Understanding & Modeling Bus Transit Driver Availability

MNTRC

MINETA NATIONAL TRANSIT
RESEARCH CONSORTIUM

RUTGERS

THE STATE UNIVERSITY
OF NEW JERSEY

MINETA TRANSPORTATION INSTITUTE
LEAD UNIVERSITY OF MNTRC

The Norman Y. Mineta International Institute for Surface Transportation Policy Studies was established by Congress in the Intermodal Surface Transportation Efficiency Act of 1991 (ISTEA). The Institute's Board of Trustees revised the name to Mineta Transportation Institute (MTI) in 1996. Reauthorized in 1998, MTI was selected by the U.S. Department of Transportation through a competitive process in 2002 as a national "Center of Excellence." The Institute is funded by Congress through the United States Department of Transportation's Research and Innovative Technology Administration, the California Legislature through the Department of Transportation (Caltrans), and by private grants and donations.

The Institute receives oversight from an internationally respected Board of Trustees whose members represent all major surface transportation modes. MTI's focus on policy and management resulted from a Board assessment of the industry's unmet needs and led directly to the choice of the San José State University College of Business as the Institute's home. The Board provides policy direction, assists with needs assessment, and connects the Institute and its programs with the international transportation community.

MTI's transportation policy work is centered on three primary responsibilities:

Research

MTI works to provide policy-oriented research for all levels of government and the private sector to foster the development of optimum surface transportation systems. Research areas include: transportation security; planning and policy development; interrelationships among transportation, land use, and the environment; transportation finance; and collaborative labor-management relations. Certified Research Associates conduct the research. Certification requires an advanced degree, generally a Ph.D., a record of academic publications, and professional references. Research projects culminate in a peer-reviewed publication, available both in hardcopy and on TransWeb, the MTI website (http://transweb.sjsu.edu).

Education

The educational goal of the Institute is to provide graduate-level education to students seeking a career in the development and operation of surface transportation programs. MTI, through San José State University, offers an AACSB-accredited Master of Science in Transportation Management and a graduate Certificate in Transportation Management that serve to prepare the nation's transportation managers for the 21st century. The master's degree is the highest conferred by the California State University system. With the active assistance of the California Department of Transportation, MTI delivers its classes over a state-of-the-art videoconference network throughout the state of California and via webcasting beyond, allowing working transportation professionals to pursue an advanced degree regardless of their location. To meet the needs of employers seeking a diverse workforce, MTI's education program promotes enrollment to under-represented groups.

Information and Technology Transfer

MTI promotes the availability of completed research to professional organizations and journals and works to integrate the research findings into the graduate education program. In addition to publishing the studies, the Institute also sponsors symposia to disseminate research results to transportation professionals and encourages Research Associates to present their findings at conferences. The World in Motion, MTI's quarterly newsletter, covers innovation in the Institute's research and education programs. MTI's extensive collection of transportation-related publications is integrated into San José State University's world-class Martin Luther King, Jr. Library.

REPORT 12-26

UNDERSTANDING & MODELING BUS TRANSIT DRIVER AVAILABILITY

Kaan Ozbay, Ph.D.
Ender Faruk Morgul, M.Sc.

July 2014

TECHNICAL REPORT DOCUMENTATION PAGE

1. Report No. CA-MNTRC-14-1140	2. Government Accession No.	3. Recipient's Catalog No.
4. Title and Subtitle Understanding & Modeling Bus Transit Driver Availability		5. Report Date July 2014
		6. Performing Organization Code
7. Authors Kaan Ozbay, Ph.D. and Ender Faruk Morgul, M.Sc.		8. Performing Organization Report MNTRC Report 12-26
9. Performing Organization Name and Address Mineta National Transit Research Consortium College of Business San José State University San José, CA 95192-0219		10. Work Unit No.
		11. Contract or Grant No. DTRT-12-G-UTC21
12. Sponsoring Agency Name and Address U.S. Department of Transportation Research & Innovative Technology Administration 1200 New Jersey Avenue, SE Washington, DC 20590 Edward J. Bloustein School of Planning and Public Policy Rutgers, The State University of New Jersey 33 Livingston Avenue New Brunswick, NJ 08901		13. Type of Report and Period Covered Final Report
		14. Sponsoring Agency Code

15. Supplemental Notes

16. Abstract

Bus transit agencies are required to hire extraboard (i.e. back-up) operators to account for unexpected absences. Incorrect sizing of extra driver workforce is problematic for a number of reasons. Overestimating the appropriate number of extraboard operators has financial implications while underestimating can lead to service disruption. It is therefore important that transit agencies properly manage extraboard operator staffing. A review of relevant literature showed that current models for extraboard management are generally agency-specific and that, in practice, extra driver assignments are usually based on the experience of the decision makers rather than the utilization of a mathematically sound modeling process.

In this study, two mathematical programming models with probabilistic constraints were developed to determine daily optimal extraboard size for bus transit (driver availability and deployment) while incorporating reliability and risk measures in the decision making process. Two distinct solution approaches were proposed. The first approach used pLEP's as the solution methodology and the second approach used second order stochastic dominance constraints. The models were tested using long-term data obtained from three Tri-County Metropolitan Transportation District of Oregon (TriMet) garage. The individual performance of both models under different cost assumptions was evaluated and then the actual historical assignments were compared with the optimal solutions obtained from these models. The results revealed possible improvements of extra driver size for one of the three garages studied. These models can be easily used in a computerized environment to assist agencies in efficient decision-making, which is also illustrated using a simulation procedure developed for comparison with observed driver assignment data.

17. Key Words Extraboard management; Stochastic modeling; Risk-averse behavior; Bus driver scheduling; Tactical workforce planning	18. Distribution Statement No restrictions. This document is available to the public through The National Technical Information Service, Springfield, VA 22161		
19. Security Classif. (of this report) Unclassified	20. Security Classif. (of this page) Unclassified	21. No. of Pages 46	22. Price $15.00

Form DOT F 1700.7 (8-72)

080614

ACKNOWLEDGMENTS

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

EXECUTIVE SUMMARY

To accommodate unplanned employee absences, caused by illness, or absenteeism while meeting the demands of scheduled service, transit agencies employ on-call backups known as "extraboard" operators. Overestimating the appropriate number of extraboard operators, however, comes at a cost. It is therefore important that transit agencies properly manage extraboard operator staffing, particularly since understaffing can lead to service issues.

This study has three major objectives:

1. To propose stochastic mathematical models that transit agencies can use to more efficiently determine the number of drivers needed for daily operations.

2. To test these models using real-world data from transit agencies and develop a series of recommendations that can be easily adopted by transit agencies wishing to reduce costs by optimizing their extraboard management.

3. To identify the potential benefits gained by using these models in different demand and supply scenarios under various reliability and customer service requirements. The models developed in this study will help transit agencies provide reliable and cost-optimized bus service. Unlike current practices, where decision makers determine the scope of the extra driver workforce using personal experience and intuition, the mathematical models developed in this study account for measures of risk and reliability with probability distributions based on historical data. Implementing these models could potentially allow agencies to realize meaningful cost reductions while properly allocating personnel.

The proposed models could also improve policies for daily transit operations, allowing agencies to better determine the minimum extra driver run hours for different levels of reliability while better understanding the relationship between social costs and operational costs. Implementing these models in a user-friendly computer tool could also lead to other improvements by creating scenarios to increase the speed and efficiency of decision making. Similar extraboard planning models were developed in Ozbay et al. (2012) and Morgul et al. (2013) for emergency evacuation operations. The major contributions of this study are as follows:

1. Unike the models in Ozbay et al. (2012) and Morgul et al. (2013), which were concerned with the number of drivers required for a successful evacuation operation, the models in this study deal with driver run hours for regular daily operations. Thus, this study provides modeling tools for daily operations as well as important insights in terms of the extraboard planning problem.

2. The models in this study are tested using real-world data; thus, the probability distributions are obtained from historical records. This provides an opportunity to compare the model estimations with the actual extraboard assignment practice of a major transit agency.

3. Social costs are defined using clearly identified measures, estimated for the case study area, such as the value of riding per hour and the average number of passengers. Social costs used in Ozbay et al. (2012) and Morgul et al. (2013), on the other hand, depended on assumptions regarding the value of life, injury costs, etc., which are very difficult to measure monetarily.

4. In this study, risk parameter effects are evaluated with a simulation tool, and comparison between cost differences and risk-neutral models are explained.

In summary, the major goal of this study is to develop and validate models to determine the optimal extraboard size for daily bus transit (driver availability and deployment) while incorporating reliability and risk measures in the decision-making process. The supply-and-demand data required for the model validation is obtained from historical data of the Tri-County Metropolitan Transportation District of Oregon (TriMet).

I. INTRODUCTION

Because of increasingly tight budgetary constraints, operational efficiency has become one of the top priorities for urban public transportation planning (Brons et al. 2005). Transit agencies must devise strategies to improve schedule reliability and on-time performance while ensuring the optimum usage of fleet size, the number of available drivers, and other limited resources. The function of cost-efficient, transit operator workforce planning is to guarantee that an adequate number of drivers is available for duty. Bus driver scheduling, or rostering, has been extensively analyzed in the literature (for a comprehensive review of developed models please see Ernst et al., 2004). This study, focuses on extraboard operator management, which deals primarily with the scheduling of backup drivers to temporarily cover unfilled assignments.

Open work or scheduled transit assignments that go unfilled may need to be evaluated, at minimum, on a daily basis (MacDorman, 1985). Therefore, timely decision making is crucial in determining the correct number of extraboard drivers. Two of the more common causes of open work are dispatcher errors and the unscheduled absence of a regularly assigned driver. During daily dispatching, a decision maker's poor judgment may result in uncovered work assignments (Perry and Long, 1984). In particular, underestimating a sufficient number of drivers jeopardizes service reliability and eventually results in missed trips. An unscheduled operator absence is the primary reason to bring on extraboard drivers (DeAnnuntis and Morris, 2008). As seen in several U.S. and European studies, average absenteeism levels for bus drivers are considerably higher than for other industry groups (Mulders et al.1982; Long and Perry, 1985; Kompier et al., 1990). A variety of reasons for bus driver absenteeism have been identified by earlier studies. These include injuries, vacations, and personal reasons. Perry and Long (1984) also noted that employee absence in transit operations is strongly related to worker compensation policies and operator job stress. In a U.S.-based study, Baker and Schueftan (1980) found that an average transit operator[1] misses approximately 12% of annual scheduled workdays, excluding vacation and holidays. Long-term absences are generally more predictable than short-term absences (i.e., 1 to 3 days), the latter being the primary reason an extraboard driver is hired (Strathman et al., 2009). Overestimating of the number of extraboard operators incurs significant costs since backup drivers are generally paid on a full-time basis, even when they might spend most of their shift on stand-by (Li and Gupta, 2012). Others have noted that absenteeism places a significant economic burden on an overall transportation budget[2] (Perry and Long, 1984).

Operator scheduling can be performed for various planning horizons. Koutsopoulos (1986) divides the transit workforce-planning framework into three levels:

Strategic planning: Systemwide decisions (i.e., overall workforce size or vacation allocations) take into account labor contracts and other policies. Planning horizons are considerably longer at this level (e.g., one year).

Tactical planning: Extra workforce numbers are determined on a daily basis, depending on the schedule requirements and garage assignments.

Operational planning: Shorter planning horizons (i.e. one hour) are considered at this level, and include determining how many extraboard operators are available when assigning specific time shifts. This level is the final duty assignment for available manpower.

In this study we consider the tactical planning problem. The models provided in the following section concern the optimum amount of extraboard work hours for a typical day.

II. LITERATURE REVIEW

The Extraboard management problem has been of interest to many researchers in both the transportation engineering and operations research fields. Most of the developed models used for crew scheduling have been influenced by demand modeling, days-off scheduling, or other mainstream modeling approaches. Using a statistical analysis based on survey data or historical crew assignment records, other studies have investigated the practice and actual performance of transit agencies in their extraboard operations.

DeAnnuntis and Morris (2007) conducted a comprehensive survey among transit agencies to determine the characteristics of extraboard management in the US. The authors noted the lack of effort on the part of agencies to record and evaluate historical rates of absenteeism. It was likewise found that automated scheduling software capable of assisting in the determination of the extraboard requirement in lieu of historical data was not commonly used by agencies. As a part of this study, the authors also reviewed the existing models and found that most of the models were tested only for specific transit systems. It is unclear if these models are applicable to other systems.

That study has reached two major conclusions: 1) In practice, extraboard sizing strongly depends upon a decision maker's personal experience, which can lead to inconsistent and inaccurate predictions. 2) Current models are not generally applicable to all transit systems, creating a major impediment to developing generic software that incorporates all models.

Strathman et al. (2009) investigated the short-term (1-3 days) unscheduled absences of transit operators using two years of attendance data from TriMet, the transit provider for Portland, Oregon. A statistical operator absence model was constructed with and without operator fixed effects (e.g., job satisfaction, absence-as-habit); and linear probability and logit estimation methods were used to determine the model parameter coefficients. The analysis showed that absence behavior is affected by personal characteristics, employment status, and seasonal factors, with the highest absenteeism probabilities occurring during winter months. The authors made suggestions for improvements to labor agreements that potentially could decrease short-term operator absences. Strathman et al. (2012) later analyzed the extraboard performance of three of TriMet's bus garages using long-term data (7 years). TriMet's extraboard assignments are distributed in a revolving system that assigns open work to operators in accordance with levels of seniority. These extraboard operators can turn down the assignment or trade it for a different shift. This variable revolving system can cause a significant volume of extraboard open work. The study noted that 31.8% of extraboard assignments were traded. On average, during the study period, 16% of the agency's scheduled service was open work that needed to be covered by extraboard operators.

From a mathematical modeling point of view, extraboard optimization problems are usually sensitive to the probability density functions that describe the unexpected open work caused by driver absences. Historical data is generally employed to define the probability distributions that are used to estimate the available number of drivers under different scenarios. In accordance with contract-defined days-off requirements, labor contracts

and fleet size are the most commonly used controlling parameters that constrain the total number of daily assigned drivers.

An early attempt to mathematically model an effective utilization of transit extraboard size was Koutsopoulos's (1990) study, which addressed tactical and operational problems for extraboard driver scheduling. At the tactical level, an integer programming formulation was developed to minimize the expected overtime (i.e., trips for which extraboard drivers are needed) using the days-off constraint for each driver in the workforce. Case study results for the Massachusetts Bay Transportation Authority (MBTA) showed the provided tactical model resulted in a 24% decrease per week in expected overtime.

Shiftan and Wilson (1994) offered a broad discussion on the relationship between service reliability and absenteeism with a focus on how they both affected optimal extraboard size. They developed a disaggregate model to understand the relationship between absenteeism and overtime and estimated the model using panel data from the MBTA. Their results suggested that reducing overtime did not adequately reduce absenteeism. Absenteeism remained unperturbed by assigning regular drivers extra time to cover most of the trips, which suggested that the only remaining solution might be enforcement via strict monitoring. Using aggregate data from the MBTA, the relationship between reliability (defined as the expected number of missed trips) and overtime was also investigated. A strong relationship was found between service reliability and overtime. The authors concluded that it is important to consider the reliability parameter in strategic extraboard management models.

In another study, Shiftan and Wilson (1995) proposed a strategic model to determine the optimal staff size while considering full-time and extra operators disjointly. To solve the workforce planning problem, the study proposed a two-stage heuristic algorithm that has an objective function of minimizing the total workforce cost for a minimum desired service reliability. The first stage estimated the optimal total staff size (by including extraboard as a function of the reliability constraint) for a given time period (e.g., a month). The second stage aimed to determine the annual hiring program and vacation day allocations. For that study, "extraboard personnel" was defined as the difference between the required work and the available manpower.

Recently, Li and Gupta (2012) developed a model that approaches the problem as a fixed job scheduling problem for which every defined job has a fixed start time and fixed end time, and each driver can be assigned to no more than one job at a time. Their model focuses on the operational level of extraboard workforce planning and assigning open work the day before the assignment. The objective function is to maximize the total duration of assigned jobs. The authors use a work-time constraint to control the maximum available time for an operator. In this study the optimization problem is NP-hard; three solution heuristics are provided and tested in a hypothetical example. Their model assumes that the amount of open work is known the day before and therefore fails to address unexpected absences that occur during the workday.

A review of the relevant literature highlights two important features of the extraboard management problem. First, that unexpected absences should be the primary focus when

developing workforce optimization problems. Scheduled absences can be anticipated, with sufficient lead time, to take necessary precautions. Unscheduled absences, however, are a major driver for open work uncertainty. The probabilistic approach based on historical performance records is therefore the best way to manage problems associated with unexpected absences. Second, during the decision-making process, the influence of personal experience—i.e., the personal intuition of the decision maker—should be minimized. This can be achieved by developing an automated process that 1) takes into account historical absenteeism trends and 2) utilizes computerized optimal extraboard size determination models. These models should rely on simple inputs to estimate the required extra driver workforce.

It's been common practice to use labor contracts as a primary constraint. However, every transit agency has its own specific regulations. Therefore, in a global tactical level probabilistic model, the probability distribution of driver availability should also take into account operators' day-off requirements.

In this present report, the research contributes to the literature by using stochastic programming methods to address the uncertainty resulting from unexpected operator absences. Developing a global model applicable to any bus transit system is one of the main objectives of this study. The models in this report deal with the hierarchy of the tactical-level work scheduling. Therefore, the decision variable selected was the optimal number of extraboard driver hours, based on the historical driver availability data. To preserve a certain degree of operational efficiency, the models also take into account the quality of service, considered as a measure of reliability. The objective function of the problems (i.e., the function this study attempts to optimize) was to minimize total costs incurred by extraboard management, which include out-of-pocket agency costs as well as social costs incurred by customers who use transit vehicles. The models were therefore designed to assess the trade-off between driver costs and improved on-time performance. Two different models were proposed to determine the optimum number of extra drivers for a predefined quality-of-service requirement. It was assumed that the quality-of-service requirement has been determined by the transit agency.

III. DATA

TriMet's daily extraboard operations data from three bus garages (Center, Merlo, and Powell) for the period between 2002 and 2010 was employed for the case studies in this report. The data consists of 8,317 garage-day observations. TriMet is one of the largest mid-size transit agencies in the US. In 2012 it served, on average, 193,800 boarding rides each weekday. An extensive summary along with the statistical analysis of the dataset can be found in Strathman et al. (2012). As seen in Table 1, for all three garages, approximately 16% of the scheduled service could not be covered by the regularly assigned drivers, which resulted in open work. The assigned extraboard driver hours were generally higher than open work hours, indicating an overstock in driver supply. To represent the time losses from extraboard hours, extraboard absence hours and extraboard report hours are given together, accounting for approximately 28% of the assigned extraboard hours.

Table 1.   TriMet Daily Extraboard Profile, September 2002 to May 2010

	Center	Merlo	Powell	Total
Platform hours of scheduled service	2,427.9	1,313.9	1,953.1	5,694.9
Platform hours of open work (% of scheduled service hours)	384.2 15.8	205.7 15.7	302.8 15.5	892.7 15.7
Extraboard operator platform hours	444.7	241.0	335.1	1,020.8
Extraboard operator absence hours (% of extraboard operator platform hours)	52.2 26.6	31.3 28.3	47.9 28.4	131.4 27.6

Source: Strathman et al., 2012.

Table 2.   TriMet Daily Reliability

	Center	Merlo	Powell	Total
Platform hours of scheduled service	2,427.9	1,313.9	1,953.1	5,694.9
Platform hours of actual service (% of scheduled service hours)	2,602.9 107.2%	1,411.6 107.4%	2,071.4 106%	6,085.9 106.9%
	Jan 2013		Jan 2012	
Bus on-time performance[3]	81.2%		81.0%	

According to the TriMet work rules, scheduled driver assignments are made for a 3-month period. Extraboard drivers cover the open work that results from regular driver absences, retirements or injuries. When open work exceeds the available extraboard assignments, regular drivers are asked to voluntarily work on their days off. Once a driver signs up to participate in the extraboard workforce, it is guaranteed that the driver will be paid for eight pay hours daily. The extraboard assignment system currently in practice in TriMet is a complex revolving system called the Red Line. In this system, seniority level is used as the decisive factor to sort extraboard operators; thus the first task in a new assignment period is given to the most senior driver. Thereafter, the red line moves through the extraboard roster, ensuring that no driver consistently receives only good or bad work assignments. In addition, extraboard drivers have the option to pass up or trade their assignments

based on certain conditions. For example, drivers can trade/pass-up assignments when their previous day's clock-out time is less than nine hours. Pass-ups can contribute to the variability in open work, particularly when there is no available extraboard driver to take on the passed-up shift. When this occurs, regular day-off drivers are called for voluntary extraboard work, but this can lead to an overstock of operators and added driver cost burden for the agency.

IV. PROBLEM FORMULATION

MODELING STRATEGY

In this report mathematical optimization models are utilized to assign the correct number of additional employees to open work. Optimization models are one of the most efficient ways to determine the best usage of limited resources under certain conditions such as budget limitations or worker contracts. We use stochastic models to account for the uncertainty and randomness of the open work. Stochastic models have the advantage of using probabilistic constraints that account for randomness of possible outcomes. Historical data can be used to determine the probabilities of expected daily open work, and necessary precautions for minimizing open work can be taken accordingly using these stochastic models.

There are stochastic optimization methods that have proven to result in to the optimum solution of a probabilistic problem. In this report we employ two such methods and determine the optimal extra-driver work hours for a daily operation based on historical driver availability data. In this context, driver availability means the expected amount of regularly assigned driver hours for a typical workday, based on historical data. There is also a scheduled amount of driver run hours included in these models, and the difference between scheduled driver run hours and available regular driver run hours is defined as open work, which needs to be covered by extra driver hours.

Stochastic models require an objective function that is a criterion with respect to which the model is optimized. Objective function in the models in this report is a linear sum of two cost terms related to driver assignment 1) driver costs and 2) social costs. Detailed descriptions of these costs are given in the next sub-section and in the Case Study section. Additionally, stochastic models are controlled by several constraints. In this report we use probabilistic constraints for representing the availability of regularly assigned drivers, which are derived from probability distributions of historical data. Probability distributions give information about what percent of time a decision maker should expect what amount of regular driver hours. For example, based on historical data, a probability distribution can tell 90% of the time that 1000 driver run hours are expected to be filled by regularly assigned drivers. If 1100 driver run hours are scheduled for that day, a decision can be made for the extra 100 driver run hour for this specific case, depending on the level of risk tolerance. These models can, for example, accommodate a risk-averse decision maker who would prefer to assign more than 100 drivers, just to be on the safe side. A second controlling parameter is the quality of service, which is a measure of reliability. Using this parameter, and depending on the service reliability preferences, a decision maker can calculate different extra driver run hours by running different scenarios. Therefore, each possible option is tried in the objective function, and the optimum result is selected as the outputs of the models. In summary, open work is estimated with the help of this probability distribution, and decision is made using service quality and cost considerations.

An important question is how these models are useful for transit agencies, or how transit agencies can make the best use of the outputs from these models. As discussed and demonstrated in detail in the Case Study section, these models need two types of inputs:

1) historical records of driver assignments and 2) costs and social impacts associated with extra drivers. As long as these inputs are provided, any transit agency can use these models to assist their decision-making process. The output of these models can be used as a supplement to the existing driver assignment strategies. Moreover, as different levels of risk tolerance yield different set of results, decision makers can make better-informed decisions based on these models. The models are not necessarily meant to replace existing strategies but to support the driver assignment process and ultimately save costs resulting from over/under estimation of extra driver need.

The following subsection provides the mathematical formulations. For more detailed information about solution approaches please see the Appendix.

MODEL STRUCTURE

Extraboard management is a challenging problem because of the uncertain nature inherent in open work. Often, the regular driver workforce might not be able to fill the run hours scheduled for daily operations. Therefore, the size of the scheduled fleet of extraboard, or backup, drivers depends on the availability of regular drivers available for duty. The probability that a regular driver will show up for assigned duty can be estimated using the historical data of scheduled driver run hours and actual regular driver run hours. The term "actual regular driver run hours" is used to mean the observed agency performance of regularly assigned operators; it can differ significantly from the scheduled driver run hours. Caused by the absence of regularly assigned full-time operators, open work hours are covered by extraboard drivers and, in some cases, by regular drivers working on their days off (Strathman, 2009). The difference between scheduled driver run hours and work hours covered by this extra workforce yields an estimate for regular driver availability.

In this report, the extraboard management problem is formulated as a stochastic problem for different levels of service quality. The term "quality of service" refers to the reliability of the service, a flexible measure that allows agency decision makers to take into account various probabilities of outcomes. The agencies' risk tolerance may be the determining factor when employing an adequate extraboard management strategy. Therefore, in this report, we present two different approaches to model decision-making, risk-tolerant behavior. Similar extraboard planning models were developed in Ozbay et al. (2012) and Morgul et al. (2013) for emergency evacuation operations. Our models modify these models to address the daily operational needs of a transit agency. Moreover, the availability of real-world data from Portland TriMet makes it possible to test the hypothetical models developed in Ozbay et al. (2012) and Morgul et al. (2013) using realistic probability distributions.

Parameters that are common in both of our developed models are given as follows:

N : Number of garages

M_i : Number of realizations of actual regular driver run hours at garage $i \in \{1,...,N\}$

R_i : Random variable representing actual regular driver run hours at garage $i \in \{1,...,N\}$. It has the following distribution: $P(R_i = R_{ij}) = p_{ij}, \ j \in \{1,...,M_i\}$

D_i: Daily demand for driver run hours at a garage i

c : Cost of unsatisfied demand per hour ($/hour)

c_i^x: Hourly extraboard driver cost ($/hour)

The decision variable used in the models is as following:

x_i: Extraboard driver run hours at a garage i

z_{ij}: Unsatisfied demand when available driver run hours are R_{ij} with probability p_{ij}

Model 1

Model 1 aims to determine the optimal number of extra driver hours for each garage, depending on the probability distributions for regular driver availabilities (i.e., the difference between scheduled run hours and extra-driver run hours). The model incorporates risk-averse behavior by way of a systemwide quality-of-service parameter. Decision makers' individual risk tolerances are not captured in this model.

Model 1 is defined as:

$$\text{Minimize } c\sum_{i=1}^{N}\sum_{j=1}^{M_i} z_{ij} p_{ij} + \sum_{i=1}^{N} c_i^X x_i \tag{1}$$

Subject to:

$$P(D - x \le R) \ge q_{service} \tag{2}$$

$$z_{ij} \ge D_i - R_{ij} - x_i, \qquad i \in \{1,...,N\}, j \in \{1,...,M_i\} \tag{3}$$

$$z_{ij} \ge 0, \qquad\qquad i \in \{1,...,N\}, j \in \{1,...,M_i\} \tag{4}$$

$$x_i \ge 0, \text{ and integer, } i \in \{1,...,N\} \tag{5}$$

where:

R : N dimensional random vector with R_i at i^{th} entry

D : N dimensional vector with D_i at i^{th} entry

x : N dimensional vector with x_i at i^{th} entry

p_{ij} : Probability that region i has R_{ij} regular drivers available. This parameter is obtained from a probability distribution $P(R_i = R_{ij}) = p_{ij}$, $j \in \{1,...,M_i\}$

$q_{service}$: Systemwide quality of service determined by decision makers

The objective function (1) has two terms that represent cost. The first term is the total expected cost for the unsatisfied demand. The second term is the driver costs created by the hired extra operator workforce. Unsatisfied demand generates an indirect cost that is quite difficult to quantify monetarily. One way to determine the costs associated with service disruptions is to consider the riding per -hour value passengers are willing to pay for a bus trip.

The second cost term in the objective function is a direct cost that arises when employing extra drivers, who derive both salaries and other benefits. The first cost parameter provides the model with flexibility so that various scenarios can be considered using different kinds of costs, such as ticket price or average hourly wages. The second cost parameter is more fixed, in accordance with the prevailing labor contracts.

The model therefore aims to optimize the number of extraboard driver hours by considering the cost burden for the operating agency. The term "quality of service" is included in the constraint (2), with the flexibility to determine the optimal extraboard scheduling. This is key for decision makers evaluating different possible situations. Here, the quality of service is defined as systemwide. When its value is equal to one, the constraint (2) will be satisfied with 100% certainty, thus becoming a deterministic constraint. Constraints (3) - (4), with the objective function force z_{ij}, are equal to the unsatisfied demand, which is $\max\{0, D_i - R_{ij} - x_i\}$. Unsatisfied demand cost is assumed to be the cost incurred by all bus line passengers who suffer because their transit vehicle was late. Since the model considers driver run hours, the average value of the passengers' is the major input in the unsatisfied demand cost.

Model 1 is solved using the p-level efficient point (pLEP) technique. Prekopa et al. introduced the pLEP algorithm for a given discrete distribution (1998), an efficient method for calculating the Pareto frontiers (p-efficient points). Employed in several studies for modeling evacuation operations, the p-lep method has been used to measure the probabilistic effects of road capacity constraints on shelter capacities for a hurricane evacuation (Yazici and Ozbay, 2007); evacuation network modeling (Yazici and Ozbay, 2010); and inventory management for disasters (Ozguven and Ozbay, 2013). For a thorough background on the development of the probabilistic constraint model, see A. Charnes et al. (1958), Miller and Wagner (1965), Prekopa (1970), Prekopa (1973), and Prekopa (1990).

Model 2

Solving Model 1 can be computationally difficult when the number of garages is large. Model 2 utilizes second-order stochastic dominance constraints, which allow decision makers to obtain a random outcome from an optimization problem that is at least as good as another random outcome. In this case, the random outcome stochastically dominates the previously given probabilistic reference outcome.

The stochastic dominance concept is defined for different orders. Figure 1 is a graphical depiction of the concept of stochastic dominance as seen previously in Seog (2010). If higher values are preferable, first-order stochastic dominance ensures that an alternative Y dominates another alternative X when the cumulative distribution function of Y (denoted by

F_Y) never lies above the cumulative distribution of $X(F_X)$, where F denotes the cumulative distribution function. However, first-order dominance can sometimes be insufficient in explaining risk-averse behavior. Therefore, higher-order stochastic dominance models are used to distinguish among the different risk attitudes of decision makers (Nie et al., 2011). In such instances, integrable random variables are used and risk-averse decisions are carried out, depending on the areas under the curve. In Figure 1b, Y dominates X in the second-order if the area under F_X is larger than the area under F_Y for any W_0, or in other words, $A1 - A2 \geq 0$.
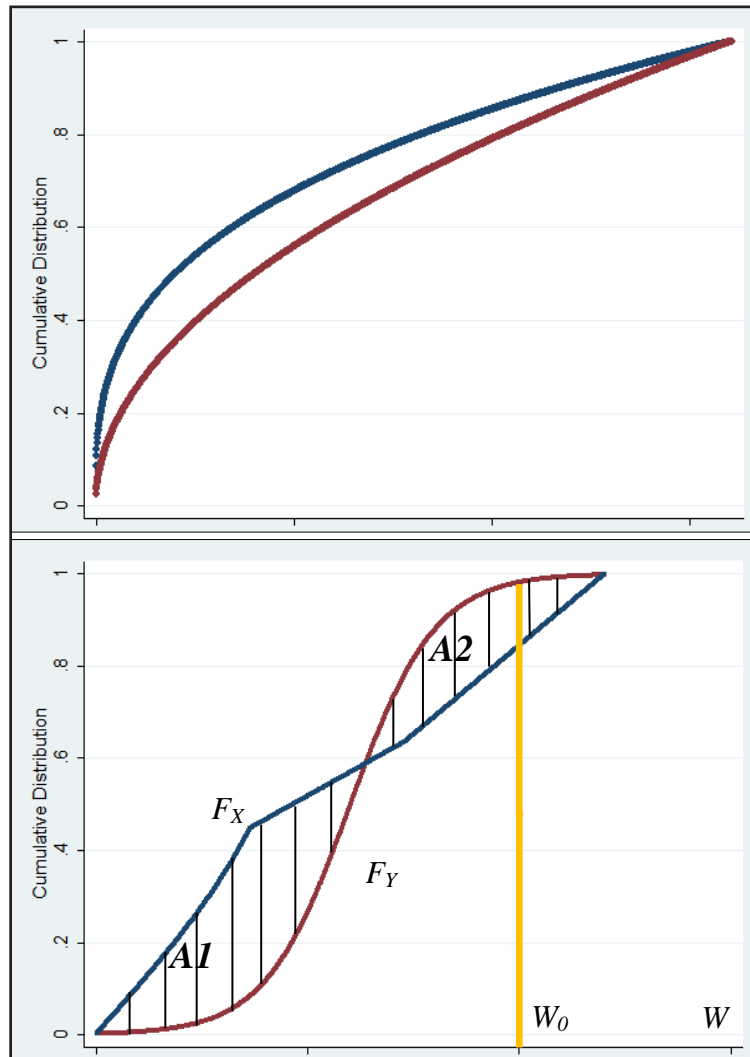
Figure 1. (a) First-Order Stochastic Dominance (b) Second-Order Stochastic Dominance

Source: Seog, 2009.

Muller and Stoyan (2002) give a comprehensive analysis of stochastic dominance relationships. If every risk-averse decision maker prefers an option Y to X (higher values are preferable for both), given that they have increasing concave utility functions, then Y dominates X, with respect to second-order stochastic dominance, and is denoted as $Y \geq_{(SSD)} X$. In this report we prefer to use $Y \geq_{(2)} X$, another commonly used notation in the

literature. The increasing convex order (icx) is a dual concept of second-order stochastic dominance, which defines an ordering relation when lower values of X and Y are preferable. $Y \leq_{(icx)} X$ is used to represent when   is less than or equal to X in the sense of the increasing convex order. It is known that $Y \leq_{(icx)} X$ if and only if $-Y \geq_{(SSD)} -X$ (see Muller and Stoyan (2002) for details).

Although a popular method used throughout the statistics and finance literature, the stochastic dominance concept remains largely unexplored within the area of transportation research. There are few examples in the transportation literature featuring the SSD concept. Two examples include the optimal path problem formulated with SSD constraints for analyzing risk-averse driver behavior (Nie et al., 2011) and an extraboard management problem for transit-based emergency evacuation operations (Ozbay et al., 2012; Morgul et al., 2013).

Dentcheva and Ruszczynski (2003, 2004) introduced stochastic optimization problems with dominance constraints and studied the optimality and duality conditions. Model 2, which uses second-order stochastic dominance constraints, is defined as:

$$\text{Minimize } c\sum_{i=1}^{N}\sum_{j=1}^{M_i} z_{ij} p_{ij} + \sum_{i=1}^{N} c_i^X x_i \tag{6}$$

Subject to:

$$-Z_i \geq_{(2)} -Y_i, \qquad i \in \{1,...,N\} \tag{7}$$

$$z_{ij} \geq D_i - R_{ij} - x_i, \qquad i \in \{1,...,N\}, j \in \{1,...,M_i\} \tag{8}$$

$$z_{ij} \geq 0, \qquad i \in \{1,...,N\}, j \in \{1,...,M_i\} \tag{9}$$

$$x_i \geq 0, \text{ and integer } i \in \{1,...,N\} \tag{10}$$

where:

$P(Z_i = z_{ij}) = p_{ij}$: Demand for driver run hours equal to z_{ij} is not satisfied with probability p_{ij}

$P(Y_i = y_{il}) = f_{il}$: Reference distribution for unsatisfied driver run hour demand at garage i, $l \in \{1,...,H_i\}$

The objective function of Model 2 is the same as the objective function of Model 1. First, the constraint (7) is introduced to find a distribution for Z_i, which is at least as good as the reference distribution. Therefore, the reference distribution should be one that is acceptable to not only the modelers but also the practitioners. Assume that the unsatisfied run hour demand for the region i has the following distribution when no extraboard drivers are hired:

$P(U = u_{il}) = f_{il}, i \in \{1,...,N\}, l \in \{1,...,H_i\}$ where H_i is the number of realizations for the unsatisfied run hour demand. Define $q_{service}^i$ for garage $i \in \{1,...,N\}$.

It corresponds to an improvement for the unsatisfied run hour demand for a garage $i \in \{1,...,N\}$.

For the reference distribution of unsatisfied run hour demand we assume:

$$P(Y = y_{il} = \lceil (1 - q_{service}^i)u_{il} \rceil) = f_{il} \tag{11}$$

Here, $\lceil a \rceil$ denotes the smallest integer greater than or equal to a. Since the unsatisfied run hour demand should come from a discrete distribution, $(1 - q_{service}^i)u_{il}$ is rounded up to the nearest integer. A similar calculation for reference distribution can be found in Noyan (2010), which uses second-order stochastic dominance constraints to model the risk aversion in an emergency response facility location and allocation problem.

Additional constraints can also be included in the models. In real-world applications there is always a budget constraint, which can be used as a model constraint as such:

$$\sum_{i=1}^{N} c_i^X x_i \leq UB_{budget} \tag{12}$$

Another constraint can be used to limit the total number of extraboard driver run hours, which is the sum of all driver run hours for all garages:

$$\sum_{i=1}^{N} x_i \leq UB_{total\ driver} \tag{13}$$

V. CASE STUDY

In this section we present an implementation of the developed models using real-world data from TriMet (see the previous section for a description of the TriMet data). This case study is useful for measuring model performance under different cost scenarios. Several assumptions for agency out-of-pocket costs and costs imposed on society are needed for a realistic interpretation. Actual observed driver hours and payments to the drivers are provided in the dataset; the hourly average payment is calculated as $22.9/hour. Note, however, that because of benefits and other driver compensations the average value may be an underestimate of the actual hourly out-of-pocket cost to the transit agency. Therefore, the proposed models are solved using three different increasing driver cost values: $22.9/hour, $50/hour, and $75/hour.

Li and Bertini (2009) estimated riding time value to be between $5/hr and $8/hr for a TriMet Bus system user. As a part of the same study of the distribution of passenger boardings and alightings they also found the number of average persons per trip to be 33.2. Therefore, a fair estimation for the unsatisfied demand cost would be around $1200 ($8/hour times 40 passengers with four full trips from a garage). To test the sensitivity of the model, we also tried alternative social cost values: $800, $2200, $4400, and $6600.

For the case study we used data from three TriMet garages: Center, Powell, and Merlo. The locations of the garages are shown in Figure 2. Center and Merlo garages are closer to the city center, while the Powell garage is located in the eastern section of downtown Portland. In this study it is assumed that cost structures, both in driver and social costs are the same for all garages. The same driver cost assumption can be considered reasonable for a realistic interpretation since all garages are operated by the same agency. Depending on the socio-demographic structures of the analysis zones, the social costs associated with unsatisfied travel demand can differ between regions. For example, a missed trip in the morning peak hour in a business district could yield a greater social cost than a missed shopping trip in the same time period. Therefore the routes and passenger characteristics should be examined carefully. The research team left this kind of social cost sensitivity analysis to future studies. In this study, we are mainly concerned with how the model performs in different hypothetical scenarios.
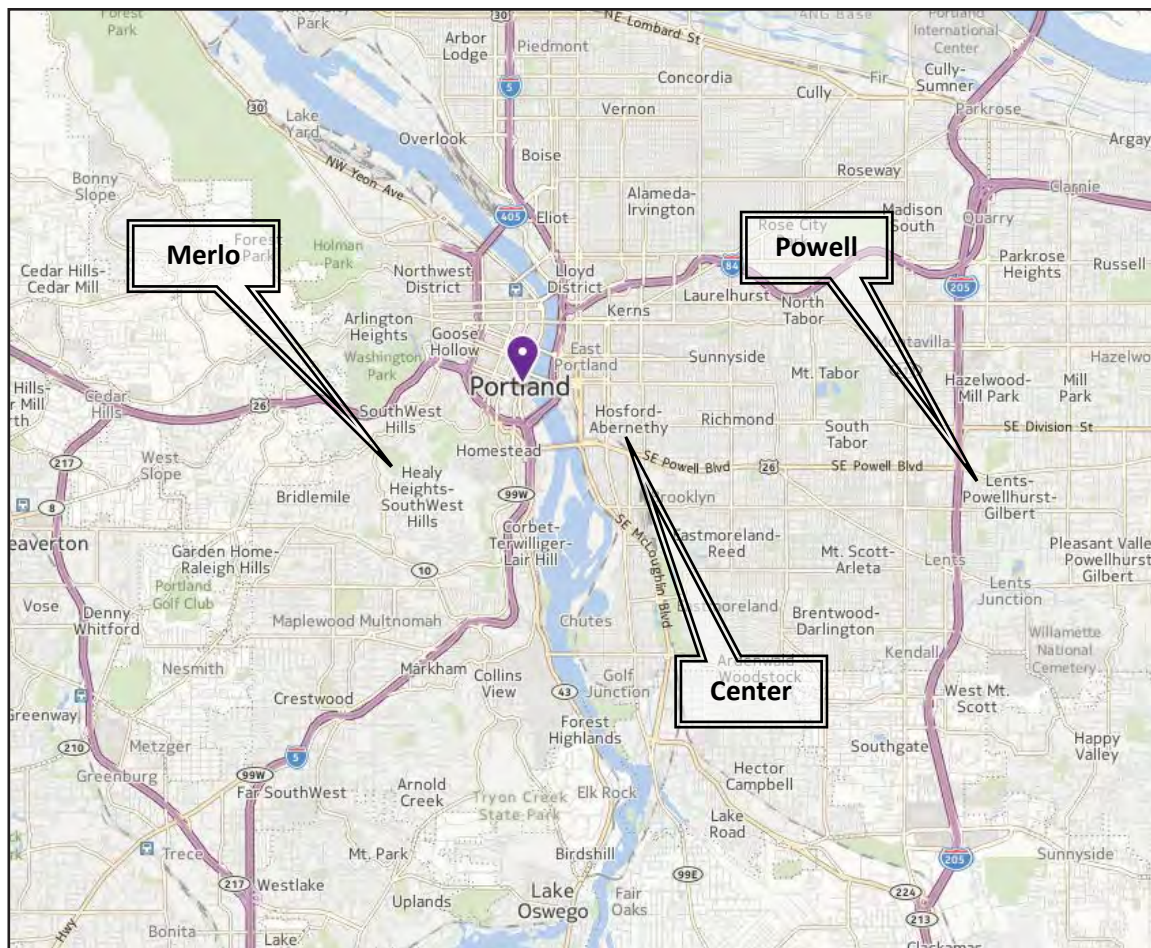
Figure 2.   Case Study Garage Locations

Source: Bing Maps, 2013.

Developed models try to estimate the optimal extraboard run hours; therefore, all the inputs are given in run hour terms. Daily run hour demands are assumed to be deterministic; an average of the 7-year data is used for the fixed demand. Only weekday demand is considered for this analysis; holidays and weekends are excluded. Figure 3 shows the optimal extra driver run hours obtained from the model outputs for the levels of 90%, 80%, and 70% quality of service.
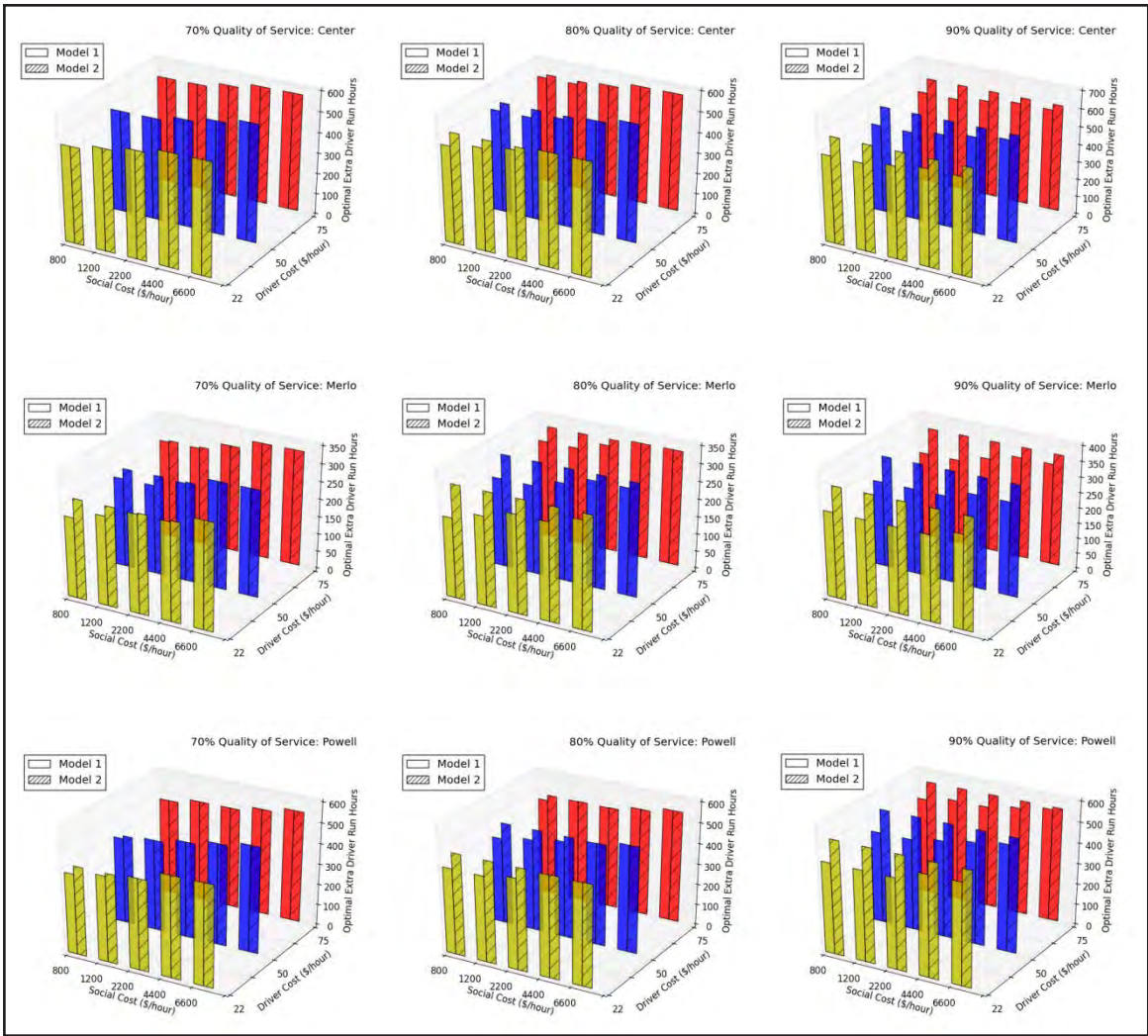
Figure 3.   Model Results for Different Cost Scenarios

For the 90% quality-of-service level, the optimal number of extra driver run hours is always greater in Model 2 than in Model 1. Compared to Model 1, Model 2 yields more conservative results for all cost scenarios. The risk-averse behavior of the two models is completely different. With the goal of reaching the highest quality of service, Model 2 uses second-order dominance constraints to handle risk while keeping the estimated extra driver run hours as high as possible. In other words, for this scenario, the quality-of-service parameter in Model 2 can be more restrictive.

Model 1 yields various outputs for different cost scenarios while Model 2 is not sensitive to the changes in cost. For the scenarios with higher social costs, the gap between the estimated total agency costs for both models (i.e. the sum of the driver costs and unsatisfied demand costs) becomes smaller, and both models yield similar results.

At the 80% quality-of-service level, the outputs of both Model 1 and Model 2 become more sensitive to the changes in costs. For two of the cost scenarios (driver cost: $22.9/hour; Social cost: $4400 and $6600), both models yield the exact same optimal result. Both of these scenarios assume a higher social cost relative to the other scenarios.

Likewise, at the 70% quality-of-service level, seven tested scenarios yield the same optimal results in both models—all of which assume higher social costs. This may mean that, for the scenarios with low quality-of-service level, the risk-neutral model may give the same optimal solution with the risk-averse models. It should be noted that in the scenarios where both models provide the same results, the social costs are significantly higher than the driver costs.

As expected when the quality-of-service levels decrease, both models estimate a higher need for extra drivers to satisfy the desire for higher quality. In this way, neither model is clearly superior to the other. The results suggest that Model 2 behaves more conservatively in almost all of the scenarios. When implemented into the real world, both models offer decision makers an opportunity to examine a variety of scenarios, ranging from the worst-case scenarios all the way to the most optimistic. The results in this report demonstrate that the developed models have the ability to respond to various cost scenarios and generate reasonable outputs.

EFFECT OF RISK CONSTRAINTS

Accounting for risk measures in the decision-making process is the main objective of the developed models. As such, the research team tested the effect of risk constraints in the models and compared the results with the risk-neutral model, which can be obtained by eliminating the probabilistic constraint from Model 1.
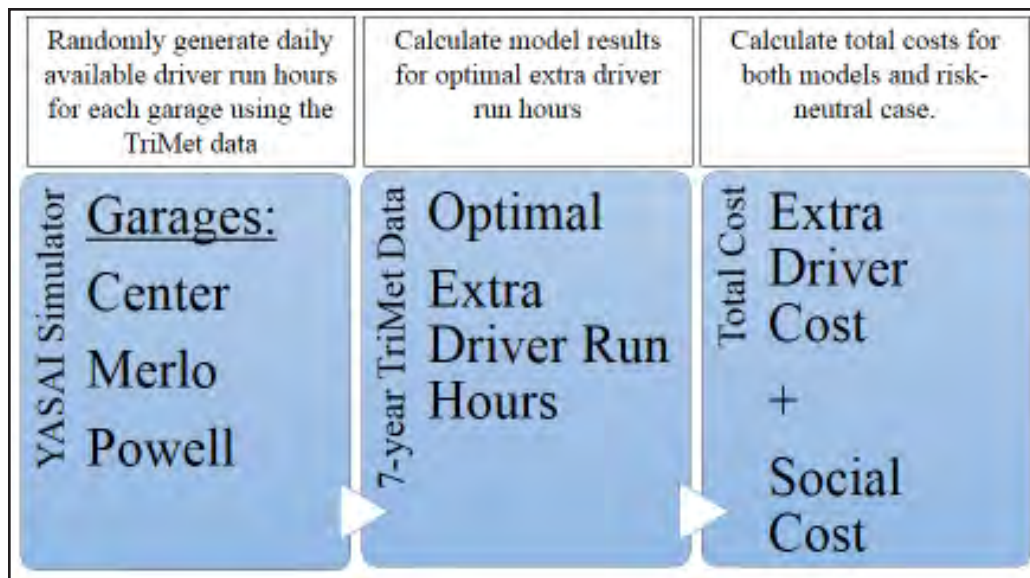
Figure 4.   Simulation Flowchart

Figure 4 shows the flowchart for the following simulation procedure. The probability distribution of actual driver hours from the 7-year TriMet data was used for this analysis. Random actual driver hours were generated in a spreadsheet using the YASAI simulation tool (Eckstein and Riedmueller, 2001). Three cases were compared: both Model 1 and Model 2 as risk-averse models, and a risk-neutral model that excluded the probabilistic constraint (Eqn.2) from Model 1. Driver cost was assumed to be $75/hour; the social cost was assumed to be $800/hour; and the extra driver run hours were calculated for the 90% quality-of-service level. There is no specific reason for selecting this cost scenario for this section, as similar results are obtained from other tested cost scenarios. Using a fixed daily demand assumption for each garage, the total daily costs were calculated for approximately 80,000 simulation draws.
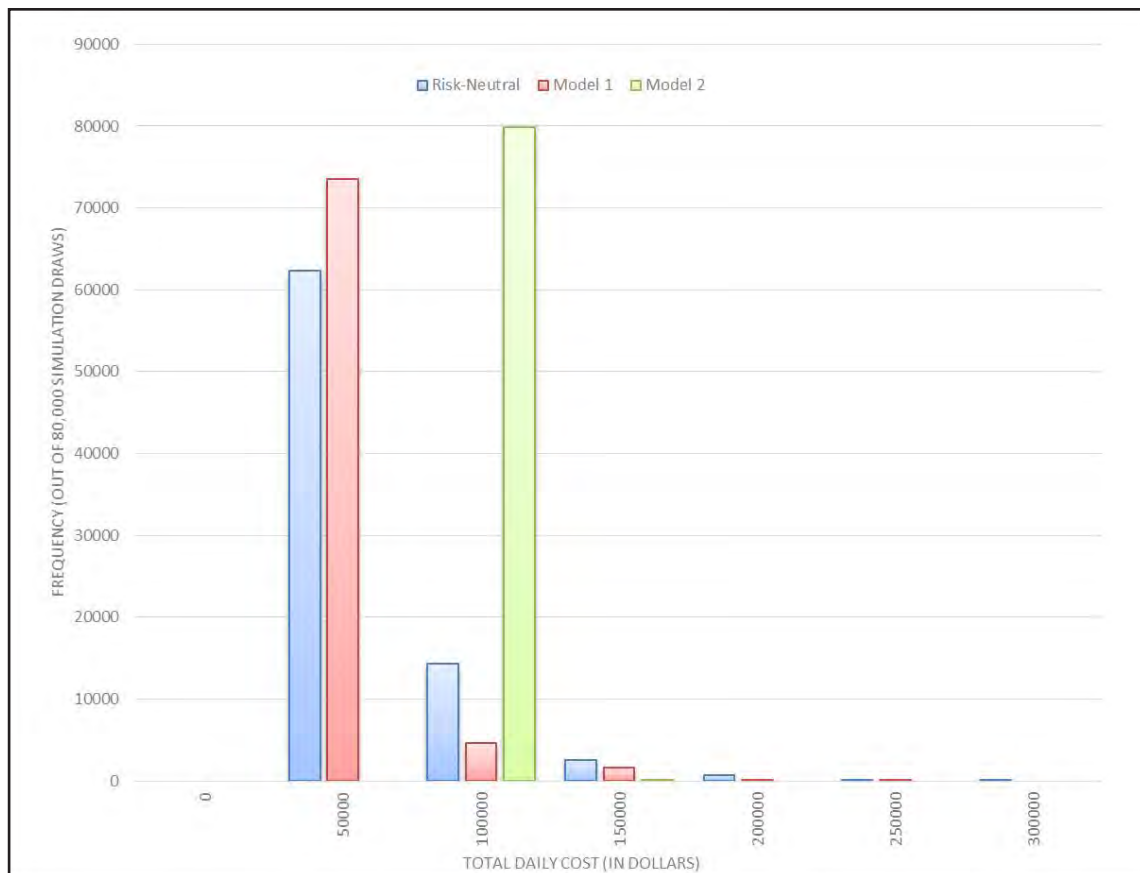
Figure 5. Histograms of Total Daily Costs for Risk-Averse and Risk-Neutral Models

Along with the risk-neutral case, the histograms of the total costs for the two risk-averse models are depicted in Figure 5. As expected, the risk-averse solutions require a higher expected total cost, estimating more extra driver run hours to meet the service quality requirement. The total costs in Model 2 are the highest among the three models for most of the simulation draws. This result is also consistent with the previously presented results. The results indicate that the variance of the total cost is smaller in risk-averse models, although expected cost is lower in the risk-neutral model.

TRIMET ACTUAL PERFORMANCE EVALUATION

In this final case, the actual extraboard assignments observed in the TriMet's 7-year data are evaluated using the developed models. Only weekday and non-holiday assignments are evaluated since the models are constructed based on the probability distribution of the driver availability for these days. To illustrate scheduling performance, the overestimated numbers are compared to the optimal solutions obtained from the presented models. According to the data, underestimating the extra drivers is less of an issue since the actual service hours are generally higher than the scheduled service hours. To solve the models with the actual daily driver run hour demands observed in the data, an automated procedure is generated using the 'gdxrww' tool in the R software package and GAMS optimization software. Table 3 shows the percentage of overestimated days using different quality-of-service levels. At all

quality-of-service levels, the Center garage extra-driver assignments are higher by at least 8.7% than the optimal solutions obtained from Model 1. This could be an indication of the extra costs associated with extra drivers that spend their workday idle. At the 90% quality-of-service level, a potentially desirable option for better service, the overestimation rates generally fall between 0-1% for all garages except Center. This shows that the observed extra driver assignments are close to the optimal solutions from the presented models. Consistent with the results in the previous section, Model 2 produces more conservative results than Model 1 for actual observed data.

Table 3.   TriMet Performance Evaluation: Overestimated Extra Drivers

Quality-of-Service Level	Center		Merlo		Powell	
	Model 1	Model 2	Model 1	Model 2	Model 1	Model 2
70%	11.40%	11.42%	7.4%	0.51%	4.50%	1.22%
80%	11.30%	6.58%	7.4%	0.20%	2.20%	0.20%
90%	8.70%	0.71%	0.6%	0.20%	0.70%	0.20%

EFFECT OF QUALITY-OF-SERVICE PARAMETER

In this study the quality-of-service parameter accounts for the different characteristics of a bus transit system that impact the reliability of the service. As previously documented in the literature, on-time performance is an important measure of service reliability. However, other factors also contribute to on-time performance, some of which, like congestion or weather, cannot be controlled by the agency (Strathman and Hopper, 1993). Equally difficult to control are the quality determinants as perceived by passengers. Based on a stated preference survey, Hensher et al. (2003) reports that bus frequency and seating availability are among the significant factors that passengers consider when evaluating system performance. Both service frequency and seating capacity can, to some extent, be related to driver availability, since driver shortage or inefficient workforce assignments can cause missed trips. At the same time, if bus frequency is high, one missed scheduled trip may not be a significant factor since passengers can simply catch a ride with the next, soon-to-arrive bus. These and other factors make one thing clear: bus transit service reliability is both complex and difficult to measure.

The quality-of-service measure presented in this study addresses the overall reliability for which driver availability is one component that can be controlled by the agency. A 100% quality-of-service assumption would not be realistic, considering both the external and internal factors that affect service quality. Agency-specific quality-of-service definitions could be the best way to incorporate reliability measure in the developed models. For practical purposes, different quality-of-service levels can be taken into account to evaluate different risk-taking scenarios. Likewise, comparative optimal solutions can be used for these various situations. For example, a high quality of service can be used for conservative practices, such as planning for a snowy day; while a low quality of service means a higher risk.

Finally, we incorporated a simulation-based analysis using a new set of 80,000 draws from actual driver-demand data and calculated the optimal driver run hours. Probability distributions for driver availability are calculated for each garage from the entire 7-year dataset. The procedure used for the simulation is also an example of automated, real-world extraboard size estimation. Daily demand for drivers, cost, quality-of-service levels, and other required inputs can be modified for the tactical crew scheduling process, which might occur the day before the assignment day. For practical purposes, the distributions can also be updated using the most recent data, and different distributions can be used for weekdays, weekends, or for holidays and other special days of the year. Figure 6 shows the cumulative distributions for optimal extra driver run hour assignments for the three garages. The graphs show that higher quality-of-service levels result in more risk-averse assignments for all cases.
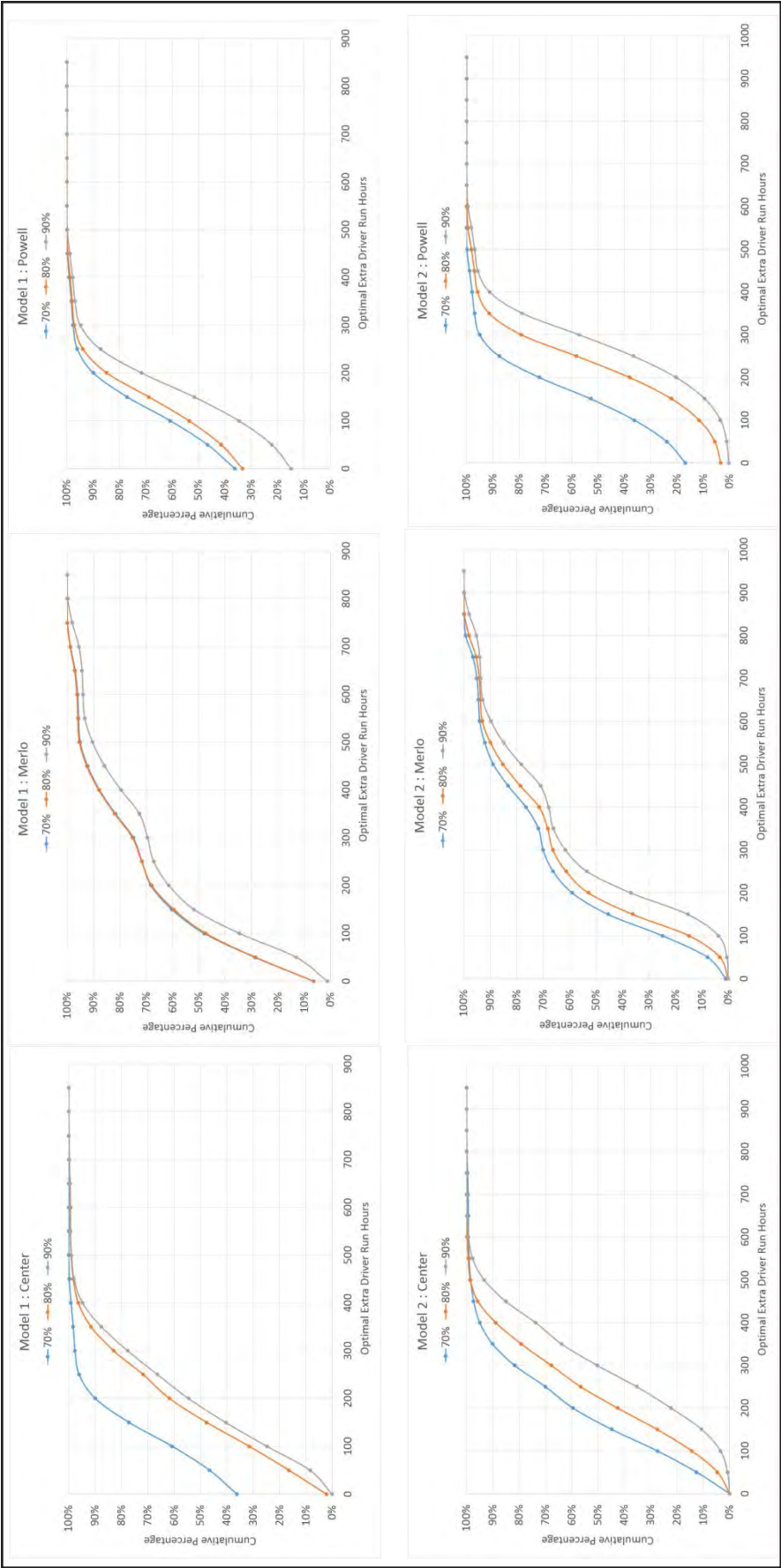
Figure 6. Simulation-Based Model Results

VI. CONCLUSION

This report presents a tactical extraboard planning and management problem for daily transportation operations. Two different stochastic programming models, with solution approaches, are developed. Using historical driver absenteeism data, the models are developed to estimate the optimum extra driver run hours. Using a quality-of-service parameter, the models also address risk tolerance. Both agency out-of-pocket driver costs and social costs associated with unsatisfied passenger demand are taken into account in the models.

The proposed models are tested using Portland's TriMet transit 7-year performance data. The results of the case study show that the models can be useful tools for real-life applications. Likewise, the output obtained from the models can help decision makers assess available options under various stochastic constraints. For larger-scale implementations, the developed models also provide the flexibility of assigning different quality-of-service (i.e., reliability) levels to different garages. Such flexibility enhances the scalability of the developed models, making them applicable for transit agencies of various sizes. The research effort presented in this report is an attempt to address the extraboard management problem from a stochastic point of view, using available driver numbers as a probabilistic parameter. This feature can adequately capture the uncertainty associated with the unexpected absenteeism of regularly assigned drivers. One potential improvement for the model's passenger demand parameter is to turn the demand into a probabilistic variable. Although this will add complexity to the calculations, modified versions of the provided solution approaches can also be developed. Another potential improvement would be to incorporate learning parameters based on historical records, with day-to-day updated model inputs (such as demand for drivers). Such a learning concept could be easily implemented in a computer application. Moreover, it could potentially generate probability distributions using the most recent data and develop various if-then scenarios using the mathematical models, ultimately yielding comparative results for advanced decision making.

VII. APPENDIX

SOLUTION APPROACHES

This study presents two solution approaches to the extraboard management problem formulated in the previous section. The two approaches differ in the following ways:

1) In the first solution approach, an algorithmic solution method is utilized to calculate the Pareto frontiers of the probability distribution of the available regular driver run hours. Probabilistic constraint of the stochastic programming model is addressed by the pLEP (p-Level efficient point) method (Prekopa, 1990).

2) In the second solution approach, a reference random probability distribution is calculated using the available regular driver run-hour data. The solution is ensured to be at least as good as the reference distribution using the second-order stochastic dominance constraints.

pLEP METHOD

Since the models in this study have to work with discrete random variables (e.g., number of available regular driver run hours), we have focused on stochastic programming problems with discrete distributions. Discrete stochastic programming problems can be formulated using a r-dimensional discrete random vector $\xi \in R^r$, $x \in R^n$ and $r \times n$ matrix T. Adding the probabilistic constraint stating that $Tx \geq \xi$ should hold with probability $p \in (0,1)$, the general formulation is as follows:

$$\text{Minimize } c^T x$$

$$\text{Subject to } P(Tx \geq \xi) \geq p \qquad \qquad \text{(A1)}$$

$$Ax \geq b$$

$$x \geq 0$$

Probability distributions of regular driver availability for scheduled run hours are assumed to be discrete. The pLEP algorithm for a given discrete distribution, which was introduced by Prekopa et al. (1998), is an efficient method for calculating the Pareto frontiers (p-efficient points). If z is a realization of ξ and F is the distribution function of ξ, then the point z is called a pLEP of the probability distribution F, if $F(z) \geq p$ and there is no $y \leq z$, where $y \neq z$ such that $F(y) \geq p$ and $p \in (0,1)$ (Prekopa, 1990).

If there are N pLEPs with $z^{(i)}$ representing the ith point, then an optimal solution of problem (A1) can be obtained by solving the following problem (Prekopa et al., 1998):

$$\text{Minimize } c^T x$$

$$\text{Subject to } Tx \geq z^{(i)}, \text{ for at least one } i \in \{1,...,N\} \qquad \text{(A2)}$$

$$Ax \geq b$$

$$x \geq 0$$

To find the optimal solution for problem (A2), A) we can solve for each pLEP and then select the one that provides the best objective function value or B) we can solve the following optimization problem:

$$\text{Minimize } c^T x$$

$$\text{Subject to } Tx \geq \sum_{i=1}^{N} \lambda_i z^{(i)}$$

$$\sum_{i=1}^{N} \lambda_i = 1 \qquad \text{(A3)}$$

$$\lambda_i \in \{0,1\}, \qquad i \in \{1,...,N\}$$

$$Ax \geq b$$

$$x \geq 0$$

Note that, as indicated in Prekopa et al. (1998), if the constraints $\lambda_i \in \{0,1\}$ in formulation (A3) are relaxed, then we obtain a relaxation of problem (A2).

If the problem has three pLEP points in a two-dimensional space, then the graphical representation of the region that T_x should fall in is given in Figure 1a.

In the case at hand, a small modification must be made to the classical pLEP formulation (Prekopa, 1990). Since in Model 1 the constraint (3) means that the demand has to be smaller than the sum of the regular and extra drivers, the formulation given in (A2) becomes as follows:

$$\text{Minimize } c^T x$$

$$\text{Subject to } P(Tx \leq \xi) \geq p \qquad \text{(A4)}$$

$$Ax \geq b$$

$$x \geq 0$$

We can modify the first constraint of this formulation to calculate the correct pLEP as follows:

$$P(\overline{T}x \geq \overline{\overline{\xi}}) \geq p \qquad\qquad (A5)$$

where $\overline{T} = -T$ and $\overline{\overline{\xi}} = -\xi$.

If, $Z^{(i)}$s are the negative of the pLEPs corresponding to (A5) then the graphical representation becomes what can be seen in Figure 2b.
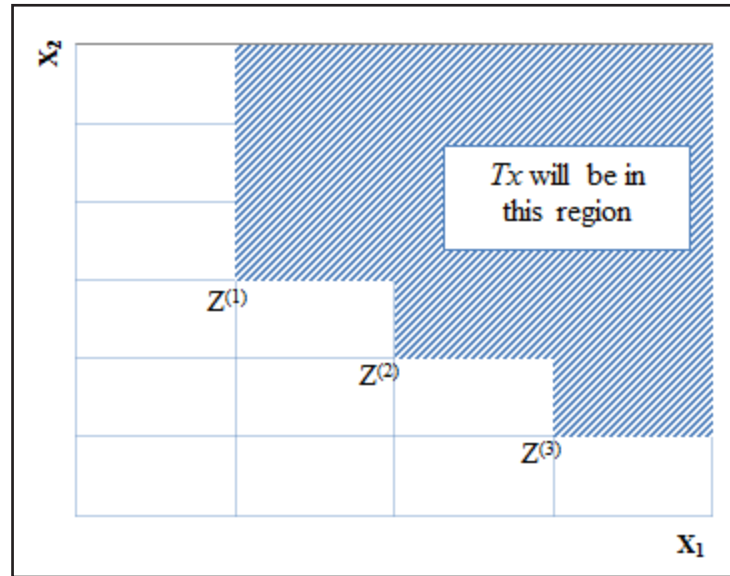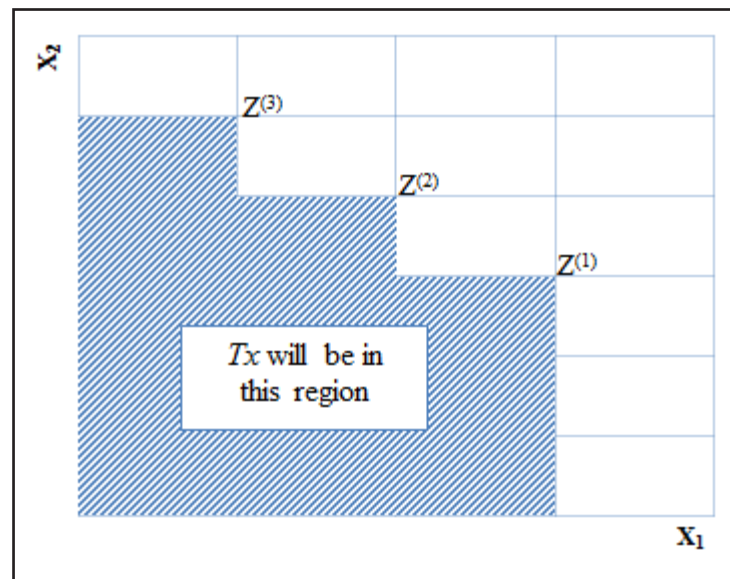

Source: Dentcheva et al., 2000.

Figure 7. (a) Theoretical P-level Efficient Points (b) Modified P-level Efficient Points

Second-order Stochastic Dominance

This study suggests a solution methodology for Model 2 that addresses the risk by using second-order stochastic dominance constraints. The dominance constraints require a reference distribution for the unsatisfied demand. Unlike in Model 1, in Model 2 we are interested in finding a distribution of unsatisfied run hour demand which is at least as good as the reference distribution. Luedtke (2008) introduced a linear formulation for SSD constrained optimization problems. Suppose W is a random variable for which higher values are preferable; the realizations of this variable are given as w_i, and Y is a reference random variable with realizations y_k such as:

$$P(W = w_i) = p_i, \quad i \in \{1,...,N\}$$
$$P(Y = y_k) = q_k, \quad k \in \{1,...,Q\}$$

Where P denotes the probability. Then $W \geq_{(2)} Y$ if and only if there exists $\pi \geq 0$ which satisfies;

$$\sum_{j=1}^{Q} y_i \pi_{ij} \leq w_i, \quad i \in \{1,...,N\}$$

$$\sum_{j=1}^{Q} \pi_{ij} = 1, \quad i \in \{1,...,N\}$$

$$\sum_{i=1}^{N} p_i \pi_{ij} = q_k, \quad k \in \{1,...,Q\}$$

Following Theorem 3.2 of Luedtke (2008), we obtain the formulation in (A6) - (A9) for the dominance constraints (7):

$$\sum_{l=1}^{H_i} y_{il} w_{ikl} \geq z_{ik}, \quad i \in \{1,...,N\}, k \in \{1,...,M_i\} \tag{A6}$$

$$\sum_{l=1}^{H_i} w_{ikl} = 1, \quad i \in \{1,...,N\}, k \in \{1,...,M_i\} \tag{A7}$$

$$\sum_{k=1}^{M_i} p_{ik} w_{ikl} = f_{il}, \quad i \in \{1,...,N\}, l \in \{1,...,H_i\} \tag{A8}$$

$$w_{ikl} \geq 0, \quad i \in \{1,...,N\}, k \in \{1,...,M_i\}, l \in \{1,...,H_i\} \tag{A9}$$

BASIC DEFINITIONS

Objective Function: In an optimization problem, the main purpose is to choose the best one of the many acceptable alternatives. The criterion with respect to which a model is optimized is known as the objective function.[4]

Pareto Frontiers: This is the framework for partially evaluating a set of "actions" with multi-dimensional outputs assuming a very weak "desirability" partial ordering which only applies only when one process is better (or at least as good) for all the outputs.[5]

Stochastically Dominance: Stochastic dominance (SD) is a fundamental concept in decision theory with uncertainty. It is a form of stochastic ordering. It describes when a particular random prospect, say a lottery, is better than another random prospect based on preferences regarding outcomes expressed in terms of monetary or utility values. In other words, "*when a decision maker can choose between two alternatives with certain prospects, he chooses one which leads to higher profits. When the actions are uncertain, decision maker will choose the one promising higher prospects, too. We assume every eligible action and its consequences can be reflected by a random variable and its outcomes. Hence it is aimed to rank random variables reflecting a decision maker's preferences. Such a ranking can be achieved by stochastic order relations.*"[6]

Stochastic Ordering: This is the concept used to determine if one random variable is "bigger" than the other one.

Risk Aversion: This is a concept used in finance or economics to describe a behavior where consumers when faced with uncertainty try to minimize this uncertainty. This behavior can be represented with the concept of second-order stochastic dominance.

Second-Order Stochastic Dominance: If there are two lotteries (gambles or probability distributions of outcomes) A and B then A is said to have a second-order stochastic dominance over lottery B, if A involves less stochastic risk than B with at least as high a mean as B.

ENDNOTES

1. The authors use the term "operator" to refer to bus drivers, train drivers and other operational personnel onboard the vehicle.

2. Baker and Schueftan (1980) estimated that more than one-fourth of the U.S. federal subsidy for transit is spent on operator absence.

3. January 2013 TriMet Monthly Performance Report. Available online at http://trimet.org/pdfs/publications/performance-statistics/2013-01.pdf, accessed on 7/2/2013.

4. Singiresu S. Rao, S. S. Rao, "Engineering Optimization: Theory and Practice", John Wiley & Sons, 2009.

5. http://www.azimuthproject.org/azimuth/show/Pareto+front, accessed on 3/19/2014.

6. Neise, F., "Risk Management in Stochastic Integer Programming: With Application to Dispersed Power Generation", Springer, 2008.

BIBLIOGRAPHY

Baker, Scott H. and Oliver Schueftan. *Study of Operator Absenteeism and Workers' Compensation Trends in the Urban Mass Transportation Industry.* Pittsburgh, PA, USA: Final Report UMTA-PA-06-0050-80-1, 1980.

Bing Maps. http://www.bing.com/maps/ (accessed December 17, 2013).

Brons, Martijn, Peter Nijkamp, Eric Pels, and Piet Rietveld. "Efficiency of Urban Public Transit: A Meta Analysis." *Transportation 32*, (2005): 1–21.

Charnes, Abraham, William W. Cooper, and G. H. Simonds. "Cost Horizons and Certainty Equivalents: An Approach to Stochastic Programming of Heating Oil." *Management Science 42*, (1958): 35–263.

DeAnnuntis, Christopher P. and William P. Morris. "Transit Extraboard Management: Optimum Sizing and Strategies." 2007. Tallahassee, FL: Florida Department of Transportation Report No: BD549-23/NCTR 777-07. http://www.nctr.usf.edu/pdf/77707.pdf (accessed October 30, 2013).

DeAnnuntis, Christopher P. and William P. Morris. "Transit Extraboard Management Optimum Sizing and Strategies." *Transportation Research Record 2072*, (2008): 110–124.

Dentcheva, Darinka, Andras Prekopa, and Andrzej Rusczynski. "Concavity and Efficient Points of Discrete Distributions in Probabilistic Programming." *Mathematical Programming*, (2000): 55–77.

Dentcheva, Darinka and Andrzej Ruszcyzynski. "Optimization with Stochastic Dominance Constraints." *SIAM Journal on Optimization 14-2*, (2003): 548–566.

Dentcheva, Darinka and Andrzej Ruszczynski. "Optimality and Duality Theory for Stochastic Optimization Problems with Nonlinear Dominance Constraints." *Mathematical Programming 99*, (2004): 329–350.

Eckstein, Jonathan and Steven T. Riedmueller. "YASAI: Yet Another A Simulation Add-In for Teaching Elementary Monte Carlo Simulation in Excel." *Piscataway, NJ: Rutgers Research Report*. 2001. Simulator available online at http://www.yasai.rutgers.edu/ (accessed December 18, 2013).

Ernst, Andreas T., Houyuan Jiang, Mohan Krishnamoorthy, and David Sier. "Staff Scheduling and Rostering: A Review of Applications, Methods and Models." *European Journal of Operational Research 153*, (2004): 3–27.

GAMS Corporation, G.D., 2013. Gams gdx facilities and tools. http://www.gams.com/dd/docs/tools/gdxutils.pdf.

Hensher, David A., Peter Stopher, and Philip Bullock. "Service quality-developing a service quality index in the provision of commercial bus contracts." *Transportation Research Part A: Policy and Practice 37*, (2003): 499–517.

Kompier, Michiel, A.J. Mulders, Helmy, Meijman, Theo F., M. Boersma, Georg Groen, R. Bullinga, "Absence Behavior, Turnover and Disability: A Study Among City Bus Drivers in the Netherlands." Work and Stress 4, (1990): 83–89.

Koutsopoulos, Harilaos Nikolaos. "A Methodology for Transit Operator Workforce Planning." *Ph.D. Thesis, Department of Civil Engineering, MIT*, (1986).

Koutsopoulos, Harilaos Nikolaos. "Scheduling of Extraboard Operators in Transit." *Transportation Science 24*, (1990): 87–104.

Li, Fei and Diwakar Gupta. "The Extraboard Operator Scheduling and Work Assignment Problem." *Industrial and Systems Engineering Dept., Univ. of Minnesota*. 2012. http://www.isye.umn.edu/labs/scorlab/pdf/OFJSWS.pdf (accessed October 30, 2013).

Li, Huan and Robert L. Bertini. "Assessment of an Optimal Bus Stop Spacing Model Using High Resolution Archived Stop-Level Data." 88th Annual Meeting of the Transportation Research Board. Washington, DC, 2009.

Long, Lyn and James L. Perry. "Economic and Occupational Causes of Transit Operator Absenteeism: A Review of Research." *Transport Reviews: A Transnational Transdiciplinary Journal 5: 3*, (1985): 247–267.

Luedtke, James. "New Formulations for Optimization under Stochastic Dominance Constraints." *SIAM Journal on Optimization 19*, 3, (2008): 1433–1450.

MacDorman, Littleton C. "Extraboard Management: Procedure and Tools." *National Cooperative Transit Research Program, Synthesis of Transit Practice, Transportation Research Board 5*, (1985).

Miller, Bruce L. and Harvey M. Wagner. "Chance-constrained Programming with Joint Constraints." *Operations Research 13*, (1965): 930–945.

Morgul, Ender Faruk, Ozlem Cavuz, Kaan Ozbay, and Cem Iyigun. "Modeling of Bus Transit Driver Availability for Effective Emergency Evacuation in Disaster Relief." *Transportation Research Record: Emergency Evacuation and Transportation System Vulnerabilities 2013*, (2013).

Mulders, Helmy, Meijman, Theo F., O'Hanlon, James, F., and Mulder, Greg. "Differential Psychophysiological Reactivity of City Bus Drivers." Ergonomics 25, (1982): 1003–1011.

Muller, Alfred and Dietrich Stoyan. *Comparison Methods for Stochastic Models and Risks.* Wiley Series in Probability and Statistics *1*, 2002.

Nie, Yu, Xing Wu, and Tito Homen-de-Mello. "Optimal Path Problems with Second Order Stochastic Dominance Constraints." *Networks and Spatial Economics*, (2011): 1–27.

Noyan, Nilay. "Alternate Risk Measures for Emergency Medical Service System Design." *Annals of Operations Research 181-1*, (2010): 559–589.

Ozbay, Kaan, Ender Faruk Morgul, Ozlem Cavus, and Cem Iyigun. "Understanding & Modeling Bus Transit Driver Availability during Emergencies for the Effective Evacuation of Populations without Automobile Access." Final Report submitter to The Transportation Coordinating Council / Federal Transit Administration (TCC/FTA) (2012).

Ozguven, Eren Erman and Kaan Ozbay. "A Secure and Efficient Inventory Management System for Disasters." *Transportation Research Part C: Emerging Technologies*, (2013): 171–196.

Perry, James L. and Lyn Long. "Extraboard Scheduling, Workers' Compensation, and Operator Stress in Public Transit: Research Results and Manegerial Implications." *Transportation Research Record 1002*, (1984): 21–28.

Prekopa, Andras. "On Probabilistic Constrained Programming." *In Proceedings of the Princeton Symposium on Mathematical Programming*, (1970): 113–138.

Prekopa, Andras. "Constributions to the Theory of Stochastic Programming." *Mathematical Programming 4*, (1973): 202–221.

Prekopa, Andras. "Dual Method for the Solution of a One-Stage Stochastic Programming Problem with Random RHS Obeying a Discrete Probability Distribution." *Zeitschrift für Operations Research 34*, (1990): 441–461.

Prekopa, Andras, Bela Vizvari, and Tamas Badics. *Programming Under Probabilistic Constraint with Discrete Random Variables.* New 43 Trends in Mathematical Programming, Kluwer Academic Publishers, 235–257. (1998).

Seog, S. Hun, *The Economics of Risk and Insurance.* John Wiley ad Sons Publication *1*, 2010.

Shiftan, Yoram and Nigel H. M. Wilson. "Absence, Overtime, and Reliability Relationships 22 in Transit Workforce Planning." *Transportation Research Part A: Policy and Practice 28-3*, (1994): 245–258.

Shiftan, Yoram and Nigel H. M. Wilson. "Public Transport Workforce Sizing Recognizing 25 the Service Reliability Objective." *Computer-aided Scheduling of Public Transport*, (1995): 248–266.

Strathman, James G. and Janet R. Hopper. "Empirical Analysis of Bus transit On-time Performance." *Transportation Research Part A: Policy and Practice 27*, (1993): 93–100.

Strathman, James G., Joseph Broach, and Steve Callas. "2009." *Transportation Research Record 2111*, (2009): 185–194.

Strathman, James G., Sung-Moon Kwon, and Steve Callas. "Extraboard Performance: TriMet Case Study." *Transportation Research Record 2274*, (2012): 61–68.

Yazici, Mustafa Anil and Kaan Ozbay. "Impact on Probabilistic Road Capacity Constraints on the Spatial Distribution of Hurricane Evacuation Shelter Capacities." *Transportation Research Record 2022*, (2007): 55–62.

Yazici, Mustafa Anil and Kaan Ozbay. "Evacuation Network Modeling via Dynamic Traffic Assignment with Probabilistic Demand and Capacity Constraints." *Transportation Research Record 2196*, (2010): 11–20.

ABOUT THE AUTHORS

KAAN OZBAY, Ph.D.

Kaan M.A. Ozbay is a professor at the Department of Civil and Urban Engineering at Polytechnic Institute of New York University and Center for Urban Science and Progress (CUSP). Dr. Ozbay's research interest in transportation covers advanced technology and sensor applications, incident and emergency management, development of real-time control techniques for traffic, traffic safety, application of artificial intelligence and operations research techniques in network optimization, development of simulation models for transit and automated highway systems, and transportation economics. He is co-editor of numerous books, including, most recently, *Dynamic Traffic Control & Guidance* published by Springer Verlag's "Complex Social, Economic and Engineered Networks" series in 2013. He has published more than 300 refereed papers in scholarly journals and conference proceedings. Prior coming to NYU, Dr. Ozbay was a tenured full Professor at the Rutgers University Department of Civil and Environmental Engineering. Since 1994, Dr. Ozbay, has been the principal investigator and co-principal investigator of 77 projects funded at a level of more than $11,00,000 by National Science Foundation, NJDOT, NYMTC, NY State DOT, New Jersey Highway Authority, USDOT, FHWA, VDOT, CUNY University Transportation Research Center (UTRC), Department of Homeland Security, and USDOT ITS Research Center of Excellence.

ENDER FARUK MORGUL, M.Sc.

Ender F. Morgul works as a graduate research assistant at the Civil and Urban Engineering Department in Polytechnic Institute of New York University. He received his B.S. from Bogazici University and M.Sc. in Civil Engineering from Rutgers University. His research interests include modeling and prediction of driver behavior, transportation economics, urban freight operations, big data and GPS-based transportation data analysis. His 2010 M.Sc. thesis investigated dynamic congestion pricing using large-scale traffic simulations.

PEER REVIEW

San José State University, of the California State University system, and the MTI Board of Trustees have agreed upon a peer review process required for all research published by MNTRC. The purpose of the review process is to ensure that the results presented are based upon a professionally acceptable research protocol.

Research projects begin with the approval of a scope of work by the sponsoring entities, with in-process reviews by the MTI Research Director and the Research Associated Policy Oversight Committee (RAPOC). Review of the draft research product is conducted by the Research Committee of the Board of Trustees and may include invited critiques from other professionals in the subject field. The review is based on the professional propriety of the research methodology.

MNTRC
★★★★★★★★★
MINETA NATIONAL TRANSIT
RESEARCH CONSORTIUM