sum_{i\,sectors} \Delta c_i \le B
\end{aligned} \tag{17}$$

The optimization problem is formulated in a manner that decouples the optimization iterations and probabilistic analysis iterations. The decoupled formulation, shown below, maximizes the benefits of second-order effects, resulting from an action, $\Delta\mathbf{c}$, invested in the economy, as calculated at worst case scenarios (i.e., at values of the coefficients A that minimize the benefits of the second order impacts). The norm of the variates of the entries of **A** in the transformed, standard normal space is constrained to be a certain target of β^t. The decoupled optimization problem is then formulated as

$$\max_{\Delta \mathbf{c}} \sum_{i\,sectors} \Delta x_i \tag{18}$$

$$s.t.\Delta \mathbf{x} = \mathbf{A}^* \Delta \mathbf{c}$$

$$\Delta \mathbf{c}_{min} \le \Delta \mathbf{c} \le \Delta \mathbf{c}_{max} \tag{19}$$

$$\sum_{i\,sectors} \Delta c_i \le B \tag{20}$$

$$where \tag{21}$$

$$\mathbf{A}^* = argmin_{\mathbf{A}} \Big(\sum_{i\,sectors} \Delta x_i (\mathbf{A}(\mathbf{u})) \,|\, \|\mathbf{u}\| = \beta^t \Big) \tag{22}$$

Eq. (18) is the objective function, which for this problem is the sum of the total outputs of all sectors of the country's economy. The decision variables are the $\Delta \mathbf{c}$ vector (sector aid in the form of increased consumer demand for the sector's product or service). Eq. (19) represents the linear system model evaluated at the values of $\mathbf{A}^*$ as determined by the analysis of Eq. (22). Eq. (20) allows the decision maker to set bounds on the amount of aid to be given to each of the economic sectors. Eq. (21) is a budget constraint. Eq. (22) is the optimization problem, based on inverse FORM and Eq. (18) is its limit state function. The inverse FORM analysis returns worst-case scenario estimates for a given policy. Eq. (22) is solved at the incumbent value of $\Delta \mathbf{c}$ given by the policy optimization problem. The policy optimization problem given in Eqs. (18) through (21) and the optimization problem, given in Eq. (22) are solved iteratively until convergence. At convergence, the policy $\Delta \mathbf{c}$ provides the best solution under the worst-case of $\mathbf{A}$. The solutions which are best under the worst case realization of the random variables are called *reliably optimal.*

4.3 Identifying Reliably Optimal Solutions

A national economy can be represented as a SoS in the form of a network of infrastructures or sectors. A network conceptualization is given which will be useful later in determining reliably optimal investment strategies. Consider an example network flow representation in Figure 4.3 of a simple national economy comprised of three economic sectors, where the flow is considered to be a transfer of goods and services measured monetarily. Assume that the system is in an equilibrium state initially, from which it will move by applying a stimulus of $\Delta\mathbf{c}$ to the economy (i.e. increased demand for finished goods). Let Δx_{ij} be the change in the flow of money from sector i to sector j resulting from the stimulus $\Delta\mathbf{c}$. The new system state can be found by the balance condition $\Delta x_{ij} = (I-a)_{ij}^{-1}\Delta c_j$ and the policy $\Delta\mathbf{c}$ determines the changes in output.

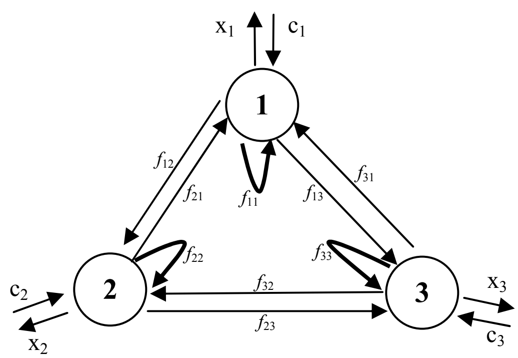


Figure 4.3 Economic Network

The framework developed in this Section 4.2 generates reliably optimal solutions by optimizing for a worst case scenario. The conditions of optimality for this approach are developed in this section and are based on the Nash equilibrium (Gibbons, 1992). Fixed-point theorems are used to show the existence of these solutions and derive their properties.

Definition 1**:** A policy $\mathbf{d}$ is said to be *first-order reliably optimal* if it is a maximizer of the program of Eqs. (18)-(22).

Theorem 1: Given a realization of the $(\mathbf{I}-\mathbf{a})^{-1}$ matrix, a policy is first-order reliably optimal if, and only if, there exists a return on investment $r^* \in \mathbb{R}^+$ such that for every sector in the economy:

$$\Delta c_i (r^* - \sum_j x_{ij} (\Delta c_i)) = 0 \quad (23)$$

$$r^* - \sum_j x_{ij} (\Delta c_i) \geq 0 \quad (24)$$

Proof 1: Assume there is a policy $\Delta\mathbf{c}$ for which Eqs. (23) and (24) are true, but policy $\Delta\mathbf{c}$ is not optimal. The first condition implies that either: 1) sector i receives no funding; or 2) the rate of return to the entire economy for the sector i funding is equal to r^*. The second condition requires that r^* be greater than or equal to the rate of return on investment at the given level of funding for all sectors. However, if $\Delta\mathbf{c}$ is not optimal, a funded sector would exist for which there is a rate of return less than r^*. An optimal strategy would require funding the sector with the higher rate of return. Thus, there cannot be a sub-optimal solution that meets these two conditions. This shows the optimality criterion for a policy, given a realization of the random parameters.

The next theorem shows that a robust optimal solution does exist, as a result of Brouwer's Fixed Point Theorem and provides the conditions for a robust optimal solution for the types of problem presented in this objective.

Theorem 2: There exists a reliably optimal solution Δc^R, such that $\mathbf{\Delta c}^*(\mathbf{u}^*(\Delta\mathbf{c}^R)) = \mathbf{\Delta c}^R$, where $\Delta\mathbf{c}^* = argmax_{\Delta\mathbf{c}}\left[\sum_i \sum_j \Delta x_{ij}(\mathbf{\Delta c}, \mathbf{u})\right]$ and $\mathbf{u}^* = argmin_{\mathbf{u}}\left[\sum_i \sum_j x_{ij}(\mathbf{\Delta c}, \mathbf{u})\right]$.

Proof 2: Since $\sum_i \sum_j x_{ij}(\mathbf{\Delta c}, \mathbf{u})$ is continuous and defined over convex, closed and bounded sets $\Delta\mathbf{c}$ and $\mathbf{u}$, $\Delta\mathbf{c}^*(\mathbf{u})$ and $\mathbf{u}^*(\mathbf{\Delta c})$ are also continuous. Further, $\Delta\mathbf{c}^*(\mathbf{u})$ and $\mathbf{u}^*(\mathbf{\Delta c})$ map $\Delta\mathbf{c}$ and $\mathbf{u}$ into itself. Therefore, Brouwer's Fixed Point Theorem assures the existence of the point $\Delta\mathbf{c}^R$ (Intriligator, 1971).

From the two proofs above, the optimality conditions for a given solution can be shown, as well as the existence of such a first-order reliably optimal solution. These are expressed mathematically as

$$c_i (r^* - \sum_j x_{ij} (c_i)) = 0$$
$$r^* - \sum_j x_{ij} (c_i) \geq 0$$
$$\mathbf{\Delta c}^*(\mathbf{u}^*(\mathbf{\Delta c}^R)) = \mathbf{\Delta c}^R \quad (25)$$

These properties will be very central in the development of an appropriate algorithm to solve the optimization of Eqs. (18) to (22). Unfortunately, it is usually impossible to derive this solution from solving Eqs. (18) to (21) and Eq. (22) in an iterative manner. A decoupled approach which allows optimization algorithms to sequentially update the policy and the worst case values of the $\mathbf{A}$ matrix will always allocate the entire investment to the sector with the largest multiplier under the most recent realization of the $\mathbf{A}$ matrix. However, upon calculating the worst case scenario under this particular single sector allocation, it is possible that another sector would have had a larger return on investment. This motivates the development of an alternative strategy to find the optimal solution, which will be developed in this subsection.

The algorithmic approach taken here is an incremental allocation approach, which is a modification to the decoupled approach. The basic idea in this optimization strategy is to more gradually move toward the fixed point. At this solution, given the worst case realization of the $\mathbf{A}$ matrix, we can have no better policy, but given the realizations of the policy, there can be no worse $\mathbf{A}$ matrix. In this approach, the economic aid is incrementally allocated to the optimal sector in each increment and the worst case scenario is found between increments given the current allocation levels. This process allows for the policy to gradually reach the fixed point solution for the OUU problem. This approach distributes the benefit more evenly and generates the largest second order effects. The algorithm for determining the alternative allocations may be implemented as follows:

- **Initialize:** Initialize random variables in the $\mathbf{A}$ matrix, using means of the random variables for initial values.

- **Step 1:** Allocate current increment of aid to the sector with the largest multiplier for the current value of $\mathbf{A}$.

- **Step 2:** Given the current allocation of effort, solve the inverse FORM problem to find the worst case scenario of the values of $\mathbf{A}$. Update values of $\mathbf{A}$ with those found from the solution of the inverse FORM problem. If the solution has not converged, go to Step 1. Otherwise, the current allocation is a candidate for optimality.

By gradually changing the policy from the optimal policy based on expected values, the risk is then able to be shared across sectors. The benefits of this risk sharing will be clearly shown in the example given in Section 4.4.

4.4 Numerical Results

4.4.1 National Economy as a System of Systems

One can represent a national economy as an interrelated system of systems through Leontief input-output coefficients. For example, a few sectors are listed in Table 4.1 comprising an interdependent system of systems. Rows 1-7 represent, in terms of the percentage of the gross domestic product (GDP), the quantities sold to each sector in columns 1-7, and demand for finished goods. Final demand figures are reported as factors of production for sectors 1-7, listed as finished goods for Consumer Consumption (C), Investment (I), Government Consumption (G), and Net Exports (X) (Roberts, 2004).

Table 4.1 Sector to Sector Economic Relationships

	Cost of Production							Final Demand				
Sectors	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**C**	**I**	**G**	**X**	**Output**
1. Agriculture	1	-	-	2.5	-	-	1.5	13.5	-	-	-	**18.5**
2. Oil	0.5	-	0.5	2	0.5	0.5	1.5	3	0.5	-	41.5	**50.5**
3. Mining	-	-	-	4	-	-	-	-	-	-	-	**4**
4. Industry	1.5	1	-	1	5	2	3.5	8	3	-	3	**28**
5. Construction	0.5	1.5	-	0.5	-	0.5	2	2	8	-	-	**15**
6. Transport	4	-	0.5	1	1	3	3	7	3	-	1.5	**24**
7. Service	0.5	3	0.5	1	0.5	0.5	2	4	-	28	-	**40**
Hhold Income	7	4	1	8	3.5	11	20					54.5
Profit & Taxes	0.5	35	0.5	3	1	3.5	2					45.5
Imports	3	6	1	5	3.5	3	4.5	16	6	4	-	52
COST (Output)	**18.5**	**50.5**	**4**	**28**	**15**	**24**	**40**	58.5	20.5	32	46	

Consider policy planning in the context of an economic aid package to this multi-sector system in which the economic impact of the package is to be maximized. The linkages of the autonomous, yet interdependent sectors are modeled by the coefficients of the Leontief input-output table and influence the aggregate GDP. Analysis of these interactions can provide measures of second and third order impacts of the economic aid on the GDP.

Table 4.2 Sector Interdependency Matrix A

Sector	1	2	3	4	5	6	7
1. Agriculture	1.071	0.007	0.009	0.104	0.041	0.012	0.055
2. Oil	0.050	1.008	0.138	0.105	0.077	0.037	0.058
3. Mining	0.019	0.006	1.005	0.153	0.054	0.017	0.019
4. Industry	0.134	0.041	0.036	1.074	0.381	0.116	0.135
5. Construction	0.041	0.035	0.016	0.031	1.020	0.030	0.062
6. Transport	0.281	0.013	0.161	0.100	0.123	1.158	0.119
7. Service	0.052	0.069	0.148	0.076	0.137	0.038	1.075
Multiplier	**1.65**	**1.18**	**1.51**	**1.64**	**1.83**	**1.41**	**1.52**

Sensitivity coefficients are computed from the sector to sector economic data in Table 4.1 by dividing each sector's cost of production by the total output. The $\mathbf{A}$ matrix, shown in Table 4.2 above, is generated from the sensitivity coefficients $\mathbf{a}$ and calculated as $\mathbf{A} = (\mathbf{I} - \mathbf{a})^{-1}$. Interdependencies among the sectors are depicted as the amount of output required by the row sector to satisfy a unit of demand for the column sector output. The sectors with greater interdependence to other sectors have larger values in their respective column in Table 4.2. For this objective, the uncertainties in the structure of the economy are represented by modeling the members of the $\mathbf{A}$ matrix as uniform random variables, centered at the mean values given in Table 4.2.

4.4.2 Single Sector Allocations

The absolute allocation policy seeks to identify a single sector which would maximize the benefit for the entire system, and allocate 100% of the aid to that critical sector. With this analysis, the decision maker can determine the best economic aid package according to the optimal results given from the problem formulation in Eqs. (18) to (22) with no constraints on the allowable funding allocations to any sector. B was set equal to 100 so that the optimal $\Delta\mathbf{c}$ values represent the percentage of the aid amount to be allocated to a sector. Initial values for the random variables were the means.

The optimal solution at the mean values was to allocate 100 percent of the funding to the construction sector. The optimization technique used was the reduced gradient method. Under the economic structure given by the mean values of the $\mathbf{A}$ matrix, focusing all effort on the construction sector created a level of economic increase equal to 183.3 percent of the total investment. This result is intuitive because the Leontief model is a linear model. The impact of a stimulus to one sector of the economy on the final production of the entire economy is found by multiplying the stimulus Δc_i by a multiplier equal to the sum of the entries of the i^{th} column of the $\mathbf{A}$ matrix. The multipliers, evaluated at the mean values of the $\mathbf{A}$ matrix entries, are given in the last row of Table 4.2.

The sectors in Table 4.2 with higher multipliers have stronger linkages, and thus generate greater secondary effects on dependent sectors. This means that if a sector with strong linkages is aided, the performance of the linked sectors will also be improved. For example, the construction sector has the greatest multiplier. Thus, the greatest secondary effects on GDP result from providing all aid to the construction sector. Inverse FORM was applied to the policy of providing all aid to the construction sector. The worst case values of the $\mathbf{A}$ matrix at a target reliability index, $\beta = 2$, are given in Table 4.3.

Table 4.3 Worst Case Matrix for Construction Only Policy

Sector	1	2	3	4	5	6	7
1. Agriculture	1.08	0.01	0.01	0.10	**0.03**	0.01	0.06
2. Oil	0.05	1.01	0.14	0.11	**0.04**	0.04	0.06
3. Mining	0.02	0.01	1.01	0.15	**0.04**	0.02	0.02
4. Industry	0.13	0.04	0.04	1.07	**0.07**	0.12	0.14
5. Construction	0.04	0.03	0.02	0.03	**1.02**	0.03	0.06
6. Transport	0.28	0.01	0.16	0.10	**0.05**	1.16	0.12
7. Service	0.05	0.07	0.15	0.08	**0.05**	0.04	1.08
Multiplier	**1.65**	**1.18**	**1.53**	**1.64**	**1.30**	**1.42**	**1.54**

The overall benefit of this plan to the economic SoS under the worst case scenario would be 130 percent of the original stimulus, as calculated by summing the worst case values in the fifth column of Table 4.3. This implies that the return on investment would be greater than 30%, with a probability of $\Phi(2) = 0.977$. An important source of uncertainty in this analysis is the extent to which other sectors are linked to the construction sector. The sectors with the greatest deviations from the means at the MPP (see Table 4.2) are those sectors with the strongest links to construction (i.e., industry, transportation, and service). As a potential investment strategy, the analysis suggests that economic data gathering efforts should focus on obtaining more information about linkages between construction and other sectors in order to reduce variability in the interaction estimates.

The solution process for Eqs. (18) to (22) continues at the worst case values returned by the solution of the inverse FORM problem. Optimization at the worst case results in a new aid package where the agriculture industry receives 100 percent of the aid package. This solution is intuitive, since at the worst case for the "construction only" package the multiplier for the construction sector drops to 1.30 and the agriculture sector has a multiplier of 1.65. Hence, at the new worst case scenario the new economic improvement for the SoS would be 165% of the aid amount, were all of it invested in the agriculture sector. With a new "agriculture only" policy, the inverse FORM problem is now solved again to generate a new worst case scenario. When inverse FORM is applied to the policy of giving all aid to the agriculture sector, the worst case scenario for the values of the **A** matrix, for $\beta = 2$, is given in Table 4.4. In this scenario, the growth in SoS GDP totals 123 percent of the total investment.

Table 4.4 Worst Case Matrix for Agriculture Only Policy

Sector	1	2	3	4	5	6	7
1. Agriculture	**1.03**	0.01	0.01	0.10	0.04	0.01	0.06
2. Oil	**0.03**	1.01	0.14	0.11	0.08	0.04	0.06
3. Mining	**0.01**	0.01	1.01	0.15	0.05	0.02	0.02
4. Industry	**0.04**	0.04	0.04	1.07	0.38	0.12	0.14
5. Construction	**0.02**	0.03	0.02	0.03	1.02	0.03	0.06
6. Transport	**0.05**	0.01	0.16	0.10	0.12	1.16	0.12
7. Service	**0.03**	0.07	0.15	0.08	0.14	0.04	1.08

When the strategy was optimized under this current worst case scenario, the optimal solution was once more to focus all effort on the construction sector. This means that cyclic behavior is observed with this approach, and thus it did not converge to a uniquely optimal solution. This cyclic behavior was the effect of the multipliers resulting from the Leontief model's linear form. The optimal strategy was for every realization to give all aid to the sector with the largest multiplier. When the strategy was optimized under this worst case scenario, the optimal solution was once more to focus all effort on the construction sector. The cycling of solutions between the construction and agriculture sectors was partially the result of initializing the random variables at their means. Were the optimization procedure initialized from a point more favorable to one of the other five sectors, this sector could have also been included in the cycling of solutions. Inverse FORM, however, allows for the comparison of solutions by calculating the worst case objective value associated with scenarios in which each sector receives 100 percent of the aid package. The worst case multipliers associated with focusing all effort on each individual sector are given in Table 4.5.

Table 4.5 Worst Case Sector Multipliers

Agriculture	1.23
Oil	1.06
Mining	1.17
Industry	1.28
Construction	1.29
Transport	1.14
Service	1.22

By comparison of worst case solutions, it can be seen that if an absolute allocation approach is used, then all effort would be directed toward the construction sector.

4.4.3 Incremental Allocation

The solution for the incremental allocation approach to the case study with the aid divided into ten equal allocations is summarized in Table 4.6. Reported are the optimal allocations and worst case multipliers at each allocation step.

Table 4.6 Allocations Based on Incremental Algorithm

iteration	Agr	Oil	Mi	Ind	Cstn	Trans	Serv	Aid to
1	1.65	1.17	1.51	1.64	1.8	1.41	1.52	Const.
2	1.65	1.17	1.51	1.64	1.2	1.41	1.52	Agric.
3	1.55	1.17	1.51	1.64	1.3	1.41	1.52	Indust
4	1.48	1.17	1.51	1.57	1.3	1.41	1.52	Indust
5	1.43	1.17	1.51	1.37	1.4	1.41	1.52	Serv
6	1.43	1.17	1.51	1.38	1.5	1.41	1.42	Mine
7	1.47	1.17	1.38	1.40	1.5	1.41	1.42	Const.
8	1.47	1.17	1.39	1.43	1.4	1.41	1.44	Agric.
9	1.43	1.17	1.40	1.44	1.4	1.41	1.44	Const.
10	**1.43**	**1.17**	**1.40**	**1.44**	**1.4**	**1.41**	**1.44**	Indust
Aid%	**20%**	**0**	**10%**	**30%**	**30**	**0**	**10%**	

For this allocation the worst case total output of all sectors is 143 percent of the original stimulus, or a worst case multiplier of 1.43. This allocation provides a substantially better worst case value than all solutions given in Table 4.6. This solution gives better performance under worst case conditions because risk is shared among several sectors. Notice that 80% of the aid was allocated to the three sectors with the strongest linkages, the agriculture, industry, and construction sectors. These sectors have the strongest linkages in the economy. These results suggest a SoS equilibrium solution may exist for certain problems, as shown by the incremental allocation method graph in Figure 4.4. The pattern observed in Figure 4.4 suggests that an incremental allocation policy may lead to an optimal policy for the SoS. The incremental allocation procedure was repeated for increasing increments and an equilibrium solution was found to exist for the problem after twenty-seven iterations.

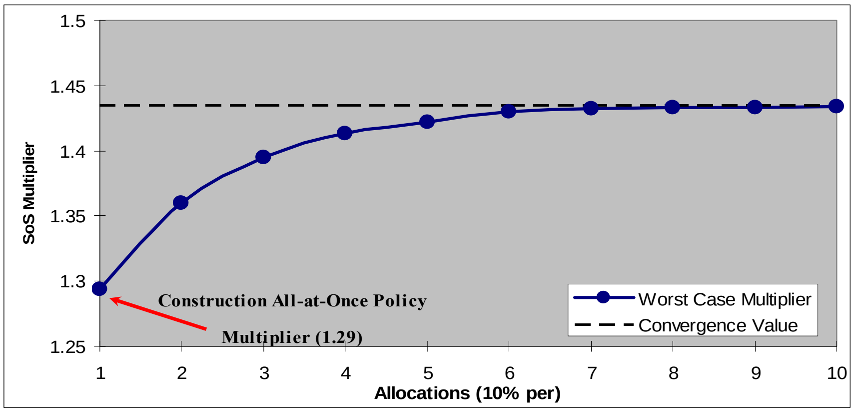


Figure 4.4 Convergence of Incremental Allocation Algorithm

As shown in Table 4.7 and Table 4.8, the sector multipliers converged as successive optimization iterations were performed. After 27 iterations, the funded sectors all had multipliers of 1.47. Under this investment policy, the percentage invested in each sector was allocated as shown in Table 4.8. At the optimal scenario, the marginal benefit of any dollar spent in any funded sector is equal. The optimal solution was accepted, based on this result. It was determined that the sector multipliers of all the funded sectors were equal for the incremental allocation policy with twenty-seven increments. The return on investment to any funded sector was equally beneficial to the economy.

Table 4.7 Incremental Policy After 10 Iterations

Agriculture	1.43 (20%)
Oil	not funded
Mining	1.40 (10%)
Industry	1.44 (30%)
Construction	1.44 (30%)
Transport	not funded
Service	1.44 (10%)
TOTAL SoS	**1.43**

Table 4.8 Incremental Policy After 27 Iterations

Agriculture	1.47 (26%)
Oil	not funded
Mining	1.47 (7%)
Industry	1.47 (26%)
Construction	1.47 (30%)
Transport	not funded
Service	1.47 (11%)
TOTAL SoS	**1.47**

4.4.4 Analysis Verification

To verify the consistency of the analysis, worst-case sector multipliers were generated using a genetic algorithm to compare with the worst-case multipliers obtained from our reduced gradient optimization technique. The values from the absolute allocation policy results of the Leontief formulation were compared. The genetic algorithm analysis contained 1000 runs, each consisting of 100 evaluations and reported the best objective function values and associated design variable values, subject to the constraint that $\beta = 2$. Table 4.9 illustrates the multiplier comparisons for the linear and non-linear formulations in which total economic investment was allocated to the agriculture sector. Both solution techniques provide similar results. The genetic algorithm multiplier of 1.26 for the Agriculture sector is close to the reduced gradient multiplier of 1.23 obtained by the Leontief based optimization.

Table 4.9 Analysis Verification

Policy Benchmarks	Optimization Technique	
	Reduced Gradient	Genetic Algorithm
Reliability Level	95%	95%
β-value	2	2
Final SoS Multiplier	1.23	1.26

The genetic algorithm results in Table 4.9 suggest the increase in GDP for the economy will be between 26%. This compares well with the corresponding worst-case estimates of GDP increase of 23% from the reduced gradient optimization technique.

The numerical results were also verified with reliability estimates derived from Monte Carlo simulation. Figure 4.5 depicts the CDF values for an economic policy in which an increase of 10% of each sector's goods and services is invested. Figure 4.5 shows the Monte Carlo and FORM estimates for worst case investment policies. The FORM method is consistent with Monte Carlo, particularly for extreme event analysis such as the worst case investment planning in this numerical example, although linearization errors due to the transformation of the uniform variables to the standard normal space are small, yet clearly present.

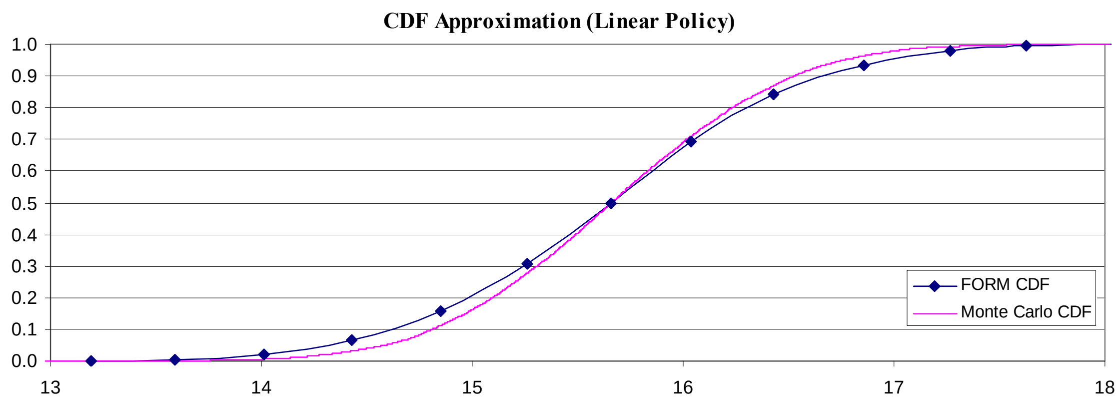


Figure 4.5 CDF Approximations of System Response

4.5 Conclusion

The methodology implemented in this chapter produces robust results and improves economic investment decision making under uncertainty. This objective integrated three techniques central to system of systems engineering. First, a linear economic input-output model was used to represent a network of economic sectors as a system of systems. Second, an inverse FORM technique was used to evaluate the probabilistic constraints in SoS optimization. Third, an efficient, decoupled, reliability-based design optimization formulation was used to produce robust policy recommendations in an uncertain decision making environment.

There are two general conclusions from this research. First, the OUU framework integrates system analysis, uncertainty modeling and decision making and is flexible enough to accommodate a variety of system models, such as the illustrated Leontief Input-Output model. Nonlinear models, such as Cobb-Douglass production functions and general equilibrium models, as well as more complex models of system behavior based on network flows, system dynamics, or agent-based models can also be integrated within the proposed OUU framework. Second, the use of FORM yielded probabilistic sensitivity measures which allow decision makers to identify the most important random variables, identify critical elements of the system, and quantify the impacts of interactions among the SoS elements. Existing literature contains many systems analysis techniques for engineering and science; however, few studies have been reported that integrate simulation, uncertainty analysis, and optimization to provide decision support for system of systems decision-makers, such as military and governmental planners. Current literature applies these approaches separately. The proposed framework integrates these techniques, by accounting for uncertainties and optimizing objectives under these uncertainties.

Additionally, the system of systems model has limitations. For instance, in a time of rapid change, an economy's structure is uncertain, thus a static, input-output, coefficient-based structure will fail to properly capture the characteristics of the dynamic environment. Haimes and Jiang

(2001) modeled this uncertainty by assigning expected values to the elements of the **A** matrix, and based decision making on maximizing a first order estimate of expected utility. As this objective has shown, this limitation may lead to oscillation among solutions, which is confusing and insufficient for making decisions under uncertainty. Using a single sector allocation approach deprives the system of benefits gained by sharing risks among the different sectors of the economy. This motivated the development of the incremental allocation method in Section 4.4.3, which mitigated the limitation and expanded the usefulness of the proposed framework for practical application.

There are several insights into SoS problems, in general, from the research performed in this objective. First, for many SoS policy optimization problems, it is unknown whether policies are stable. In some systems of systems, perturbations of some solutions may result in large divergence in the trajectory of the system's state variables, resulting in chaotic behavior and defeating the purpose of controls. The emphasis on determining a best "worst case" optimal policy serves as an example for formulating other SoS problems with a focus on robustness.

A second insight for SoS problems, in general, is that interdependent problems can often posess multiple types of uncertainty. A potential extension of the OUU framework will be to account for multiple types of uncertainty. Currently the uncertainty associated with system interdependence is accounted for through random variables. It is likely for uncertainty in many system interdependencies to be epistemic in nature, and requiring non-probabilistic representations (i.e., fuzzy sets, evidence theory, etc.). An important benefit of using the proposed framework is that it can help decision makers incorporate coupling effects between systems in a SoS.

Another general insight is system performance can be modeled and analyzed by representing the SoS as a network. The Leontief input-output model is just one representation that can be easily constructed, readily understood by a wide range of policy planners, and effectively implemented on a short time scale. The overall framework developed in this objective can serve as a black box approach to accomodate more detailed models. The computational framework is based on analytical reliability approximations and greatly reduces the expense of the OUU process. The ability to rapidly construct informative and understandable models for decision making under uncertainty and perform OUU in a computationally affordable manner is particularly valuable when planners must make decisions under uncertainty and time pressure.

CHAPTER V

MODELING AND SIMULATING STOCHASTIC HUMAN-PHYSICAL NETWORKS WITH HYBRID DYNAMICS USING AGENT-BASED MODELS

5.1 Introduction

In a coupled human-physical network, user choices change the state of the physical system and the resulting physical conditions are the basis of subsequent user choices. The interdependent nature of the relationship between humans and their operational environment is inherently complex. Failure to properly account for the interactions between them can lead to unintended events with unanticipated consequences. A human-physical network analysis framework that integrates user choice and system physics under uncertainty is developed in this objective. This framework extends the general OUU approach from the previous objective. The framework developed in this objective replaces the system modeling and uncertainty analysis loops in the OUU approach as a more detailed method for SoS decision-makers to model and analyze performance in systems where stochastic and hybrid dynamics are present. Only the uncertainty associated with the stochastic nature of the physical system is captured in this objective using stochastic agent-based simulation. Uncertainty due to random network demands, as well as, uncertainty in SoS objective priorities comprise the third OUU loop (decision analysis/optimization) and are addressed in later chapters.

The application and SoS example for this objective and the remaining objectives in this dissertation is in the transportation domain. This objective uses the transportation network as the example and considers the problems for associated with freeway and surface streets, multi-modal transit issues and the impacts of the coupled nature of such stochastic networks with multiple decision makers. The purpose of properly simulating the operating environment and selecting the most appropriate models to characterize its major elements is to support insightful optimization and lead to effective policy-making. Both are most effective for the broader transportation system when aspects of and impacts to as many facilities as possible are considered. The result of the integrated approach presented in this objective is a suggested optimal control policy for the specified operating environment. The example network is a hybrid SoS, meaning both discrete and continuous elements govern network performance. The framework process is applied to an example transportation SoS comprised of a multi-modal flow network and multiple decision makers. Success of the policy is measured in terms of modal choice, network efficiency, revenue and reliability.

To illustrate the framework, an example human-physical network is presented. A stochastic simulation represents a multi-modal transportation system in which users first decide whether to travel by bus or auto, and then auto drivers choose between a surface street and freeway route. Iterative analysis between the physical system model and a logit-based user choice model generate equilibrium modal flow times and aggregate mode splits representing the proportion of users preferring each mode under the given operational conditions. Equilibrium network parameters are used to estimate network failure probabilities associated with given operational policies (i.e., toll, bus fare, signal timing) under given conditions (i.e., demand) and the number of users who fail to experience a prescribed level of service on the physical network. Resulting trends and sensitivities serve as insights to inform SoS policies aimed to regulate or incentivize preferred user behavior and shape policy decisions such as pricing modal access to promote a desired aggregated flow across the multi-modal network.

5.2 The Framework

The Stochastic Human-Physical Network Analysis shown in Figure 5.1 is the step-by-step process developed in this research to characterize and analyze networks with multiple flow systems, multiple modes of flow, multiple decision-makers, multi-disciplinary network properties. The benefit of this framework its ability to facilitate representative reduced models for faster network optimization of policy objectives. This section describes the framework and its components.

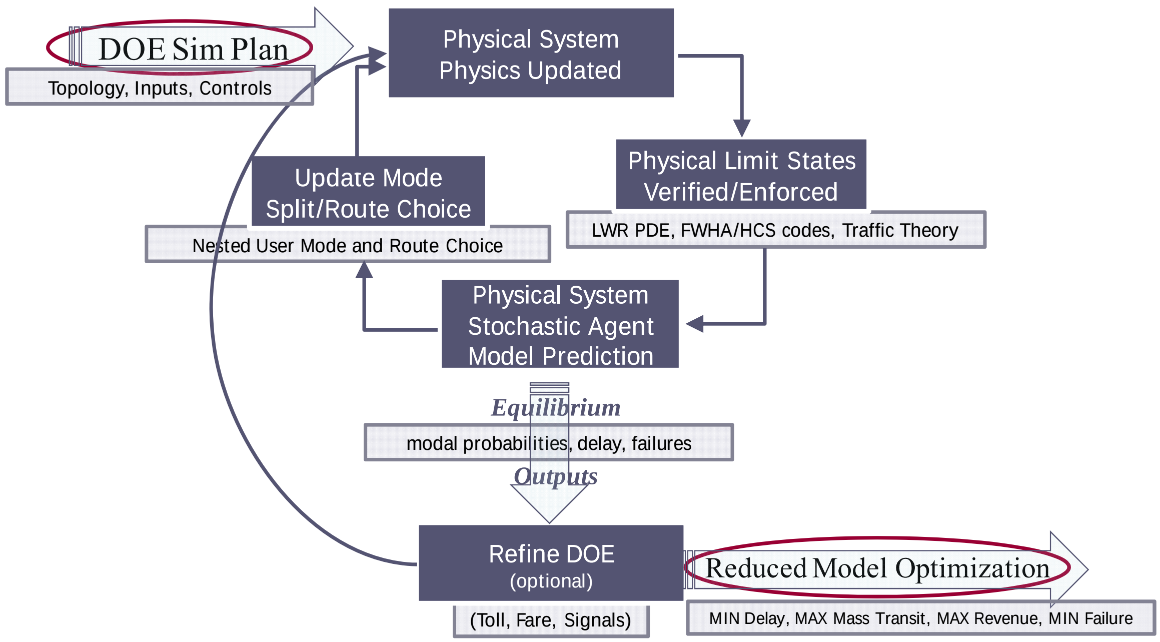


Figure 5.1 SoS Analysis Framework

Design of Experiments

The initial step in the framework requires understanding of the operational environment being studied. The network simulation process must sufficiently characterize the network topology, inputs relevant to performance and available controls desired to be used. Depending on the number of input variables, an appropriate experimental design must be followed when conducting network simulations. Latin hypercube designs (LHDs), as described in (Prescott, 2009), use a relatively small number of design points to investigate a relatively large number of factors. These designs were initially introduced by (McKay et al, 1979); however, since their introduction, LHDs have found extensive use in computer experiments and are also a popular design type for Kriging (Beers and Kleijnen, 2005). A carefully selected, space-filling fractional factorial design will be used in this research to efficiently explore the network performance under various control policies and across a range of operating conditions.

Physical System Simulation

The physical simulation represents the performance of the physical networks in the SoS and comprises the following three steps in Figure 5.1: Physical System Physics Updated; Physical Limit States Verified/Enforced; Physical System Stochastic Agent Model Prediction). A detailed physics model that captures many of the complexities of the network is used, while imposing many of the realistic physical limitations of the actual networks (i.e., capacities, flow rates, densities, user types, laws, regulations, etc). The simulation is coded to enforce the natural laws of physics and governing traffic theory and safety regulations on the network over an analysis period. These constraints ensure the physical boundaries of time and space are not violated in the simulated environment.

The physical system is simulated as a stochastic agent simulation. Not every network can be represented with a stochastic agent simulation or requires this level of detail in analysis; however, when necessary, stochastic network simulation using agent-based modeling give some important benefits. One benefit of using a stochastic agent simulation is the ability to propagate uncertainty across the myriad to network attributes. Another benefit is the network can represent the flow of individual agents, each following simple rules and whose collective behaviors represent a comprehensive view of the network. The disadvantage of using a high-fidelity stochastic simulation is that model evaluations under these conditions can be computationally expensive.

User Choice Modeling

Network users decide on their mode of transportation and their route of travel and this activity comprises the Update Mode Split/Route Choice step in Figure 5.1. Utility functions are used to model user mode and route preferences at various mode and route costs (i.e., bus fares and freeway tolls). Certain inherent preferences for a given operating environment are also reflected in the user models. For example, surveys of the users in a densely populated urban center with expensive parking fees and fuel prices may indicate an inherent preference for bus travel. Appropriate user choice models must be employed to predict the portion of the usership which prefer each network mode and route, based on utility functions and user choice models. Logit-based user choice models are well-researched and shown to be valid for many these type of problems (Sheffi, 1985). Users are assumed rational decision makers who prefer the mode and route with the greatest utility based on a combination of travel time, mode cost and route cost.

Equilibrium Analysis

The computational models for user choice and system physics are coupled through an iterative scheme designed to achieve stochastic network equilibrium (SUE). Daganzo and Sheffi (1977), (Daganzo, 1979) and (Powell and Sheffi, 1982) introduced the first stochastic user equilibrium formulations. Network equilibrium is defined by a convergence criterion in this study that is consistent with the formal proof presented in (Powell and Sheffi, 1982). Initial demand values are assigned in the network simulation which evenly distributes users across the modes and routes according to the user choice models. As simulations are performed, mode and route travel times become inputs to the logit-based models and new mode and route preferences are obtained. These updated demands are adjusted in the simulation and the physical system simulation is repeated until a convergence criterion is met. Equilibrium is based on the mode and route load values from the previous iteration changing less than a prescribed amount. In this study, when user preferences change by < 2% from the previous iteration, then network equilibrium is assumed and the equilibrium output values reported.

Refine Design of Experiments (DOE)

This step in the framework is the opportunity for the network level decision maker to refine or adjust the network control policy. If the initial policy values (i.e., toll price, bus fare, signal timing, etc) for the system of systems are satisfactory, then this step is skipped and the reported equilibrium outputs are used to derive appropriate reduced models from which policy optimization can be obtained. If the initial policy values (i.e., toll price, bus fare, signal timing) for the network are not satisfactory, then the new policy values are reflected as an update in to system physics. The framework steps are repeated as described above until network equilibrium values are once again obtained.

5.3 Example Network Simulation

In traffic engineering, the concept of traffic control is giving way to the broader philosophy of Advanced Traffic Management Systems (ATMS), whose purpose is not only to move vehicles, but also to optimize the utilization of transportation resources to improve the movement of people and goods without impairing the community (Yang and Koutsopoulos, 1996). One of the most important analytical tools of traffic engineering is computer simulation. Computer simulation is more practical than a field experiment for the following reasons:

- It is less costly
- Results are obtained quickly
- Some measures of effectiveness are more easily obtained over field studies
- No disruption on traffic operations, which can accompany a field experiment
- Significant physical changes to the facility can be explored
- Evaluation and analysis of the operational impacts of various controls are possible

5.3.1 Traffic Software Integrated System (TSIS)

Transportation network control research is widely performed, but typically as deterministic analysis and focused on a single facility or system (Bayen et al, 2004). Current stochastic simulation in transportation analysis does not explicitly consider user choice (McTrans, 2008). Therefore, a computational approach that facilitated the coupled human-physical relationship was required. The software used to simulate the example network was a simulation package called the Traffic Software Integrated System (TSIS). TSIS was chosen because it was a customizable, stochastic agent simulation with the features and capabilities that fit the needs of this research.

TSIS is an integrated development environment designed to perform traffic operations analysis. TSIS is a toolbox, built using component architecture, which allows the user to define and manage traffic studies, define traffic networks and create inputs for traffic simulation analysis, execute traffic simulation models, and interpret the results of those models (TSIS, 2008). This concept of a single integrated simulation system with flexibility and ease of use and that can optimize the efficiency of all computations was conceived by the Federal Highway Administration (FHWA) in the mid-1970s (TSIS, 2008). FHWA has since supported a series of projects to implement this design and to develop and update the software. TSIS-CORSIM 6.1, the latest version of the software, was used for this research. The analysis engine for TSIS is written in FORTRAN and the system interface is coded in C^{++}.

5.3.2 Description of the Stochastic Network Simulation (CORSIM)

To test the effect of control policies on trip patterns, it was necessary to analyze a given area that contained a substantial portion of the routes that the trip-makers follow. TSIS was used to create a corridor simulation (CORSIM) to represent the physical environment. CORSIM simulated the operational environment consisting of a multi-modal, multiple flow route simulation that was capable of representing many of the features and complexities of traffic flow in large urban areas containing surface street networks and freeways and bus transit within reasonable computer usage requirements.

Figure 5.2 illustrates the integrated transportation network as a surface street network with a two-way bus line surrounded by a freeway network and two subnetwork interfaces.

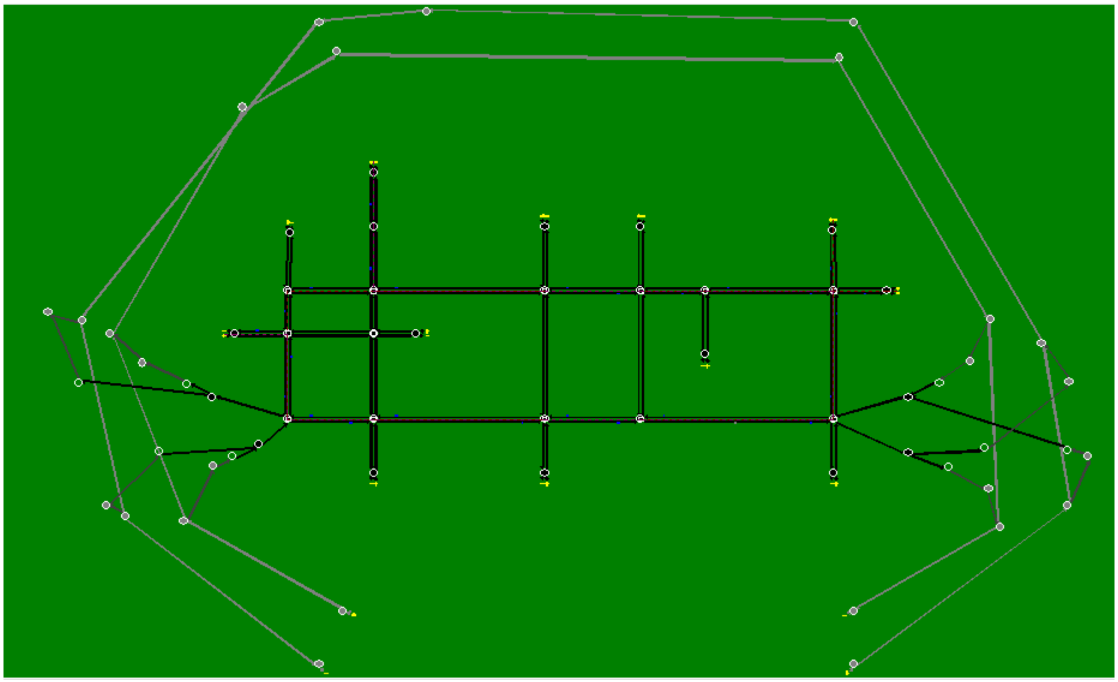


Figure 5.2 Integrated Transportation Network Topology

The multi-network model designed for this research had many standard features of the prevailing freeway geometries, such as multiple-lane freeways, on/off ramps, connectors to other freeways, variations in grade, lane additions and lane drops and auxiliary lanes to facilitate lane-changing or freeway entry and exit. The CORSIM network had 100 surface street links, 12 Surface Entry Nodes, 2 Freeway Entry Nodes, 2 interface nodes for switching subnetworks, a 3 lane freeway system with a capacity of 2200 vehicles per lane per hour, and a 2.5 mile East and West bound bus line.

Agent-based approaches are very suitable for this domain (Davidsson, 2005). CORSIM is a stochastic agent simulation consisting of an integrated set of two microscopic simulation models that represent the entire traffic environment. Each of the component models of CORSIM simulates a different subnetwork. NETSIM is a simulation model that represents traffic on surface streets; FREESIM is a simulation model to represent traffic on freeways. These microscopic simulation models propagate movements of individual agents (vehicles) across the network. Agents move about the network according to specific rules that reflect stochastically determined agent decisions constrained by system safety and traffic theory boundaries. An important characteristic of this simulated network is the assumption that no catastrophic events occur during the analysis period. The flow of traffic is in accordance with natural physical laws and US transportation guidelines (McTrans, 2008). Exogenous events such as wrecks, non-working control devices, work zones or upstream drivers are not considered congestion factors in this study.

The interface nodes represent points at which vehicles leave one sub network and enter another (see Figure 5.3). Nodes of this type were distinguished from other nodes in the network for the purposes of defining the boundaries of the subnetworks in order to assessing both subnetwork and full network performance metrics. Entry interface links received traffic from the adjoining subnetwork and exit interface links carried traffic exiting the subnetwork to adjoining subnetworks. The interfacing of adjoining subnetworks was accomplished by defining the network specific demand in the simulation code as the route demands generated by the user choice model.

Figure 5.3 CORSIM Subnetwork Interface

CORSIM is a stochastic model, which means that randomness is a factor in assigning driver and vehicle characteristics and to decision making processes. The MOEs that are obtained from a simulation are the result of a specific set of random outcome. For example, one random number seed may result in three very conservative drivers driving side by side on a three-lane roadway blocking more aggressive drivers behind them. The resulting MOE would reflect a lower average speed then has been observed in the real world. In order to mitigate the potential for MOEs generated from a single run to be misleading, the network runs were simulated several times using different sets of random number seeds. The resulting distribution of MOEs reflected a more accurate representation of the network performance.

The stochastic elements of the simulation greatly enhance the model's ability to reflect a broader sense of the impacts to the network under various conditions. As with real traffic conditions, the effects of control strategies on network performance in a stochastic simulation depend on agent behavior and the evolution of the network over a time period. CORSIM applies time step simulation to describe network operations. A time step is one second. Each vehicle is a distinct object that is moved every second. Each variable control device (such as traffic signals) and each event are updated every second.

Each vehicle was identified by type. Up to nine different types of vehicles (auto, carpool, truck, bus, etc.) with different operating and performance characteristics were specified. Furthermore, driver behavioral characteristics (i.e., passive or aggressive) were assigned to each vehicle. Based on which driver type (out of nine types) was assigned, the associated kinematic properties such as speed and acceleration, as well as its moving or queued status were determined. Also assigned stochastically were turn movements, queue discharge headways, and bus dwell times at bus stops. As a result of such specific behavioral attributes, each agent's behavior was intended to reflect real-world processes and adhere to the governing physical laws and regulations imposed by transportation guidelines (i.e., Highway Capacity Manual). Three views of CORSIM simulations are shown in Figure 5.4, Figure 5.5 and Figure 5.6 and represent respectively, the integrated FREESIM-NETSIM network, the NETSIM network only, and close-up view of intersection activities.

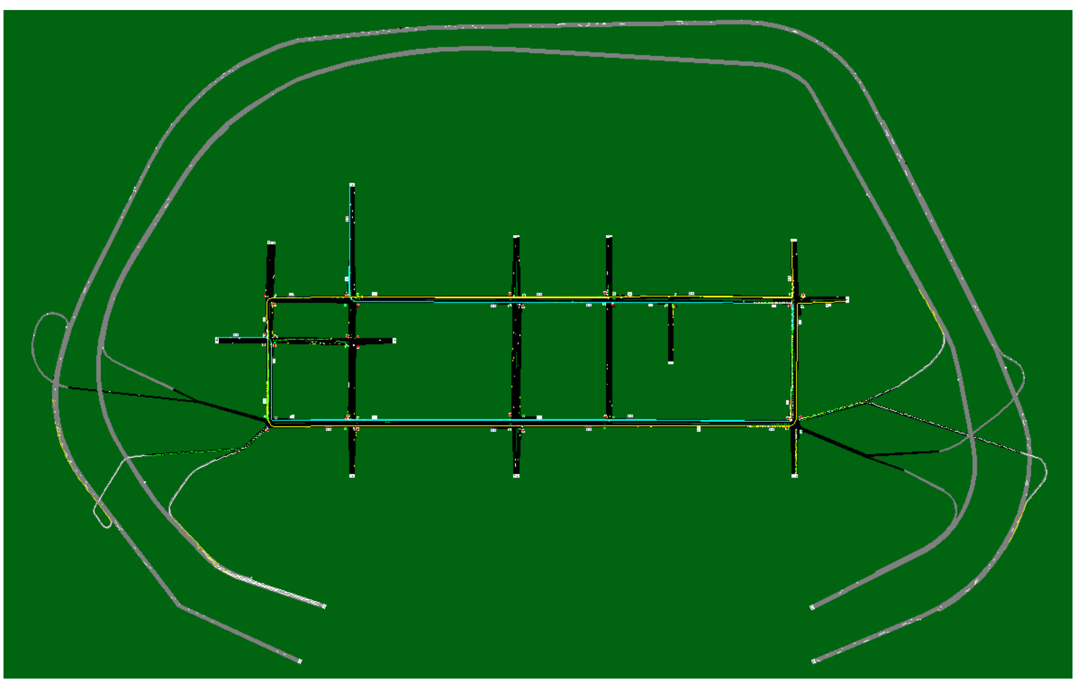


Figure 5.4 Integrated NETSIM-FREESIM Network

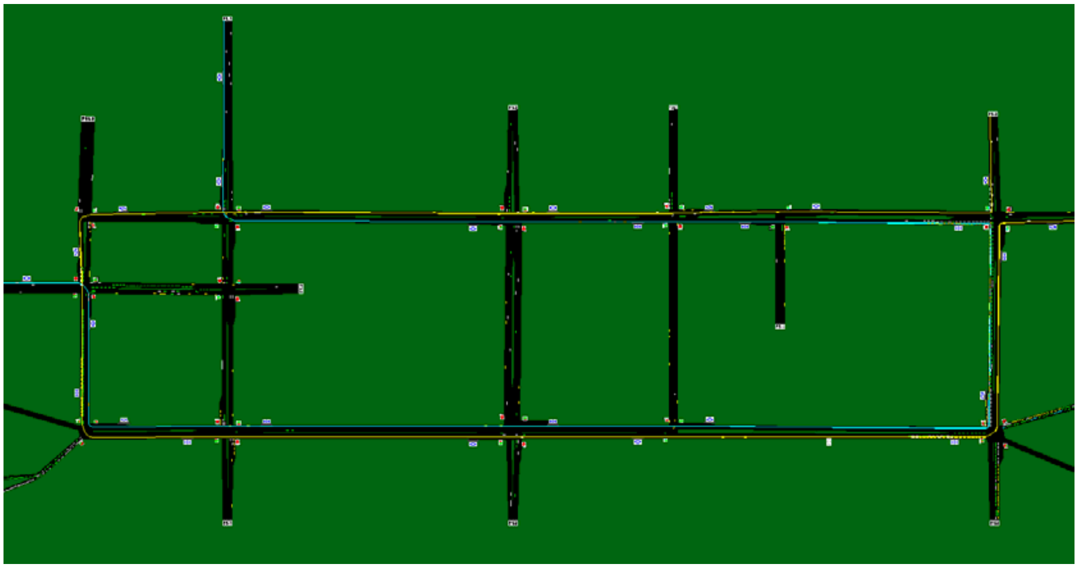


Figure 5.5 NETSIM Network Only

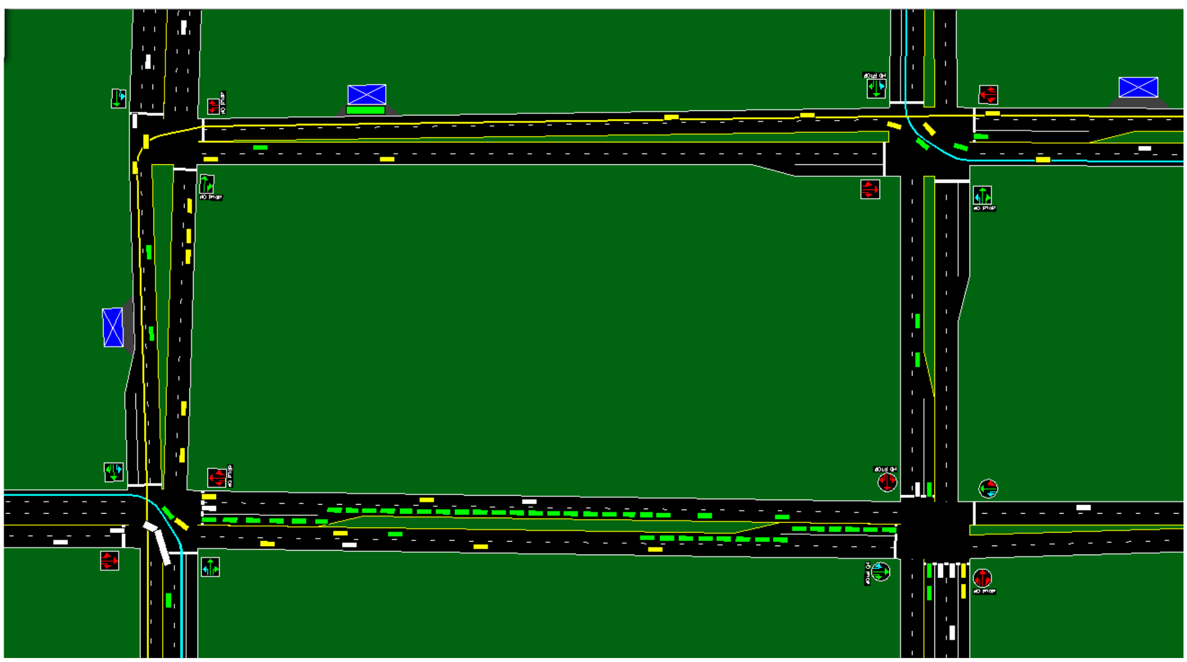


Figure 5.6 Intersection Activities (Close-up)

Each time a vehicle is moved, its position (both lateral and longitudinal) on the link and its relationship to other vehicles nearby were recalculated, as were its speed, acceleration, and queued or moving status. Agents moved according to car-following logic and in response to traffic control devices. For example, buses service passengers at bus stops with movements which differ from those of private vehicles. Congestion can result in queues that extend throughout the length of a link and block the upstream intersection, thus impeding traffic flow. CORSIM accumulated data every time step. At the end of each time period the accumulated data was used to produce and report Measures of Effectiveness.

5.3.3 Measures of Effectiveness

Measures of Effectiveness are standards against which the capability of a solution to meet the needs of a problem may be judged (Sproles, 2000). The CORSIM contains a comprehensive set of MOEs, defined for each subnetwork separately (TSIS, 2008). There were 21 NETSIM MOEs and 16 FREESIM MOEs. MOEs were also measured for the combined CORSIM network. Some of the MOEs particularly significant to the research performance objectives were: Total Time, Vehicle Minutes per Mile, Delay, Bus Travel Time, and Phase Failures.

The MOEs provide insight into the effects of the applied strategies on the traffic stream, and they also provide the basis for optimizing that strategy. The Travel Time (TT) MOEs were the basis for evaluating whether the network had reached equilibrium. The logit-based utility functions were evaluated and the physical system modal and route splits were updated until the mode and route splits for consecutive simulation runs were within the convergence criterion (Sheffi, 1981). The importance of the reaching network equilibrium was that it meant, at a macro level, that users were not incentivized to change (prefer) a different mode or route, under the prescribed loading and policy (modal access cost, route access cost and signal timing).

5.4 Framework Applied to the Example Network

This section illustrates how the research framework was applied in the context of the CORSIM network described in the previous section.

5.4.1 Network Description

Consider the CORSIM transportation system previously described located in a geographic locale in which user surveys have shown an inherent preference for bus travel over auto commutes due to high fuel prices and expensive, scarce available parking in the area. Among auto routes, freeway travel is preferred to surface street routes based on the faster free flow speed on the freeway network. User choice is a function of the cost of travel measured in terms of time and money. Individual mode and route choice follows utility functions defined for each based on user survey results. It is assumed that half of the user decision is simply which travel option is faster. The other half is the financial cost associated with obtaining the faster travel option (time value of money reflected in the utility function).

5.4.2 Objectives

Network decision-makers seek to control the flow such that the tradeoff between progression on each mode, governed by the flow physics (i.e., LWR PDE and queuing principles) optimizes a given system-level measure of performance (i.e., delay, network failure). Many metering schemes simply transfer the travel cost between modes; however, an effective metering scheme should optimize system-wide metrics (i.e., minimize overall user delay). Planners and managers for the example transportation system of systems are assumed to be cooperative and willing to implement policies deemed optimal for the entire network. These policy makers seek to use available operational controls to promote travel behavior that optimizes: 1) network delay; 2) mass transit ridership; 3) total revenue; 4) network reliability. These objectives are described below and the simulation data pertaining to each of these objectives was collected.

Network Delay

Delay is measured as a function of how much longer the trip lasts as compared to an unimpeded trip at free flow speed. The aggregate time above free flow speed is reported as delay.

Total delay is comprised of travel delay and control delay, both of which are captured as simulation outputs. Travel delay is a straight forward time-distance calculation across each CORSIM link. A detailed description of control delay based on the diagram in Figure 5.7 is shown below.

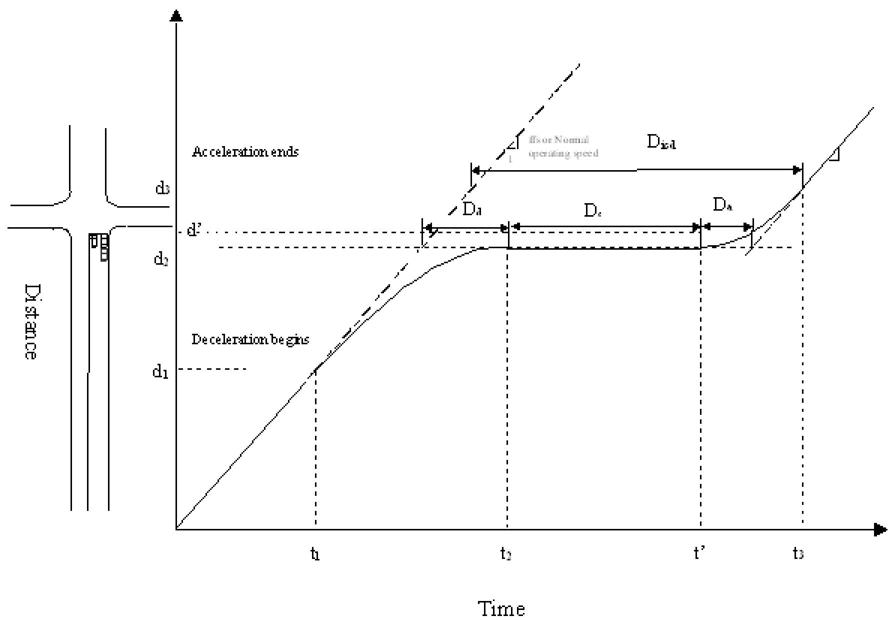


Figure 5.7 Delay Diagram

D_{cntl}: Control delay; D_s: Stopped delay; D_d: Delay incurred while decelerating in approaching the stop light or the end of the queue; D_a: Delay incurred while accelerating to gain full operating speed after signal turns green. As Figure 5.7 indicates, $D_{cntl} = D_d + D_s + D_a$ and the computations for intersection control delay and total delay are:

$$D_{cntl} = (t_3 - t_1) - \frac{d_3 - d_1}{V}$$

$$D_{total} = D_{cntl} + (\frac{d_3 - d_1}{V} - \frac{d_3 - d_1}{ffs})$$

Where V is the normal operating speed of the vehicle before it slows in response to the intersection control and ffs is the free flow speed of the vehicle. Vehicles are delayed by both intersection control and high volume. When traffic volume is light, ffs approaches V. However, at high demand, V can be considerably smaller than ffs.

Mass Transit Ridership

For this example, when time and cost are equal, mass transit is heavily preferred. The mass transit option in the CORSIM problem is bus travel with East and West bound bus lines available for a given bus fare. The SoS level objective is to maximize the number of bus riders for the benefits of public revenue and reduced auto congestion.

Total Revenue

For this example, revenue is defined as public proceeds gained through toll and bus fare. Tolls are charged to each vehicle that enters the freeway at two locations represented as interface nodes between the surface street and freeway subnetworks. The values for the toll and bus fare are part of the network control policy.

Network Reliability

For this example, reliable network performance is defined in terms of level of service (LOS) to the transportation network commuters. LOS guidelines are published by transportation management authorities to indicate the degree to which the traffic situation is satisfactory. According to transportation guidelines, a failing level of service for a network is reached when the mean number of phase failures for the network exceeds 10%. For this network, each simulation analysis period contained 600 green phases across 100 links. Thus, a mean number of failed phases per link that is greater than 6 phases constituted a failing network service level.

5.5 Experimental Design

Knowledge of the importance of certain main effects and interactions was emphasized in (Xu, 2004) as critical to the choice of experimental design. One priority for the design of experiments (DOE) in this study was to efficiently estimate important effects within the interior of the design space. Emphasizing this priority, (Sacks et al, 1989) suggested a good design tended to fill the design space rather than to concentrate on the boundary. Latin hypercube designs are found to be more accurate than random sampling and stratified sampling to estimate the means, variances, and distribution functions of an output (Lian and Liou, 2005). Thus, a fractional factorial design developed in (Xu, 2004) was selected as an appropriate DOE for this research problem. This selected design achieves efficient coverage of points in the interior design space where capturing non-linear interactions is preferred. This design is preferred to alternative techniques which stretch the same number of points to span the extremes in the input space and provide less coverage in the interior. The chosen design is a space filling design, appropriate for the operational constraints on the input space (infeasible policies such as free bus, free toll, and shut-off freeway). Future work can explore the tradeoff of competing DOEs on resulting model performance. The chosen DOE is depicted in Table 5.1.

Table 5.1 Experimental Design for Hybrid System Simulation

Run #	Toll ($)	Mass Transit Fare ($)	Signal 1 (% Fwy Green)	Signal 2 (% Fwy Green)	Network Volume (vph)
		0 = Low	1 = Med	2 = High	
1	0	0	0	0	0
2	0	0	1	1	0
3	0	0	2	2	0
4	0	1	0	1	2
5	0	1	1	2	2
6	0	1	2	0	2
7	0	2	0	2	1
8	0	2	1	0	1
9	0	2	2	1	1
10	1	0	0	1	1
11	1	0	1	2	1
12	1	0	2	0	1
13	1	1	0	2	0
14	1	1	1	0	0
15	1	1	2	1	0
16	1	2	0	0	2
17	1	2	1	1	2
18	1	2	2	2	2
19	2	0	0	2	2
20	2	0	1	0	2
21	2	0	2	1	2
22	2	1	0	0	1
23	2	1	1	1	1
24	2	1	2	2	1
25	2	2	0	1	0
26	2	2	1	2	0
27	2	2	2	0	0

The input space was divided into bins classified Low-Medium-High for each of the five input variables. Stochastic sampling was used to select inputs values for the simulation in accordance with the experimental design. Tolls range from $0 to $12; Mass Transit Fares range from $1 to $15; Signal Timings (% of each cycle length freeway bound flow is green) range from 0% to 100% freeway green. Node volumes range from 50 to 650 vehicles/hour (see Table 5.2). NETSIM and FREESIM entry node volumes are set such that initial subnetwork volumes are equal then adjust iteratively as user choices update.

Table 5.2 Input Variable Ranges

LOW	Toll	Mass Transit Fare	Timing Signal 1	Timing Signal 2	Node Volume
min	0	1	0	0	50
max	4	5	25	25	250
	$	$	(% Fwy Green)	(% Fwy Green)	vehicles/hour
MED	**Toll**	**Mass Transit Fare**	**Timing Signal 1**	**Timing Signal 2**	**Node Volume**
min	4	5	25	25	250
max	8	10	75	75	450
	$	$	(% Fwy Green)	(% Fwy Green)	vehicles/hour
HIGH	**Toll**	**Mass Transit Fare**	**Timing Signal 1**	**Timing Signal 2**	**Node Volume**
min	8	10	75	75	450
max	12	15	100	100	650
	$	$	(% Fwy Green)	(% Fwy Green)	vehicles/hour

The 1/9 factorial design reduced the five variable, three level problem from 243 problem instances to 27 carefully chosen problem instances. This made the problem feasible to simulate. For each of the five factors, a latin hypercube sampling scheme was employed to determine the actual value within the Low-Medium-High bin to assign to each variable. A LHS scheme was generated in Matlab and applied to the input space. The distribution of these LHS sampling values across the 27 runs is shown in Figure 5.8.

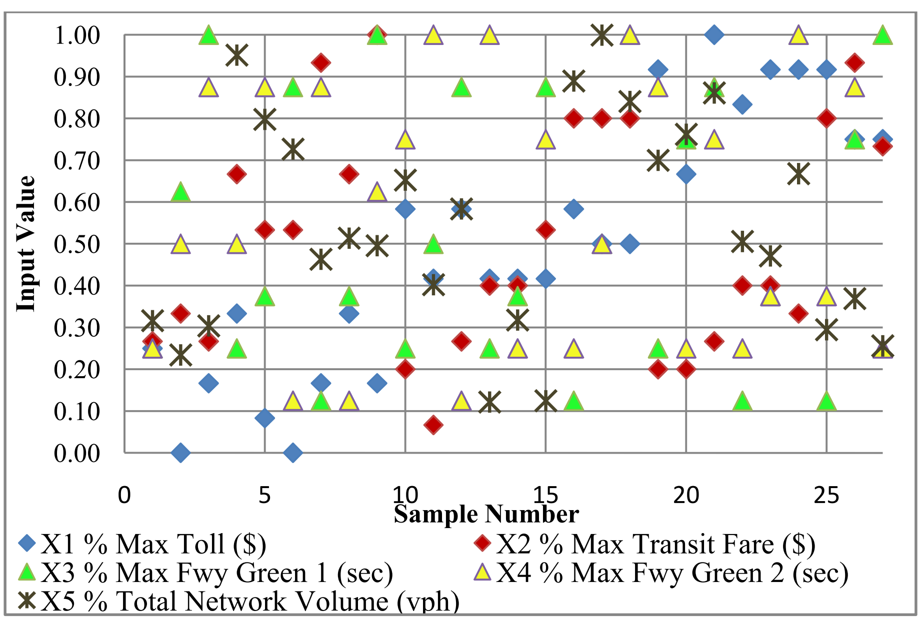


Figure 5.8 Treatment Combinations for the 3^{5-2} Design with Latin Hypercube Sampling

5.6 Example Network Model Simulation

Once the input space was designed and the CORSIM was built to reflect the example operating environment, the 27 high fidelity model simulations were performed. The model simulations were performed on a Dell XPS 1530 with a 2.5 GHz Intel processor. The average CPU run time to simulate a 60 minute analysis period was 20 minutes. Adjustments to the CORSIM inputs and controls were made by editing the user interface. The settings were formatted as record type entries. Multiple runs were performed and aggregated statistics were reported for MOEs.

The stochastic elements (i.e., randomized start points, distribution based values) propagated the uncertainty across the network. For example, if a mean number of vehicles are assigned to turn left at a given intersection, an individual vehicle turn is randomly determined. The randomized sequence ensures the aggregate number of left turning vehicles is equal to the prescribed mean at the end of the analysis period. The variation in vehicle flow caused by the uncertainty leads to varied downstream flow across the network and the collective impacts change the network performance. For this reason, 10 replications for each policy simulation were performed; identical simulations were run 10 times and cumulative metrics reported. The stochastic elements in the CORSIM also varied based on random number seeds (i.e., driver type assignment) and distribution parameters (i.e., Poisson arrivals at entry nodes).

For each problem instance, a consistent procedure was followed. Initial physical system values were specified for network volume and signal timing. Initial user choice values were set for mode and route split. Initial settings were consistently 50% for the bus mode, 25% each for auto (surface street) and auto (freeway). The analysis period for each simulation was consistent at 3600 seconds. The physical limit states were enforced within the simulation to ensure physical laws, traffic theory and domain specific safety codes were not violated. Next, simulation outputs were reported, utility functions evaluated and mode and route volumes assessed to determine whether convergence was reached. If convergence was not achieved, the updated mode and route splits were reflected in the CORSIM settings and the simulation was repeated. Once convergence was reached, the associated travel times were reported as the Equilibrium Travel Times for the given control policy and the relevant MOEs (i.e., delay, mode/route distribution, failures) were exported to an Excel file for analysis.

5.6.1 User Choice Modeling

The systematic incorporation of user decisions is a pivotal feature of the framework and the nexus of the coupled human-physical network analysis. The ability to predict human choice is inexact, but some models have been shown to be helpful in modeling decisions under certain circumstances. Decision-making is at the heart of any user choice model. Decision-making is defined in (Dilts and Pence, 2006) as cognitive choices resulting from a combination of bounded rationality and perspective. Typically, individual decisions incorporate "conceptual lenses" based on the setting, using one or a combination of several decision models. Sheffi (1985) adds the behavioral mechanism underlying traffic models is a choice or decision-making process in which users choose a travel path. The iterative updating of the mode split and route choice in this research requires a decision-making model to be applied.

Logit-based models are commonly used to represent choices between two mutually exclusive options, such as a commuter deciding to drive to work or to use public transit (Wen and Koppelman, 2001). The underlying premise with logit-based discrete choice models is the degree of individual preference for an alternative is proportional to how distinct the user perceives the alternative is superior (Sheffi, 1985). Because discrete alternative decisions were relevant to this research and logit-based models were widely used in the domain, a logit-based user choice model was chosen for this problem. The model computed mode and route preferences based on the perceived user utility associated with each mode and route alternative. The next section presents the utility functions used in the logit-based user choice model.

5.6.2 Utility Functions

The following utility functions represented user preference among mode and route choices.

$$U_{FW} = TVM \cdot Toll + \alpha(TT_{FW}) + C_{FW}$$
$$U_{SS} = \alpha(TT_{SS}) + C_{SS}$$
$$U_{Bus} = TVM \cdot Fare + \alpha(TT_{Bus}) + C_{Bus}$$

where,
C_i = calibration value for the i^{th} modal disutility
α = sensitivity value for % of user decision that is based on Travel Time
TVM = time value of money in dollars per minute

For this operating environment, the expressions can be thought of as (dis)utility functions with negative values indicating a reduction in (dis)utility for the user. It can be assumed that surveys of the usership within the operating environment were conducted to understand inherent mode and route preferences among the usership. The surface street route was used as a reference model; this route was the slowest but free. Relative to the base utility function, coefficients and calibration values were assigned for this example:

$C_{SS} = 0$	reference utility function
$C_{FW} = -1$	represents greater utility from a higher free flow speed
$C_{Bus} = -3$	utility from savings in parking and fuel costs
TVM = 0.25	$0.25 perceived worth for each minute saved
$\alpha = 0.5$	50% of decision based solely on which travel time is less

The resulting expressions were the utility functions used in the logit model for updating mode and route splits during the simulations:

$$U_{FW} = 0.25(Toll) + 0.5(TT_{FW}) - 1$$
$$U_{SS} = 0.5(TT_{SS})$$
$$U_{Bus} = 0.25(Fare) + 0.5(TT_{Bus}) - 3$$

5.6.3 Logit Model

Explicitly modeling user choice in an iterative scheme with the system physics is a centerpiece of the analysis framework. This section formulates logit-based mode and route choice models. Travel time on the network modes and routes are inputs to the user choice model; user's choice of mode and route are inputs to the physical model. This coupling facilitates modeling the interdependent systems in a way that reflects more operational realities and impacts than previous approaches. The general form of the user choice model represents the user demand for an alternative *i*, at a given travel time *t (*as a function of link flow *q*), as expressed below.

$$\frac{q_i^{n+1}}{\sum_{i=1}^{n} q_i^{n+1}} = \frac{e^{-\alpha\, t(q_i)}}{\sum_{i=1}^{n} e^{-\alpha\, t(q_i)}}$$

Substituting the utility functions defined in the previous section into the expression above, modal and route decision formulae were obtained. The following expressions were used to compute the proportion of users who prefer each mode and route.

- Proportion of network users who prefer to pay the policy fare and ride the bus:

$$P(Bus) = \frac{e^{-\mathbf{U}_{bus}}}{e^{-\mathbf{U}_{bus}} + e^{-\mathbf{U}_{auto}}}$$

- Proportion of auto users who prefer to pay the policy toll and drive the freeway route:

$$P(FW \mid auto) = \frac{e^{-\mathbf{U}_{FW}}}{e^{-\mathbf{U}_{FW}} + e^{-\mathbf{U}_{SS}}}$$

- Proportion of auto users who prefer to drive the surface street route at no extra cost:

$$P(SS \mid auto) = \frac{e^{-\mathbf{U}_{SS}}}{e^{-\mathbf{U}_{SS}} + e^{-\mathbf{U}_{FW}}}$$

5.7 Network Equilibrium Results

For each of the 27 policy simulations, the steps of the framework were followed until the convergence criteria was met. The number of iterations needed to reach convergence varied with each simulation. The primary factor in how quickly the simulation reached equilibrium was the value of α in the utility functions. The sensitivity factor, α, represents the proportion of a user's mode and route decisions based on travel time. Low values for α resulted in the utility functions incrementing slowly; a $\alpha = 1$ resulted in a solution for mode and route loading after only one evaluation. For the sake of illustration, the value of $\alpha = 0.5$ was used in the utility functions in this problem. Another key factor in the number of iterations required to converge was the values of the policy variables. Policies with balanced signal timing and comparable toll and fare prices tended to distribute the usership most quickly (typically converging in less than 4 iterations). Policies with disparate toll and fare prices and unbalanced signal timings tended to take much longer (as many as 10 iterations) to converge to an equilibrium load distribution of users across the network.

The equilibrium results for the 27 policy simulations are listed in Table 5.3. The input and output variables for each simulated policy are defined below:

Inputs

X_1- % of the \$12 Maximum Toll

X_2- % of the \$15 Maximum Bus Fare

X_3- % of the 120 sec cycle that Signal 1 is Green

X_4- % of the 120 sec cycle that Signal 2 is Green

X_5- 0.001 x Total Network Vehicles per Hour

Outputs

Y_1- % of Total Users Traveling by Bus

Y_2- % of Total Users Traveling by Freeway

Y_3- % of Total Users Traveling by Surface Street

Y_4- 0.01 x Mean Vehicle Delay in seconds

Y_5- Mean # of Phase Failures per Link

Table 5.3 Equilibrium Results for High-Fidelity Simulation

	X1 % Max Toll ($)	X2 % Max Transit Fare ($)	X3 % Max Fwy Green 1 (sec)	X4 % Max Fwy Green 2 (sec)	X5 Total Network Volume (vph) x 1000	Y1 P(Bus)	Y2 P(FW)	Y3 P(SS)	Y4 Control Delay (sec/veh) x 100	Y5 Mean Phase Failures/Link
1	0.25	0.27	0.25	0.25	4.91	0.33	0.56	0.11	0.23	0.25
2	0.00	0.33	0.63	0.50	3.63	0.15	0.77	0.08	0.22	0.12
3	0.17	0.27	1.00	0.88	4.72	0.13	0.83	0.04	1.28	3.93
4	0.33	0.67	0.25	0.50	14.77	0.48	0.28	0.24	1.17	9.94
5	0.08	0.53	0.38	0.88	12.39	0.15	0.78	0.07	1.26	8.80
6	0.00	0.53	0.88	0.13	11.28	0.17	0.76	0.08	1.12	8.58
7	0.17	0.93	0.13	0.88	7.19	0.04	0.88	0.08	0.92	3.65
8	0.33	0.67	0.38	0.13	7.98	0.16	0.72	0.13	0.58	2.59
9	0.17	1.00	1.00	0.63	7.70	0.18	0.50	0.32	1.17	5.38
10	0.58	0.20	0.25	0.75	10.13	0.85	0.09	0.05	0.55	2.27
11	0.42	0.07	0.50	1.00	6.25	0.21	0.71	0.08	1.25	3.59
12	0.58	0.27	0.88	0.13	9.06	0.63	0.28	0.09	0.92	5.38
13	0.42	0.40	0.25	1.00	1.89	0.14	0.76	0.10	0.69	1.00
14	0.42	0.40	0.38	0.25	4.94	0.34	0.50	0.16	0.24	0.28
15	0.42	0.53	0.88	0.75	1.94	0.20	0.60	0.19	0.24	0.15
16	0.58	0.80	0.13	0.25	13.81	0.27	0.31	0.42	1.27	10.68
17	0.50	0.80	0.50	0.50	15.52	0.44	0.13	0.43	1.22	10.79
18	0.50	0.80	1.00	1.00	13.05	0.01	0.21	0.78	1.86	13.04
19	0.92	0.20	0.25	0.88	10.87	0.98	0.00	0.02	0.96	5.84
20	0.67	0.20	0.75	0.25	11.83	0.89	0.05	0.07	0.69	4.72
21	1.00	0.27	0.88	0.75	13.36	0.98	0.00	0.02	1.22	9.19
22	0.83	0.40	0.13	0.25	7.85	0.67	0.13	0.20	0.79	3.48
23	0.92	0.40	0.38	0.38	7.32	0.76	0.10	0.14	0.32	0.85
24	0.92	0.33	1.00	1.00	10.37	0.75	0.10	0.15	1.98	11.35
25	0.92	0.80	0.13	0.38	4.572	0.25	0.37	0.38	0.431	0.97
26	0.75	0.93	0.75	0.88	5.731	0.17	0.61	0.22	0.652	1.49
27	0.75	0.73	1.00	0.25	3.972	0.12	0.54	0.33	0.968	2.50

5.8 Conclusion

This objective produced a decision support framework for evaluating control strategies for human-physical systems in which network flows and human decisions are coupled. An integrated traffic system software toolbox was used to customize a multi-modal, multi-route network with multiple decision-makers. A stochastic agent simulation represented the network physics; discrete choice models represented user mode and route preferences. An iterative analysis produced network equilibrium outputs.

Several general insights for SoS problems were made. First, the interdependent nature of user choice and system physics can be investigated through an iterative experimental design. Second, an operating environment for a SoS can be modeled and control strategies investigated through stochastic agent simulation. Such microscopic perspectives of SoS behavior offer opportunities for bottom-up performance analysis that may differ from aggregated macroscopic models. Third, it is reasonable to expect a SoS analysis to explicitly consider the coupled relationship between the physical state of the network and the user choices that impact and result from various states of the operational environment. Finally, the data used and results generated in this objective are illustrative and not tied to a specific number problem. However, the process and findings are generalizable to investigate and determine control policies that optimize network objectives for a SoS. No single policy will suffice for all situations, but ideally, a control scheme should be specially tailored to the network it serves to include the inherent user preferences within the defined operational environment.

CHAPTER VI

SURROGATE MODELING, MODEL VALIDATION AND SENSITIVITY ANALYSIS FOR STOCHASTIC HUMAN-PHYSICAL NETWORKS WITH HYBRID DYNAMICS

6.1 Introduction

For this objective, computationally inexpensive surrogate models are developed to faithfully predict more detailed simulation outputs. The computational expense of detailed simulations, such as the transportation network previously described, limit the ability to exhaustively study human-physical networks. Less expensive approximation models, often called surrogate models or response surface approximations, can represent network performance and reduce the computational expense of repeating the more detailed simulation over and over for every possible combination of policy settings. Several methods are available for developing response surface approximations, including the development of regression models with reduced degrees of freedom (Friedman, 1991) and (Schumaker, 2007), and Gaussian process interpolation (Bichon et al, 2008); (Jones et al, 1998); (Kennedy and O'Hagan, 2001); (Bayarri et al, 2002); (Simpson et al, 2001); (Kaymaz, 2005); (Kennedy et al, 2006); (Oakley and O'Hagan, 2002) and (McFarland et al, 2008). Using the transportation network example, surrogate models are developed and compared for use in this study. Gaussian process models and Quadratic Response Surface models are developed to handle the continuous variables (i.e., tolls and fares in dollars, signal timing in minutes). In order to determine which models are the most appropriate for predicting the responses, a quantitative model evaluation method called a Predicted Residual Sum of Squares (PRESS) test is performed. (Allen, 1971) introduced PRESS and it is a commonly used technique to compare candidate models. Results from the PRESS test indicate the most appropriate model for the four continuous output variables is a Quadratic Response Surface model. To evaluate the predictive strength of the developed models, each model is compared to high-fidelity simulation outputs and statistical tests are performed on the fitted models.

A Binary Logistic Regression model is developed to address the discrete variable with categorical responses (i.e., pass/fail network state). This model is used to support probabilistic reachability analysis to assess the likelihood of the network reaching a failed network state. Reachability is an important topic in classical control theory (Abate et al, 2008) aimed at determining the probability of reaching a given system state from a given set of initial conditions and subject to a given control. Finally, sensitivity analysis is performed to demonstrate the effects of varying control values on output metrics of interest and the SoS objectives.

6.2 Surrogate Modeling

Computational expense precludes evaluating high fidelity simulations for the many possible policy combinations. For the purpose of analysis, reduced models produced from response surface methods can be used to generate interpolating predictions. Gaussian Process and Quadratic Response Surface models were developed as surrogates for the four continuous responses. This problem also contained a discrete output variable representing the binary response of whether the network reached a failing level of service. A Binary Logistic Regression Model was developed as the surrogate for the discrete response.

6.2.1 Gaussian Process and Quadratic Response Surface Models

Response surface models were developed to serve as surrogates for the physical disciplinary model. Gaussian Process (GP) models and a Quadratic Response Surface (QRS) model were developed and compared as potential surrogates for predicting the responses for the four continuous output variables for the continuous output variables (pBus, pFW, pSS and Delay). Both methods have their advantages. GP models, as noted in (Kennedy and O'Hagan, 2001), can account for its own uncertainty and quantify the model error associated with the surrogate prediction. QRS models are popular because they are usually low-order polynomial alternatives to computationally expensive simulation codes (Lian and Liou, 2005).

Constant trend, linear trend and quadratic trend GP Models were developed and compared with a Quadratic Response Surface Model. For the purposes of demonstration, model predictions for the output variable, Delay, are presented. Five GP model predictions were computed. Four predictions were interpolations and one prediction was a training point for the GP model. Figure 6.1 illustrates the GP model predictions for delay at each point. The vertical lines in Figure 6.1 depict the variance value in the GP model prediction. As expected, a perfect prediction was made for the training point. This is illustrated in Figure 6.1 by the blue circle exactly covering the observed value (a blue star) and the blue vertical whisker of length zero (indicating the prediction variance is zero). For the other four points, the observed and predicted values and prediction variance is shown. As Figure 6.1 shows, predictions near the design space boundary were greater than one standard deviation off; interior point predictions were within one standard deviation.

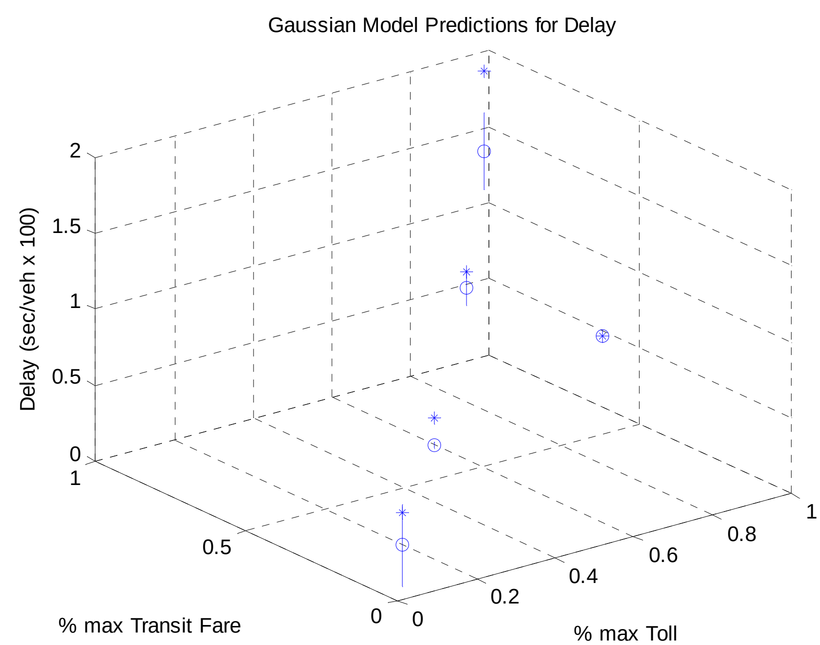


Figure 6.1 GP Model Prediction

A QRS model was also developed, using the same five points. Estimates from the QRS model were generated and compared to candidate GP models (GP_0, GP_1, GP_2), representing underlying trend functions (constant, linear and quadratic respectively). QRS predictions for delay in Figure 6.2 (shown as red star) were generally closer to the observed delay values than the GP predictions.

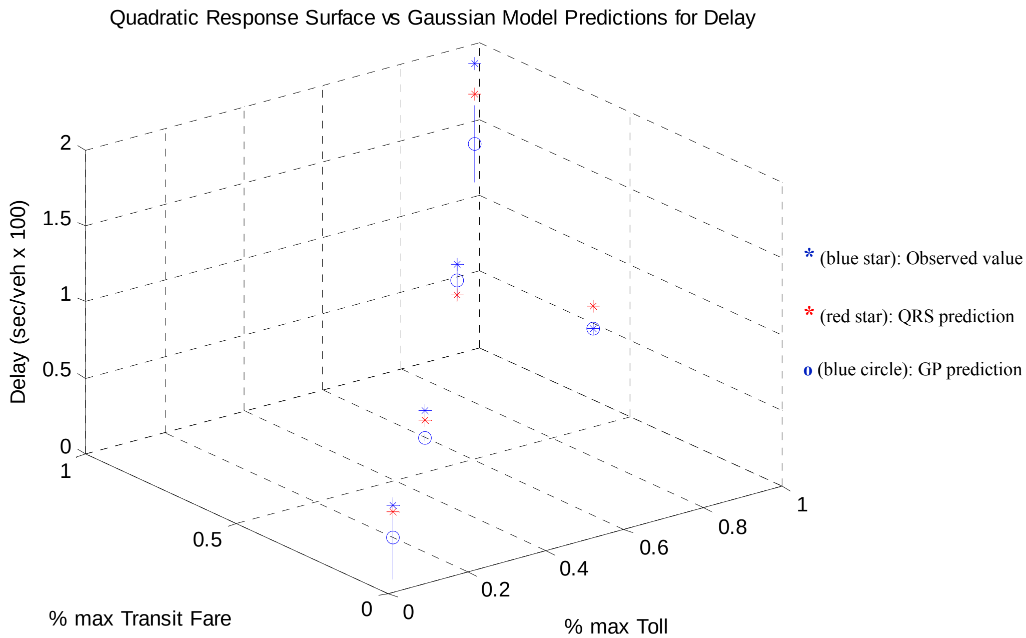


Figure 6.2 GP Model and Quadratic Response Surface Predictions

A second test was performed to compare delay predictions on the same operational policy for additional network demands ranging from 4,000 to 14,000 vehicles/hour (Note: in this research, minimum demand is 1,000 vph and maximum demand is 15,000 vph). Numerical results from this test are shown below.

Inputs: X_1= % Max Toll; X_2= % Max Fare; X_3= % Max Fwy Grn1; X_4= % Max Fwy Grn2; X_5= Demand (vph) x 1000

X1	X2	X3	X4	X5
.83	.40	.13	.25	7.85
.05	.05	.10	.10	4.0
.95	.95	.95	.05	14.0
.40	.40	.40	.40	8.0
.75	.75	.75	.75	10.0

The output predictions for the GP1 model and associated variance were

	X1	X2	X3	X4	X5
Output:	.79	0.50	1.95	.55	0.95
GP Var:	0.00	0.054	0.052	0.004	0.023

The output predictions for the QRS model were

	X1	X2	X3	X4	X5
Output:	.942	.45	1.75	.49	0.75

The QRS model and GP1 (linear trend function) produced similar results. For comparison, the absolute differences in Delay predictions (seconds per vehicle) were calculated and reported

	X1	X2	X3	X4	X5
Absolute Difference:	15.2	5.0	20.2	6.0	20.0

Differences in network delay estimate more than 20 seconds per vehicle could be significant. Therefore, it was necessary to quantitatively compare the models in order to determine which of the four candidate models (GP_0, GP_1, GP_2, QRS) best approximates each response variable. A quantitative model evaluation method called a Predicted Residual Sum of Squares (PRESS) test was performed and described in the next section.

Predicted Residual Sum of Squares, or PRESS (Allen, 1971) is an effective technique to evaluate candidate models. The procedure for performing PRESS on a sample of size n is as follows:

1. Individually omit each i^{th} observation; recalculate the fitted model for remaining *n-1* data
2. Calculate the prediction error for the i^{th} observation and square the difference
3. Repeat the process for all *n* observations and compute the sum of squares
4. Compare sum of squares value to other candidate models, with lowest value preferred

A PRESS test was performed for each of the candidate surrogate model predictions for pBus, pFW, pSS and Delay. Three GP models (constant, linear and quadratic) and the QRS model were compared for predictive accuracy. Following the procedure described above, each model was compared to the output values of the 27 high fidelity simulations and the sum of squares was reported.

Table 6.1 depicts the PRESS results. The Quadratic Response Surface model ranked the best for the each model with the Gaussian Process with linear trend function next with the exception of the Surface Street Model for which the Gaussian Process model with constant trend ranked second to the QRS Model.

Table 6.1 PRESS Test Results

	RESPONSE VARIABLE PRESS VALUE			
MODEL (trend)	**Bus**	**Freeway**	**Surface**	**Delay**
GP (quadratic)	2.641	2.025	1.212	7.204
GP (constant)	2.608	1.804	0.882	5.055
GP (linear)	1.968	1.484	1.03	3.509
Quadratic Resp Surf	0.676	0.471	0.442	2.866

Based on the PRESS test results, the most appropriate surrogate model for the each of the continuous response variables is the quadratic response surface model. Thus, the QRS models are the most appropriate low-fidelity models for estimating pBus, pFW, pSS and Delay for policy analysis and optimization in the next objective.

6.2.2 Probabilistic Reachability

Reachability is an important topic in classical control theory (Abate et al, 2008). Reachability analysis in discrete, continuous or hybrid systems seeks to partition states into two categories: those that are reachable from the initial conditions, and those that are not (Mitchell et al, 2001). The concept of probabilistic reachability centers around determining the probability of reaching a given system state from a given set of initial conditions and subject to a given control. For deterministic problems, reachability is a yes/no problem evaluating whether starting from a given set of initial states the system will reach a certain set or not. In stochastic problems, the different trajectories originating from each initial state have likelihoods of reaching the set (Abate, 2007).

System evolution in stochastic human-physical networks is influenced by control policy, so a SoS priority is to choose appropriate controls to minimize the probability that the state of the system will enter the failed state.

Support for reachability analysis in optimally controlling deterministic problems has been pointed out in (Hedlund and Rantzer, 2002) and (Lygeros, 2004). Connections between reachability, and safety for deterministic hybrid systems (mostly applied to air traffic management) has been stressed in (Mitchell et al., 2005) and (Lygeros et al, 1999), and (Tomlin et al, 1998). Reachability for stochastic hybrid systems, such as the class of problem presented in this research, is a recent focus of research. Bujorianu and Lygeros (2003) address theoretical issues regarding the measurability of the reachability events. However, even the most recent approaches consider the problem of reachability analysis for continuous time stochastic hybrid systems without any control input.

This research extends the current approaches by including stochastic hybrid systems with controls and addressing the complex dynamical evolution in a computationally efficient way. The stochastic processes are modeled with a high fidelity simulation and the physical system evolves in response to network controls. Additionally, uncertain initial conditions and model parameters contribute to the system evolution which may or may not reach a failed system state. Furthermore, analysis is performed under uncertainty to optimize weighted combinations of policy objectives, so the impacts of stochastic demand and varying objective weights are considered in assessing reachability.

The reachability computation in (Tomlin, 1998) and (Tomlin et al, 2000) follows an iterative, two stage algorithm in which an outer iteration computes reachability over the discrete switches and an inner iteration runs a separate continuous reachability problem in each of the discrete modes to compute the estimates. The approach to network reliability used for this problem is an adaptation to the procedure described in (Tomlin, 1998) and (Tomlin et al, 2000). The reachability computation is a function of two factors: the stochastic nature of the simulation and the impact of a prescribed network policy on the physical system. Uncertainty is propagated throughout the TSIS simulation. For example, agent behavior (i.e., turn movement) is initiated based on random seeds, agent arrivals follow a Poisson process and agent attributes (i.e., driver type assignment) follow a probabilistic distribution. Depending on the stochastic system evolution, successive simulations for the same network policy could result in different network states. Likewise, control policy significantly influences the system physics which, in turn, influences user mode and route choice and the evolution of the physical system. Given the computational expense to perform many high-fidelity simulations, an appropriate and cheaper model was developed to predict the probability of the network reaching a failed state.

6.2.3 Binary Logistic Regression Model

Over the last decade, logistic regression has become a standard method of analyzing model relationships with discrete responses (Hosmer and Lemeshow, 2000). It is appropriate for data in which there is a binary (success/failure) response variable, such as the discrete response variable in this study problem, network failure. Unlike linear regression, where one estimates the relationship between predictor variables and an outcome variable, logistic regression estimates the conditional probability that a dichotomous outcome occurs. (Hilbe, 2009). The general form for the logistic model is logit(π) = log(odds) = $\alpha + \beta X$, where

$$\pi = P(Y = 1 \mid X) \text{ and is given by } \frac{e^{\alpha+\beta X}}{1+e^{\alpha+\beta X}}$$

The first step to generate the binary logistic regression model was to classify the simulations based on whether the network reached a failed state. Network failure was defined using transportation guidelines for levels of service (LOS). If the mean number of phase failures per link on the network exceeded 10%, the network is considered to have reached the failed state.

The next step was to estimate (infer) parameter values for the logistic regression model using the Minitab toolbox.

As previously discussed the simulation inputs were values for toll, fare, signal timing and demand. Since each stochastic simulation had 10 replications, binary logistic regression was performed on the 270 high fidelity simulation scenarios to determine the maximum likelihood coefficients for the model. The regression model coefficients with standard errors and the Z and p-values for each predictor are below.

Predictor	Coefficient	Std Error	Z	P
Constant	-0.1458	0.603	-0.24	0.809
% Max Toll ($)	-0.8422	0.479	-1.76	0.078
% Max Transit Fare ($)	-0.0420	0.554	-0.08	0.940
% Max Fwy Green (sec)	3.3640	0.637	5.28	0.000
Network Volume (vph x 1000)	-0.6285	0.036	-1.73	0.084

Using a 0.1 significance level, the p-values for the regression coefficients showed three of the four model coefficients to be significant at the 0.10 level (% Max Toll; % Max Fwy Grn; Network Volume). Depending on preference, the non-significant factors can be dropped and the simpler model used for prediction. For the purposes of illustration in this dissertation, the complete model is used. The selected predictive model for π, the probability the network reaches a failed state under a given policy is

$$\pi = \frac{e^{-0.84X_1 - 0.42X_2 + 3.36X_{3,4} - 0.63X_5 - 0.146}}{1 + e^{-0.84X_1 - 0.42X_2 + 3.36X_{3,4} - 0.63X_5 - 0.146}}$$

The listed predictors are X_1= % Max Toll; X_2= % Max Fare; $X_{3,4}$= % Max Fwy Grn (combined); X_5= Demand (vph) x 1000 and ε = an intercept term.

6.3 Model Validation

The strategy of model verification and model validation according to (AIAA, 1998) is the assessment of error and uncertainty in a computational simulation. Yet, depending on the domain, there are philosophical debates over their definitions. A common convention is for verification to focus on the mathematical accuracy of an implemented procedure in a computer simulation, while validation determines the degree to which a computer model represents the real world from the perspective of the intended model applications (AIAA, 1998) and (DOE, 2000).

In this research, the high fidelity TSIS model is approximated by candidate, low fidelity models. The data evaluated is synthetic; no field experiments are performed. So, the high fidelity model performance is the closest representation of the true physical environment and validation is meant in a broad sense. One set of data was used for building and evaluating candidate surrogate models. A separate set of data was evaluated by the high fidelity simulation and output values used to compare with the surrogate predictions. Quantitative assessments such as goodness-of-fit and hypothesis tests are used for prediction testing. Qualitative assessments such as graphical analysis are also used to confirm agreement between surrogate predictions and the high fidelity simulation evaluations. For this reason, a general interpretation of validation is adopted which considers the described assessments to be satisfactory for the intended application.

6.3.1 Binary Response Model

Goodness-of-Fit Test

Statistical software has increased the popularity of goodness-of-fit testing for predictive models; however the methods used for models with binary outcomes face some challenges. Limitations in assessing fitted models for binary responses involve the choice of cutting points, size of subgroups and disparate covariate values, and each is an area of current research (Hosmer et al, 1997). For this reason, subjective assessment should accompany quantitative measures (i.e., p-values) in assessing goodness-of-fit for this type of model. Despite the limitations, such methods provide an analytical basis for evaluating a binary response model. Pearson Chi-square and Hosmer-Lemeshow (H-L) methods are currently leading methods in literature and used in this study to assess the fitted model for pFail. These tests are widely used goodness-of-fit test and only available for binary responses, such as pFail, with only pass/fail responses. To evaluate the fit of the pFail model, Pearson Chi-square and Hosmer-Lemeshow test statistics are evaluated. The hypothesis test, as shown in (Hosmer and Lemeshow, 1989), evaluates a goodness-of-fit test statistic which is distributed as chi-square and tests for evidence of a lack of fit. Both test results are reported, but for the purposes of illustration, only the H-L procedure is formulated.

H_0: The model is a good fit $\quad\quad$ H_A: The model is NOT a good fit

Test Statistic: $\hat{C} = \sum_{k=1}^{g} \frac{(O_k - E_k)^2}{E_k(1 - E_k/n_k)} \sim X_{g-2}^2$

where, n_k = number of observation in the k^{th} group
O_k = observed number of cases in the k^{th} group
E_k = expected number of cases in the k^{th} group

In the H-L procedure, the observations are sorted in increasing order of their estimated event probability. The observations are then divided into groups and the Hosmer-Lemeshow goodness-of-fit statistic is obtained by calculating the Pearson chi-square statistic based on the observed and expected frequencies. To illustrate, Figure 6.3 depicts the observed and expected frequencies from the H-L test applied to the simulation results used in the binary logistic regression procedure to generate the selected predictive model for π previously described. Hosmer and Lemeshow (1989) suggests comparison of observed to expected frequencies within each group can be useful to indicate regions where the model may or may not perform satisfactorily.

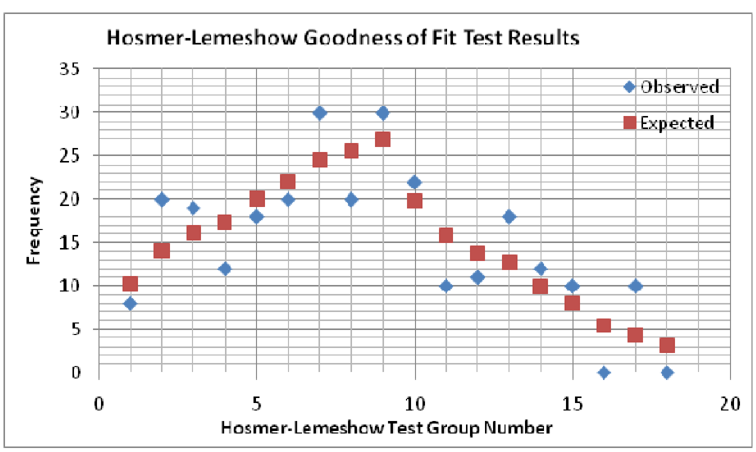


Figure 6.3 Goodness-of-Fit Test

To evaluate the fit of the reduced model, a quantitative comparison of the prediction results and simulation results was performed. As a technique to validate the fitted model, the H-L procedure was applied to simulation results from seven high fidelity simulations obtained for the purposes of comparing the high and low fidelity models. The H-L and Pearson method results are reported below.

Goodness-of-Fit Test Results

Method	**χ^2 test statistic**	**DF**	**P-value**
Pearson	3.057	2	0.210
Hosmer-Lemeshow	3.057	5	0.691

Goodness-of-fit tests are an exception to the widely held paradigm that low p-values are desirable. The alternative hypothesis states the fit is NOT good. So, at the 0.1 significance level, p-values $\geq$ 0.1 increasingly provide less and less evidence against a good model fit.

Both the Pearson and Hosmer-Lemeshow methods suggest the reduced model is a good fit for predicting pFail, particularly the H-L method. Given the limitations of the validation methods for binary response models, the consistent p-values from two of the leading methods is reason to conclude the developed model is an appropriate surrogate for π, the probability of the system evolution reaching a failed level of service for a given policy.

Hypothesis Test about the Mean Failure Probability

In addition to confirming how well the reduced model fit the high fidelity model, a hypothesis test was also performed to determine whether a hypothesized value for the true pFail, equal to the reduced model pFail, could be shown to be statistically different from a pFail generated from a set of high fidelity simulations. To investigate the predictive strength of the pFail surrogate model, a set of validation data was obtained from the high fidelity model from which comparisons were based. The accuracy of the estimate depends on the number of simulations (Haldar and Mahadevan, 2000), so 100 network simulations were evaluated for a base control policy with % Max Toll, % Max Fare and % Max Fwy Grn all equal to 50%. The stochastic agent simulation with demand fixed at the mean value of 7,000 was compared to the reliability model for the same base policy and network demand values. Output values for mean phase failures per link from 100 high fidelity replications are listed in Table 6.2. Recall, the transportation level of service threshold for classifying the network as "failed" was a mean greater than 10%. In this case, network runs with means ≥ 6 were classified as failed. The resulting point estimate for pfail was 0.04 with a variance of 0.00038.

Table 6.2 High Fidelity Results for Mean Phase Failures

Run	Mean Fails						
1	4.45	26	3.65	51	3.25	76	3.85
2	3.45	27	4.3	52	4.5	77	3.55
3	3.8	28	4.1	53	3.7	78	3.55
4	4.9	29	3.7	54	5.9	79	4.15
5	4.25	30	2.2	55	5.6	80	8.25
6	4.55	31	3.55	56	7.95	81	5.05
7	4.5	32	4.35	57	3.3	82	4.25
8	4.7	33	3.4	58	4.5	83	5.9
9	5.65	34	3.7	59	4.9	84	2.8
10	4.55	35	5.1	60	3.7	85	4.4
11	9.1	36	3.8	61	4.8	86	3.95
12	4.3	37	3.3	62	3.3	87	2.65
13	4.95	38	5.35	63	3.35	88	4.65
14	5.4	39	3.05	64	4.95	89	3.7
15	4.55	40	3.15	65	4.1	90	3.9
16	5.05	41	2.1	66	6.85	91	4.9
17	3.6	42	4.6	67	4.1	92	4.75
18	3.85	43	5.7	68	4	93	3.4
19	3.45	44	4.95	69	3.75	94	2.8
20	4.6	45	2.8	70	3.7	95	3.3
21	4.45	46	5.2	71	3.95	96	3.5
22	5.15	47	3.6	72	5.35	97	3.55
23	2.75	48	4.6	73	2.95	98	4.95
24	4.5	49	3.5	74	4.3	99	3.55
25	4.65	50	4.05	75	3.35	100	4.1

By comparison, the pFail from the surrogate model evaluated at the base policy and demand of 7,000 was equal to 0.029. This result implies the probability of the network reaching a failed LOS, under this policy, is 0.029.

A two-tail hypothesis test of the mean pFail was performed, using the reliability model value as the assumed mean pFail. The test results indicated, at the 0.05 significance level, the true mean pFail was not statistically different from 0.029, based on the high fidelity sample with

$n = 100$, $\hat{p} = 0.04$ and $\sigma^2 = 0.00038$. The hypothesis test to determine whether the true pFail is equal to 0.029 is shown below.

H_0: $p = 0.029$ H_A: $p \neq 0.029$ $\alpha = 0.05$

Test Statistic: $t = \frac{\hat{p}-p}{\sqrt{\frac{(\hat{p})(1-\hat{p})}{n}}} = 0.561$

Critical value: $t_{\alpha/2,\, n-1} = 1.98$
p-value: 0.78

Conclusion: At the 0.05 significance level, the true pFail can be assumed to be equal to 0.029.

Additionally, computations were performed to determine the minimum sample size required to invalidate the hypothesis test at the .05 significance level. When $n = 1{,}250$, the value for the test statistic exceeds the critical value and the conclusion is to reject the null and infer the true pFail is not equal to 0.029. Based on the computational expense required to perform 1,250 high fidelity simulations, decision makers should evaluate the tradeoff between accuracy and expense. For a 95% level of confidence, it is recommended that decision-makers use the reduced model when it is infeasible to perform more than 1,250 simulation runs. If enough resources are available to perform more than 1,250 runs then the high fidelity simulation results are considered more appropriate.

6.3.2 Continuous Response Models

This section describes the validation of the surrogate models for the four continuous input variables, whose models were developed, compared and selected section 6.2. Output values were computed as expected performance and compared with observed performance values from the high fidelity model. As previously discussed, a priority for the experimental design was to focus on the non-linear interactions in the interior of the design space more than at the boundaries. This is because the expected operational conditions for this problem were assumed non-extreme. More moderate operational conditions were typical; thus, more simulations in these areas would improve the efficacy of corresponding surrogates for their intended applications. An appropriate set of inputs was generated as a validation data set. Table 6.3 contains these inputs and the observed and expected outputs.

Table 6.3 High Fidelity Results for Validation Trials

		Policy			pBus		pSS		pFW		Delay	
	Volume	Toll	Fare	FwyGrn	Expected	Observed	Expected	Observed	Expected	Observed	Expected	Observed
1	2000	1	0.8	0.2	0.112	0.124	0.537	0.491	0.350	0.385	0.432	0.485
2	5000	0.75	0.4	0.7	0.397	0.415	0.177	0.189	0.427	0.396	1.053	0.998
3	6000	0.25	0.6	0.6	0.161	0.150	0.071	0.101	0.768	0.749	0.600	0.541
4	7000	0.75	0.2	0.8	0.526	0.531	0.146	0.160	0.328	0.309	1.544	1.397
5	8000	1	0.4	0.5	0.773	0.777	0.147	0.154	0.080	0.069	0.848	1.001
6	9000	0.5	0.2	0.6	0.523	0.531	0.106	0.121	0.370	0.348	0.907	0.879
7	10000	0.75	0.5	0.8	0.453	0.439	0.245	0.229	0.302	0.332	1.240	1.300

Graphical Analysis

The first assessment was a visual inspection of the plotted results. Figure 6.4 depicts the observed and expected results for the network mode and routes. The strong evidence in the model agreement was expected given the network outputs were equilibrium values prior to building the surrogate to which it was compared.

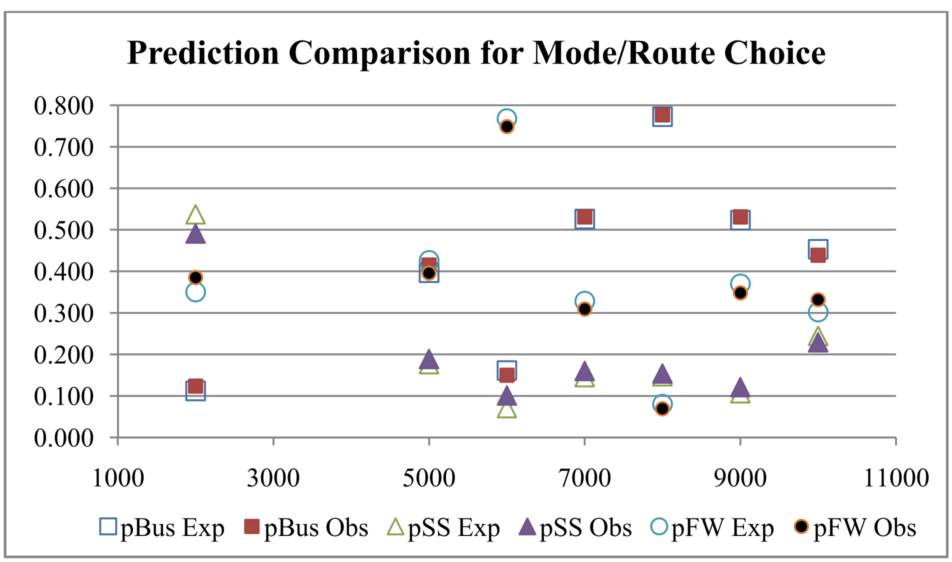


Figure 6.4 Mode and Route Predictions

By contrast, the observed delay results in Figure 6.5 do not appear to match as closely. The delay outputs matched well (R^2=0.9706), but not quite as well as mode and route outputs (R^2> 0.9810). One reason for this relative difference is the how the output values are obtained in the high fidelity simulation. The mode and route proportions are the results of iterative convergence, but the delay is simply the corresponding metric value for delay, at convergence.

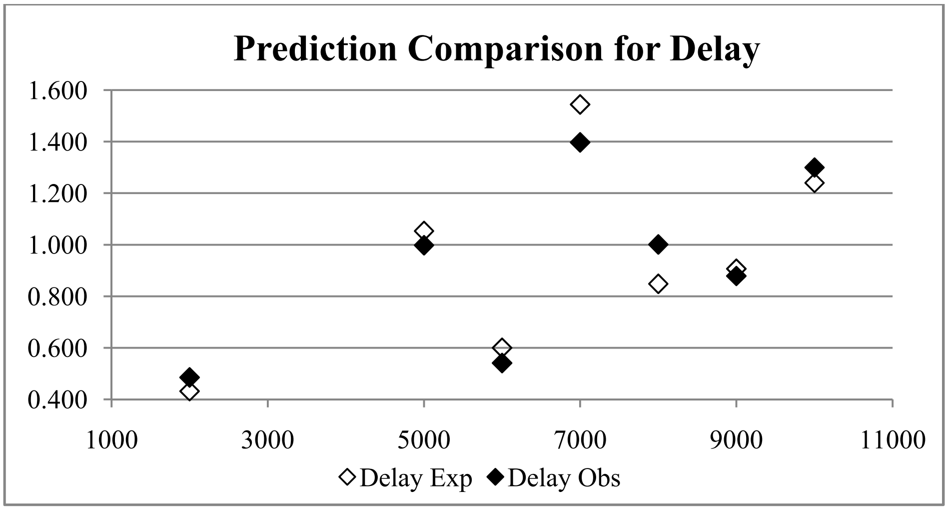


Figure 6.5 Delay Predictions

The stochastic nature of the simulation impacted the delay metric value in the seven validation runs. The equilibration requirement for the mode and route values caused the validation runs to tend more closely to their training run values. By extension, the surrogate model results for mode and route compared well with the validation run results.

Goodness-of-Fit Test

In addition to graphical inspection, residual analysis comparing the sum of squares for the error with the total variability as shown in (Haldar and Mahadevan, 2000) was performed as a prediction test. Additionally, a χ^2 goodness-of-fit test was performed to evaluate the fitted model. Table 6.4 lists the R^2 and p-values for each model.

Table 6.4 Prediction Test Results for Continuous Responses

pBus		pSS		pFW		Delay	
$\chi 2$ p-value	R^2	$\chi 2$ p-value	R^2	$\chi 2$ p-value	R^2	$\chi 2$ p-value	R^2
0.9999	0.9990	0.9998	0.9810	0.9999	0.9950	0.9979	0.9706

These results support a good model fit with the outputs from the validation data set. It is important to note that the R^2 metric in Table 6.4 is specific to the validation data set. As a coefficient of determination, it represents the proportion of variability in the high fidelity outputs that is explained by the surrogate model. Because the seven high fidelity run results do not span the input space as thoroughly as the 27 simulations (upon which the surrogate models were based), the predictive strength is not assumed to be representative of the entire input space. This metric is an indicator of the predictive strength across the space to the extent that other regions in the design space perform similarly to the validation set. Depending on available computational resources, tests should be performed to verify the model fit in other areas of the non-linear space.

Of note, is the difference in the statistical test results between the 27 point training data set and the 7 point validation data set. The relatively lower performance for the broader space as compared to the localized interior space suggests the fit may be worse near the boundaries of the design space. This is expected because the LHS design was intentionally chosen to capture more interior points. The 27 run results are more representative of the interior and the edges of the design space. The relatively lower R^2 values for the 27 point set (see Table 6.5), while definitely suggesting strong fit, also illuminate the vulnerability of model predictions for boundary values using the surrogate quadratic response surface models.

6.4 Final Reduced Models

The result of the stochastic simulation and reduced model building is a set of the most appropriate surrogate models to use in reliability analysis and policy optimization in the next objective. The reduced models determined to be the most appropriate for response predictions are presented below, along with the R^2 value for each model.

Inputs Variables: (control policy and network demand)

X_1= % Max Toll; X_2= % Max Fare; $X_{3,4}$= % Max Fwy Grn (total); X_5= Demand (vph) x 1000

Output Variables: (for a given control policy)

pBus: Proportion of users to choose the bus mode

pSS: Proportion of user to choose the auto mode and surface street route

pFW: Proportion of user to choose the auto mode and freeway route

Delay: Average seconds vehicles drive longer than free flow speed to travel each mile x 0.01

Network Failure: Probability the network reaches a failed LOS state

The model form is a linear combination of the variables with coefficients and intercepts term, ε, listed in Table 6.5. The R^2 values for the four continuous models represent the proportion of the variability in the stochastic agent simulation that is explained by the reduced model.

Table 6.5 Final Surrogate Models

Response Model Coefficients

Input variable	pBus (R^2=.94)	pSS (R^2=.96)	pFW (R^2=.92)	Delay (R^2=.88)	pFail
ε	0.165	0.664	0.370	-0.465	-0.146
X_1	0.296	0.561	-0.953	1.459	-0.840
X_2	-0.237	-0.627	0.697	-0.174	-0.420
$X_{3,4}$	-0.222	-0.900	0.991	0.061	3.360
X_5	0.047	-0.107	0.028	0.084	-0.630
X_1^2	0.332	0.357	0.058	0.147	---
X_2^2	0.015	0.240	0.002	0.653	---
$X_{3,4}^2$	-0.060	0.448	-0.383	1.617	---
X_5^2	0.001	0.005	-0.003	-0.004	---
X_1X_2	-0.813	-0.504	0.428	-1.186	---
$X_1X_{3,4}$	-0.052	-0.661	0.569	0.415	---
X_1X_5	0.037	0.039	-0.035	-0.139	---
$X_2X_{3,4}$	0.951	0.839	-1.283	-2.130	---
X_2X_5	-0.061	0.025	0.007	0.162	---
$X_{3,4}X_5$	-0.052	0.031	-0.005	0.022	---

The relatively lower R^2 values as compared to prediction test results in section 6.3.2 indicate the ability for the QRS models to fit the entire design space is slightly weaker than for the primarily interior point validation data. Therefore, boundary point predictions carry an implicit variability for which decision-makers should be mindful.

6.5 Sensitivity Analysis

6.5.1 Overall Metric Trends

A macro-level view of metric and objective value sensitivities to changes in the controls was performed. Figure 6.6, Figure 6.7, and Figure 6.8 provide a high level view of trend in the outputs.

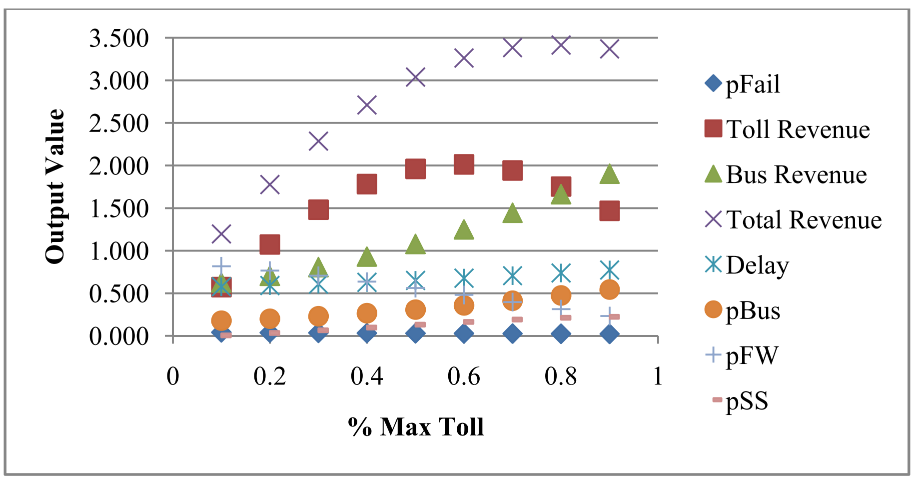


Figure 6.6 Overall Metric Sensitivity to Toll Changes

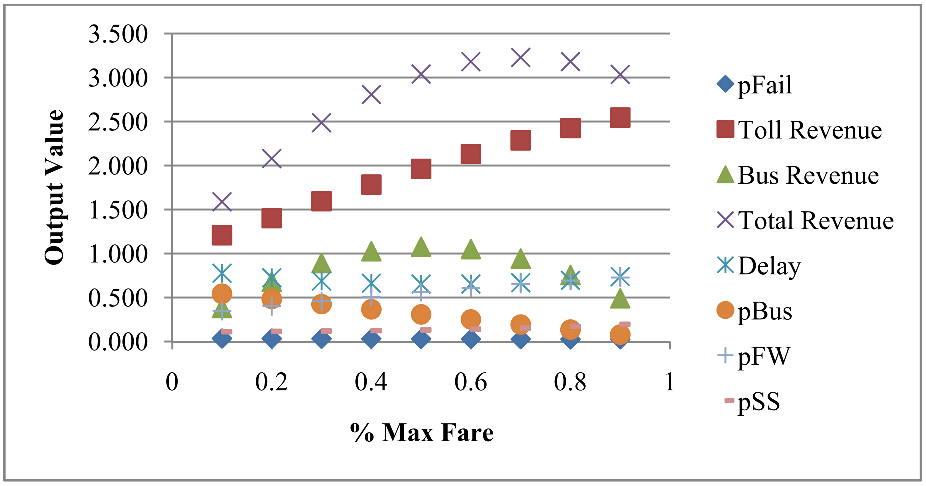


Figure 6.7 Overall Metric Sensitivity to Fare Changes

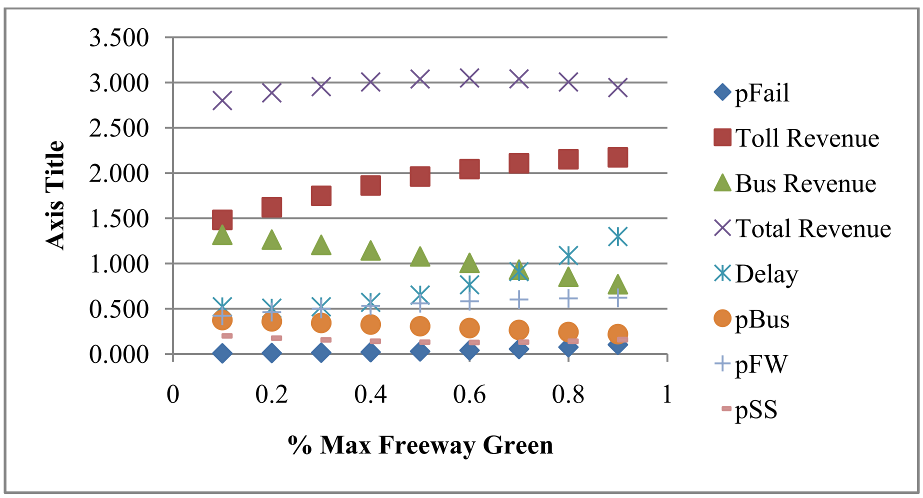


Figure 6.8 Overall Metric Sensitivity to Freeway Green Changes

6.5.2 Sensitivity of Equilibrium User Preferences—An Example

To illustrate the impacts of control pricing on user choice, figures are presented to depict the impact of increasing and decreasing tolls, using the equilibrium solutions for the 27 high fidelity simulation policy scenarios. Figure 6.9, Figure 6.10, and Figure 6.11 depict the equilibrium mode and route preferences obtained from the high fidelity simulations. Appendix B contains the graphical analysis of the sensitivity of mode and route choice to toll pricing.

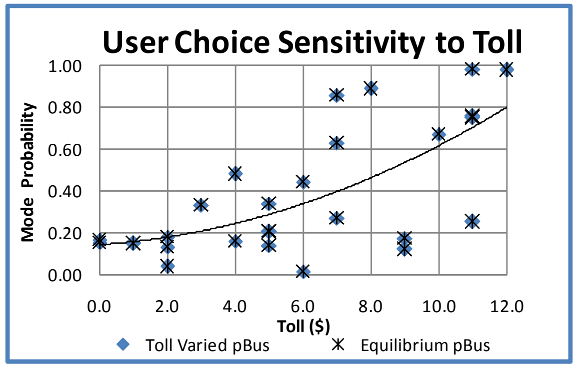


Figure 6.9 Equilibrium pBus for 27 Scenarios

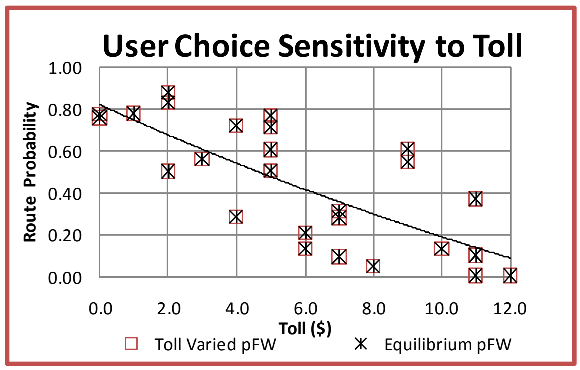


Figure 6.10 Equilibrium pFW for 27 Scenarios

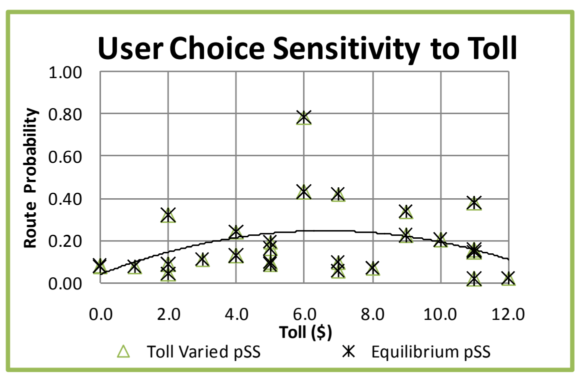


Figure 6.11 Equilibrium pSS for 27 Scenarios

6.5.3 Sensitivity Analysis for Objective Values

This section depicts the objective value for each of the four SoS objectives at various control policies. The family of optimal control policies is in Appendix C. Each depicted curve, in this section, is the result of fixing two controls at the base policy and changing each control, one at a time, in increments of 10%. Recall, the base policy sets each control at 50% of its maximum. The base policy results for each objective are shown in Table 6.6.

Table 6.6 SoS Metric and Objective Values for Base Policy

%Toll	%Fare	%Fwy Grn	pFail	Toll Revenue	Bus Revenue	Total Revenue	Delay	pBus	pFW	pSS
50%	50%	50%	0.029	$23,537	$8,086	$31,622	65.164	0.308	0.560	0.132

- pFail = 0.029 (2.9% expected likelihood of the network reaching a failed LOS state)
- Total Revenue = $31,622 in combined Toll and Fare revenue
- Delay = 65.2 (Average delay per vehicle of 65.2 seconds)
- pBus = 0.308 (30.8% of users prefer the bus at a $6 fare)

Objective: Network Reliability

Figure 6.12 depicts the expected change in pFail for varying control policies. The pFail for the base policy and at the mean demand is 0.029. As toll and fare increase, pFail decreases. Higher prices promote more surface street travel which results in congestion and increases pFail. As freeway green increases, pFail also increases. Favorable freeway green time promotes more freeway bound traffic. However, the 3-lane freeway has a capacity of 2200 veh/ln/hr, so network demands sufficiently clog freeways and spill over congestion to adjacent streets increasing pFail.

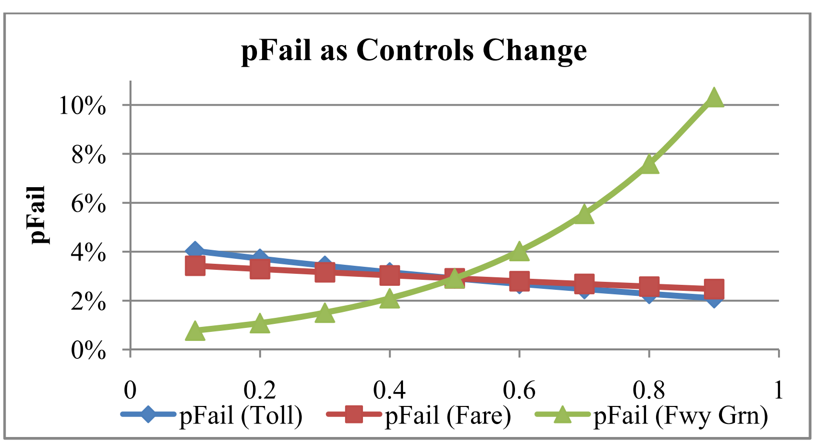


Figure 6.12 pFail Sensitivity to Control Changes

Objective: Total Revenue

Figure 6.13 depicts the expected change in Total Revenue for varying control policies. The total revenue for the base policy and at the mean demand is $31,622. Total revenue positively increases with each control. Total revenue is the most sensitive to the revenue producing controls, toll and fare. The value piques around 80% of the max toll and fare, then falls as reductions in ridership causes less total revenue.

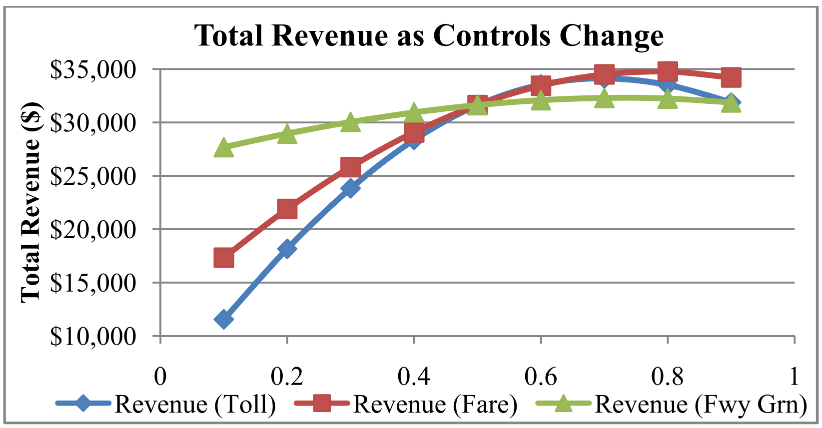


Figure 6.13 Total Revenue Sensitivity to Control Changes

Objective: Network Delay

Figure 6.14 depicts the expected change in Delay for varying control policies. Total revenue for the base policy and at the mean demand is $31,622. Delay increased with the toll, even more with freeway green. Heavy loads slowed freeway flow; related congestion on adjacent surface streets slowed local streets. Increasing bus fares remained competitive relative to tolls, but at 75% of max many bus riders switched modes and joined the congestion reflected by the increasing delay.

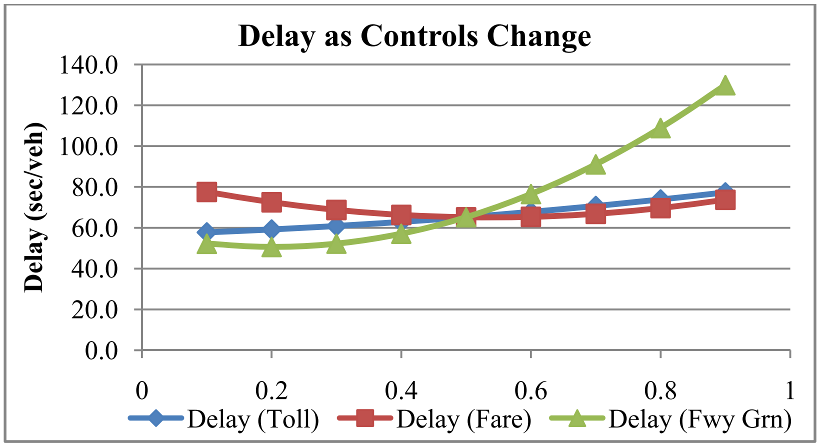


Figure 6.14 Delay Sensitivity to Control Changes

Objective: Mass Transit Ridership

Figure 6.15 depicts the expected change in pBus for varying control policies. The pBus for the base policy and at the mean demand is 0.308. Bus travel increases as tolls prices increase, due to many riders switching modes. Similarly, mass transit ridership steadily falls as fare increase. Freeway green increases cause pBus to decrease. This is a result of the increased access to load the freeway which attracts bus travelers to switch modes. As depicted in the figure, pBus is not as sensitive to freeway green as to the fare, since it is an indirect factor and fare is a direct factor.

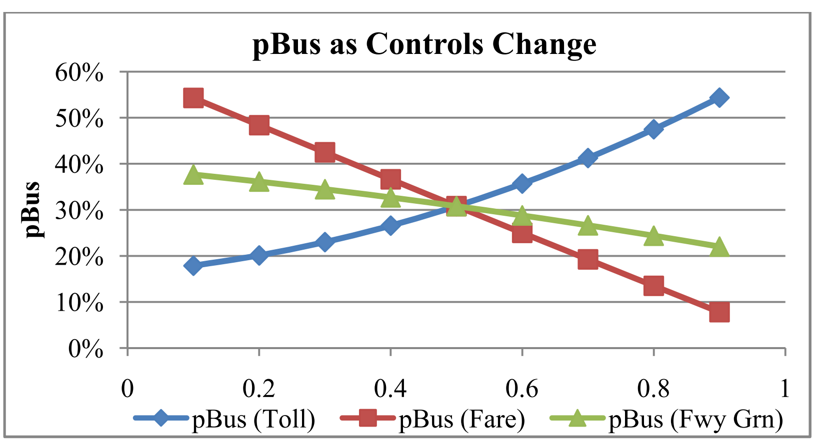


Figure 6.15 pBus Sensitivity to Control Changes

6.6 Conclusion

Solutions designed to effectively control network flow may inadvertently cause unintended consequences, particularly when the network performance is conditioned on user choice and uncertainty is propagated across the network. Therefore, capturing the SoS performance over the largest possible set of operational conditions is preferred. Yet, evaluating these aspects with a detailed stochastic simulation is limited by computational expense; therefore, surrogate models are needed. This objective illustrated how to develop and validate computationally inexpensive surrogate models to predict the outputs from a more detailed simulation, and use these models to perform probabilistic reachability analysis and sensitivity analysis.

This objective also offers some general insights for SoS problems, in general. First, prediction testing and surrogate model validation for SoS problems should be both quantitative and qualitative. Graphical inspection of predictions combined with statistical tests offer the greatest potential for properly assessing candidate surrogate models. As in this example, prediction testing may show one model to be very accurate in some places, but another model to have an overall better fit. In these circumstances, considerations such as regions of interest and intended applications become important in model selection.

Second, when analyzing coupled human-physical SoS with hybrid dynamics, the sensitivity of user preferences to changes in control policies appear to be the most clear when performed on equilibrated scenario results. The numerous trajectories that stochastic system behavior can follow complicate the understanding of the degree of the impact of changes in control policy. However, standardizing the analysis to the equilibrium results and isolating each control individually produces consistently proportional comparison values form which impacts can be judged.

Third, although the data used in this research is not intended to solve a specific number problem, the generalizable findings from the example problem suggests that for human-physical

SoS, the capacity of the faster mode is strongly linked to the likelihood of network failure. The user preference for the faster route was quickly hampered by the route capacity. Even when the cost to access the faster route (i.e., toll) was quite low and the opportunity to access the faster route (i.e., freeway green) was quite high, the network failures were significantly increased. The findings suggested additional capacity on the freeway may relieve some of the pressure on the adjacent system. When possible, SoS researchers should investigate the impacts of control policies on operations with an eye on augmenting SoS design in order to minimize the adverse impacts on the rest of the SoS.

This objective developed and validated computationally inexpensive surrogate models to predict high-fidelity simulation outputs, and used these models to perform probabilistic reachability analysis and sensitivity analysis. Such efficient models enable analysis techniques such as Monte Carlo simulation to propagate uncertainty, generate output statistics and facilitate policy optimization. The next objective applies this technique to the stochastic hybrid system in order to determine optimal policies under uncertainty. This involves determining optimal control policies by solving the network flow problem as a multi-objective optimization problem using the results from the interdependent computational models.

CHAPTER VII

UNCERTAINTY PROPAGATION AND COOPERATIVE/NON-COOPERATIVE POLICY OPTIMIZATION IN STOCHASTIC HUMAN-PHYSICAL NETWORKS WITH HYBRID DYNAMICS

7.1 Introduction

Analysis of stochastic human-physical networks with hybrid dynamics is often nested. The numerous system conditions generate many sample simulations and evaluating each sample is computationally expensive. In the hybrid system problem in this dissertation, every policy combination required enough simulation to equilibrate the integrated discrete choice and agent-based simulation. The nested analysis is particularly difficult for probabilistic reachability analysis. There are two layers of uncertainty in the problem. The sampling-based approach creates uncertainty in the initial conditions. The stochastic nature of the simulation introduces uncertainty in the potential trajectories the system could follow that lead to a failed system state. For this objective, system metrics are evaluated under uncertainty and output statistics are computed. The stochastic approach facilitates policy optimization which includes a broad set of possible network demands.

This objective also performs stochastic policy optimization for cooperative and non-cooperative decision-makers. Policy optimization is performed and compared for the network, given two cases. The hierarchical nature present in many SoS is captured in these cases by including two types of decision makers. The simulation agents are decision makers and represent the humans systems in the network that make travel path decisions based on the rules of the agent-based simulation. SoS decision makers are the other type.

SoS decision-making can be centralized under an agreement by the systems to adhere to a single policy, typically for the benefit of the whole. In the first case, centralized policy optimization is performed for cooperative system leaders that are willing to adhere to control policies set by a central authority. Centralized SoS policy optimization generates normative policies which suggest "what *should* happen" to optimize the network performance.

SoS decision-making can also be decentralized such that independent system decisions optimize single system priorities. In the second case, a de-centralized optimization is performed for non-cooperative system leaders who seek to myopically optimize the objectives that most benefit their constituent system. Decentralized SoS policy optimization generates policies from exploratory analysis of "what *could* happen" to network performance and is formulated as a game theory problem. Game theory is an important tool in making decisions in a system of systems context since SoS have managerial independence with operational interdependence (Sage and Cuppan, 2001).

In section 7.2, uncertainty is propagated and the system objectives of network delay, mass transit ridership, total revenue and network reliability are evaluated. Surrogate models are used to generate output statistics for these objectives over a stochastic demand using Monte Carlo simulation. Probabilistic reachability is evaluated and an expectation for pFail is reported. Sections 7.3 and 7.4 examine choice in optimal behavior by formulating a game theory problem.

7.2 Uncertainty Propagation

Various methods have been proposed in literature for uncertainty propagation with Monte Carlo simulation and moment-based methods as the most common (DeLaurentis, 2000), (Oberkampf et al, 1998). The surrogate models developed in the previous objective were specifically intended for use with MCS to evaluate the moments for each of the four continuous output variables.

7.2.1 Continuous Responses

Output statistics in Table 7.1 were computed for the four continuous variables using 10,000 MCS evaluations.

Table 7.1 Output Statistics for Continuous Response Variables

	pBus	pSS	pFW	Delay
Mean	0.3386	0.1745	0.4869	0.6582
Variance	0.0078	0.0003	0.0096	0.0211
Skewness	0.6517	1.7769	-1.0943	-0.6699
Kurtosis	2.8217	6.1175	3.5939	2.7948

Probability density functions were generated to provide insights into the distribution of the variables. Figure 7.1, Figure 7.2, Figure 7.3, and Figure 7.4 depict the histogram, probability density function and a "best fit" theoretical distribution.

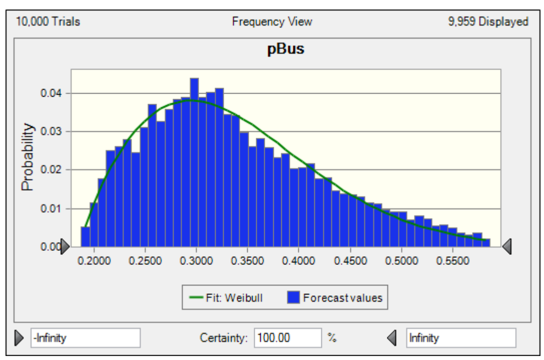


Figure 7.1 pBus Distribution from MCS

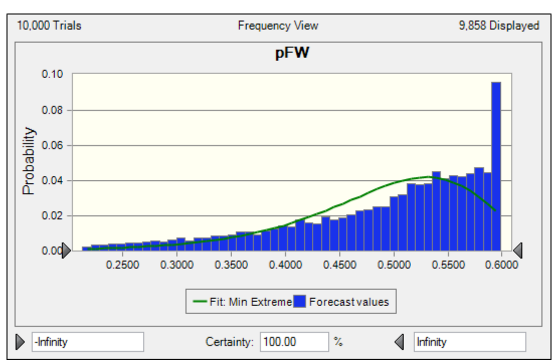


Figure 7.2 pFW Distribution from MCS

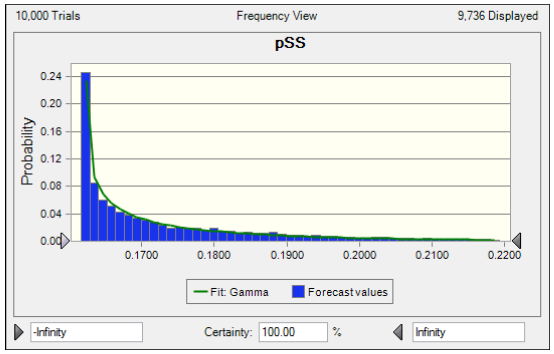


Figure 7.3 pSS Distribution from MCS

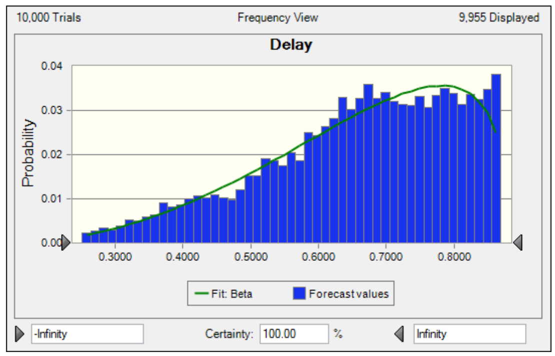


Figure 7.4 Delay Distribution from MCS

7.2.2 Probabilistic Reachability

For this problem, reachability is a key point of interest, and is made more difficult when stochastic demand is considered. The analysis of pFail shows two distinct uncertainties which influence the likelihood of reaching a failed network state. One uncertainty that impacts pFail is the system evolution due to the stochastic nature of the agent simulation. Another uncertainty is the impact of the random demand on the value of pFail. As reported in previous objective, the value for pFail = 0.029, for the base policy evaluated at a demand of 7,000. By contrast, the expectation of pFail from 10,000 Monte Carlo evaluations was 0.06 (see Table 7.2). The impact of the higher potential demand on the network was shown to increase the likelihood of the network reaching a failed state.

Table 7.2 Deterministic and Stochastic Reliability Results

	Demand	pFail
Deterministic	D = 7000	0.029
10,000 MCS	D ~ Traingle (1, 15)	0.06

An important reason for doing uncertainty propagation is to facilitate stochastic policy optimization. This section propagated uncertainty across the network and used MCS to generate output statistics for the continuous variables and an expectation for the probability of network failure. The values computed the MCS will be used by a genetic algorithm in the policy optimizer in sections 7.3 and 7.4.

7.3 Policy Optimization with Cooperative Decision Makers

Centralized SoS policy optimization generates normative policies which suggest "what *should* happen" to optimize the network performance. Centralized decision making involves a central SoS authority or an agreed consensus by constituent systems to adhere to the prescribed policies and can be implemented in many ways. Consider a problem in which three systems agree to mutually support the decisions of a SoS central authority. In this problem, the three players are: a state turnpike authority (STA), department of public transit (DPT), and a local government. The STA sets the toll price. The DPT sets the bus fare. The city engineer for the local government sets the signal timing that determines the freeway green time. The optimization problem seeking to determine the optimal control policy is formulated below.

max E[*Obj Fctn*]
w.r.t.
Toll, Fare, Fwy Grn Policy
s.t.
$0 \leq \%\ max\ Toll \leq 1$
$0 \leq \%\ max\ Fare \leq 1$
$0.2 \leq \%\ max\ FwyGrn \leq 0.8$
$pFail \leq 0.1$

The toll and fare controls were constrained to a range between 0 and 1, representing the price range as defined by the problem (toll: $0 - $12; fare: $1 - $15). The freeway green control was constrained to between 0.2 and 0.8 to represent a policy assumption for minimum and maximum percentage of route access dedicated to freeway bound traffic. This assures neither the surface street nor the freeway access can be completely shut down. A pFail < 0.1 constraint assured network reliability, as previously defined.

7.3.1 Objective Function and Weighting

A weighted objective function representing varying objective priorities is used to evaluate a family of policies. The objective function for the mathematical program is below.

$$Obj\ Fctn = w_1\left(\frac{pBus}{pBus_{ref}}\right) + w_2\left(\frac{Delay}{Delay_{ref}}\right) + w_3\left(\frac{Revenue}{Revenue_{ref}}\right)$$

As previously presented, the problem has four SoS objectives (mass transit, delay, revenue and network reliability), but only mass transit, delay and revenue have weights, as shown.

Objective Weights	Reference Values
w_1 = pBus weight	pBus = 1
w_2 = Delay weight	Delay = 2.11
w_3 = Revenue weight	Total revenue = 4.63

The network reliability objective is not weighted; rather, it is a chance constraint and imposed to assure network reliability. A 10% pFail threshold conditions all of the weighted analyses. The three weights range from w=0, implying no priority for an objective, to w=1 for exclusive priority. The reference values for each objective were determined based on single metric optimal solutions evaluated at the mean demand (see Table 7.3).

Table 7.3 Control Policies for Individually Optimized Objectives

Optimal Policy	%Toll	%Fare	%Grn	pFail	Toll Rev	Bus Rev	Total Rev	Delay	pBus	pFW	pSS
BASE POLICY	**0.5**	**0.5**	**0.5**	**0.029**	**1.961**	**1.078**	**3.039**	**0.652**	**0.308**	**0.560**	**0.132**
max pBus	1	0	0.2	0.009	0	0	0	0.768	1	0	0
min Delay	0	0	0.2	0.020	0	0	0	0.053	0.442	0.558	0
max Delay	1	0	0.8	0.063	0.684	0	0.684	2.114	0.902	0.098	0
max Revenue	1	0.844	0.751	0.038	2.485	1.794	4.630	0.882	0.304	0.355	0.341

The pBus = 1 optimizes the max pBus objective and benchmarks the highest possible mass transit ridership. The delay = 2.11 optimizes the max delay objective and benchmarks the control value for delay. The total revenue = 4.63 optimizes the max total revenue objective and benchmarks the control value for total revenue. An interesting observation from the results of optimization based on single objectives was that a policy of a maximum toll and a maximum mass transit fare was a sub-optimal policy. The results in the table above confirm the graphical analysis in Chapter 6 that total revenue (pBus + pFW revenue) reached a point at which the impact of reduced ridership offset the revenue such that total revenue began decreasing. Weights can also be decision variables. For certain SoS problems, it may also be important to solve for the optimal weights with respect to a given system-wide performance metric.

7.3.2 Deterministic and Stochastic Optimization Results

The deterministic optimal policies were first order approximations of the true design variable values. The stochastic optimal policies were the expectations of the design variables after 10,000 Monte Carlo simulations. Table 7.4 and Table 7.5 show the weight scheme and resulting optimal policies (conditioned on pFail < 10%).

Table 7.4 Deterministic Weighted Policy Results

WEIGHTS			POLICY			SoS OBJECTIVES			
w1 (pBus)	***w2*** (Delay)	***w3*** (Revenue)	**% Toll** (Deterministic)	**% Fare** (Deterministic)	**% Fwy Grn** (Deterministic)	**pFail** (Deterministic)	**Total Revenue** (Deterministic)	**Delay** (Deterministic)	**pBus** (Deterministic)
1.00	0.00	0.00	0.89	0.00	0.33	0.01	0.00	0.86	1.00
0.75	0.00	0.25	1.00	0.25	0.20	0.01	1.75	0.64	1.00
0.75	0.25	0.00	0.71	0.00	0.20	0.01	0.00	0.53	1.00
0.50	0.00	0.50	1.00	0.25	0.20	0.01	1.75	0.64	1.00
0.50	0.25	0.25	1.00	0.25	0.20	0.01	1.75	0.64	1.00
0.50	0.50	0.00	0.71	0.00	0.20	0.01	0.00	0.53	1.00
0.25	0.00	0.75	1.00	0.76	0.78	0.04	4.24	0.99	0.37
0.25	0.25	0.50	1.00	0.74	0.56	0.02	4.10	0.74	0.38
0.25	0.50	0.25	1.00	0.33	0.20	0.01	2.17	0.62	0.90
0.25	0.75	0.00	0.00	0.00	0.20	0.02	0.00	0.05	0.44
0.00	0.00	1.00	1.00	0.84	0.75	0.04	4.28	0.88	0.31
0.00	0.25	0.75	1.00	0.88	0.62	0.02	4.24	0.73	0.26
0.00	0.50	0.50	1.00	0.89	0.54	0.02	4.17	0.68	0.24
0.00	0.75	0.25	1.00	0.88	0.45	0.01	4.06	0.65	0.24
0.00	1.00	0.00	0.00	0.00	0.20	0.02	0.00	0.05	0.44

Table 7.5 Stochastic Weighted Policy Results

WEIGHTS			POLICY			SoS OBJECTIVES			
w1 (pBus)	***w2*** (Delay)	***w3*** (Revenue)	**% Toll** (Stochastic)	**% Fare** (Stochastic)	**% Fwy Grn** (Stochastic)	**pFail** (Stochastic)	**Total Revenue** (Stochastic)	**Delay** (Stochastic)	**pBus** (Stochastic)
1.00	0.00	0.00	1.00	0.00	0.20	0.01	0.00	0.77	1.00
0.75	0.00	0.25	1.00	0.26	0.20	0.01	1.80	0.64	0.99
0.75	0.25	0.00	1.00	0.01	0.20	0.01	0.07	0.76	1.00
0.50	0.00	0.50	1.00	0.46	0.20	0.01	2.75	0.60	0.73
0.50	0.25	0.25	1.00	0.31	0.20	0.01	2.07	0.63	0.92
0.50	0.50	0.00	0.76	0.00	0.20	0.01	0.00	0.57	1.00
0.25	0.00	0.75	1.00	0.70	0.66	0.03	4.10	0.87	0.42
0.25	0.25	0.50	0.96	0.54	0.26	0.01	3.17	0.61	0.59
0.25	0.50	0.25	1.00	0.44	0.20	0.01	2.67	0.61	0.76
0.25	0.75	0.00	0.00	0.00	0.20	0.02	0.00	0.05	0.44
0.00	0.00	1.00	1.00	0.78	0.61	0.02	4.19	0.77	0.35
0.00	0.25	0.75	1.00	0.79	0.53	0.02	4.13	0.70	0.34
0.00	0.50	0.50	1.00	0.78	0.44	0.01	4.01	0.65	0.34
0.00	0.75	0.25	1.00	0.72	0.33	0.01	3.74	0.63	0.40
0.00	1.00	0.00	0.00	0.00	0.20	0.02	0.00	0.05	0.44

Since the objective weights form a partition of unity, explicit values for w_3 are not evaluated (represented as zeroes in the lower triangles of each of the corresponding tables). This approach was inclusive of all four objectives in the optimal solutions and facilitated the creation of 3-D plots of the 15 point evaluations and resulting optimal policy values for % max toll, % max fare and % max freeway green. The optimal policy plots in Figure 7.5, Figure 7.6, and Figure 7.7 depict w_1 and w_2 and each optimal control value. Values for w_3 are implied as $1 - (w_1 + w_2)$.

Optimal Toll Results

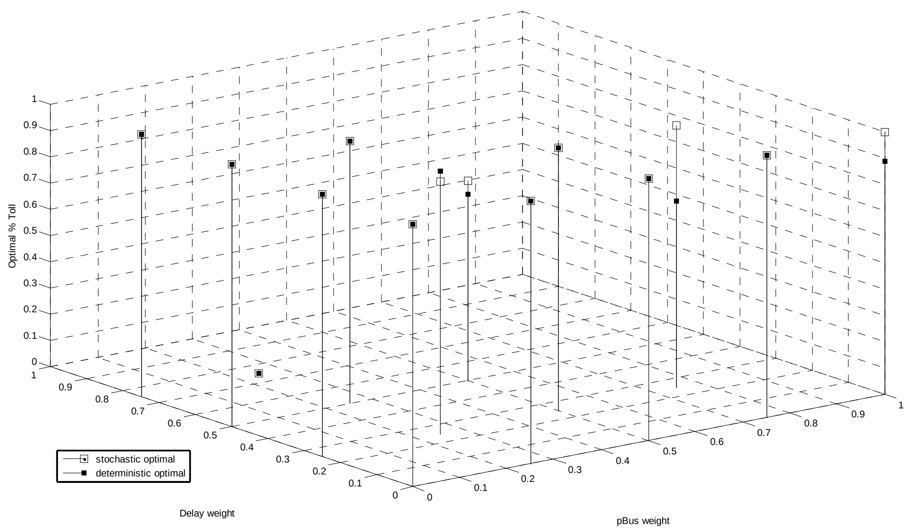


Figure 7.5 Optimal Toll Policies

Stochastic Optimal Toll Policy Results

w2 \ w1	0.00	0.25	0.50	0.75	1.00
0.00	1.00	1.00	1.00	1.00	1.00
0.25	1.00	0.96	1.00	1.00	0
0.50	1.00	1.00	0.76	0	0
0.75	1.00	0.00	0	0	0
1.00	0.00	0	0	0	0

Deterministic Optimal Toll Policy Results

w2 \ w1	0.00	0.25	0.50	0.75	1.00
0.00	1.00	1.00	1.00	1.00	0.89
0.25	1.00	1.00	1.00	0.71	0
0.50	1.00	1.00	0.71	0	0
0.75	1.00	0.00	0	0	0
1.00	0.00	0	0	0	0

Optimal Fare Results

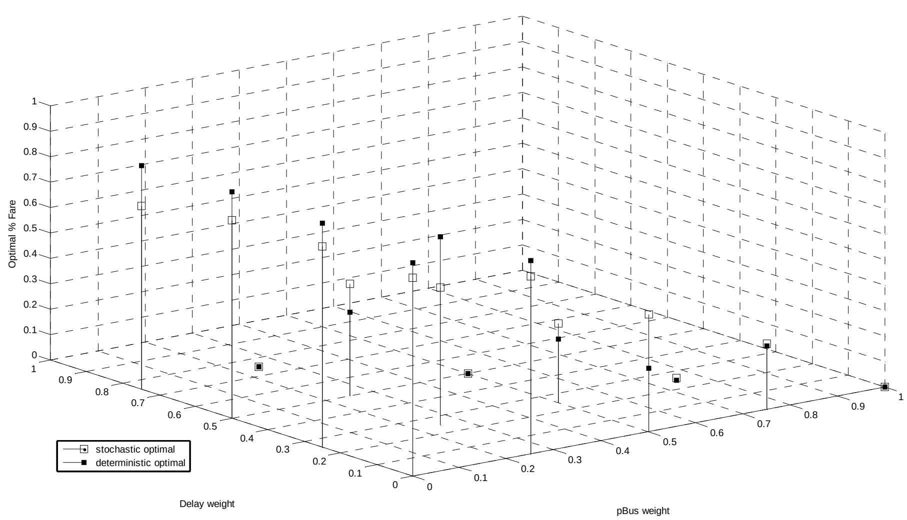


Figure 7.6 Optimal Fare Policies

Stochastic Optimal Fare Policy Results

w2 \ w1	0.00	0.25	0.50	0.75	1.00
0.00	0.78	0.70	0.46	0.26	0.00
0.25	0.79	0.54	0.31	0.01	0
0.50	0.78	0.44	0.00	0	0
0.75	0.72	0.00	0	0	0
1.00	0.00	0	0	0	0

Deterministic Optimal Fare Policy Results

w2 \ w1	0.00	0.25	0.50	0.75	1.00
0.00	0.84	0.76	0.25	0.25	0.00
0.25	0.88	0.74	0.25	0.00	0
0.50	0.89	0.33	0.00	0	0
0.75	0.88	0.00	0	0	0
1.00	0.00	0	0	0	0

Optimal Freeway Green Results

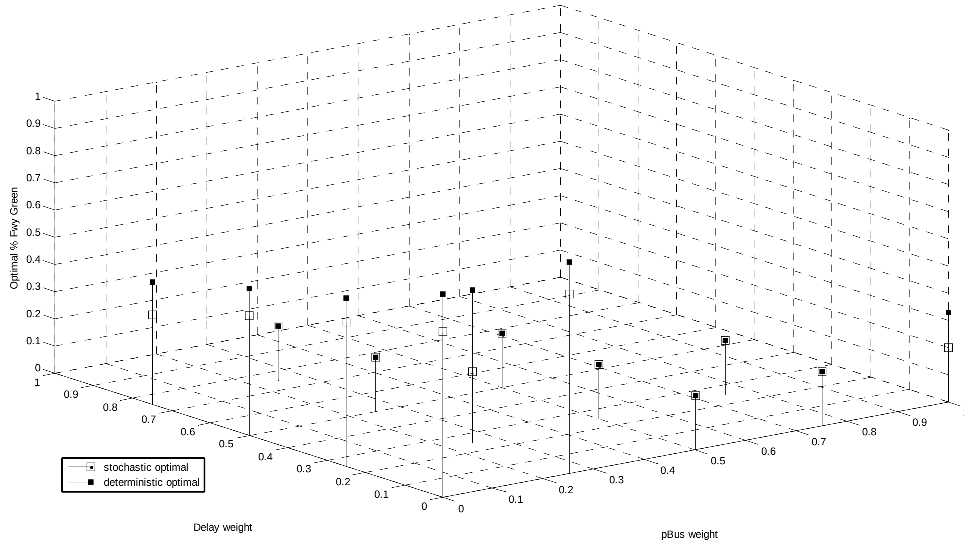


Figure 7.7 Optimal Freeway Green Policies

Stochastic Optimal Freeway Green Policy Results

w2 \ w1	0.00	0.25	0.50	0.75	1.00
0.00	0.61	0.66	0.20	0.20	0.20
0.25	0.53	0.26	0.20	0.20	0
0.50	0.44	0.20	0.20	0	0
0.75	0.33	0.20	0	0	0
1.00	0.20	0	0	0	0

Deterministic Optimal Freeway Green Policy Results

w2 \ w1	0.00	0.25	0.50	0.75	1.00
0.00	0.75	0.78	0.20	0.20	0.33
0.25	0.62	0.56	0.20	0.20	0
0.50	0.54	0.20	0.20	0	0
0.75	0.45	0.20	0	0	0
1.00	0.20	0	0	0	0

An important factor in the differences observed in the deterministic and stochastic optimal policies was the way the two methods handled network reliability. The deterministic approach solved for optimal control policies at a constant mean network demand of 7,000 vehicles/hour. The deterministic approach simply enforced a 10% limit on the pFail value, but did not ensure the actual probability the system avoided a failed state. The stochastic approach solved for optimal control policies for triangle distributed demands with a mean equal to 7,000 vehicles/hour. The inclusion of the network demands across the spectrum of potential operational conditions is an advantage of the stochastic policy optimization approach. The stochastic approach allowed for a safeguard against exceeding the pFail threshold of 10%, which provides stronger assurance to those who manage and operate the transportation network.

7.4 Policy Optimization with Competitive Decision Makers

Decentralized SoS policy optimization is needed because the normative analysis performed in centralized optimization is too idealistic for many practical situations. System cooperation within an SoS is vulnerable to any single system breaking the alliance when the myopic interests justify. A decentralized approach generates policies from exploratory analysis of "what *could* happen." Unlike the centralized approach which employs a single mathematical program, this approach is treated as a multi-player game theory problem and formulated as a combination of multiple, coupled mathematical programs.

Game theory studies situations where players choose different actions in an attempt to maximize their returns. This study of the interactions of decision makers is central to formulating and solving system of systems problems. It provides a formal modeling approach to social situations in which decision makers interact. A game consists of a set of players and a set of rewards for each player for each combination of strategies selected by the players. A game has an equilibrium strategy, if and only if, there is a strategy in which no single player can be made better off by switching strategies unilaterally (Gibbons, 1992).

The urban transportation planning problem presented in this dissertation formulated user behavior as macro-analysis of a discrete choice model coupled with an agent-based simulation. However, a game theoretic could have also been used with individual users or classes of users represented as players in the game. For example, the user equilibrium traffic flow pattern solution would be the pattern of traffic flow where no traveler is able to reduce their own travel time by unilaterally switching mode or route in the network (Wardrop, 1952), (Sheffi, 1985). The following section describes the SoS problem as a game theory problem, formulated as a three player game.

7.4.1 Problem Description

For this problem, a central authority is not presumed to oversee the SoS. Systems individually decide what is best for them conditioned on what is believed the responses of the other decision making entities will be. The situation is defined such that each system observes the decisions of the other two and responds by optimizing their control in response. For the purposes of this study, there are three players— state turnpike authority (STA), department of public transit (DPT), and a local government. The STA sets the toll price. The DPT sets the bus fare. The city engineer for the local government sets the signal timing that determines the freeway green time. This three player game of observe and respond continues as depicted in Figure 7.8 until system policies reach an equilibrium from which no system is benefited by adjusting their policy.

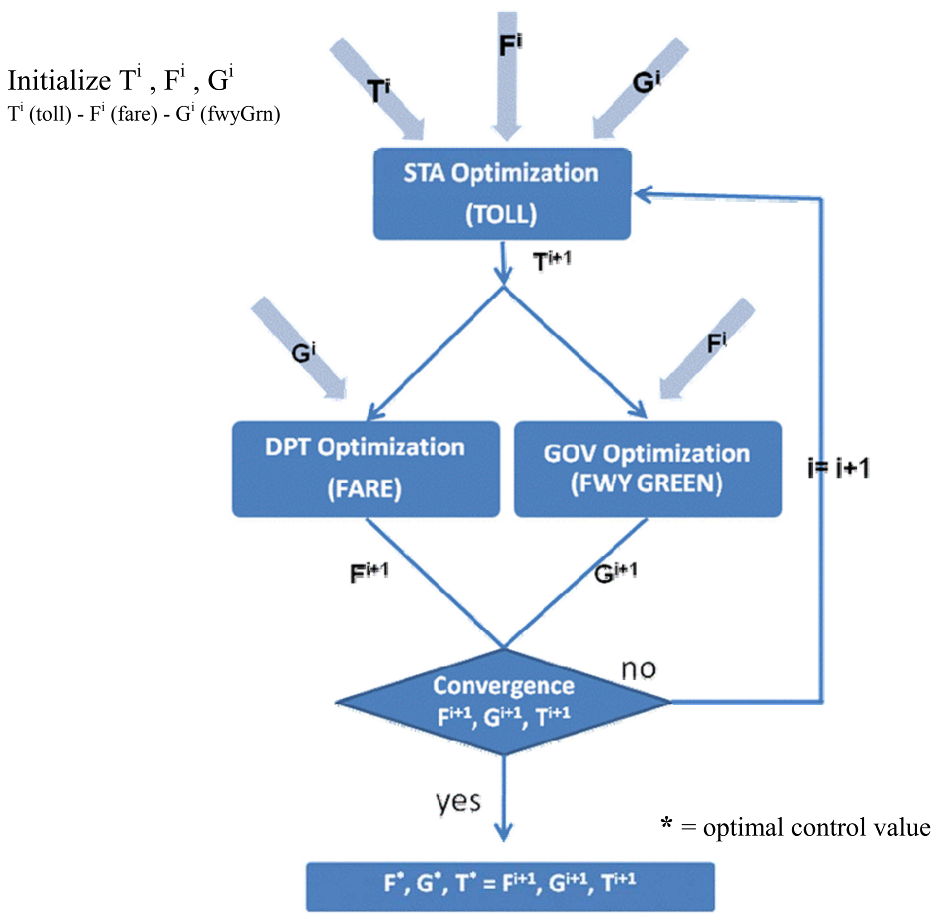


Figure 7.8 Decentralized Policy Optimization Flow Chart

System preferences are represented by weights in the objective functions. Table 7.6 contains the weights for each system, as well as weights for a centralized scheme that is proportional to the system preferences. The policies will be compared to determine how the SoS objectives differed based on system strategies. The weight table and problem formulation for each system is below.

Table 7.6 Weights for SoS Objective Priorities

Optimzation Weights	w_1	w_2	w_3	Σw_i
DeCentralized (STA)	**0**	**0.25**	**0.75**	**1**
DeCentralized (DPT)	**0.45**	**0.1**	**0.45**	**1**
DeCentralized (GOV)	**0.3**	**0.7**	**0**	**1**
Centralized (proportional)	**0.25**	**0.35**	**0.4**	**1**

State Turnpike Authority (STA):

max E[0.25(Delay) + 0.75(Total Revenue)]
w.r.t.
Toll, Fare, Fwy Grn Policy
s.t.
$0 \leq \%\ max\ Toll \leq 1$
$0 \leq \%\ max\ Fare \leq 1$
$0.2 \leq \%\ max\ FwyGrn \leq 0.8$
$pFail \leq 0.1$

Department of Public Transit (DPT):

max E[0.45(pBus) + 0.10(Delay) + 0.45(Total Revenue)]
w.r.t.
Toll, Fare, Fwy Grn Policy
s.t.
$0 \leq \%\ max\ Toll \leq 1$
$0 \leq \%\ max\ Fare \leq 1$
$0.2 \leq \%\ max\ FwyGrn \leq 0.8$
$pFail \leq 0.1$

Local Government (GOV):

max E[0.30(pBus) + 0.7(Delay)]
w.r.t.
Toll, Fare, Fwy Grn Policy
s.t.
$0 \leq \%\ max\ Toll \leq 1$
$0 \leq \%\ max\ Fare \leq 1$
$0.2 \leq \%\ max\ FwyGrn \leq 0.8$
$pFail \leq 0.1$

7.4.2 Assumptions

The presented decentralized approach includes a couple of simplifying assumptions on the uncertainties in the problem and how they are addressed in the solution. One simplifying assumption for this problem is that each system knows every system policy decision. Another assumption is the STA is the first mover and sets the initial toll from which the DPT and city government simultaneously respond. The iterative process continues until equilibrium is reached.

Robust analysis is not considered in this approach. All users are assumed to have no risk aversion and only seek to optimize the expectation of their objective function. These priorities are represented by the weights in the system problem formulations. The STA prioritizes efficient travel (to attract ridership) and toll revenue. The DPT equally prioritizes mass transit ridership and fare revenue, as they go hand-in-hand. The DPT has a very limited emphasis on efficient travel. Local government prioritizes mass transit ridership and efficient travel, because they are both public concerns. Viable mass transit is a public service; efficient travel promotes public satisfaction. Both serve to bolster public sentiment in the ability of the local government to meet their travel needs.

7.4.3 Results

The decentralized solution was obtained through an iterative process in which the STA set the toll and DPT and GOV responded with decisions for fare and green time. Since it is assumed that all players see each other, the three systems again set their respective policy in light of the other system decisions. The policy optimization problems for STA, DPT and GOV were set up in separate Excel worksheets. The start point for the process was a base policy where each control was set at the median (Toll, Fare and Fwy Grn all 0.5). The preference weights from the previously shown optimization weight table were applied to each system. The first round consisted of STA setting an optimal. This solution was the expected value for toll that optimized the mathematical program for STA. The process repeated itself for DPT and GOV with each system updating the control values for the other systems and then determining the optimal value for their control. The game theoretic reached equilibrium after three rounds of policy adjustments. Convergence was assumed when the round three was unchanged from round two for all three systems. The control policy at convergence was % max Toll = 0.79; % max Fare = 0.71; and % Fwy Grn = 0.5. The decentralized solution was the control policy from which no system could improve their situation by changing.

Table 7.7 SoS Policy Optimization Summary

	Optimal Policy	%Toll	%Fare	%FwyGrn	pFail	Toll Rev	Bus Rev	Total Rev	Delay	pBus	pFW	pSS
Deterministic	**Centralized**	1	0.68	0.45	**0.015**	**1.741**	**2.113**	**3.854**	**0.683**	**0.444**	**0.249**	**0.308**
Deterministic	**Decentralized**	1	0.84	0.38	**0.011**	**2.350**	**1.600**	**3.950**	**0.638**	**0.272**	**0.336**	**0.393**
Stochastic	**Centralized**	1	0.51	0.2	**0.007**	**0.559**	**2.381**	**2.940**	**0.603**	**0.668**	**0.080**	**0.253**
Stochastic	**Decentralized**	0.79	0.71	0.5	**0.021**	**2.460**	**1.480**	**3.940**	**0.682**	**0.297**	**0.445**	**0.258**

Table 7.7 summarizes the SoS policy optimization results. It contains deterministic and stochastic optimal policies for both the centralized and decentralized approaches. For each optimal policy, the corresponding metric values are reported. The centralized values were the solutions to the problem where a central authority sets the policy and each system obliges. In order to create comparable policies the weights for the centralized policies were set at the values in the optimization weight in the previous section. These weights are the normalized weights that are proportional to the preference structure expressed in the objective function of each system. Using these weights, the deterministic and stochastic optimal policies were determined for the network operating a demand equal to the mean.

The decentralized approach required a judgment be made assumed all three systems were risk neutral systems. The decentralized control values are the optimal solutions for the controls when determined individually by each system authority and equal to the converged solution of the game theoretic described in this section.

A key issue examined in the decentralized policy optimization is how competition affects the optimal solutions. The results show several differences between the two approaches. The centralized approach eliminated the impacts of competitive pressure. The central authority was assumed to take actuarially fair bets based solely on expected values and set the policy for each system. The decentralized approach facilitated competitive system interactions. When compared with the proportionally weighted centralized policy, some interesting observations were made. Table 7.8 summarizes the impacts of competitive pressure. Table entries depict neutral, slight or significant change in policy and objective values when competition was considered.

Table 7.8 Competitive Pressure Impacts

↓↑ Competitive Pressure	%Toll	%Fare	%FwyGrn	pFail	Total Rev	Delay	pBus
Deterministic Optimal	↔ (0%)	↑(+24%)	↓ (-16%)	↔ (0%)	↑(+3%)	↓ (-7%)	↓ (-39%)
Stochastic Optimal	↓ (-21%)	↑(+39%)	↑(+150%)	↑(+200%)	↑(+34%)	↑(+13%)	↓ (-56%)

The impact of competitive pressure for the deterministic optimal controls was neutral to toll, increasing to fare and decreasing to freeway green. The impact of competition for the deterministic problem objectives was neutral to pFail, slightly increasing to total revenue, slightly decreasing to delay and significantly decreasing to pBus. Regarding the stochastic problem optimal controls, the impacts were all significant and decreasing to toll, increasing to fare and increasing to freeway green. The impacts to the objective values were significant increases to pFail and total revenue, slight increases to delay and a significant decrease to pBus.

The results prompted understanding of the operational environment and insights into the reason for some of these impacts. The competitive pressure resulted in an increase in total revenue due primarily to the increase in freeway travelers. This is especially clear in the stochastic case where the pFW increased from 8% to 45% in response to the toll being lowered by more than 20%. It appears the decision by the STA to lower the toll attracted many of the bus travelers. The response by the DPT to the 40% drop in pBus was an increase in fare among the faithful bus riders in order to boost revenue. The GOV appears to make decision that value operations over revenue. In both cases the competitive pressure only had slight impact on the network delay (local government's highest priority). If a delay of approximately 60 seconds/vehicle is acceptable, then GOV maintained their priority by increasing freeway green time in order to preserve the low delay. The DPT appears to be most negatively impacted in the decentralized solution. In summary, the competitive pressure increased total revenue in both cases, but the revenue split was more disparate from the centralized solution. The local government, unaffected by bus and toll revenue, has to make operational decisions to keep delay under control that exacerbate the revenue difference.

7.5 Conclusion

This objective investigated policy optimization for stochastic human-physical network with hybrid dynamics. The system objectives of network delay, mass transit ridership, total revenue and network reliability were used to evaluate the different approaches. Output statistics for control policies were evaluated over a stochastic demand using Monte Carlo simulation. Policy optimization was performed and compared for cooperative and non-cooperative network systems. The SoS policy optimization indicated that "what should happen" when system leaders are willing to implement a control policy (deemed optimal for the entire SoS by a central authority) can significantly differ from "what could happen" when system leaders optimize the objectives that most benefit their constituent system.

This objective illuminated several insights into SoS problems. First, virtually all SoS problems are impacted by uncertainty. Properly capturing its impacts begins at the very beginning of the modeling process. The complexity of SoS problems of any realistic size requires careful handling of what is known, what is not known, and even what may be unknown unknowns. Second, probabilistic reachability analysis showed an increase in pFail when stochastic demand was considered. This suggests practitioners should exercise caution when employing deterministic SoS models evaluated to infer network performance. Third, the humans-in-the-loop for SoS problems, introduce aspects for which the rationality caveat imposed in this research are not sufficient. This simplified decision models are only generalizable to problems for which similar rationality can be assumed. This was evidenced by the numerous contingencies that emerged when formulating the game theory problem, but were beyond the scope of the illustrative focus of the approach presented in this objective.

A final insight for SoS, in general, is that competitive pressures within interdependent networks are an important consideration for those who operate and manage SoS. Competitive pressures have both direct and indirect impacts which may be difficult to observe or anticipate, depending upon the complexity of the system relationships. Also, SoS policies agreed upon by cooperative system leaders are vulnerable to competitive decisions by systems to leverage opportunities that benefit an individual system, despite the potential detriment to the SoS as a whole.

CHAPTER VIII

SUMMARY AND FUTURE NEEDS

8.1 Summary of Contributions

While particular views vary, it is widely agreed that System of Systems (SoS) is a new and critical discipline for which design and analysis techniques are incomplete. Both computational and physical systems science have often ignored the responses of the humans who will ultimately be the end users of modern engineered systems and systems of systems. Because humans both affect and react to their environment, the nature of their interdependent relationship is inherently complex. SoS exhibit hybrid dynamics in the sense that network evolution is governed by the interaction of physical and informational processes. Systems often form stochastic human-physical networks in which decision are made under uncertainty by both individuals and system managers. Failure to properly account for the nature of the interactions between these systems can lead to unintended events with unanticipated consequences. This dissertation has motivated and developed a SoS approach for evaluating existing or potential operational policies for human-physical systems in which network flows and human decisions are coupled and stochastic and hybrid dynamics are present.

This dissertation research did not directly address SoS architecture or design issues; rather it accepted the architecture and design as given and sought to improve the overall performance of the existing SoS through identifying optimal operational controls. In the context of an economic example, a preliminary optimization under uncertainty approach was developed to produce optimal investment policies for economic stimuli. In the context of a transportation example, this research illustrated how current approaches may simply transfer the travel cost between modes and why effective policies should optimize across broader system-wide metrics. An integrated discrete choice and agent-based simulation approach for stochastic human-physical networks with hybrid (continuous and discrete) dynamics was developed. The stochastic human-physical analysis framework facilitated the integration of system simulation, uncertainty analysis and optimization under uncertainty. The dual impacts of user decisions on physical system performance and the state of the physical system on subsequent user choices were shown to converge to stochastic user equilibria. Surrogate models to approximate the agent-based stochastic simulation were developed and validated with statistical tests. Using these surrogates, uncertainty was propagated and stochastic and deterministic policy results were computed and compared. Monte Carlo simulation was used to propagate uncertainty and an objective weighting scheme was employed to determine optimal control policies that captured varying SoS decision maker priorities.

Stochastic policy optimization was illustrated for an urban transportation SoS problem as a multi-objective optimization problem and both centralized and decentralized approaches were formulated and compared. Optimal policies were obtained and compared for both cooperative and competitive strategies to assess the impact of competitive pressure on resulting optimal policies. The stochastic hybrid system example illustrated the difference between what "should" happen and what "could" happen in application. Through the integration of system simulation, uncertainty analysis and optimization under uncertainty, an SoS approach was shown to provide decision support to those who manage and operate human-physical networks, including networks with stochastic and hybrid dynamics.

8.2 Future Needs

Although this dissertation research is a contribution to the body of knowledge in this field, this line of research contains both incremental and substantial needs which provide many opportunities for future work.

8.2.1 Incremental Needs

Four areas of future work should be considered for advancing the research presented in this dissertation. One area involves implementing a different experimental design and performing boundary analysis. This research intentionally focused on the interior of the design space from the LHS design to the region of validation points. The computational complexity dictates some narrowing of the areas of interest in order to feasibly study stochastic hybrid systems. Future work should investigate differences in results between boundary points where extreme events occur and the results presented in this dissertation. Optimally controlling the typical conditions is an important endeavor, but performance at the edges of the design space would prove equally important and quite interesting.

A second extension for this research is to expand the analysis to include robust policy optimization. Expectations of the variables were the predominant measures for this work, but adding variance to the objective function would incorporate how the users and system managers view risk. A weighted combination of $E[X] + \sigma^2 x$ in the objective function would facilitate not only stochastic hybrid system performance but also risk profiles for the various decision makers.

A third worthwhile area to study in future work is how a more sophisticated discrete choice model would change the optimal policy. This research used a macroscopic model to assign proportions of the users to each mode and route. If users could be classified and a detailed model employed, then optimal policy could be tailored to the area of operations. Computational expense is a limiting factor for this extension, but advanced analysis of how human decisions are made and the impacts of those decisions in the physical system would be very insightful.

8.2.2 Substantial Needs

Given the ever-increasing size of SoS problems, more work to develop computationally affordable approaches to integrate multi-disciplinary system models and uncertainty analysis to provide optimal control policies to SoS with decision-makers is clearly needed and timely. There are substantial needs relevant to this effort on which future work should orient. One such need is the further development of the foundations of a modern systems science that integrates the computational, physical, and human behavioral aspects of modern systems. A modern systems science must combine, in a computationally affordable way, the models of physical, computational, and human behavioral processes to capture the emergent behaviors stemming from the interactions of these three types of processes. This has proven difficult because these systems exhibit stochastic hybrid dynamics, and these models must be capable of reproducing the behavior of a dynamic system that has continuous time (and possibly stochastic) dynamics, interrupted by discrete (and possibly stochastic) events. Agent-based models are capable of representing human logic and computational processes. Differential equations and difference equations are capable of modeling physical processes with continuous time dynamics. These modeling languages must be efficiently integrated to derive a modeling tool for systems of systems. These detailed capabilities would be a substantial extension to the work presented in this dissertation.

A second substantial need in this research area involves the employment of surrogate models. When the phenomena are complex, a compromise must be found between completeness and simplicity, particularly when models are the basis for uncertainty analysis. Because surrogate models are less detailed than the computational models whose behavior it is attempting to replicate, the question of whether the model is accurate enough for its intended purpose must be addressed. Techniques to verify that surrogate models are predicting the evolution of the system to a required level of accuracy, at a given time horizon, are needed. Also, for surrogate models deemed accurate enough, the uncertainty analysis must be able to incorporate a variety of types of uncertainty arising from a lack of data, uncertainty in computational models, and basic randomness in model inputs and system evolution.

Finally, in order to properly validate the appropriateness of an SoS approach for other applications and domains, the approach presented in this dissertation should be applied to a real world problem. Analysis of a real-world operational environment, especially an area for which historical knowledge can be used to validate the results, would serve to validate the approach for certain problem instances under certain operational conditions.

The broadest understanding of our interdependent world is required in order to achieve meaningful advances. Interdisciplinary research and cross-domain collaboration must continue to increase to provide the methods and tools to face the significant challenges posed by multidisciplinary systems with hybrid dynamics. In providing decision support to those who manage SoS, it is clear that all three aspects of physics, computation, and human behavior must be integrated into a single analysis. Critical to this effort is the establishment of a systems science which integrates these processes into a modeling language where all three of these processes can interact with each other to jointly describe the evolution of the system. These models must be reduced to a form that replicates the system behavior without the need for a computationally expensive simulation, and then they must be exploited to optimally control the networked systems in accordance with priorities set by numerous decision-makers whose strategies may be cooperative, competitive or a bit of both.

The transportation and economic examples featured in this dissertation are among the many important SoS examples in which computational, physical and behavioral processes are intertwined. SoS in areas such as telecommunications, space travel, national security, disaster response and even social networks continue to increase in our world as advances in technology and communication foster interdependent systems from micro-networks to global networks. The nexus of human decisions and their physical consequences will no doubt continue to be heavily investigated, both academically and practically, for years to come.

APPENDIX A. MATLAB CODE FOR PRESS TEST

Models produced from Response Surface Methods can be used to generate interpolating predictions. Predicted Residual Sum of Squares, or PRESS (Allen, 1971) is an effective technique to evaluate candidate models. The procedure for performing PRESS on a sample of size *n* is as follows:

1) Individually hold out each *i* th observation and recalculate and evaluate the fitted model for the n-1 remaining data.
2) Calculate the prediction error for the *i* th observation and square the difference.
3) Repeat the process for all *n* observations and compute the sum of squares
4) The value for sum of squares can be compared to other candidate models with the lowest value preferred.

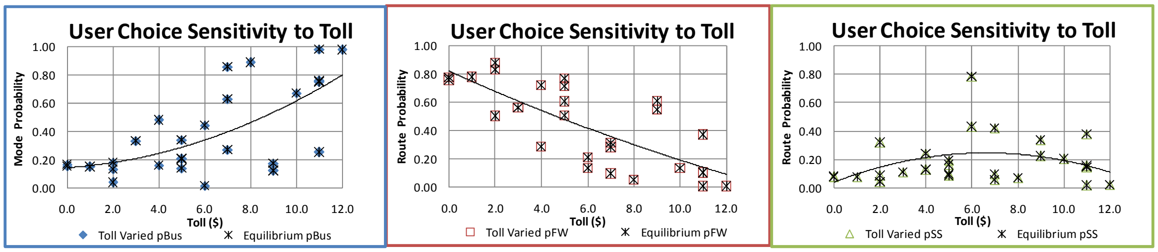
A PRESS test was performed for each of the candidate Gaussian Process models. Separate GP models were generated for the four continuous output variables of interest (P(bus); P(FW); P(SS); Delay). Three trend functions (constant, linear and quadratic) were used to assess the predictive accuracy of each GP model. The MATLAB code below was used to produce PRESS values for each of the candidate models.

```
training_points4;

nsams = size(train_pnts,1)-1;
ndims = size(train_pnts,2)-1;
order = 1;
trdim = order*ndims + 1;        % num entries in trend fn

SSresid=0
for i=1:24
    if i==1
        Delete_pnts=train_pnts(i+1:24,:)
        Delete_vals=train_vals(i+1:24,:)
    else if i==24
            Delete_pnts=train_pnts(1:23,:)
            Delete_vals=train_vals(1:23,:)
        else
            Delete_pnts=train_pnts(1:i-1,:)
            Delete_vals=train_vals(1:i-1,:)
            Delete_pnts=[Delete_pnts; train_pnts(i+1:24,:)]
            Delete_vals=[Delete_vals; train_vals(i+1:24,:)]
            model=build_gp(nsams,ndims,Delete_pnts,Delete_vals,order,trdim);
            SSresid=SSresid+[eval_gp(train_pnts(i,:),model,false)-
train_vals(i)]^2;
        end
    end
end

PRESS=SSresid
```

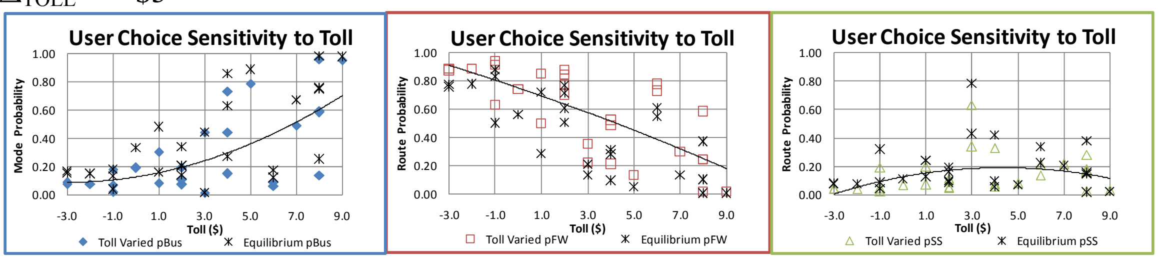
Appendix B. CONTROL SENSITIVITY ANALYSIS: AN EXAMPLE

Mode and Route Sensitivity to Decreasing Toll Prices

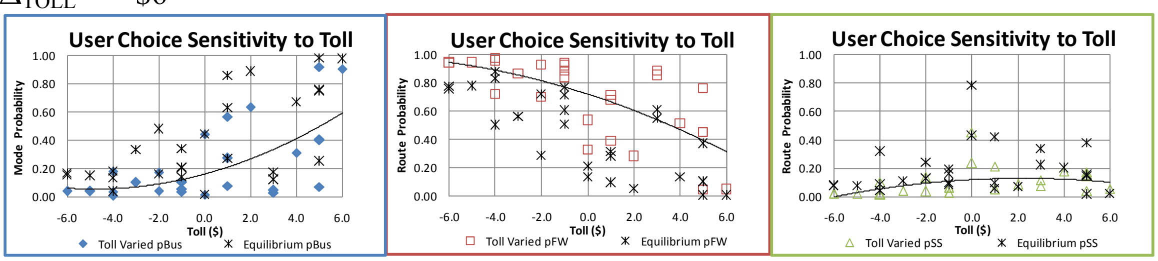
[**Current Policy**]. TOLL = $6



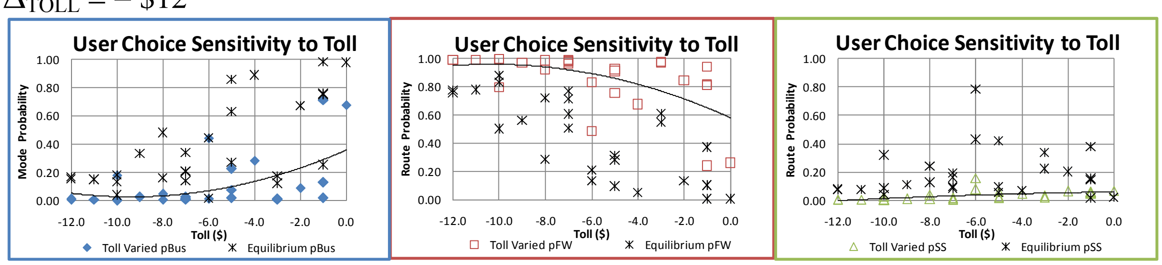
$\Delta_{TOLL} = -\$3$



$\Delta_{TOLL} = -\$6$

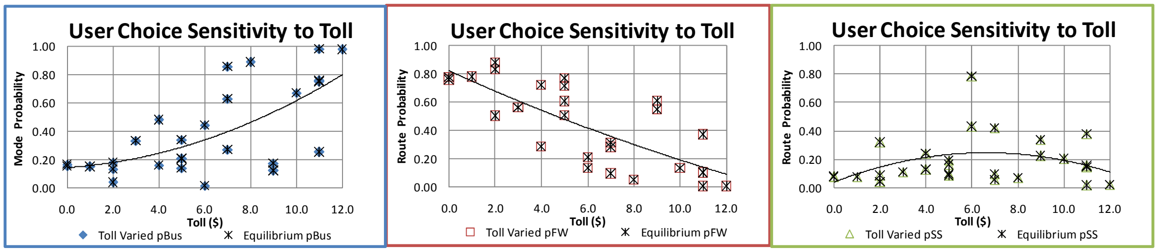


$\Delta_{TOLL} = -\$12$

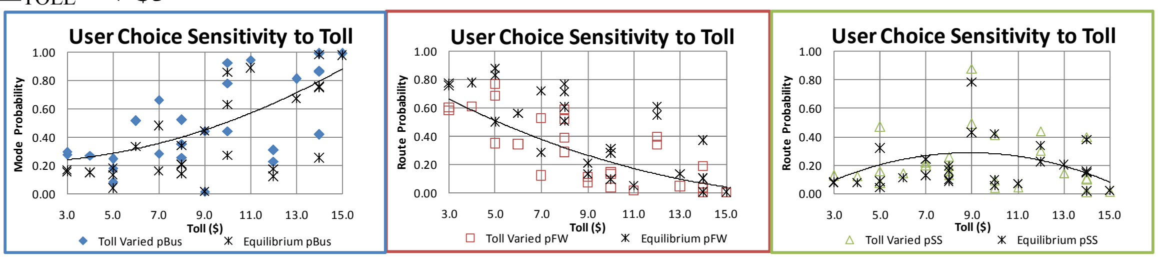


Mode and Route Sensitivity to Increasing Toll Prices

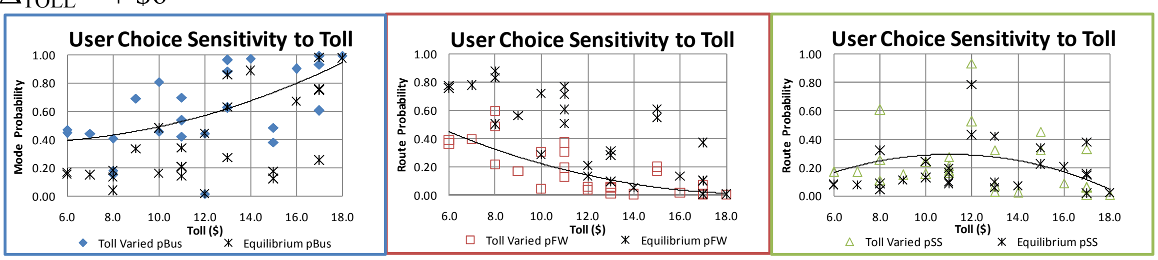
[**Current Policy**]. TOLL = \$6



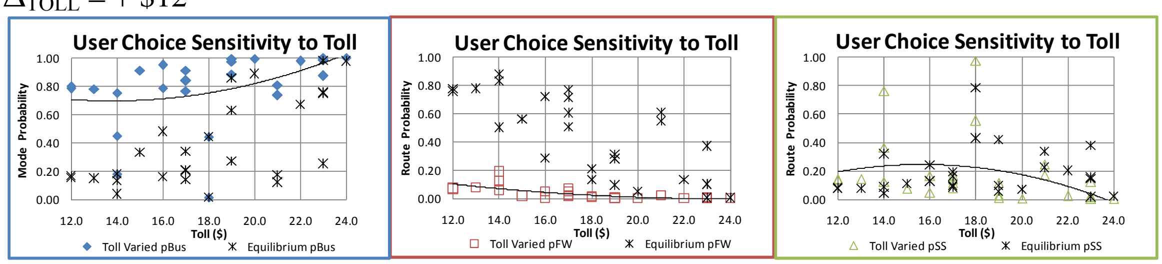
$\Delta_{\text{TOLL}} = +\$3$



$\Delta_{\text{TOLL}} = +\$6$



$\Delta_{\text{TOLL}} = +\$12$



Appendix C. SOS METRIC AND OBJECTIVE RESULTS FOR FAMILY OF POLICIES

%Toll	%Fare	%Fwy Grn	pFail	Toll Revenue	Bus Revenue	Total Revenue	Delay	pBus	pFW	pSS
10%	50%	50%	0.040	$6,865	$4,688	$11,553	57.7	0.179	0.817	0.004
20%	50%	50%	0.037	$12,878	$5,276	$18,154	59.1	0.201	0.767	0.032
30%	50%	50%	0.034	$17,786	$6,038	$23,824	60.8	0.230	0.706	0.064
40%	50%	50%	0.032	$21,386	$6,975	$28,361	62.9	0.266	0.636	0.098
50%	**50%**	**50%**	**0.029**	**$23,537**	**$8,086**	**$31,622**	**65.2**	**0.308**	**0.560**	**0.132**
60%	50%	50%	0.027	$24,169	$9,370	$33,539	67.8	0.357	0.480	0.163
70%	50%	50%	0.025	$23,300	$10,829	$34,129	70.7	0.413	0.396	0.191
80%	50%	50%	0.023	$21,043	$12,462	$33,505	73.9	0.475	0.313	0.212
90%	50%	50%	0.021	$17,620	$14,270	$31,890	77.3	0.544	0.233	0.223

%Toll	%Fare	%Fwy Grn	pFail	Toll Revenue	Bus Revenue	Total Revenue	Delay	pBus	pFW	pSS
50%	10%	50%	0.034	$14,486	$2,850	$17,337	77.5	0.543	0.345	0.112
50%	20%	50%	0.033	$16,817	$5,079	$21,896	72.5	0.484	0.400	0.116
50%	30%	50%	0.032	$19,127	$6,692	$25,818	68.7	0.425	0.455	0.120
50%	40%	50%	0.030	$21,378	$7,692	$29,071	66.3	0.366	0.509	0.125
50%	**50%**	**50%**	**0.029**	**$23,537**	**$8,086**	**$31,622**	**65.2**	**0.308**	**0.560**	**0.132**
50%	60%	50%	0.028	$25,566	$7,876	$33,442	65.3	0.250	0.609	0.141
50%	70%	50%	0.027	$27,432	$7,069	$34,501	66.8	0.192	0.653	0.155
50%	80%	50%	0.026	$29,099	$5,669	$34,767	69.6	0.135	0.693	0.172
50%	90%	50%	0.025	$30,532	$3,680	$34,212	73.7	0.078	0.727	0.195

%Toll	%Fare	%Fwy Grn	pFail	Toll Revenue	Bus Revenue	Total Revenue	Delay	pBus	pFW	pSS
50%	50%	10%	0.008	$17,792	$9,889	$27,681	52.2	0.377	0.424	0.200
50%	50%	20%	0.011	$19,464	$9,485	$28,950	50.6	0.361	0.463	0.175
50%	50%	30%	0.015	$20,989	$9,050	$30,039	52.2	0.345	0.500	0.156
50%	50%	40%	0.021	$22,352	$8,583	$30,935	57.1	0.327	0.532	0.141
50%	**50%**	**50%**	**0.029**	**$23,537**	**$8,086**	**$31,622**	**65.2**	**0.308**	**0.560**	**0.132**
50%	50%	60%	0.040	$24,526	$7,556	$32,082	76.5	0.288	0.584	0.128
50%	50%	70%	0.056	$25,298	$6,996	$32,294	91.1	0.267	0.602	0.131
50%	50%	80%	0.076	$25,828	$6,405	$32,233	108.9	0.244	0.615	0.141
50%	50%	90%	0.103	$26,091	$5,782	$31,873	129.9	0.220	0.621	0.159

REFERENCES

Abate, A. (2007). Probabilistic Reachability for Stochastic Hybrid Systems: Theory, Computations, and Applications. Technical Report No. UCB/EECS-2007-132. University of California at Berkeley.

Abate, A., Prandini, M. Lygeros, J. and Sastry, S. (2008). Probabilistic reachability and safety for controlled discrete time stochastic hybrid systems. Volume 44, Issue 11, November 2008.

Agarwal, H., and Renaud, J. (2004). "A Unilevel Method for Reliability-based Design Optimization," in *Proceedings of 45th AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics and Material Conference*, AIAA-2004-2029.

Agresti A. (2002). Categorical data analysis (2nd edition). New York: Wiley.

AIAA (American Institute of Aeronautics and Astronautics). (1998), Guide for the Verification and Validation of Computational Fluid Dynamics Simulations, AIAA-G-077-1998, Reston, VA, American Institute of Aeronautics and Astronautics.

Alekseev, A. and Navon, I. (2003). "Calculation of uncertainty propagation using adjoint equations," International Journal of Computational Fluid Dynamics**,** 17(4) pp. 283-288.

Alexandrov, N. and Lewis, R. (2000). "Algorithmic Perspectives on Problem Formulation in MDO", AIAA Paper 2000-4718, 8th AIAA/USAF/NASA/ISSMO Symposium on MA&O, Long Beach, CA, 9/5-9/00.

Allen, D (1971). Mean square error of prediction as a criterion for selecting variables. *Technometrics* 13, 469-475.

Ambartzumian R, Der Kiureghian A, Ohanian V, Sukiasian H. (1997). "Multinormal Probability by Sequential Conditioned Importance Sampling." In: *Advances in Safety and Reliability, Proc. ESREL '97*, vol. 2, p. 1261–1268.

Arthur, W. , Holland, J., LeBaron, B., Palmer, R. and Tayler, P. (1997). *The Economy as a Complex Evolving System II*, Santa Fe Institute Studies in the Sciences of Complexity, eds. Arthur, W. B., Durlauf, S. & Lane, D. (Addison–Wesley, Reading, MA), Proceedings Vol. 27, pp. 15–42.

Asarin, E., Dang, T. and Girard, A. (2003). Reachability analysis of nonlinear systems using conservative approximation. In O. Maler and A. Pnueli, editors, Hybrid Systems: Computation and Control, Lecture Notes in Computer Science 2623, pages 20–35. Springer Verlag.

Axelrod, R., 1997, *The Complexity of Cooperation: Agent-Based Models of Competition and Collaboration* (Princeton Univ. Press, Princeton, NJ).

Bartolomei, J. "Qualitative Knowledge Construction for Engineering Systems: Extending the Design Structure Matrix Methodology in Scope and Procedure." (2007). Ph.D. Dissertation. MIT.

Bayarri, M., Berger, J., Higdon, D., Kennedy, M., Kottas, A., Paulo, R., Sacks, J., Cafeo, J., Cavendish, J., Lin, C. and Tu, J. (2002). "A framework for validation of computer models." Technical Report 128, National Institute of Statistical Sciences, Research Triangle Park, NC, 2002.

Bayen, A., Raffard, R., and Tomlin, C. (2004). "Network Congestion Alleviation Using Adjoint Hybrid Control: Application to Highways," Proceedings, Hybrid Systems: Computation and Control (R. Alur, G. Pappas, Eds.), LNCS 2993, pp. 95-110.

Bazarra, M., Jarvis, J., and Sherali, H. (1990). *Linear Programming and Network Flows*, New York: John Wiley & Sons.

Beers, W. and Kleijnen, J. (2005). Robustness of Kriging when interpolating in random simulation with heterogeneous variances: Some experiments. *European Journal of Operational Research*, 165(3), 826-834.

Bellemans, T., De Schutter, B. and De Moor, B. (2002). "Models for Traffic Control," Journal A, v43, no. 3-4, pp. 13-22.

Bellman, R. 91957). Dynamic Programming, Princeton University Press, Princeton, NJ, 1957.

Berkson, J. (1944). Application of the logistic function to bio-assay. Journal of the American Statistical Association 39, 357-365.

Bertsimas, D. and Tsitsiklis, J. (1997). Introduction to Linear Optimization, Athena Scientific, Belmont, Massachusetts.

Bichon, B., Eldred M., Swiler, P., Mahadevan, S., and McFarland, J. (2008). "Efficient Global Reliability Analysis for Complex Engineering Applications" *AIAA J.* Under Review

Birge, J. and Louveaux, F. (1997). Introduction to Stochastic Programming, Springer.

Bliss, C. (1934). The Method of Probits. Science 79, 38-39.

Boulding, K. (1956). "General Systems Theory: The Skeleton of Science," *Management Science*, 2, pp.197-208.

Brannick, M., "Logistic Regression." (2007). [http://luna.cas.usf.edu/~mbrannic/files/regression/Logistic.html] accessed April 2008.

Braun, R. (1996). "Collaborative Optimization: An Architecture for Large-Scale Distributed Design," Ph.D. Thesis, Stanford University.

Braun, R., and Kroo, I. (1996). "Development and applications of the collaborative optimization architecture in a multidisciplinary design environment," In: Alexandrov N, Hussaini MY (eds) Multidisciplinary Design Optimization: State-of-the-Art.

Brooks and Sage, "System of Systems Integration and Test." (2005). Information Knowledge Systems Management 5 p. 261.

Buckley, W. (1967). *Sociology and Modern Systems Theory*, Prentice-Hall. 1967.

Bürger, H. (1991). Mathematik Oberstufe: Lehrplankommentar. Wien: Österreichischer Bundesverlag, [translated quote in Ossimitz (2003)].

Bujorianu, M. and Lygeros, J. (2003). Reachability questions in piecewise deterministic markov processes. In O. Maler and A. Pnueli, editors, Hybrid Systems: Computation and Control, Lecture Notes in Computer Science 2623, pp. 126–140. Springer Verlag.

Carlock, P. and Fenton. R. (2001). "System of Systems (SoS) Enterprise Systems for Information-Intensive Organizations," Systems Engineering, Vol. 4, No. 4, pp. 242–261.

Cassidy, M. (2002), "Critique of an On-Ramp Metering Scheme and Broader Related Issues," Institute of Transportation Studies.

Cassidy, M. (2003), "Freeway On-Ramp Metering, Delay Savings and the Diverge Bottleneck," Transportation Research Record: Journal of the Transportation Research Board, vol. 1856, 1-5.

Cassidy, M., Anani, S., Haigwood, J. (2002), "Study of Freeway Traffic Near an Off-Ramp," Transportation Research: Part A, 36, 563–572.

Cayford, R., Wei-Hua, L., and Daganzo, C. (1997). The NETCELL simulation package: Technical description. California PATH Research Paper. Institute of Transportation Studies. University of California Technical Paper.

Chen, C., Jia, Z. and Varaiya, P. (2001). "Causes and cures of highway congestion," IEEE Control Syst. Mag., vol. 21, no. 6, pp. 26-32.

Chen, X., Hasselman, T. and Neill, D. (1997). "Reliability-based Structural Design Optimization for Practical Applications," *Proceedings of the 38th IAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics, and Material* Conference, Kissimmee, Florida, AIAA-97-1403.

Cheng, I., Cruz, J. and Paquet, J.G. (1974). “Entrance Ramp Control for Travel Rate Maximization in Expressways,” Transportation Research, vol. 8, pp. 503-508.

Chiralaksanakul, A. and Mahadevan, S. (2004). “Multi-disciplinary design optimization under uncertainty,” Proceedings, Tenth AIAA/ISSMO Multi-disciplinary Analysis and Optimization Conference, Paper No. AIAA-2004-4311.

Chiralaksanakul, A. and Mahadevan, S. (2007). “Decoupled Approach to Multi-disciplinary Design Optimization Under Uncertainty,” *Optimization Engineering*, 8:21-42.

Chiralaksanakul, A., and Mahadevan, S. (2004). “Reliability-Based Design Optimization Methods,” DETC04/DAC-57456, Proceedings of DETC'04, Salt Lake City, UT, Sept. 28-Oct. 2.

Cooper, J., Ferson, S., Ginzburg, L. (1996). “Hybrid Processing of Stochastic and Subjective Uncertainty Data,” Risk Analysis vol.16 iss.6, pg.785.

Cramer, E., Dennis, J., Frank, P., Lewis, R., (1994). “Problem Formulation for Multi-disciplinary Optimization,” *Society for Industrial and Applied Mathematics*. Vol. 4, No. 4.

Crossley, W. (2004). "System of Systems: An Introduction of Purdue University Schools of Engineering's Signature Area", Engineering Systems Symposium, MIT Engineering Systems Division, Cambridge, MA.

Cruse, T., Wu, Y., Dias, J., and Rajagopal, K. (1988). “Probabilistic Structural Analysis Methods and Applications,” Computers & Structures, 30, pp. 163-170.

Daganzo, C. (1979). Multinomial Probit: The Theory and Its Application to Demand Forecasting. Academic Press, New York.

Daganzo, C. and Sheffi, Y (1977). On Stochastic Models of Traffic Assignment. Transportation Science. vol. 11, No.3, August 1977, pp. 253-274.

Daganzo, C. (2008). “Existence of urban-scale macroscopic fundamental diagrams: Some experimental findings,” Transportation Research B.

Daganzo, C.F., Laval, J., and Muñoz, J., (2002), “Ten Strategies for Freeway Congestion Mitigation with Advanced Technologies,” California Partners for Advanced Transit Highways (PATH), Institute for Transportation Studies.

DAU. Defense Acquisition University. (2006). Defense Acquisition Guidebook. Vers. 1.06. DAU.

Davenport, T., Leibold, M., and Voelpel, S. (2006). *Strategic Management in the Innovation Economy*. Publicis Corporate Publishing and Wiley-VCH.

Davidsson, P., Henesey, L., Ramstedt, L., Törnquist, J., and Wernstedt, F. (2005). An Analysis of Agent-based Approaches to Transport Logistics. Transportation Research Part C: Emerging Technologies. Vol 13(4). pps 255-271.

DeLaurentis D. and Marvris D. (2000). "Uncertainty Modeling and Management in Multi-disciplinary Analysis and Synthesis," 38th AIAA Aerospace Sciences Meeting, Paper No. AIAA 2000-0422, 10-13.

DeLaurentis, D., Sindiy, O., Fry, D., and Ayyalasomayajula S. (2006). Modeling Framework and Lexicon for System of Systems Problems. IEEE Transactions on Systems, Man, and Cybernetics, Part A: Systems and Humans.

DeLaurentis, D. (2005). “Understanding Transportation as a System of Systems Design Problem,” 43rd AIAA Aerospace Sciences Meeting, Reno, Nevada, Jan. 10-13, 2005. AIAA-2005-0123.

Denardo, E.V. (1965). Sequential Decision Processes, PhD Dissertation, Northwestern University.

Der Kiureghian, A., Zhang, Y. and Li, C. (1994). Inverse Reliability Problem. Journal of Engineering Mechanics, ASCE, 120(5): 1154-1159.

Dilts, D., and Pence, K. (2006). "Impact of Role in the Decision to Fail: An Exploratory Study of Terminated Projects," *Journal of Operations Management*, 24, 378-396.

Ditlevsen, O. (1979). "Narrow Reliability Bounds for Structural Systems." J. Struct. Mech., 7 (4) , 453-472.

Ditlevsen, O. and Madsen, H. (1996). *Structural Reliability Methods,* J. Wiley & Sons, New York, 384 pp. ISBN 0-471-96086-1.

DOD. Department of Defense Directive 4630.5, "Interoperability and Supportability of Information Technology (IT) and National Security Systems (NSS)," Nov 2003. [http://www.dtic.mil/whs/directives/corres/pdf/d46305_011102/d46305p.pdf] accessed Mar 2004.

DOE. (2000). Department of Energy. Advanced Simulation and Computing (ASCI) Program Plan, 01-ASCI-Prog-01, Sandia National Laboratories, Albuquerque, New Mexico.

DOE. (2008). Department of Energy. "Multiscale Mathematics and Optimization for Complex Systems." National Laboratories Announcement, LAB 08-13.

Dreyfus, S. (1996). Dynamic Programming and the Calculus of Variations, Academic Press, New York.

Du, X., and Chen, W. (2000). "Towards a Better Understanding of Modeling Feasibility Robustness in Engineering Design," ASME J. Mech. Des., 122, pp. 385-394.

Du, X., and Chen, W. (2004). "Sequential Optimization and Reliability Assessment Method for Efficient Probabilistic Design," *J. Mech. Des.*, 126, pp. 225–233.

Du, X., and Sudjianto, A. (2003). "Reliability-Based Design with the Mixture of Random and Interval Variables," DETC03/DAC-48709, Proceedings of DETC'03, Chicago, IL, Sept. 2-6.

Dunnett, C., and Sobel, M. (1954). A bivariate generalization of Student's t-distribution with tables for certain special cases. *Biometrika*, 41, 153–169.

Ferson S., Joslyn C., Helton J., Oberkampf W., and Sentz K. (2004). Summary from the epistemic uncertainty workshop:consensus amid diversity, Reliability Engineering and System Safety, 85 (2004) 355–369.

Forrester, J. (1961). Industrial Dynamics. Productivity Press.

Forrester, J. (1968) Urban Dynamics. MIT.

Francois, C. (1999) "Systemics and Cybernetics in a Historical Perspective," Systems Research and Behavioral Science Syst Res. 16, 203-219.

Friedman, J. (1991). "Multivariate Adaptive Regression Splines," Annals of Statistics 19.

Geroliminis, N., and Daganzo, C., (2007), "Existence of Urban-scale Macroscopic Fundamental Diagrams: Some Experimental Findings," October, VWP-2007-5.

Gharajedaghi, J. (2005). Systems Thinking: Managing Chaos and Complexity, Butterworth-Heinemann.

Gibbons, R. (1992). Game Theory For Applied Economists. Princeton University Press, 45-47.

Girard, A., Julius, A. and Pappas, G. (2006). Approximate simulation relations for hybrid systems. In International Federation of Automatic Control Conference on Analysis and Design of Hybrid Systems, Alghero, IT.

Gunawan, S., and Azarm, S. (2005). "A Feasibility Robust Optimization Method Using Sensitivity Region Concept," ASME J. Mech. Des., 127, pp. 858-865.

Haimes, Y. and Jiang, P. (2001). "Leontief Based Model of Risk in Complex Interconnected Infrastructures," Journal of Infrastructure Systems, pp. 1-12.

Haldar, A., Mahadevan, S. (2000). Probability, Reliability and Statistical Methods in Engineering Design, J. Wiley & Sons, New York, NY.

Haldar, A., Mahadevan, S. (2000). Reliability Assessment Using Stochastic Finite Element Analysis, Wiley, New York.

Haldar, A., Mahadevan, S. (2002). *Probability, Reliability and Statistical Methods in Engineering Design,* J. Wiley & Sons, New York, NY, 304 pp., ISBN-10: 0-471-33119-8.

Ham, C., Qu, Z., and Johnson, R. (2000). "Robust Fuzzy Control for Robot Manipulators" *IEEE Proc. Control Theory & Appl.* vol. 147, no. 2, pp. 212-216.

Hasofer, A, and Lind, N. (1974). "Exact and Invariant Secnond Moment Code Format," Journal of the Engineering Mechanics Division, ASCE, Vol. 100, No. EM1, pp. 111-121.

Hedlund, S. and Rantzer, A. (2002). Convex dynamic programming for hybrid systems. IEEE Transactions on automatic Control, AC-47(9):1526–1540.

Hilbe, J. (2009). Logistic Regression Models. Boca Raton, FL. Chapman & Hall/CRC Press.

Hohenbichler, M., and Rackwitz, R. (1987). "First-Order Concepts in Systems Reliability." *Structural Safety*; Vol. 1 pp. 177-188.

Holmes, D. (1994). "A Collection of Stochastic Programming Problems," Technical Report 94-11, National Science Foundation Grant DDM-9215921.

Hoogendoorn, S. and Bovy, P. (2001). "State-of-the-art of vehicular traffic modeling," *J. Syst. Control Eng*., vol. 215, pp. 283-303.

Hosmer D. and Lemeshow S. (1989). *Applied Logistic Regression*. New York: John Wiley and Sons.

Hosmer, D. and Lemeshow, S. (2000). *Applied logistic regression*. Edition: 2. New York: John Wiley and Sons.

Hosmer, D., Hosmer, T., Cessie, S. and Lemeshow, S. (1997). A Comparison of Goodness-of-Fit Tests for the Logistic Regression Model. Statistics in Medicine. vol 16, 965-980.

Intriligator, M. (1971). Mathematical Optimization and Economic Theory. Prentice-Hall, Inc: Englewood Cliffs, NJ.

Isukapalli, S., Roy, A. and Georgopoulos, P. (1998). "Stochastic response surface methods (SRSMs) for uncertainty propagation: Application to environmental and biological systems." *Risk analysis*; 18(3): 351-363.

Jansen, B., Swinkels, P., Teeuwen, G., Antwerpen de Fluiter, B., and Fleuren, H. (2004). "Operational planning of a large-scale multi-modal transportation system." *European Journal of Operational Research*. vol 156(1), July. 41-53.

Jin, W. and Zhang, M., (2001), "Evaluation of On-ramp Control Algorithms," California PATH Working Paper, UCB-ITS-PWP-2001-14, California Partners for Advance Transit and Highways.

Jones, D., Schonlau, M., and Welch, W. (1998). Efficient global optimization of expensive blackbox functions. Journal of Global Optimization, 13(4):455–492.

Jongen, J., Ruckmann, J. and Stein, O. (1998). Generalized semi-infinite optimization: A first order optimality condition and examples, Mathematical Programming 83 (1998) 145-158.

Jung, D. and Lee, B. (2002). "Development of a Simple and Efficient Method for Robust Optimization," Int. J. Numer. Methods Eng., 53, pp. 2201-2215.

Kaymaz, I. (2005). Application of kriging method to structural reliability problems. Structural Safety, 27(2):133–151.

Keating, C., Rogers, R. Unal, R., Dryer, D., Sousa-Poza, A., Safford, R., Peterson, W. and Rabadi, G. (2003). "System of Systems Engineering," Engineering Management Journal, Vol. 15, No. 3, pp. 36-45.

Kennedy, M. and O'Hagan, A. (2001). "Bayesian calibration of computer models." *J. R. Statist. Soc. B*, Vol. 63, Part 3, pp. 425-464.

Kennedy, M., Anderson, C., Conti, S. and O'Hagan, A. (2006). Case studies in gaussian process modelling of computer codes. Reliability Engineering and System Safety, 91:1301–1309.

Kodiyalam, S. (1998). *Evaluation of Methods for Multi-disciplinary Design Optimization (MDO), Phase I.* NASA/CR-1998-20716.

Kortanek, K. (2001). On the 1962-1972 decade of semi-infinite programming: a subjective view, in: M.A. Goberna, M.A. Lopez (Eds.), Semi-infinite programming. Recent Advances, Nonconvex Optimization and Its Applications 57, Kluwer, Dordrecht, pp. 3-41.

Kotov, V. (1997). "Systems of Systems as Communicating Structures," Hewlett Packard Computer Systems Laboratory Paper HPL-97-124, pp. 1–15.

Kotsialos, A. and Papageorgiou, M., (2001). "Efficiency versus fairness in network-wide ramp metering," in *Proceedings 4th IEEE Conference on Intelligent Transportation Systems*, pp. 1190-1195.

Kurzhanski, A. and Varaiya, P. (2002). On reachability under uncertainty. SIAM Journal of Control and Optimization, 41(1):181–216.

Lee, K., and Park, G., (2002). "Robust Optimization in Discrete Design Space for Constrained Problems," AIAA J., 40(4), pp. 774-780.

Leontief, W., (1953). "Domestic Production and Foreign Trade: The American Capital Position Re-Examined." Proceedings of the American Philosophical Society; 97: 332-349.

Lian, Y. and Liou, M. (2005). "Multiobjective Optimization Using Coupled Response Surface Model and Evolutionary Algorithm. " AIAA Journal vol.43 no.6 (1316-1325).

Liang, J., Mourelatos, Z., and Tu, J., (2004). "A Single-Loop Method for Reliability-Based Design Optimization," *in Proceedings of the ASME Design Engineering Technical Conferences.*

Lighthill, M., and Whitham, G., (1956). "On Kinematic Waves II. A Theory of Traffic Flow on Long Crowded Roads," Proceedings of the Royal Society of London, 229(1178):317-345.

Liu, P., and Der Kiureghian, A. (1986). Multivariate distribution models with prescribed marginals and covariances, Probabilistic Engineering Mechanics, 1(2), 105-112.

Luce, D. and Raiffa, H. (1957). *Games and Decisions: Introduction and Critical Survey*, Dover ISBN 0-486-65943-7.

Luskasik, S. (1998). "Systems, Systems of Systems, and the Education of Engineers," Artificial Intelligence for Engineering Design, Analysis, and Manufacturing, Vol. 12, No. 1, pp. 55-60.

Lygeros, J. (2004). On reachability and minimum cost optimal control. Automatica, 40 - 6:317–927, 2004.

Lygeros, J., Tomlin, C. and Sastry, S. (1999). Controllers for reachability specifications for hybrid systems. Automatica, 35(3):349–370.

Madsen, H., and Hansen, P. (1992). "A Comparison of Some Algorithms for Reliability-based Structural Optimization and Sensitivity Analysis," *Proceedings of the 4th IFIP WG 7.5 Conference, Munich, Germany, R. Rackwitz and P. Thoft-Christensen*, Eds., Springer-Verlag, Berlin, pp. 443–451.

Mahadevan, S. and Dey, A. (1997). "Adaptive Monte Carlo Simulation for Time-Variant Reliability Analysis of Brittle Structures," *AIAA Journal*, 35(2), pp.321-326.

Mahadevan, S. and Rebba, R. (2005). "Validation of reliability computational models using Bayes networks," *Reliability Engineering and System Safety*, 87(2), 223–232.

Mahadevan, S., and Dey, A. (1998). “Ductile System Reliability Analysis Using Adaptive Importance Sampling,” *Structural Safety,* Vol. 20, pp. 137-154.

Mahadevan, S., and Smith, N. (2006). “Efficient First-Order Reliability Analysis of Multi-disciplinary Systems,” *International Journal of Reliability and Safety*, Vol. 1, Nos. 1-2, pp. 137-154.

Mahadevan, S., Zhang, R., and Smith, N. (2001). “Bayesian networks for system reliability reassessment,” *Structural Safety*, 23 231-251.

Maier, M. (1998). "Architecting Principles for System of Systems," Systems Engineering, Vol. 1, No. 4, pp. 267-284.

Maier, M. (1994). “Heuristic Extrapolation in System Architecture,” in *Proceedings of the 4th International Symposium of the National Council on System Engineering*, NCOSE, Vol. 1, pp. 525- 532.

Manthorpe, W. (1996). "The Emerging Joint System of Systems: A Systems Engineering Challenge and Opportunity for APL," John Hopkins APL Technical Digest, Vol. 17, No. 3, pp. 305–310.

Mas-Collel, A., Whinston, M., and Green, J. (1995). *Microeconomic Theory*, Oxford University Press.

McDonald, M. and Mahadevan, S. (2007). “Design Optimization with System Reliability Constraints,” *ASME Journal* of Mechanical Design.

McDonald, M. and Mahadevan, S. (2008). “Design Optimization with System Reliability Constraints,” ASME Journal of Mechanical Design, published online.

McDonald, M. and Mahadevan, S. (2008). “Design Optimization with Discrete and Continuous Design and Random Variables,” ASME Journal of Mechanical Design, published online.

McFarland, J. (2008). Uncertainty Analysis For computer Simulations Through Validation and Calibration. Ph.D. Dissertation. Vanderbilt University. p.67.

McFarland, J., Mahadevan, S., Romero, V. and Swiler, L. (2008). “Calibration and uncertainty analysis for computer simulations with multivariate output,” AIAA Journal vol. 46 no. 5.

McKay, M., Conover, W., and Beckman, R. (1979). "A Comparison of Three Methods for Selecting Values of Input Variables in the Analysis of Output from a Computer Code". Technometrics 21: 239–245.

McTrans Highway Capacity Software. (2008). [http://mctrans.ce.ufl.edu/hcs/]. Accessed April, 2009.

Mehta, S., Cruse, T., and Mahadevan, S. (1992). “System Certification”, Reliability Technology – 1992. The Aerospace Division, Ed. T. A. Cruse, ASME, 28.

Midgley, G. (2003). Systemic Practice and Action Research, Vol. 16, No. 2.

Mitchell, I., Bayen, A. and Tomlin, C. (2001). Validating a Hamilton-Jacobi Approximation to Hybrid System Reachable Sets? LNCS 2034, pp. 418-432, 2001. Springer-Verlag Berlin Heidelberg.

Mitchell, I., Bayen, A., and Tomlin, C. (2005). A time-dependent hamilton-jacobi formulation of reachable sets for continuous dynamic games. IEEE Transactions on Automatic Control, 50(7):947–957.

Morgenstern, O. and von Neumann, J. (1947). The Theory of Games and Economic Behavior, Princeton University Press.

Muñoz, J. and Daganzo, C. (2000), “Experimental Characterization of Multi-lane Freeway Traffic Upstream of an Off-ramp Bottleneck,” California PATH Program, University of California.

Nagurney, (2003). A. Innovations in Financial and Economic Networks, Edward Elgar Publishing.

Nagurney, A., and Toyasaki, F. (2003) “Supply chain supernetworks and environmental criteria.” *Transportation Research D*, 8, 185-213.

Nagurney, A., Dong, J., and Zhang, D. (2002) “A supply chain network equilibrium model.” *Transportation Research E*, 38, 282-303.

Nagurney, A., Dong, J. (2001). Supernetworks: Decision-Making for the Information Age. Edward Elgar Publishers, Cheltenham, England.

Nash, J. (1950). "Equilibrium Points in n-person Games" Proceedings of the National Academy of the USA 36(1):48-49.

NET, (1996). “System Wide Adaptive Ramp Metering Algorithm – High Level Design,” Final Report prepared by NET for Caltrans and FHWA.

Nowak, A. and Collins, K. (2000). *Reliability of Structures,* McGraw-Hill Companies, Inc., New York, NY.

Oakley, J. and O’Hagan, A. (2002). Bayesian inference for the uncertainty distribution of computer model outputs. Biometrika, 89:769–784.

Oberkampf, W., Diegert, K., Alvin, K., and Rutherford, B. (1998). “Variability, Uncertainty, and Error in Computational Simulation,” AIAA*/ASME Joint Thermophysics and Heat Transfer Conference* ASME-HTD-Vol. 357-2. pp. 259-272.

Oberkampf, W., Deland, S., Rutherford, B., Diegert, K. and Alvin, K. (1999). “A New Methodology for the Estimation of Total Uncertainty in Computational Simulation,” *Proceedings of* 40th *AIAA/ASME/ASCE/AHS/ASC Structures, Structural dynamics, and Materials Conference and Exhibit*, pp. 3-3060.

Ossimitz, G. (2003). “Systems Thinking and System Dynamic Modeling,” *5th International Conference on Technology in Mathematics Teaching*, Klagenfurt, Austria.

OUSD-ATL. (2004). “SoS and FoS FAQ,” Office of the Under Secretary of Defense for Acquisition, Technology, and Logistics, [http://www.acq.osd.mil/ dpap/Docs/FAQs%20%20SoS%20&%20FoS.doc], accessed March, 2004.

Page, Scott (2005). “Agent Based Models." The New Palgrave Dictionary of Economics 2nd edition. L. Blume and S. Durlauf (eds.), Pal-grave Macmillan, Basingstoke.

Pahl, G. and Beitz, W. (1996). Engineering Design, A Systematic Approach, Second Edition. Springer-Verlag, London.

Pandey, M. (1998). “An Effective Approximation to Evaluate Multinormal Integrals.” Structural Safety; Vol. 20 pp. 51–67.

Papageorgiou, M. and Kotsialos, A., (2000), “Freeway Ramp Metering: an Overview,” Procs. IEEE Intelligent Transportation Systems, pp. 228-238.

Papageorgiou, M., Diakaki, C., Dinopoulou, V., Kotsialos, A., and Wang, Y. (2003). “Review of Road Traffic Contol Strategies,” Proceedings of the IEEE, Vol., 91, No. 12.

Parkinson, A., Sorensen, C., and Pourhassan, N. (1993). “A General Approach for Robust Optimal Design,” ASME J. Mech. Des., 115, pp. 74-80.

Payne, H. and Thompson, W. (1974). “Allocation of freeway ramp metering volumes to optimize corridor performance.” IEEE Transactions on Automatic Control (1974) (19).

Pei, R. (2000). "Systems of Systems Integration (SoSI) – A Smart Way of Acquiring Army C4I2WS Systems," Proceedings of the Summer Computer Simulation Conference, pp. 574–579.

Powell, W. and Sheffi, Y. (1982). The Convergence of Equilibrium Algorithms with Predetermined Step Sizes. Transportation Science. vol. 16, No. 1.

Prescott, P. (2009). “Orthogonal-column Latin hypercube designs with small samples.” Computational Statistics and Data Analysis 53, 1191-1200.

Prietula, M., Gasser, L. and Carley, K., eds. (1998). *Simulating Organizations: Computational Models of Institutions and Groups* (MIT Press, Cambridge, MA).

Putko, M., Newmann, P., Taylor, A. and Green, L. (2001). *Approach for uncertainty propagation and robust design in CFD using sensitivity derivatives***.** AIAA CFD Conference*,* vol. 15.

Rackwitz, R. and Fiessler, B. (1978). "Structural Reliability under Combined Load Sequences," Comput. Struct., vol. 9, pp 489 - 494.

Rackwitz, R., Fiessler, B. (1976). Note on Discrete Safety Checking When Using Non-Normal Stochastic Models for Basic Random Variables. Load Project Working Session, MIT, Cambridge, MA.

Rechtin, E. (2000). The Art of Systems Architecting. New York: CRC Press.

Richards, P. (1956). "Shock waves on the highway," Operations Research, 4(1):42-51.

Roberts, J. (2004). WP6: Recovery from Economic Collapse: Insight from Input-Output Models and the Special Case of a Collapsed Oil Producer. Overseas Development Institute: London, England.

Rockafellar, R. (2007). Coherent Approaches to Risk in Optimization Under Uncertainty. Informs 2007 Annual Conference, Seattle Washington.

Rockafeller, R. (2001). "Optimization Under Uncertainty," lecture notes, University of Washington.

Rosenblatt, M. (1952). Remarks on a Multivariate Transform, Annals of Mathematical Statistics, Vol. 23, No. 3, pp. 470-472.

Ross, T., Booker, J., Parkinson, W. (2002). *Fuzzy logic and probability applications: Bridging the Gap*. Taylor and Francis.

Royset, J., Der Kiureghian, A., and Polak, E. (2001). "Reliability-Based Optimal Design of Series Structural Systems." Journal of Engineering Mechanics*,* ASCE, Vol. 127, No. 6, 607-614.

Royset, J., Der Kiureghian, A, and Polak, E. (2001). "Reliability-Based Optimal Structural Design by the Decoupled Approach," Reliability and Structural Safety, Vol. 73, pp. 213-221.

Sabbagh, K. (1996). Twenty-First Century Jet: The Making and Marketing of the Boeing, 777, Scribner, New York, NY

Sacks, J., Welch, W., Mitchell, T., and Wynn, H. (1989). "Design and Analysis of Computer Experiments," Statistical Science, Vol. 4, No. 4, pp. 409–435.

Sage, A., and Cuppan, C. (2001). "On the Systems Engineering and Management of Systems of Systems and Federations of Systems," Information, Knowledge, Systems Management, Vol. 2, No. 4, pp. 325–45.

Schumaker, L. (2007). Spline Functions: Basic Theory, 3rd ed. Cambridge Mathematical Library.

Schwartz, S. and Tan, H. (1977). "Integrated Control of Freeway Entrance Ramps by Threshold Regulation," Proceedings IEEE Conference on Decision and Control, pp. 984-986.

Sheffi, Y. (1981). Aggregation and Equilibrium with Miltinomial Logit Models. Proceedings, International Symposium on Equilibrium and Supply Models. University of Montreal, Montreal.

Sheffi, Y. (1985). *Urban Transportation Networks*, Prentice-Hall. Englewood Cliffs, NJ.

Simon, H. (1957). *Models of Man: Social and National.* Wiley, New York.

Simpson, T., Mauery, T., Korte, J., and Mistree, F. (2001). Kriging models for global approximation in simulation-based Multi-disciplinary design optimization. AIAA Journal, 39(12): 2233–2241.

Smith, N. and Mahadevan, S. (2005). "Integrating System-Level and Component-Level Designs Under Uncertainty." *Journal of Spacecraft and Rockets*, Vol. 42, No.4, pp. 752-760.

Smith, N., and Mahadevan, S. (2007). "Experience with Reliability-Based Design Optimization for Multi-disciplinary Systems," submitted to *Structural and Multi-disciplinary Optimization.*

Sproles, N. (2000). Coming to Grips with Measures of Effectiveness. Inc. Syst Eng 3: 50.58, John Wiley & Sons. 2000.

Stocki, R., Kolanek, K., Jendo, S., and Kleiber, M. (2001). "Study on Discrete Optimization Techniques in Reliability-Based Optimization of Truss Structures," Comput. Struct., 79, pp. 2235-2247.

Stoebner, A. M. and Mahadevan, S. (2000). "Robustness in Reliability-Based Design," *Proceedings of the 41st* AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics, and Materials Conference, Atlanta, GA.

Stursberg, O. and Krogh, B. (2003). Effcient representation and computation of reachable sets for hybrid systems. In A. Pnueli O. Maler, editor, Hybrid Systems: Computation and Control, Lecture Notes in Computer Science 2623, pp. 482–497. Springer Verlag.

Tabac, D. (1972). "A Linear Programming Model of Highway Traffic Control," 6th Annual Princeton Conference on Information Science and Systems, Princeton, NJ, pp. 568-570.

Tarko, A. (2003). "Highway Traffic Operations," Chapter 64, Section 7, Civil Engineering Handbook, 2nd ed., Eds. Chen, W. and Liew, J., CRC Press LLC.

Tomlin, C. (1998). Hybrid Control of Air Traffic Management Systems. PhD thesis, Department of Electrical Engineering, University of California, Berkeley.

Tomlin, C., J. Lygeros, and S. Sastry. (2000). "Controller design for hybrid systems." Proceedings of the IEEE, vol. 88, no. 7, July 2000.

Tomlin, C., Lygeros, J. and Sastry, S. (1998). Synthesizing controllers for nonlinear hybrid systems. In T. Henzinger and S. Sastry, editors, Hybrid Systems: Computation and Control, Lecture Notes in Computer Science 1386, pp.360–373. Springer Verlag.

Tong, C. (2008). "Refinement Strategies for Stratefied Sampling Methods," Reliability Engineering and System Safety.

Torng, T., and Yang, R. (1993). "Robust Structural System Design Using a System Reliability-Based Design Optimization Method," *in Probabilistic Structural Mechanics: Advances in structural reliability method*, P. Spanos and Y. Wu, eds. Springer-Verlag, Berlin, pp. 534-549.

TRB. (2005). Transportation Research Board Executive Committee Report. "Critical Issues in Transportation," Transportation Research Board of the National Academies.

Troncale, L. R. (1985). "The Future of General Systems Research: Obstacles, Potentials, and Case Studies." Systems Research 2(1): 43-84.

TSIS. (2008). Transportation System Integrated Software. version 6.1. McTrans Center. University of Florida.

Tu, J., Choi, K., and Park, Y. (2001). "Design Potential Method for Robust System Parameter Design," *AIAA J.*, 39, pp. 667–677.

USCG (2004). "About the FCS Program," The US Coast Guard, [http://www.boeing.com/defensespace/ic/fcs/bia/about.html], accessed March, 2004.

Valerdi, R., Ross, A., Rhodes, D. (2007). *A Framework for Evolving System of Systems Engineering*, CrossTalk-The Journal of Defense Software Engineering.

Wang, C. (1972). "On a Ramp-Flow Assignment Problem," Transportation Science, vol. 6, pp. 114-130.

Wang, J., and May, A. (1973). "Computer Model for Optimal Freeway On-Ramp Control," Washington, DC Highway Research Board, Highway Research Record, 469, pp. 16-25.

Wang, L. and Kodiyalam, S. (2002). "An Efficient Method For Probabilistic and Robust Design With Non-normal Distribution*" AIAA-2002-1754.* 43rd AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics, and Materials Conference, Denver, Colorado, Apr. 22-25.

Wardrop, J. (1952). "Some Theoritical Aspects of Road Traffic Research," Proceedings, Institution of Civil Engineers Part 2, 9, pp. 325-378.

Wattleworth, J., (1965), "Peak-Period Analysis and Control of a Freeway System," Washington, DC Highway Research Board, Highway Research Record, 157, pp. 1-21.

Wen, C. and Koppelman, F. (2001). The Generalized Nested Logit Model. Transportation Research Part B.

Wiener, N. (1938). "The homogeneous chaos", American Journal of Mathematics, vol. 60, pp. 897-936.

Winston, W. (1991). Operations Research, Applications and Algorithms, PWSKENT, Second Edition.

Wu, Y., Millwater, H., and Cruse, T. (1990). "Advanced Probabilistic Structural Analysis Method for Implicit Performance Functions," AIAA J., 28(9), pp. 1663-1669.

Wu, Y. and Wang, W. (1998). "Efficient Probabilistic Design by Converting Reliability Constraints to Approximately Equivalent Deterministic Constraints," *J. Integr. Des. Process Sci.*, 2, pp. 13–21.

Wu, Y., Shin, Y., Sues, R., and Cesare, M. (2001). "Safety Factor Based Approach for Probabilistic-Based Design Optimization," *in Proceedings of 42nd AIAA Structural Dynamics and Materials Conference*. AIAA-2001–1522.

Xu, H. (2004). A Catalogue of Three-Level Fractional Factorial Designs. Department of Statistics, University of California.

Yang, Q. and Koutsopoulos, H. (1996). A Microscopic Traffic Simulator for evaluation of dynamic traffic management systems. Transportation Research Part C: Emerging Technologies. Volume 4, Issue 3, June. pp. 113-129.

Yang, R. and Gu, L. (2004). "Experience with Approximate Reliability-Based Optimization Methods," *Struct. Multidiscip. Optim.,* 26, pp. 152–159.

Yazarel, H. and Pappas, G. (2004). Geometric programming relaxations for linear system reachability. In Proceedings of the 2004 American Control Conference, Boston, MA.

Youn, B. and Choi, K. (2004). "Selecting Probabilistic Approaches for Reliability-Based Design Optimization," AIAA J., 42(1), pp. 124-131.

Yu, J. and Ho, W. (2000). "Modified Sequential Programming for Feasibility Robustness of Constrained Design Optimization," DETC00/DAC-14531, Proceedings of DETC'00, Baltimore, MD, Sept. 10-13.

Yuan, L. and Kreer, J. (1971), "Adjustment of Freeway Ramp Metering Rates to Balance Entrance Ramp Queues," Transportation Research, vol. 5, pp. 127-133.

Zang, T., Hemsch, M., Hilburger, M., Kenny, S., Luckring, J., Maghami, P., Padula, S., and Stroud, W. (2002). *Needs and opportunities for uncertainty-based Multi-disciplinary design methods for aerospace vehicles*, NASA/TM-2002-211462 technical report series, Langley Research Center, Hampton, Virginia.

Zhang, Y. and Der Kiureghian, A. (1994). "Two Improved Algorithms for Reliability Analysis." Proceedings of the 6th IFIP WG 7.5 Working Conference on Reliability and Optimization of structural systems, R. Rackwitz, G. Augusti and A. Borri Eds., Chapman & Hall, New York, NY, pp. 297-304.

Zhao, Y. and Ono, T. (1999). "A General Procedure for First/Second-Order Reliability Method (FORM/SORM)," Structural Safety, 21, pp. 95-112.

Zou, T. and Mahadevan, S. (2006). "Versatile Formulation for Multiobjective Reliability-Based Design Optimization," Paper No. MD-05-1272, *ASME Journal of Mechanical Design*, in press.