CALIFORNIA PATH PROGRAM
INSTITUTE OF TRANSPORTATION STUDIES
UNIVERSITY OF CALIFORNIA, BERKELEY

Developing Optimal Planning and Management Strategies for a Robust Highway System

Yafeng Yin
Samer Madanat

California PATH Research Report
UCB-ITS-PRR-2005-35

This work was performed as part of the California PATH Program of the University of California, in cooperation with the State of California Business, Transportation, and Housing Agency, Department of Transportation, and the United States Department of Transportation, Federal Highway Administration.

Final Report for Task Order 5329

November 2005

ISSN 1055-1425

CALIFORNIA PARTNERS FOR ADVANCED TRANSIT AND HIGHWAYS

Developing Optimal Planning and Management Strategies for a Robust Highway System

Yafeng Yin and Samer M. Madanat
California PATH Program
University of California at Berkeley

Final Report for Task Order 5329

October 5, 2005

ACKNOWLEDGEMENTS

This work was performed by the California PATH Program at the University of California at Berkeley in cooperation with the State of California Business, Transportation and Housing Agency, Department of Transportation (Caltrans). The contents of this report reflect the views of the authors, who are responsible for the facts and the accuracy of the data presented herein. The contents do not necessarily reflect the official views or policies of the State of California.

The authors thank their colleagues, Xiao-Yun Lu and Kenneth Kuhn for their contributions to this work and their support of this project. The authors acknowledge Nathan Loebs of Caltrans' Division of Research and Innovation for his support of the project.

ABSTRACT

The report attempts to deliver a proof of concept that optimal planning and management strategies can be formulated through applying robust optimization methodology such that limited resources could be allocated more rationally, and reliability of a highway network improved more efficiently.

The report focuses on two applications: the network design problem under demand uncertainty, and fleet allocation for freeway services patrols. Their corresponding decision-makings are formulated as several optimization models. By solving these models robust optimal strategies can be obtained. Numerical examples and simulation tests are presented to demonstrate the validity and usefulness of the proposed models.

The report has proved that through applying robust optimization methodology, robust optimal improvement strategies can be obtained to improve substantially the capacity of a highway system against high-consequence scenarios incurred by fluctuations in travel demand or irregular incidents.

Key words: highway planning and management, robust optimal strategies

EXECUTIVE SUMMARY

Fluctuations in travel demand and irregular incidents give rise to recurrent and non-recurrent congestion delay on highway segments and thus make the system performance unstable and unreliable. The report attempts to deliver a proof of concept that optimal planning and management strategies can be formulated through applying robust optimization methodology such that limited resources could be allocated more rationally, and reliability of a highway network improved more efficiently.

Robust optimization is a modeling methodology to solve optimization problems in which the data are uncertain and only known to belong to some uncertainty set. The approach is to seek optimal (or near optimal) solutions that are not overly sensitive to any realization of uncertainty.

The first-half of the report addresses the network design problem under demand uncertainty, namely, subject to a given budget, determining which links from a network need improvement (capacity increase) and deciding how much budget should be allocated to the links so as to maximize the improvement of the performance of the network. Three models, sensitivity-based, scenario-based and min-max, are proposed for determining robust optimal improvement schemes that can improve the performance of a network efficiently while at the same time allowing it to remain “close” to the designated level under any realization of the uncertain demand. Numerical examples and simulation tests are presented to demonstrate the validity and usefulness of the proposed models. It is suggested that if decision makers aim to achieve a mean-variance tradeoff, the sensitivity-based and scenario-based models should be used with particular caution placed on the minimization of the sensitivity of total travel time with respect to demand perturbations or minimization of the variance of total travel times under various demand scenarios. If fluctuations of travel demand are believed to be non-significant, the sensitivity-based model is more appropriate to use, because it is simpler and requires much less computation efforts. Otherwise, the scenario-based model should be used, and additional efforts are needed to generate the demand scenarios and determine the corresponding probabilities of occurrence. On the other hand, if decision makers are more concerned with the worst-case scenarios, the min-max model should be applied.

The scone-half of the report is to investigate how to make use of freeway service patrols (FSP) to mitigate the impacts incurred by incidents. As one component of traffic incident management systems, freeway service patrols (FSP) facilitate quick removal of incidents through faster response and reduced clearance time. One of the key issues in determination of the deployment strategy is how to allocate tow trucks among patrol beats to maximize the effectiveness of the FSP services. The report presents a min-max bi-level programming model to determine a robust optimal fleet allocation that minimizes the maximal system travel time that incidents may incur. A heuristic iterative solution algorithm is proposed to solve the model. Both the model and the algorithm are demonstrated and validated through a numerical example. It is found that the robust optimal fleet allocation always performs better against the worst-case scenarios. The performance difference becomes more significant with higher levels of capacity uncertainty or network congestion, suggesting that the model may contribute more or make more difference in the situations of high frequencies of incidents or high levels of network congestion

The report has proved that through applying robust optimization methodology, robust optimal improvement strategies can be obtained to improve substantially the capacity of a highway system against high-consequence scenarios incurred by fluctuations in travel demand or irregular incidents.

TABLE OF CONTENTS

SECTION		PAGE
ACKNOWLEDGEMENTS		iii
ABSTRACT		v
EXECUTIVE SUMMARY		vii
LIST OF TABLES		xiii
LIST OF FIGURES		xv
1	INTRODUCTION	1
2	ROAD NETWORK DESIGN UNDER DEMAND UNCERTAINTY	5
2.1	Background	5
2.2	Sensitivity-Based Model	7
2.3	Scenario-Based Model	17
2.4	Min-Max Model	23
3	FLEET ALLOCATION OF FREEWAY SERVICE PATROLS	31
3.1	Background	31
3.2	Model Formulation	33
3.3	Solution Algorithm	38
3.4	Numerical Example	40
4	CONCLUSION	45
REFERENCE		47

LIST OF TABLES

TABLE		PAGE
TABLE 2-1	Network Characteristics of the Example Network	12
TABLE 2-2	O-D Travel Demand for the Example Network	13
TABLE 2-3	Values of Components of Objective Function of Sensitivity-Based Model	13
TABLE 2-4	Improvement Schemes from the Sensitivity-Based Model	14
TABLE 2-5	Robustness Test Results for the Sensitivity-Based Model	15
TABLE 2-6	Improvement Schemes from the Scenario-Based Model	21
TABLE 2-7	Robustness Test Results for the Scenario-Based Model	21
TABLE 3-1	Network Characteristics of the Example Network	41
TABLE 3-2	O-D Travel Demand for the Example Network	42

LIST OF FIGURES

TABLE **PAGE**

FIGURE 1-1 Disruptive Events with Combinations of Impacts and Frequencies 1

FIGURE 2-1 The Example Network 12

FIGURE 2-2 Average Travel Time versus Standard Deviation for the Sensitivity-Based Model 16

FIGURE 2-3 Average Travel Time versus Standard Deviation for the Scenario-Based Model 22

FIGURE 2-4 Convergence of the Iterative Algorithm for the Min-Max Model 28

FIGURE 2-5 System Performance Difference between Robust and Deterministic Improvement Schemes 29

FIGURE 3-1 An Illustrative Example of FSP Impacts 36

FIGURE 3-2 An Example Network 41

FIGURE 3-3 Convergence of the Iterative Algorithm for the Min-Max Model 42

FIGURE 3-4 System Performance Differences between Uniform and Optimal (Total Travel Time of Uniform Minus That of Optimal) 43

1. INTRODUCTION

The potential sources of disruption to highway traffic operations are numerous, ranging from irregular and random incidents, like earthquakes, terrorist attacks, floods, adverse weathers, traffic accidents, breakdowns, signal failures, roadwork etc, to regular fluctuations of travel demand in times of day, days of the week, and seasons of the year. The scales, impacts, frequencies and predictability of these disruptive events will of course vary enormously, with natural or man-made disasters at one extreme, and routine events that happen every now and then at the other, as illustrated in Figure 1-1.

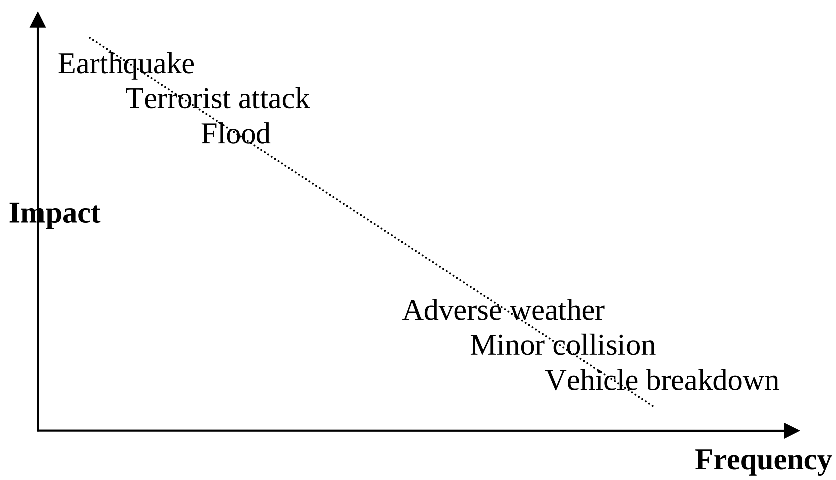


FIGURE 1-1 Disruptive Events with Combinations of Impacts and Frequencies

While little can be done about their scales, frequencies or predictability, particularly where natural disasters and accidents are concerned, it remains possible to design and manage highway networks so as to minimize the disruption such events can cause. It is not realistic to expect the performance of a highway network under catastrophic disasters is the same as that under minor incidents. Therefore, the required functionalities and ways to ensure the functionalities for those two extreme types of disruptive events would differ significantly. The research focuses on improving performance reliability of a highway network at its routine operations, addressing everyday incidents that would be placed in the lower right-hand corner in Figure 1-1.

Various policies and activities can be adopted to improve highway network reliability. In order to facilitate delivering a proof of concept, we build our modeling framework to consider road improvement and incident management. Other policies, such as providing information through advanced traveler information systems (ATIS) can be accommodated by the proposed modeling framework without much difficulty.

With a limited budget available to highway network management, engineers and planners sometimes need to decide on which links improvement works, such as maintenance and rehabilitation or road expansion, should be implemented in order to maintain or improve the effectiveness of a highway system. It is quite often that the decisions are made without considering the impacts of disruptive events on the system performance. As a result, the system performance may deteriorate significantly upon the onset of those disruptive events. In some rare cases, in order to deal with the impacts, sensitivity analyses have been performed to evaluate sensitivities of the decisions to the uncertainties (onsets of disruptive events). However, such practices are intrinsically posteriori or reactive, and provide no direct mechanism for controlling the sensitivities.

On the other hand, traffic incident management is emerging as a proven solution to ensure highway reliability. It is a planned and coordinated process to detect, respond to, and remove traffic incidents and restore traffic capacity as safely and quickly as possible. As one component of incident management, incident response teams and freeway service patrols (FSP) facilitate the quick removal of incidents through fast response and clearance times. Given a limited budget, setting up FSP beats that tow trucks patrol on and assigning FSP resources to beats so as to maintain sufficient service intensity are very important to success of the system. However, these decisions are often made in a heuristic manner. For example, criteria for funding allocations in the FSP program in California have been based on population, urban freeway lane miles, and vehicle hours of congestion delay (Skabardonis et al., 1998). A wiser allocation of available funding could be made from a systematic perspective such that the effectiveness of the FSP system can be maximized.

In summary, the decision problem that we tackle in the research is, with a given budget, choosing links from a network to implement road improvement works and determining the corresponding investment magnitudes, and/or choosing links to implement FSP program and determining the corresponding service intensity so as to improve performance reliability of the network at most. Our approach is not to propose some reliability index and then maximize it to obtain the optimal improvement scheme. In contrary, we attempt to gracefully trade off effectiveness vs. guaranteed robustness and reliability. Our objective is still maximization of effectiveness, but the fundamental idea is to seek a robust optimal solution that tolerates changes of travel demand and network supply caused by disruptive events, up to a given bound known *a priori*. The solution is therefore neither careless (without considering uncertainty at all) nor overly conservative. Because the improvement scheme is robust, we expect that the resulting highway network will remain "close" to its *designated* performance under any realization of the uncertainties (onsets of disruptive events). In this sense, the network will be more reliable and predictable.

We apply robust optimization to determine robust planning and management strategies for a highway network. Robust optimization is a modeling methodology to solve optimization problems in which the data are uncertain and only known to belong to some uncertainty set. The approach is to seek optimal (or near optimal) solutions that are not overly sensitive to any realization of uncertainty. Recent reviews on this topic can be found in Mulvey et al. (1995), Ben-Tal and Nemirovski (2002) and El Ghaoui (2003) among others.

The most widely known approach to deal with uncertain or perturbed data is sensitivity analysis. Sensitivity analysis is an a posteriori or reactive tool for infinitesimal uncertainty. It simply measures the sensitivity of a solution to changes in the input data and provides no direct mechanism for controlling this sensitivity. Stochastic programming and robust optimization on the other hand are proactive methods. However, unlike the stochastic approach, which assumes the uncertainty is random, with a known distribution, robust optimization assumes the data of the problem are unknown but bounded (e.g., via intervals of confidence for the data). No underlying stochastic model of the data is assumed to be known, although such knowledge may be of use to obtain reasonable uncertainty sets. A robust feasible solution is one that tolerates changes in the problem data, up to a given bound known a priori, and a robust optimal solution is a robust feasible solution with the best possible value of the objective function. By carefully constructing and efficiently solving the robust counterpart of the original problem, it is possible to obtain solutions that gracefully trade off performance vs. guaranteed robustness and reliability.

Previous studies on robust optimization have mainly focused on solving uncertain linear, conic quadratic and semi-definite programming problems. Successful applications of robust optimization can be found in many areas, such as finance, telecommunication and structural engineering. Relatively few applications have been done in the transportation field. More importantly, to our best knowledge, no study has been done on the subject of bi-level programming model, the modeling structure that we will use to determine robust planning and management policies.

The reminder of the report is organized as follows. Chapter 2 focuses on robust network design under demand uncertainty while Chapter 3 addresses the robust optimal fleet allocation problem for FSP services. Conclusions and recommendations for further research are offered in the last chapter.

2. ROAD NETWORK DESIGN UNDER DEMAND UNCERTAINTY

2.1 BACKGROUND

As the gap between investment needs and available funds for road network improvements continues to grow, a critical issue facing state departments of transportation today is how to allocate limited resources so as to obtain the best return for their expenditure. Often, road improvement decisions are made without adequately taking into account the impacts of disruptive events. As a result, road system performance may deteriorate significantly upon the onset of those disruptive events. In some cases, in order to deal with the impacts, sensitivity analysis has been performed to evaluate the sensitivities of the decisions to the uncertainties (FHWA, 2003). However, such practices are intrinsically a posteriori or reactive, and thus provide no direct mechanism for controlling these sensitivities.

This chapter is concerned with development of robust improvement schemes for road networks under demand uncertainty. Demand uncertainty here refers to day-to-day or within-day fluctuation of travel demand in the daily operations of road networks, rather than the changes in demand patterns after major disruptive events. The problem of interest is the following: subject to a given budget, determine which links from a network need improvement (capacity increase) and decide how much budget should be allocated to the links so as to maximize the improvement of the performance of the network. More specifically, we attempt to determine a robust optimal improvement scheme that maximizes the improvement of the effectiveness of the network, while at the same time ensuring it remains "close" to the designated level under any realization of demand uncertainty. The implementation of such a scheme will improve the network reliability.

Solving for improvement schemes is a component of the network design problem. Although a vast, growing body of research on network design has been developed in the past two decades [see the paper by Yang and Bell (1998) for a recent survey of this topic], most of the studies were conducted without considering the impacts of uncertainties. Barnhart (2000) presented a scenario-based stochastic model, called the average plan model, to incorporate uncertainty into deterministic transportation planning models. The objective was to find a solution that is average in the sense that it is closer to the solution of very high probability events, as opposed to infrequent events. The average plan model was applied to a network design problem for the distribution of crops in Mexico, considering the uncertainty in demand of commodities that should be transported. Yin and Ieda (2002) investigated a network design problem with stochastic travel time. They demonstrated a new reliability version of Braess' paradox where reducing the variability of travel time (by implementing incident management) or increasing the capacity (by road expansion) of some links may actually lead to a less reliable network. They further proposed a bi-level optimization model through which the reliability paradox can be avoided and optimal improvement can be achieved. Waller *et al.* (2001) examined the impact of demand uncertainty on the evaluation of network improvements. They concluded that using expected demand tends to overestimate performance of the network and could lead to erroneous choice of improvements. Several potential actions were discussed to deal with the problem, the simplest of which is demand inflation that yields benefits not only selecting improvements with lower expected total system travel time but also significant reductions in the variance associated with these measures. However, as suggested in their paper, a well-defined theory is needed for selecting a suitable inflation level. Waller and Zilliaskopoulos (2001) applied the demand inflation method in a dynamic network design problem to address demand uncertainty. They also suggested a two-stage stochastic formulation with recourse for the problem. Chen *et al.* (2003) developed a multiobjective bi-level mean-variance model to determine the optimal toll and capacity in a build-operate-transfer roadway subject to demand uncertainty. The objective of the model was to maximize the mean profit as well as the variance of the profit. A simulation-based multiobjective genetic algorithm was developed to solve the model. Their model and solution algorithm are readily applicable to the problem of interest in this chapter.

This chapter presents three alternate models, sensitivity-based, scenario-based and min-max, for determining robust optimal improvement schemes for road networks under demand uncertainty. These models have simple structures and are computationally tractable, yet efficient. The following three sections of this chapter introduce these three models in sequence, covering model formulation, solution algorithm, a numerical example and model validation.

2.2 SENSITIVITY-BASED MODEL

2.2.1 FORMULATION PREPARATION

Consider a network $G=(N,A)$, where *N* is the set of nodes, and *A* is the set of links. Let *W* be the set of all origin-destination (O-D) pairs in the network, R_w be the set of routes between O-D pair $w \in W$ and q_w be the demand between O-D pair *w*. To represent demand uncertainty, we assume that there exists some perturbation in the demand and the perturbed demand $q(\varepsilon)$ is given by the following equation:

$$q_w(\varepsilon)=q_w+\varepsilon_w \tag{2-1}$$

where ε_w is a perturbation associated with the travel demand q_w.

Let f_r^w be the flow on route $r \in R_w$, $w \in W$, and v_a be the traffic flow on link $a \in A$. We thus have the following flow conservation equations:

$$v_a=\sum_{w\in W}\sum_{r\in R_w} f_r^w \delta_{ar}^w, a \in A \tag{2-2}$$

$$\sum_{r\in R_w} f_r^w = q_w, w\in W \tag{2-3}$$

$$f_r^w \geq 0, r\in R_w, w\in W \tag{2-4}$$

where δ_{ar}^w =*1* if route *r* between O-D pair *w* uses link *a*, and 0 otherwise. Denote the travel time for each link $a \in A$ as $t_a(v_a,c_a)$, which is assumed to be an increasing/decreasing and strictly convex function of link flow v_a on that link/link capacity c_a. Consequently, the route travel time is:

$$t_r^w=\sum_{a\in A} t_a(v_a)\delta_{ar}^w, r\in R_w, w\in W \tag{2-5}$$

where t_r^w is the travel time on route $r \in R_w$ between O-D pair $w \in W$.

Because the modeling framework we propose is intended for planning purposes, various static network traffic models are applicable to describe drivers' reaction to an improvement scheme and to evaluate the resulting system performance. In this chapter, we adopt the stochastic user equilibrium (SUE) model for the following two major reasons: firstly, the SUE model has been proved to be a large-demand approximation to the mean of more general stochastic models that explicitly represent drivers'' information acquisition in a stochastic environment (Clark and Watling, 2004); Secondly, in contrast to the *directional* differentiability of deterministic user equilibrium link flows, the perturbed Logit-based and Probit-based SUE link flows are both continuously differentiable functions in the perturbation parameters (Meng *et al.*, 2004). Following the work by Daganzo (1983) and Cantarella (1997), the SUE assignment can be expressed as a fixed-point (FP) problem in the link flow space, over the non-empty, compact and convex set of feasible link flow patterns. The FP problem is written as:

$$v_a = \sum_{w \in W} \sum_{r \in R_w} q_w \delta_{ar}^w p_r^w(t^w), a \in A \tag{2-6}$$

where $p_r^w(t^w)$ is the probability of drivers choosing route $r \in R_w$ and t^w is a vector of travel times of all routes between the O-D pair *w*.

Moreover, the Logit-based SUE model is employed herein to facilitate the presentation of the robust optimization schemes, because of its closed-form analytical expression and associated computational advantage. In order to overcome the main shortcoming of the Multinomial Logit, which is that it yields unrealistic choice probabilities for overlapping or correlated routes, we apply the C-Logit model proposed by Cascetta *et al.* (1996). The C-Logit model retains a closed analytical form, but does not suffer from the independence of alternatives assumption. Note that other Logit models that overcome this weakness, such as the cross-nested Logit model by Prashker and Beckhor (1998) and the paired combinatorial Logit model by Koppelman and Wen (2000) may also be applied here. With the C-Logit model, we have:

$$p_r^w(t^w) = \frac{\exp(-\theta_0 t_r^w - \theta_1 CF_r^w)}{\sum_{k \in R_w} \exp(-\theta_0 t_k^w - \theta_1 CF_k^w)}, r \in R_w, w \in W \tag{2-7}$$

where CF_r^w is the commonality factor for route $r \in R_w$, representing the degree of similarity (overlapping) of route *r* with other routes in the set of R_w. Cascettta *et al.* (1996) suggested several ways to specify the commonality factor that give similar results. One of those specifications is given as below:

$$CF_r^w = \sum_{a \in A} \delta_{ar}^w w_{ar}^w \ln N_a^w, r \in R_w, w \in W \tag{2-8}$$

where N_a^w is the number of routes, connecting O-D pair *w* that share link *a* and w_{ar}^w is the proportional weight of link *a* for route $r \in R_w$, specified as the fraction of total route travel time which can be attributed to link *a*:

$$w_{ar}^w = t_a / t_r^w \tag{2-9}$$

2.2.2 MODEL FORMULATION

The traditional network design model prescribes optimal improvement schemes that minimize the total travel time. In this chapter, we attempt to determine a scheme with which the resulting system performance (more specifically, the total travel time) is insensitive to changes in travel demand. In other words, in addition to minimization of the total travel time, our objective includes minimization of the sensitivity of the total travel time with respect to the uncertain demand.

With a limited budget available for road improvement, a sensitivity-based model for a robust scheme can be written as:

$$\min_{c^+} Z = \alpha \cdot T\big|_{\varepsilon=0} + (1-\alpha)\cdot(\nabla T_\varepsilon\big|_{\varepsilon=0})^T \cdot \nabla T_\varepsilon\big|_{\varepsilon=0} \tag{2-10}$$

Subject to:

$$\sum_{a\in A} h_a(c_a^+) \le B \tag{2-11}$$

$$0 \le c_a^+ \le c_a^{\max}, \quad a \in A \tag{2-12}$$

$$v_a = \sum_{w\in W}\sum_{r\in R_w} q_w \delta_{ar}^w p_r^w\left(t^w(c^0 + c^+)\right), a \in A \tag{2-13}$$

where Z is the value of the objective function; c^+ is a vector having as elements the continuous capacity increases of all links; α is a parameter, $0 \le \alpha \le 1$; $T\big|_{\varepsilon=0}$ is the total travel time when the demand perturbation equals 0 and $T = \sum_{a\in A} v_a t_a(v_a)$; $\nabla T_\varepsilon = [\partial T/\partial\varepsilon_1, \Lambda, \partial T/\partial\varepsilon_w, \Lambda]^T$, which is the vector of derivatives of the total travel time with respect to the demand perturbation and $\nabla T_\varepsilon\big|_{\varepsilon=0}$ represents the values of derivatives when the demand perturbation equals 0; c_a^+ is the continuous capacity increase of link *a*; $h_a\left(c_a^+\right)$ is the construction cost function that is generally assumed to be non-negative, increasing and differentiable; *B* is the available budget; $c_a^{\max}$ is the upper limit of the capacity increase and c^0 is the vector of the original link capacities.

The model (Equations 2-10 through 2-13) can be treated as a bi-level programming model, in view of that the problem has the structure of a leader-follower game. The upper level (Equations 2-10 through 2-12) represents the leader's behavior that minimizes the objective function in the feasible set defined by Equations 2-11 and 12, and the lower level problem (Equation 2-13) is a SUE traffic assignment problem, representing the follower's response to the leader's decision. On the other hand, the model can also be solved as a single-level mathematical program with equilibrium constraints (MPEC) (Davis, 1994 and Meng *et al.*, 2004). In this case, the decision variables include the continuous capacity increase c_a^+ and the link traffic flow v_a as well.

The model can be considered a first-order approximation to a mean-variance model because the variance of total travel time $\sigma_T^2 \approx \nabla T_\varepsilon^T \cdot \mathrm{Cov}(\varepsilon) \cdot \nabla T_\varepsilon$ where $\mathrm{Cov}(\varepsilon)$ is the covariance matrix of the demand perturbations (Chen *et al.*, 2002). The parameter α represents practitioners' caution towards demand uncertainty. A larger value of α would be used when travel demand is believed to be stable. If the fluctuation of travel demand is a concern, a smaller value should be used.

The sensitivity-based model is a direct extension of the traditional network design model, which is in fact a special case of the former when $\alpha = 1$. The model is deterministic (in terms of the way it handles demand uncertainty) and thus is computationally simple if compared with previous stochastic network design models (e.g., Chen *et al.*, 2003). Its disadvantage is that it works well only if demand fluctuation is infinitesimal.

2.2.3 SOLUTION ALGORITHM

For calculating the value of the objective function *Z*, we note that $\left.\frac{\partial T}{\partial \varepsilon_w}\right|_{\varepsilon=0} = \frac{\partial T}{\partial q_w}$. We then have:

$$\frac{\partial T}{\partial q_w} = \sum_a \left(t_a + v_a \frac{\partial t_a}{\partial v_a} \right) \frac{\partial v_a}{\partial q_w} \tag{2-15}$$

where $\partial v_a / \partial q_w$ is the derivative of the SUE link flow with respect to the travel demand, which can be calculated by using the sensitivity analysis method for SUE link flows developed by Meng *et al.* (2004) and Ying and Miyagi (2001). The results are presented below without proof.

Define the function:

$$H_a = v_a - \sum_{w \in W} \sum_{r \in R_w} q_w \delta_{ar}^w p_r^w (t^w), a \in A \tag{2-16}$$

Thus the SUE FP problem (Equation 2-6) is equivalent to $H_a = 0, a \in A$. According to the implicit function theorem, gradients of SUE link flows are:

$$\nabla v_q = -\nabla H_v^{-1} \cdot \nabla H_q \qquad (2\text{-}17)$$

where ∇v_q is the gradient matrix of SUE link flows with the dimension of $|A| \times |W|$; ∇H_v and ∇H_q are the respective Jacobian matrices of a set of Equation 2-16 with respect to the link flow and the demand perturbation. The elements of these two matrices are:

$$\frac{\partial H_a}{\partial q_w} = -\sum_{r \in R_w} \delta_{ar}^w P_r^w \qquad (2\text{-}18)$$

$$\begin{aligned}\frac{\partial H_a}{\partial v_b} = \delta_{ab} + \frac{\partial t_b}{\partial v_b} \sum_{w \in W} q_w \Bigg\{ & \sum_{r \in R_w} \delta_{ar}^w \delta_{br}^w P_r^w (\theta_0 + \theta_1 \frac{\ln N_b^w - CF_r^w}{t_r^w}) \\ & - \left(\sum_{r \in R_w} \delta_{ar}^w P_r^w \right) \left(\sum_{r \in R_w} \delta_{br}^w P_r^w (\theta_0 + \theta_1 \frac{\ln N_b^w - CF_r^w}{t_r^w}) \right) \Bigg\}\end{aligned} \qquad (2\text{-}19)$$

A number of algorithms can be applied to solve the sensitivity-based model, including the sensitivity-analysis-based iterative algorithms by Yang *et al.* (1994) and Meng *et al.* (2004) respectively for solving bi-level programming models, and the successive quadratic programming (SQP) algorithm by Davis (1994) for solving the single-level models with equilibrium constraints. These algorithms may require the gradients of the objective function with respect to link capacity (∇Z_c). Calculation of ∇Z_c involves evaluation of the second-order derivatives $\partial^2 v_a / \partial q_w \partial c_b$, which is difficult to obtain directly through the sensitivity analysis on the SUE link flows. One remedy, though time-consuming, is to compute the derivatives $\partial^2 v_a / \partial q_w \partial c_b$ by finite difference approximations. Another possibility is to apply the genetic-algorithm-based approach of Yin (2000) since this approach does not require the gradient information at all.

2.2.4 NUMERICAL EXAMPLE

A numerical example is now provided to illustrate the proposed model. The example road network shown in Figure 2-1 has 13 nodes, 19 links and 4 O-D pairs, adopted from Nguyen and Dupuis (1984). The Bureau of Public Road link travel time function was used

$$t_a(v_a) = t_a^0 \left(1 + 0.15 \cdot \left(\frac{v_a}{c_a} \right)^4 \right) \qquad (2\text{-}20)$$

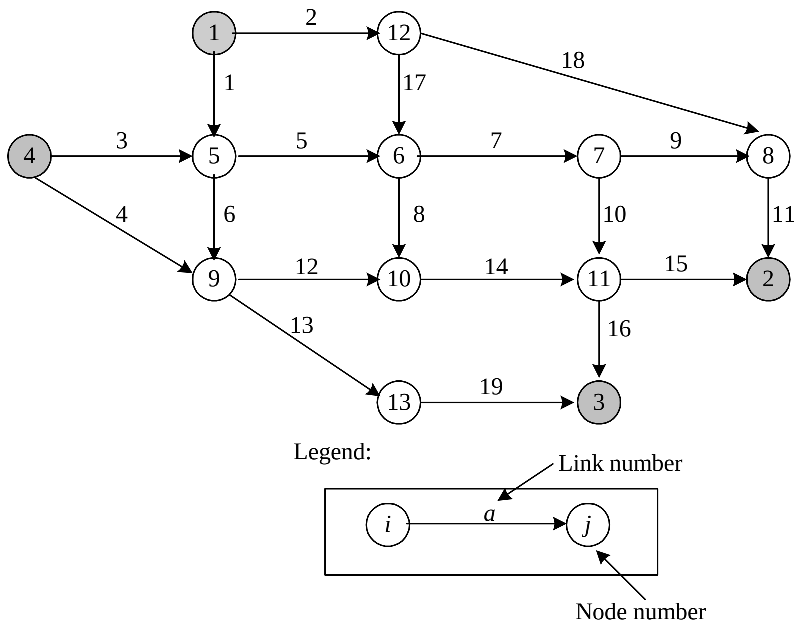


FIGURE 2-1 An Example Network

TABLE 2-1 Network Characteristics of the Example Network

Link a	t_a^0	c_a
1	7.0	800
2	9.0	400
3	9.0	200
4	12.0	800
5	3.0	350
6	9.0	400
7	5.0	800
8	13.0	250
9	5.0	250
10	9.0	300
11	9.0	550
12	10.0	550
13	9.0	600
14	6.0	700
15	9.0	500
16	8.0	300
17	7.0	200
18	14.0	400
19	11.0	600

And the network characteristics and O-D demand are given in Tables 2-1 and 2 respectively. In the example, the link construction cost function was assumed to be $h_a\left(c_a^+\right) = c_a^+$, and no upper limit was set for the capacity increase. The total budget for the improvement project was 200; the parameters θ_0 and θ_1 for the SUE assignment were 0.05 and 0.5 respectively. A SQP subroutine with finite-differencing derivatives in Matlab was used to solve the example as well as two self-programmed subroutines: one to implement the SUE assignment to obtain the SUE link flows and the other to conduct the sensitivity analysis to estimate the derivatives ∇T_q.

TABLE 2-2 O-D Travel Demand for the Example Network

O/D	2	3
1	400	800
4	600	200

Table 2-3 presents the values of two components of the objective function (total travel time T and its sensitivity index $\nabla T_q^T \cdot \nabla T_q$) under varying values of α. As expected, we observe that the sensitivity index decreases but the total travel time increases as α decreases, which clearly shows a tradeoff between the optimality (minimization of travel time) and the robustness (minimization of sensitivity) for the nominal demand. This result suggests that to improve robustness, a certain degree of optimality has to be given up. Table 2-4 presents the corresponding optimal schemes under varying values of α.

TABLE 2-3 Values of Components of Objective Function of the Sensitivity-Based Model

α	T $(\times 10^4)$	$\nabla T_q^T \cdot \nabla T_q$ $(\times 10^3)$
0.0	9.9588	9.4989
0.1	9.9248	9.5217
0.3	9.8421	9.7203
0.5	9.8207	9.8519
0.7	9.8125	9.9713
0.9	9.8091	10.215
1.0	9.8075	10.344

TABLE 2-4 Improvement Schemes from the Sensitivity-Based Model

Scheme	C_2^+	C_3^+	C_5^+	C_6^+	C_{15}^+	C_{16}^+
SEN-0.0	71.4	0.0	0.0	114.7	0.0	13.9
SEN-0.1	53.5	0.0	10.1	109.8	0.0	26.6
SEN-0.3	4.4	0.0	54.9	84.6	0.0	56.1
SEN-0.5	0.0	0.0	60.7	59.6	0.0	79.7
SEN-0.7	0.0	0.0	62.3	40.2	0.0	97.6
SEN-0.9	0.0	14.6	58.2	28.1	0.0	99.1
SEN-1.0	0.0	22.2	55.8	22.9	0.0	99.1

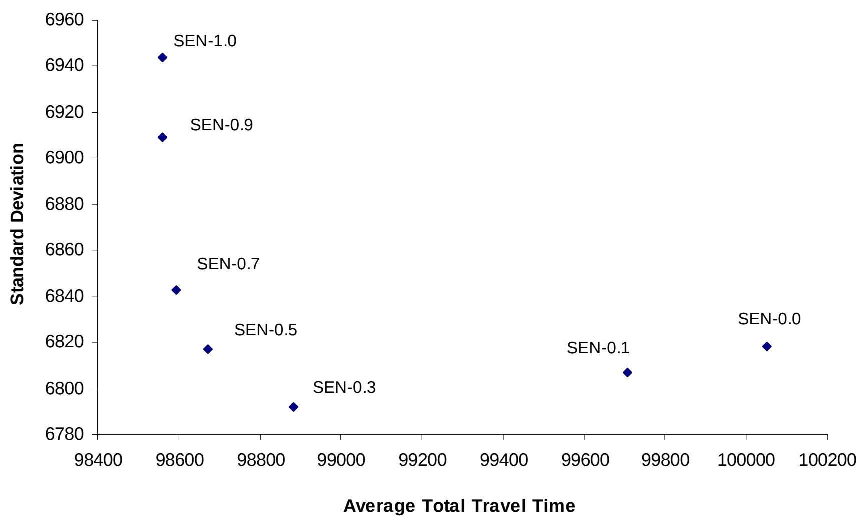
Note: 1) Scheme naming rule, for instance, SEN-0.5 stands for the improvement scheme obtained by using the sensitivity-based model with α equal to 0.5. 2) $c_a^+ = 0$ for the other links.

The effectiveness and robustness of the resultant improvement schemes can be tested by simulation. That is, samples can be drawn from specified distributions of travel demand, and for each sample a SUE assignment is conducted to evaluate the improvement schemes. The mean and variance of the resulting total travel times for all samples for each scheme will be calculated and then used for comparison of effectiveness and robustness. Although the sensitivity-based model is intended for use in the situation where the demand fluctuation is minor, we conducted simulation tests to examine how the resultant improvement schemes behave if confronted with higher levels of demand fluctuation.

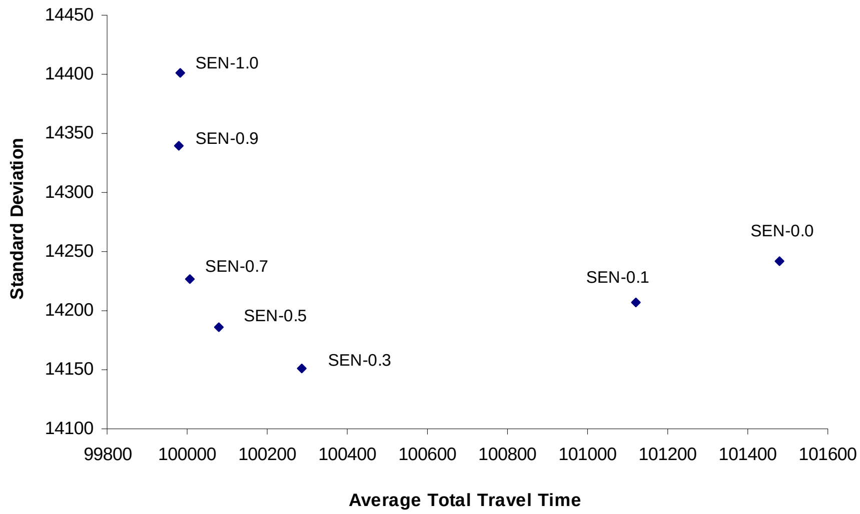
Simulation tests were performed with two types of demand distributions. One is a truncated normal distribution. For O-D pairs 1-2, 1-3, 4-2 and 4-3, the demand follows the following independent distributions, respectively: $300 \le N(400,2500) \le 500$, $600 \le N(800,2500) \le 1000$, $400 \le N(600,2500) \le 800$ and $100 \le N(200,900) \le 300$. The other is a uniform distribution. The O-D demand was assumed to be independently and uniformly distributed in the following intervals respectively: [300, 500], [600, 1000], [400, 800] and [100, 300]. Note that for both sets of distributions, the average travel demand for each O-D pair is the same, which is the nominal demand we used to obtain the optimal improvement schemes.

TABLE 2-5 Robustness Test Results for the Sensitivity-Based Model

Normal Distribution							
Scheme	SEN-0.0	SEN-0.1	SEN-0.3	SEN-0.5	SEN-0.7	SEN-0.9	SEN-1.0
Mean ($\times 10^4$)	10.005	9.971	9.888	9.867	9.859	9.856	9.856
S.D. ($\times 10^3$)	6.818	6.807	6.792	6.817	6.843	6.909	6.944
Uniform Distribution							
Scheme	SEN-0.0	SEN-0.1	SEN-0.3	SEN-0.5	SEN-0.7	SEN-0.9	SEN-1.0
Mean ($\times 10^4$)	10.148	10.112	10.029	10.008	10.001	9.998	9.998
S.D. ($\times 10^3$)	14.242	14.207	14.151	14.186	14.227	14.339	14.401



(a) Normal Distribution



(b) Uniform Distribution

FIGURE 2-2 Average Travel time versus Standard Deviation for the Sensitivity-Based Model

Means and standard deviations of total travel times from the tests with 1000 samples respectively are reported in Table 5. Figure 2 presents a graphic representation of the same set of results. From Table 5 and Figure 2, it is interesting to observe that:

- There exists a tradeoff between effectiveness (average travel time) and robustness (standard deviation). Therefore, for a network design problem with demand uncertainty, there may not be an unambiguously optimal improvement scheme that is able to yield the lowest expected total travel time and the largest reduction in the variance at the same time. Hence, we may end up with a set of Pareto-optimal (non-dominated) schemes, which are optimal in the sense that no improvement can be achieved in one objective without degradation in the other. With these non-dominated schemes, decision makers can then select one that appeals to them based on their preferences.

- Even confronted with higher levels of demand fluctuation, the sensitivity-based model is able to generate improvement schemes that are more robust than that from a deterministic network design model (the scheme of SEN-1.0), but only by giving up effectiveness to some extent. However, the schemes of SEN-0.0 and SEN-0.1 are dominated by SEN-0.3 and SEN-0.5 in both simulation tests, suggesting that higher caution placed on demand uncertainty (smaller α) does not necessarily result in a more robust scheme. Therefore, when facing high levels of demand fluctuation, the performance of the schemes from the sensitivity-based model may not be pre-ascertainable and robust schemes have to be determined through trial and error.

2.3 SCENARIO-BASED MODEL

The shortcoming of the sensitivity-based model with high levels of demand uncertainty motivates us to apply another approach, the scenario-based optimization, to the problem of interest. Scenario-based optimization represents uncertainty via a limited number of discrete uncertainty scenarios associated with strictly positive probability of occurrence, and attempts to solve the optimization problem across these scenarios for solutions that are near-optimal with respect to the population of all possible realizations of uncertainty. Scenario-based optimization has been widely used in different domains, such as electric utilities and telecommunications (e.g., Mulvey *et al*., 1995 and Laguna, 1998).

2.3.1 MODEL FORMULATION

Consider a network $G = (N, A)$ where travel demand is uncertain. To represent the uncertainty, we introduce a set of scenarios $\Omega = \{1,2,3,\Lambda\ ,S\}$. For each scenario $s \in \Omega$, the probability of occurrence is π_s and the demand between O-D pair *w* is q_w^s.

It is assumed that under each demand scenario, network traffic condition can be represented by the SUE model, for which the notation and formulation are the same as those in the sensitivity-based model, with an additional scenario index *s* where appropriate. For instance, under demand scenario *s*, the SUE assignment can be expressed as:

$$v_a^s = \sum_{w \in W} \sum_{r \in R_w} q_w^s \delta_{ar}^w p_r^{w^s}, a \in A \qquad (2\text{-}21)$$

where v_a^s is the traffic flow on link *a* under demand scenario *s* and $p_r^{w^s}$ is the probability of drivers choosing route $r \in R_w$ under demand scenario *s*.

With a set of demand scenarios, we now seek a robust improvement scheme that may achieve a tradeoff between the effectiveness and the robustness of a road network. In a similar manner as the sensitivity-based model, the scenario-based model for determining a robust improvement scheme can be written as:

$$\min_{c^+} Z = \beta \sum_{s\in\Omega} \pi_s T_s + (1-\beta)\sum_{s\in\Omega} \pi_s \left(T_s - \overline{T}_s\right)^2 \tag{2-22}$$

Subject to:

$$\sum_{a\in A} h_a(c_a^+) \le B \tag{2-23}$$

$$0 \le c_a^+ \le c_a^{\max}, \quad a \in A \tag{2-24}$$

$$v_a^s = \sum_{w\in W}\sum_{r\in R_w} q_w^s \delta_{ar}^w p_r^{w^s}, a \in A \tag{2-25}$$

where β is a parameter, $0 \le \beta \le 1$; T_s is the total travel time under demand scenario *s*, $T_s = \sum_{a\in A} v_a^s t_a^s(v_a^s)$, and $\overline{T}_s$ is the expected total travel time and $\overline{T}_s = \sum_{s\in\Omega} \pi_s T_s$.

It is easy to see that the first component of the objective function (Equation 2-22) is the average total travel time while the second represents the variance of the total travel times across all demand scenarios. Thus the model is apparently a mean-variance model and the parameter β reflects the trade-off between mean (effectiveness) and variance (robustness). Similarly, the scenario-based model can be treated as either a bi-level programming model or an MPEC.

There is no doubt that the scenarios in Ω are only one possible set of realizations of the uncertain demand. Two important questions about scenario-based optimization are: 1) how many scenarios should be included in order to find a solution that is robust across the population of all possible realizations of uncertainty, and 2) how to specify these scenarios and their associated probabilities. Intuitively, the more scenarios we include, the more robust solution we are likely to obtain. However, as the number of scenarios increases, the problem may become prohibitively large. This is actually one of the major shortcomings of the scenario-based model. Fortunately, prior studies (e.g., Mulvey *et al.*, 1995 and Laguna, 1998) have shown that relatively small samples will be able to produce near-optimal polices. In regard to the other question, as Mulvey *et al.* (1995) pointed out, although scenario-based optimization does not provide a means by which the scenarios can be specified, variance reduction methods, such as importance sampling in stochastic simulation can be applied to generate the representative scenarios. There is an extensive literature on random sampling. Borcherds (2000) and Linderoth *et al*. (2004) provide an illustrative introduction and describe the empirical behavior of various sampling methods respectively.

2.3.2 SOLUTION ALGORITHM

Compared with the sensitivity-based model, the scenario-based model will require a higher computational effort. However, the model has a structure that allows it to be decomposed into sub-problems corresponding to scenarios. Therefore the computational complexity of the model only increases linearly as the number of scenarios increases, and thus a relatively efficient algorithm can be developed to solve the model

In this chapter, we treat the model (Equations 2-22 through 25) as a bi-level programming model and apply the sensitivity-analysis-based iterative algorithms to solve it. In order to implement these algorithms, the values and gradients of the objective function at iterative points are required, based on which a linear programming (LP) or a quadratic programming (QP) sub-problem can be defined and solved. To calculate the values of the objective functions at iterative points, the SUE assignment will be performed for each demand scenario respectively, for a total of *S* times. To calculate the gradients of the objective function ∇Z_c, we have:

$$\frac{\partial Z}{\partial c_a} = \sum_{s\in\Omega}\left(\beta + 2(1-\beta)(T_s - \overline{T}_s)\right)\pi_s \frac{\partial T_s}{\partial c_a}, a \in A \tag{2-26}$$

It is easy to show that:

$$\frac{\partial T_s}{\partial c_b} = \sum_a \left(t_a^s + v_a^s \frac{\partial t_a^s}{\partial v_a^s} \right) \frac{\partial v_a^s}{\partial c_b} \tag{2-27}$$

where $\partial v_a^s / \partial c_b$ is the derivative of the SUE link flow under demand scenario *s* with respect to the link capacity c_b, which can be calculated through sensitivity analysis, in a similar manner as that presented earlier for the sensitivity-based model. In addition, the Jacobian matrix of a set of Equation 2-16 with respect to the link capacity, denoted as ∇H_c, should be calculated. Its element are:

$$\frac{\partial H_a}{\partial c_b} = -\frac{\partial t_b}{\partial c_b} \sum_{w\in W} q_w \left\{ \sum_{r\in R_w} \delta_{ar}^w \delta_{br}^w P_r^w (\theta_0 + \theta_1 \frac{\ln N_b^w - CF_r^w}{t_r^w}) - \left(\sum_{r\in R_w} \delta_{ar}^w P_r^w \right) \left(\sum_{r\in R_w} \delta_{br}^w P_r^w (\theta_0 + \theta_1 \frac{\ln N_b^w - CF_r^w}{t_r^w}) \right) \right\} \tag{2-28}$$

And then,

$$\nabla v_c = -\nabla H_v^{-1} \cdot \nabla H_c \tag{2-29}$$

where ∇v_c is the gradient matrix of the SUE link flows with the dimension of $|A| \times |A|$.

The general framework of the proposed decomposition algorithm is summarized as follows:

Step 0: Initialization. Determine an initial value $c^{+(n)}$. Set iteration counter: *n*=0;

Step 1: Solve the lower-level SUE assignment problem for the given $c^{+(n)}$ for each demand scenario *s* and then conduct the sensitivity analysis respectively to obtain ∇v_c^s;

Step 2: Calculate *Z* and ∇Z_c;

Step 3: Formulate a LP or a QP approximation to the original upper-level constrained optimization problem at the points of $c^{+(n)}$ (the details are described in Yang *et al*. ,1994 and Meng *et al*., 2004), and then solve the resulting problem to obtain a search direction $x^{(n)}$;

Step 4: Do a line search to obtain the step length $\lambda^{(n)}$;

Step 5: Compute $c^{+(n+1)} = c^{+(n)} + \lambda^{(n)} x^{(n)}$.

Step 6: Convergence check. If $\max_a \left(\left| c_a^{+(n+1)} - c_a^{+(n)} \right| \right) \le \delta$, then stop, where δ is a predetermined error tolerance. Otherwise, let *n*=*n+1*, go to Step 1

2.3.3 NUMERICAL EXAMPLE

The scenario-based model was applied to the same example network to obtain robust improvement schemes. The demands were assumed uncertain but bounded within the following intervals: [300, 500], [600, 1000], [400, 800] and [100, 300] for O-D pairs 1-2, 1-3, 4-2 and 4-3 respectively. In order to investigate the impacts of the number of scenarios (denoted by *S*), we considered four sizes: 5, 10, 20 and 50. The representative demands were determined as the points that equally divide the demand intervals into $S+1$ segments (that is, 6, 11, 21 or 51 segments respectively), and the corresponding probabilities of occurrence were all assumed to be $1/S$, the assumption also made by Kouvelis and Yu (1997) and others.

The SQP algorithm based on the sensitivity analysis method was applied to solve for the improvement schemes. Table 2-6 reports the optimal schemes under each set of demand scenarios with different values of β .

TABLE 2-6 Improvement Schemes from the Scenario-Based Model

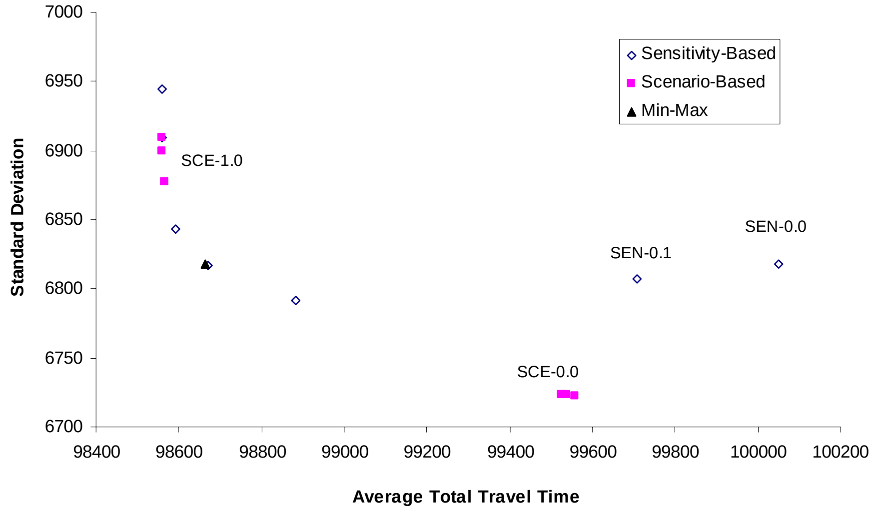
Scheme	C_3^+	C_5^+	C_6^+	C_{15}^+	C_{16}^+
SCE-5-0.0	0.0	13.4	127.5	59.2	0.0
SCE-5-1.0	13.4	60.3	24.7	0.0	101.7
SCE-10-0.0	0.0	16.5	125.1	58.4	0.0
SCE-10-1.0	10.8	61.3	24.8	0.0	103.2
SCE-20-0.0	0.0	18.2	123.6	58.2	0.0
SCE-20-1.0	3.3	64.0	24.0	0.0	108.7
SCE-50-0.0	0.0	18.7	123.3	58.1	0.0
SCE-50-1.0	3.6	63.9	24.2	0.0	108.3

Note: Scheme naming rule, for instance, SCE-5-1.0 stands for the improvement scheme obtained by using the scenario-based model with 5 scenarios and β equal to 1.0.

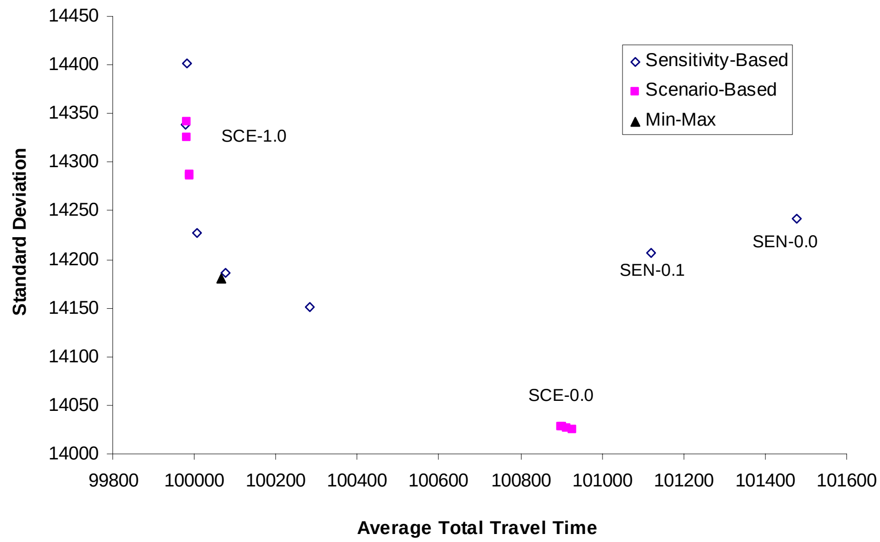
The effectiveness and robustness of the resultant improvement schemes were examined by the aforementioned simulation tests. The resultant means and standard deviations of total travel time from the tests are reported in Table 2-7, and further illustrated by Figure 2-3. For comparison purposes, we included the schemes obtained by using the sensitivity-based model.

TABLE 2-7 Robustness Test Results for the Scenario-Based Model

Normal Distribution								
Scheme	SCE-5-0	SCE-5-1.0	SCE-10-0	SCE-10-1.0	SCE-20-0	SCE-20-1.0	SCE-50-0	SCE-50-1.0
Mean ($\times 10^5$)	9.956	9.856	9.954	9.856	9.953	9.857	9.953	9.857
S.D. ($\times 10^4$)	6.722	6.909	6.723	6.900	6.723	6.877	6.723	6.877
Uniform Distribution								
Scheme	SCE-5-0	SCE-5-1.0	SCE-10-0	SCE-10-1.0	SCE-20-0	SCE-20-1.0	SCE-50-0	SCE-50-1.0
Mean ($\times 10^5$)	10.093	9.998	10.091	9.998	10.090	9.999	10.090	9.999
S.D. ($\times 10^4$)	14.025	14.341	14.027	14.326	14.028	14.286	14.028	14.286



(a) Normal Distribution



(b) Uniform Distribution

FIGURE 2-3 Average Travel Time versus Standard Deviation for Three Models

Several observations on Table 2-7 and Figure 2-3 are summarized as follows:

- The results confirm that there is a tradeoff between effectiveness and robustness. The schemes obtained with emphasis on effectiveness ($\beta = 1$) and those on robustness ($\beta = 0$) are non-dominated optimal solutions.

- In face of high levels of uncertainty, the scenario-based model is able to generate schemes that are more robust than those from the sensitivity-based model.

- The schemes resulting from a smaller number of scenarios have similar performances to those from larger numbers of scenarios, suggesting that relatively small samples may be able to produce near-optimal polices.

- When the number of scenario becomes large (more than 20 for the numerical example), there are no performance improvement even if we increase the sample size. However, the threshold value is dependent on the underlying distribution of the uncertainty and how the uncertainty scenarios are generated and the corresponding probabilities of occurrence are determined.

- In the simulation tests with the truncated normal distribution, the schemes obtained by using the identical probability of occurrence (that is to say, using the uniform distribution) are proved to be robust. This implies that with the objective of minimization of variance, using distorted probabilities of scenario occurrence may still result in robust schemes.

2.4 MIN-MAX MODEL

Both the sensitivity-based and the scenario-based models use variance (or approximation of variance) as a proxy for solution robustness and attempt to establish a mean-variance tradeoff. However, variance gives equal weight to deviations above and below the mean, and the focus on the mean-variance tradeoff often fails to address the risks associated with extreme outcomes (List *et al.*, 2003). In real-world applications, decision makers tend to be risk-averse and are more concerned with the worst-case scenarios. In their mind, as long as the system performance achieves a certain acceptable level, it does not matter that much how it changes above that level. Therefore, it would be desirable to have an improvement scheme that performs better in the worst cases, even though the average performance is poorer. More specifically, we intend to determine an improvement scheme that minimizes the total travel time for the worst-case demand scenario among all of the realizations of the uncertain demand, which is bounded by an uncertainty set known *a priori*.

Such a min-max concept is the basic notion behind a stream of recent research in the area of robust optimization (see Ben-Tal and Nemirovski, 2002 and El Ghaoui, 2003 among others). Previous studies have mainly focused on finding robust counterparts to LP, QP, semi definite programs and dynamic programming algorithms. Successful applications of robust optimization can be found in many areas, such as finance, telecommunication and structural engineering. Relatively few applications can be found in the field of transportation (see Ordonez and Zhao, 2004 for an example). Moreover, to the best of our knowledge, no robust counterpart to bi-level programming or MPEC has been successfully formulated and solved.

2.4.1 FORMULATION

Denote travel demand between all O-D pairs as a vector *q*, which is assumed to be unknown but bounded by an uncertainty set *Q*. The uncertainty set should be closed and convex, specified by practitioners based on their knowledge on the uncertainty associated with travel demand. No underlying stochastic model of travel demand is assumed to be known, although such knowledge may be of use to obtain a reasonable uncertainty set. Below are three typical sets that are meaningful in confining uncertain travel demand:

1) Box

It is assumed that the travel demand in each O-D pair independently varies within an interval of $Q_w = \left[q_w^{\min}, q_w^{\max}\right]$. This interval could be the confidence interval of the estimated demand obtained through surveys or by using an O-D estimation model. As a result, the whole uncertainty set will be $Q = Q_1 \times Q_2 \times \Lambda \times Q_{|W|}$, which is a box centered at the nominal (average) travel demand.

2) Polyhedron

A polyhedron is the solution set of a finite number of linear equalities and inequalities that travel demand of different O-D pairs have to satisfy. For example, in addition to interval constraints, we may have extra constraints such as $\sum_{w \in O_i} q_w \le D_i^{\max}$, where O_i could be the set of O-D pairs whose origin is node *i* and $D_i^{\max}$ is the maximum possible number of trips generated from node *i*, specified by a trip generation model considering car ownership and population etc. Alternatively, O_i could be the set of O-D pairs across the cordon line *i* and $D_i^{\max}$ is the maximum number of trips observed crossing the cordon line *i* during a certain period of time. As another example, one could observe traffic volumes at selected locations to obtain their upper and lower bounds, and then construct the following constraints to confine the uncertain demand: $v_a^{\min} \le \sum_{w \in W} q_w p_a^w \le v_a^{\max}$, where p_a^w is the proportion of travel demand q_w in link volume v_a, which should be determined exogenously (but do not need to be precise) when specifying the uncertainty set. Note that the box uncertainty set is a special case of the polyhedral set.

3) Ellipsoid

An ellipsoidal set is defined as $Q = \left\{ q \in R^{|W|} \middle| q = q^0 + M \cdot u, \|u\|_2 \le 1 \right\}$, where q^0 is the nominal travel demand and the center of the ellipsoid, and $M \in R^{|W| \times |W|}$, which is symmetric and positive definite. The ellipsoid uncertainty set is a compromise between the flexibility of modeling diverse uncertainties (with the box and polyhedral sets as special cases) and the computational complexity of the resultant robust counterpart (Ben-Tal and Nemirovski, 2002). An ellipsoid may be given parametrically by observation data of moderate size. Due to the absence of an efficient way to obtain O-D information, it is difficult to specify such an ellipsoidal uncertainty set in real-world applications. However, it is feasible to use or develop an O-D estimation model to estimate the likelihood region of the O-D matrix based on a large set of archived loop traffic data, and then determine an ellipsoid that approximates the likelihood region.

It should be stressed that the uncertainty set will affect the optimality and robustness of the resultant optimal improvement scheme. However, when applying the min-max concept, it is not necessary to include into the uncertainty set all the possible realizations of travel demand (Ben-Tal and Nemirovski, 2002).

Once the uncertainty set is determined, we aim to find a robust solution that tolerates changes in travel demand up to the given bound. Thus the robust counterpart of the original network design problem can be formulated as:

$$\min_{c^+} \max_{q \in Q} \sum_a v_a t_a(v_a) \qquad (2\text{-}30)$$

Subject to:

$$\sum_{a\in A} h_a(c_a^+) \le B \tag{2-31}$$

$$0 \le c_a^+ \le c_a^{\max}, \quad a \in A \tag{2-32}$$

$$v_a = \sum_{w\in W}\sum_{r\in R_w} q_w \delta_{ar}^w p_r^w \left(t^w(c^0 + c^+)\right), a \in A \tag{2-33}$$

It is easy to see that by solving the robust counterpart, we will be able to obtain an improvement scheme that results in the minimal total travel time under its corresponding worst-case demand scenario (note that the worst-case demand scenarios could be different for different feasible schemes).

2.4.2 SOLUTION ALGORITHM

We propose a heuristic algorithm to solve the above robust optimization model. The algorithm involves an iteration procedure to solve two inner optimization problems to obtain move directions and generate a sequence of solutions until a convergence criterion is met. The algorithm is described as follows:

Step 0: Initialization. Set outer iteration counter *n*=0. Determine an initial value $c^{+(n)}$.

Step 1: Direction finding.

Step 1.1: Solve the following inner problem for the given $c^{+(n)}$ to identify its corresponding worst-case demand scenario:

$$\max_{q^{(n)}} \sum_a v_a t_a(v_a) \tag{2-34}$$

subject to:

$$q^{(n)} \in Q \tag{2-35}$$

$$v_a = \sum_{w\in W}\sum_{r\in R_w} q_w \delta_{ar}^w p_r^w \left(t^w(c^0 + c^{+(n)})\right), a \in A \tag{2-36}$$

Step 1.2: With the resultant optimal solution $q_w^{(n)}$, formulate a traditional deterministic network design problem as below and solve it for a search direction $x^{(n)}$:

$$\min_{x^{(n)}} \sum_a v_a t_a(v_a) \tag{2-37}$$

subject to:

$$\sum_{a\in A} h_a(x_a^{(n)}) \le B \tag{2-38}$$

$$0 \le x_a^{(n)} \le c_a^{\max}, \quad a \in A \tag{2-39}$$

$$v_a = \sum_{w \in W} \sum_{r \in R_w} q_w^{(n)} \delta_{ar}^w p_r^w \left(t^w (c^0 + x^{(n)}) \right), a \in A \qquad (2\text{-}40)$$

Step 2: *Move*. Compute $c^{+(n+1)} = c^{+(n)} + \lambda^{(n)} \left(x^{(n)} - c^{+(n)} \right)$, where $\lambda^{(n)}$ is the step length.

Step 3: *Convergence check*. If $\max_a \left(\left| c_a^{+(n+1)} - c_a^{+(n)} \right| \right) \le \delta$, then stop, where δ is a predetermined error tolerance. Otherwise, let *n* = *n+1*, go to Step 1.

In this algorithm, the step size $\lambda^{(n)}$ could be determined *a priori*, which should be small enough and satisfy $\lim_{n \to \infty} \lambda^{(n)} = 0$ and $\sum_n \lambda^{(n)} = \infty$. It also could be obtained by conducting a local line search to ensure the successive decease of the objective function value. Such an iterative scheme follows the general framework of iterative descent methods, and has been widely applied to solve different types of transportation optimization problems. Although the scheme cannot be always guaranteed theoretically to converge to a local optimum solution, actual applications normally show good convergence and results.

The first inner problem is again a bi-level programming problem or an MPEC, and thus can be solved efficiently by the sensitivity-analysis-based iterative methods. In each iteration, one has to solve a localized linear approximation to the original upper-level constrained optimization problem (Yang *et al.* 1994).

When the uncertainty demand set is an ellipsoid of $Q = \left\{ q \middle| q = q^0 + M \cdot u, \|u\|_2 \le 1 \right\}$, the problem turns out to be a quadratically constrained quadratic program (QCQP) as follows: $\max_{\|u\| \le 1} \nabla T_q^{\ T} \cdot M \cdot u$, where ∇T_q is the vector of derivatives of the total travel time with respect to the travel demand, calculated by using Equations (2-15)-(19). This QCQP can be analytically solved; the optimal objective value is $\sqrt{(M \cdot \nabla T_q)^T \cdot (M \cdot \nabla T_q)}$ and the optimal solution is $M \cdot \nabla T_q \Big/ \sqrt{(M \cdot \nabla T_q)^T \cdot (M \cdot \nabla T_q)}$.

When the uncertainty set is a box or a polyhedron, the localized approximation becomes a simple LP. An important note for the case of the box uncertainty set is that, although in many circumstances the worst-case demand scenario resulting from the first inner problem is the one with each individual O-D demand equal to its upper bound, this is not always true due to the presence of Fisk's paradox that both origin to destination and total travel times decrease as a result of an increase in demand inputs (Fisk, 1979).

The iterative algorithm essentially solves a sequence of two inner bi-level programming problems. Therefore, the computation effort of the algorithm for solving the robust counterpart only increases in a polynomial manner.

2.4.3 NUMERICAL EXAMPLE

We used the same example network and setting to illustrate the min-max model and the iterative algorithm. We assumed a polyhedral uncertainty set that includes interval constraints [300, 500], [600, 1000], [400, 800] and [100, 300] for O-D pairs 1-2, 1-3, 4-2 and 4-3 respectively, and a total demand constraint $q_{1-2} + q_{1-3} + q_{4-2} + q_{4-3} \le D$. The maximum total demand *D* can be viewed as a variable to represent the level of demand uncertainty for the numerical example. To be consistent with the interval constraints, *D* was set to change from 1400 to 2600.

We first examine the convergence of the iterative algorithm. Figure 2-4 plots the convergence gap $\max_a\left(\left|c_a^{+(n+1)} - c_a^{+(n)}\right|\right)$ against the iteration number, where the maximum total demand *D* is 1900. It can be observed that the algorithm has a fairly fast convergence; convergence is achieved in about 40 iterations. For other cases, we observed similar or faster convergence speeds.

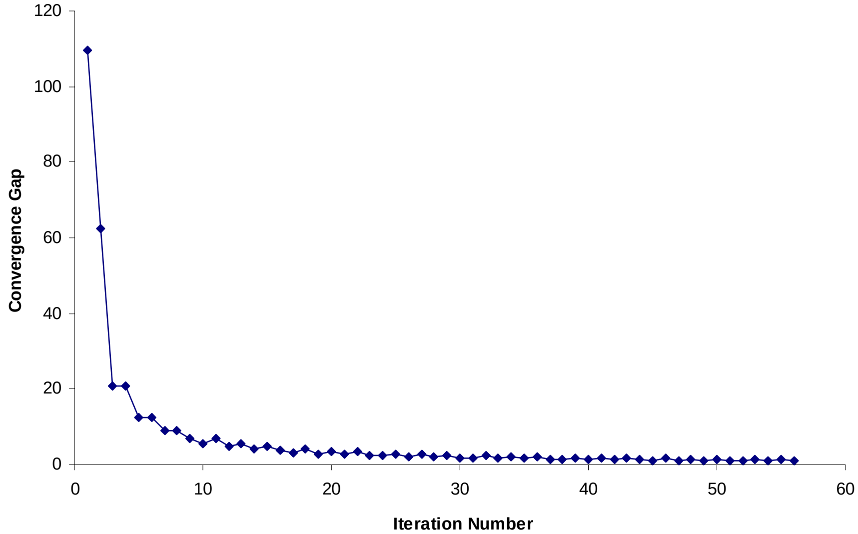


FIGURE 2-4 Convergence of the Iterative Algorithm for the Min-Max Model

For each experiment, we computed the system performance differences between the robust and deterministic improvement schemes under the worst-case and best-case demand scenarios respectively. The deterministic improvement scheme was obtained by using the nominal demand of 300, 600, 400 and 100 for O-D pairs 1-2, 1-3, 4-2 and 4-3 respectively while the robust schemes were obtained by solving the robust counterpart with varying values of maximal total demand *D*. The system performance differences under the worst scenario were calculated as the worst-case objective value (total travel time) of the deterministic solution minus the optimal objective value of the robust solution, while the difference under the best scenario was the optimal objective value of the deterministic solution minus the objective value of the robust solution with the nominal demand. Intuitively, the worst-case performance differences should be positive while the best-case ones negative, if the models are properly formulated and solved.

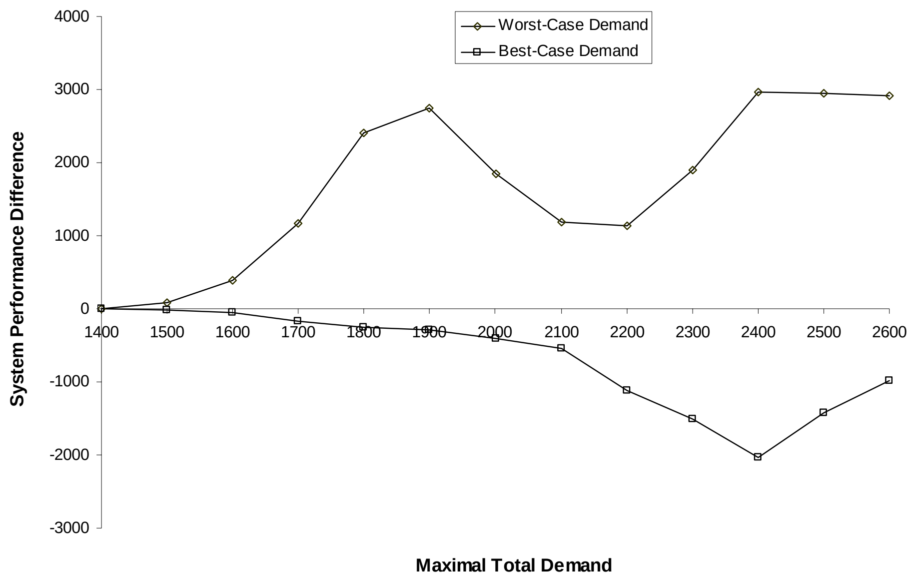


FIGURE 2-5 System Performance Difference between Robust and Deterministic Improvement Schemes

Figure 2-5 plots how demand uncertainty affects the system performance differences. When the maximal demand *D* is 1400, the only feasible demand scenario was the nominal one. Therefore, there is no performance difference. As the maximum demand (the uncertainty level) increases, the robust solutions are obviously superior to the deterministic solution in the aspect of guarding against the worst-case demand scenario. Interestingly, the performance differences do not always increase as the uncertainty level increases. This observation is consistent with the one reported by Ordonez and Zhao (2004) for transportation networks without congestion and route choice.

The robust solutions do not lose much “nominal” optimality, when the maximum demand *D* is less than 2100. After that, the loss increases then decreases. The loss does not increase monotonically with the uncertainty level. In all cases, the gains under the worst-case scenarios by the robust solutions can compensate the losses under the best-case (nominal) scenario.

To fairly compare this model with the sensitivity-based and the scenario-based models, we also solved the min-max model with the interval constraints only. The resultant min-max scheme is $c_5^+ = 57.6$, $c_6^+ = 51.3$, $c_{15}^+ = 10.2$, $c_{16}^+ = 80.9$ and $c_a^+ = 0$ for the other links. The mean and variance of the robust scheme was also examined by the aforementioned simulation tests, and reported in Figure 3. It can be seen that although the min-max model is not designed for achieving a mean-variance tradeoff, it does generate a non-dominated optimal solution. Compared with the schemes obtained from the former models, the min-max scheme offers middle-range performance in terms of both mean and variance.

3. OPTIMAL FLEET ALLOCATION OF FREEWAY SERVICE PATROLS

3.1 BACKGROUND

Traffic incident management is a planned and coordinated process to detect, respond to, and remove traffic incidents and restore traffic capacity as safely and quickly as possible. It has emerged as a proven solution to ensure highway efficiency and reliability (PB Farradyne, 2000). As one component of incident management systems, freeway service patrols (FSP) facilitate quick removal of incidents through faster response and reduced clearance time.

FSP typically operate as follows. The freeways are divided into disjoint beats, each 10-20 miles long with a certain number of tow trucks patrolling on. These trucks travel back and forth along the beat, stopping to clear incidents in a first-reach-first-serve manner. The tow trucks would remove the vehicles stalled in the freeways and provide services such as changing flat tires and providing a needed gallon of gasoline. If they cannot get the vehicles operational in a few minutes they will tow them off the freeway to a designated area (Petty, 1997). FSP systems have been deployed extensively across the U.S., such as in Chicago, Los Angeles and the San Francisco Bay Area. Reviews on these practices can be found in Morris and Lee (1994), Petty (1997) and David and Ogden (1998) among others. Note that the way FSP systems work is different from the incident-response dispatch system, where trucks are placed at certain depots, waiting for the dispatch commands. Once an incident is detected or reported, the dispatch center will dispatch a truck to the incident location. In contrast, in FSP systems tow trucks spontaneously detect, respond and clear the incidents.

Previous research has examined the benefits of FSP. For example, Skabardonis et al. (1998) evaluated the FSP system on a 7.8 mile section of I-10 freeway (Beat 8) in Los Angeles, and reported that the services reduce incident duration in the order of 15 minutes and the B/C ratio is greater than 5 where benefits calculated include delay and fuel savings. Levinson et al. (2003) carried out a stated preference analysis to investigate the utility that FSP provide to an individual and found that the B/C ratio for the Los Angeles FSP is in the range of 6.2-6.3. Moore II et al. (2004) examined the prevailing assumption that FSP may reduce the likelihood of secondary accidents and concluded that secondary accidents on Los Angeles freeway are much less frequent than generally reported and avoiding secondary accidents provides only a small incentive to deploy FSP. However, the expected benefits associated with reducing already low secondary rates may be sufficient to justify the program.

It is well recognized that the deployment strategy of FSP services is the key to the success of the program. Design of the strategy involves determination of patrol beats, fleet size, allocation of the fleet among beats and hours of operations etc. In practice, the deployment is often made based upon engineering experience and judgment. In view of this, previous investigations have been conducted to develop simulation, statistical and optimization models to help the decision making.

Pal and Sinha (2002) and Ozbay and Bartin (2003) have developed simulation models that can be used for evaluating various FSP system configurations. Certainly, if a small number of alternatives can be predetermined, such simulation models can also be adopted to select the best deployment or expansion FSP strategy.

Davies et al. (2003) developed a tool to determine the B/C ratio for providing new FSP service to a freeway section or enhancing the existing service. Given the number of tow trucks on the section, the tool uses statistical models, derived from analysis of over 120 existing beats with 680,000 assists in California, to estimate the delay, fuel and emission savings per assisted incident, and consequently calculate the B/C ratio. The tool can be used to help decision making on where to implement the next service patrol. Khattak et al. (2003) developed another tool for the same purpose. The tool allows users to obtain statewide rankings for a freeway section based on three index criteria, and estimate the B/C ratio of implementing FSP on that section. The estimates are made primarily based upon statistical data and user inputs. Both of these decision-support tools focus on the facility level and localized impacts, lacking a systems perspective.

This chapter does not attempt to address all of the issues associated with the deployment strategy of FSP. It is only concerned with how to assign tow trucks to each beat to maximize the effectiveness of FSP services, given the setup of FSP beats and the fleet size. In the current practices, the allocation is made in a heuristic manner. Uniform allocation is sometimes adopted, or the fleet is allocated proportionally to the criteria like traffic volume, vehicle miles traveled and incident rates etc.

The most directly relevant research from the literature is Petty (1997) and Ozbay et al. (2004). Petty (1997) proposed a model for determining where to place tow trucks so as to maximize the expected reduction in congestion, based on traffic theory in combination with marginal benefit analysis. Ozbay et al. (2004) developed a mixed-integer programming model to determine the number of service vehicles assigned to each depot given the locations of depots, and the distribution of incident occurrences. Both studies assume a prior knowledge of incident occurrence distributions, and do not consider the interaction among system performance, incident occurrence, drivers' spontaneous responses to incidents, and the service intensities of FSP on various beats.

3.2 MODEL FORMULATION

3.2.1 DEFINITION OF THE PROBLEM

We consider a FSP fleet allocation problem for a general traffic network. Given a limited number of FSP tow trucks and the setup of FSP beats, the decision to make is to assign trucks to each beat to maximize the effectiveness of FSP services. Since the length of each FSP beat is fixed, more tow trucks patrolling on beats imply higher service intensity, quicker removal of incidents and thus less incident-induced delay. Since the impacts of the same incidents vary significantly from location to location, where to place the trucks is critical to the effectiveness of the services.

There exist a variety of ways to represent the effectiveness of FSP services, such as the expected reduction in congestion. Considering the fact that decision makers are mostly risk averse and more concerned with the high-consequence scenarios, this study attempts to determine the number of tow trucks patrolling on each beat to minimize the maximal system travel time that incidents may incur. In other words, we try to obtain a fleet allocation that achieves the least system total travel time under its corresponding worst-case scenario (note that the worst-case scenarios could be different for different fleet allocation plans).

3.2.2 TIME-INDEPENDENT MODELING FRAMEWORK

A time-independent modeling framework is used in this chapter. There are two reasons for adopting such a static framework. Firstly, as aforementioned, FSP work in a way different from the real-time dispatch systems, and thus the FSP fleet allocation is a "once for all" decision rather than a real-time one. Different patterns of service intensity or vehicle allocation could be determined for different times of day. Secondly, for each feasible vehicle allocation it is favorable to use a time-dependent analytical or simulation traffic model to evaluate the effectiveness of the alternative. However, such a model is always too complicated to be incorporated in the optimization procedure. Moreover, since the model developed in this chapter is intended for the planning purpose, details of traffic dynamics may not be the major concern at such a macroscopic level.

3.2.3 BASIC SETTINGS

Consider a network $G=(N,A)$, where *N* is the set of nodes, and *A* is the set of links. Let W be the set of all origin-destination (O-D) pairs in the network, R_w be the set of routes between O-D pair $w \in W$ and q_w be the demand between O-D pair *w*. Denote the total number of FSP beats as *I*, and B_i is the set of links that beat *i* comprises. Let f_r^w be the flow on route $r \in R_w$, $w \in W$, and v_a be the traffic flow on link $a \in A$. We thus have the following flow conservation equations:

$$v_a = \sum_{w \in W} \sum_{r \in R_w} f_r^w \delta_{ar}^w, a \in A \tag{3-1}$$

$$\sum_{r \in R_w} f_r^w = q_w, w \in W \tag{3-2}$$

$$f_r^w \geq 0, r \in R_w, w \in W \tag{3-3}$$

where $\delta_{ar}^w = 1$ if route *r* between O-D pair *w* uses link *a*, and 0 otherwise. Denote the travel time for each link $a \in A$ as $t_a(v_a, c_a)$, which is assumed to be an increasing/decreasing and strictly convex function of link flow v_a on that link/the capacity c_a of that link. Consequently, the route travel time is:

$$t_r^w = \sum_{a \in A} t_a(v_a) \delta_{ar}^w, r \in R_w, w \in W \tag{3-4}$$

where t_r^w is the travel time on route $r \in R_w$ between O-D pair $w \in W$.

Incidents may reduce capacities of freeways for a certain period of time. With a static modeling framework, to represent the impacts of incidents, we assume the capacity of each link varies within a certain range. Mathematically, for each link *a* we have:

$$c_a^0 - \varepsilon_a c_a^0 \le c_a \le c_a^0 + \varepsilon_a c_a^0 \tag{3-5}$$

where c_a^0 is the nominal link capacity and ε_a is the coefficient of link capacity uncertainty (variability), whose value depends on the characteristics of that link, such as frequency and severity of incidents and geometry. The value can be calibrated using historical incident data including locations, types and durations.

If the capacity of each link varies independently, the uncertainty set of link capacity pattern for the whole network will be a box. Recall that we are concerned with the worst case incurred by incidents. With a box uncertainty set, if Braess' paradox (1968) is not present, it is straightforward to identify the worst-case capacity pattern where each link has its minimal capacity. However, it is rare, if not impossible, that such a case would ever occur in reality. In other words, a box uncertainty set is too conservative. To be more realistic, we define an ellipsoidal set to confine the pattern of uncertain link capacity for a general network, written as:

$$C = \left\{ c \in R^{|A|} \middle| \sum_{a=1}^{|A|} (\varepsilon_a c_a^0)^{-2} (c_a - c_a^0)^2 \le 1 \right\} \tag{3-6}$$

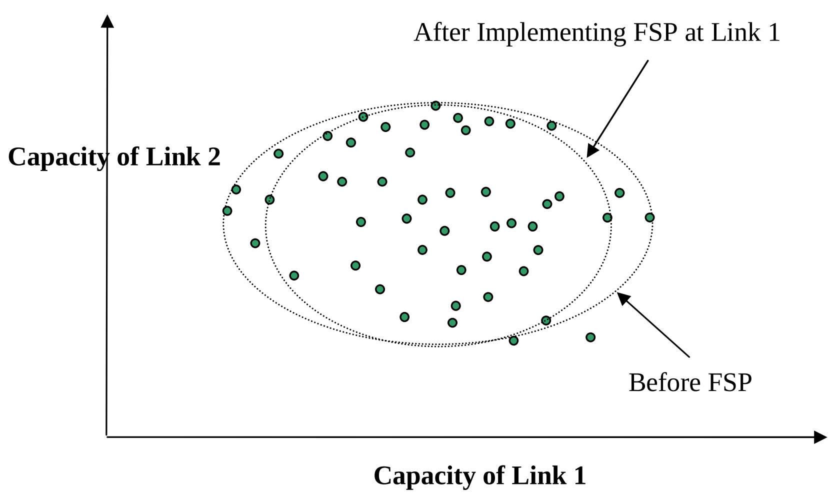
where $|A|$ is the dimension of set *A* (the total number of links). The set can be also written as: $C = \left\{ c \in R^{|A|} \middle| c = c^0 + M \cdot u, \|u\|_2 \le 1 \right\}$, where *M* is a diagonal matrix whose element is $\varepsilon_a c_a^0$. Note that such an ellipsoidal uncertainty has been widely used in a recent stream of research on robust optimization (Ben-Tal and Nemirovski, 2002; El Ghaoui, 2003). An ellipsoid may be given parametrically by observation data of moderate size.

We further assume that drivers have perfect information and always choose their routes with minimal travel times, and therefore network traffic flows would achieve a user equilibrium condition (Beckmann et al., 1956). For any realization of the uncertain link capacity pattern from the ellipsoidal set (caused by certain incidents), we have a corresponding network equilibrium problem, which can be solved to *approximately* estimate the resulting system travel time incurred by the incidents.

3.2.4 REPRESENTATION OF FSP IMPACTS

Since FSP trucks continuously patrol on beats looking for incidents to assist, they would be able to respond to incidents more quickly and thus reduce incident durations. Incidents may reduce freeway capacities, and consequently locations of incidents could become bottlenecks. Essentially, FSP are able to reduce durations of activation of the bottlenecks. With the current time-independent modeling framework, it is impossible to exactly replicate such impacts of FSP. Rather, we assume that FSP may shorten the range of the capacity variation of the link that the trucks are patrolling on. In other words, the impacts of the services are represented as reducing the variability of link capacity.

With an ellipsoidal uncertainty set of link capacity across the network, what the FSP service really change is the geometry of the set, as illustrated in Figure 3-1 for an illustrative two-link example, where the outer ellipse represents the uncertainty set before TSP while the inner ellipse is the set after implementing FSP at link 1. Therefore, different FSP fleet allocation may change the uncertainty set differently, leading to different worst-case scenarios realized from the corresponding sets. More precisely, the set *C* should be written as $C(z)$, where *z* is the vector of fleet allocation.



Legend: dots represent actual observations of link capacities

FIGURE 3-1 An Illustrative Example of FSP Impacts

It is easy to know that the intensity or frequency of the FSP tow trucks on each beat is given as:

$$h_i = \frac{z_i}{t_i} \qquad (3\text{-}7)$$

where h_i is the service intensity at beat *i*, z_i is the number of tow trucks assigned to beat *i*, and t_i is the round trip time of beat *i*. This round trip time is endogenously or exogenously calculated by considering round trip length, actual truck patrol speed, layover times at ends of trips and incident clearance time averaged across the trips made within the time period of interest.

We further represent the relationship between service intensity of FSP and link capacity variability. For $\forall a \in B_i$, the set of links that beat *i* comprises, we assume the following relationship:

$$\varepsilon_a = s(h_i) = s'(z_i) \qquad (3\text{-}8)$$

Where *s* or *s'* is a decreasing function of service intensity or number of assigned tow trucks, which could be continuous or discrete. The relationship is intuitively correct, and it is feasible to calibrate it from empirical data reported in Davies et al. (2004) and Dowling et al. (2004).

3.2.5 FORMULATION

As below we present a model to determine an optimal tow-truck fleet allocation that minimizes the maximal system travel time that incidents may incur. With the above considerations, the optimization problem can be written as below:

$$\min_{z} \max_{c \in C(z)} \sum_{a} v_a t_a(v_a, c_a) \qquad (3\text{-}9)$$

subject to:

$$\sum_{i}^{I} z_i \le Z \qquad (3\text{-}10)$$

where v_a is obtained by solving the following lower-level problem:

$$\min \sum_{a} \int_0^{v_a} t_a(\varpi, c_a) d\varpi \qquad (3\text{-}11)$$

subject to:

$$\sum_k f_k^w = d_w, \forall w \tag{3-12}$$

$$f_k^w \geq 0, \forall k, w \tag{3-13}$$

where Z is the number of total available trucks. In summary, the upper level problem also has the following two definition constraints respectively:

$$C = \left\{ c \in R^{|A|} \middle| \sum_{a=1}^{|A|} (\varepsilon_a c_a^0)^{-2} (c_a - c_a^0)^2 \leq 1 \right\} \tag{3-14}$$

$$\varepsilon_a = s'(z_i) \tag{3-15}$$

and the lower level problem satisfies the following definition constraint:

$$v_a = \sum_{w \in W} \sum_{k \in R_w} f_k^w \delta_{a,k}^w, a \in A \tag{3-16}$$

This is a min-max bi-level programming model. The upper level problem represents planners' behavior, determining tow truck allocation to minimize the maximal total travel time incurred by incidents. The lower level problem represents drivers' route choice behaviors, affected by the allocation decision from the upper level and capacity reductions caused by incidents.

Note that the variable z_i, the number of allocated trucks should be integer, and thus the above model should be an integer programming model. However, due to the computation difficulty, this chapter treats them as real numbers. This simplification does not necessarily impair the applicability of the model. In actual application, one could use the model to obtain optimal service intensity, and then marginally adjust layover times to provide the intensity with an integer number of vehicles, and thus eventually determine the fleet allocation.

3.3 SOLUTION ALGORITHM

The model (9)-(13) is non-convex, and thus only local optima can be found. There is no available solution algorithm for the model. Therefore, we propose a heuristic iterative algorithm to solve the model. Although the algorithm cannot be guaranteed theoretically to converge to a local optimum solution, actual applications normally show good convergence and results.

The iterative algorithm views the model (3-9)-(13) as a master problem with a slave problem. The master problem is written as:

$$\min_{z} J(z) \tag{3-17}$$

subject to:

$$\sum_{i}^{I} z_i \leq Z \tag{3-18}$$

where $J(z)$ is a non-convex function, defined by the optimal objective function of a bi-level programming slave model.

For solving the master problem, many efficient algorithms proposed in the literature of operations research could be applied. Unfortunately, since it is difficult to derive analytically the gradient $\nabla_z J(z)$ and only values of $J(z)$ can be made available, we could only use the sequential simplex method (Nelder and Mead, 1965) or apply iterative descent methods with finite differencing derivatives, such as the sequential quadratic programming algorithm (SQP, Han, 1976).

The slave bi-level programming model defining $J(z)$ is given as below:

$$J(z) = \max_{c} \sum_{a} v_a t_a(v_a, c_a) \tag{3-19}$$

subject to:

$$c \in C(z) \tag{3-20}$$

where v_a is obtained by solving the following lower-level problem:

$$\min \sum_{a} \int_0^{v_a} t_a(\varpi, c_a) d\varpi \tag{3-21}$$

subject to:

$$\sum_{k} f_k^w = d_w, \forall w \tag{3-22}$$

$$f_k^w \geq 0, \forall k, w \tag{3-23}$$

A number of algorithms can be applied to solve the slave problem, such as those proposed by Yang *et al.* (1994) and Chiou (2005). In this chapter, we apply the sensitivity-analysis-based iterative method by Yang *et al.* (1994). The basic idea of the algorithm is to formulate local linear approximation of the upper-level objective function using the derivative information from sensitivity analysis for equilibrium flows (Tobin and Friesz, 1998), and solve the resultant linear programming problems for a descent search direction. Therefore, the algorithm is in fact a sequence of linear approximation to the original problem.

In this study, with the ellipsoidal capacity uncertainty set, the local linear approximation to the original bi-level model turns out to be a quadratically constrained quadratic program (QCQP) as follows:

$$\max_{\|u\| \leq 1} \nabla_c T^T \cdot M \cdot u \tag{3-24}$$

where $\nabla_c T$ is the gradient of the total travel time with respect to link capacity, calculated by conducting the sensitivity analysis for user equilibrium flows. This QCQP can be analytically solved; the optimal objective value is $\sqrt{(M \cdot \nabla_c T)^T \cdot (M \cdot \nabla_c T)}$ and the optimal solution is $M \cdot \nabla_c T \Big/ \sqrt{(M \cdot \nabla_c T)^T \cdot (M \cdot \nabla_c T)}$.

3.4 NUMERICAL EXAMPLE

A numerical example is now presented to illustrate the proposed model. The example road network shown in Figure 3-2 has 13 nodes, 19 links and 4 O-D pairs, adopted from Nguyen and Dupuis (1984). The Bureau of Public Road link travel time function was used

$$t_a(v_a) = t_a^0 \left(1 + 0.15 \cdot \left(\frac{v_a}{c_a} \right)^4 \right) \tag{3-25}$$

And the network characteristics and O-D demand are given in Tables 3-1 and 3-2 respectively.

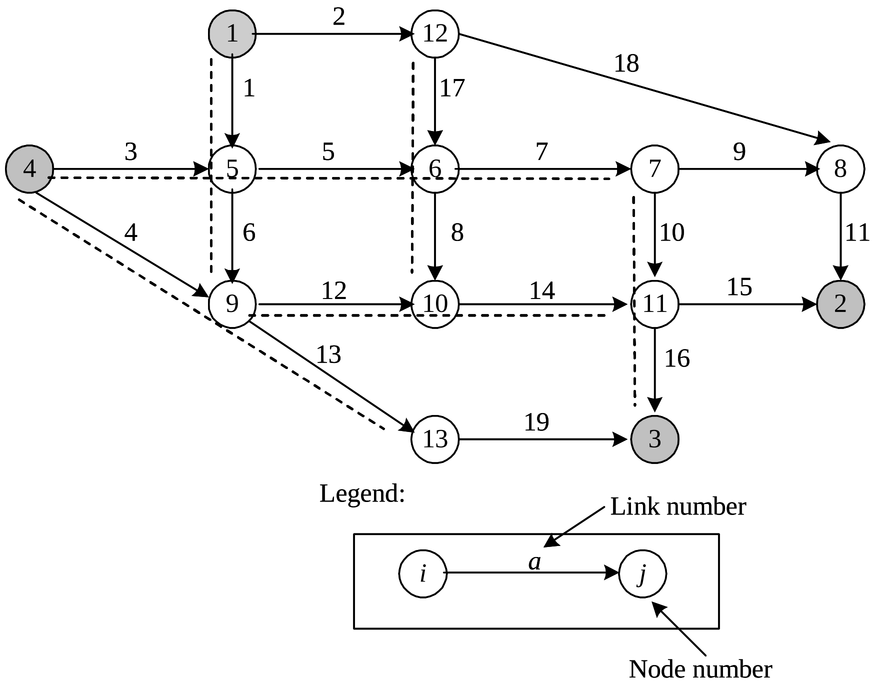


FIGURE 3-2 An Example Network

TABLE 3-1 Network Characteristics of the Example Network

Link a	t_a^0	c_a^0	ε_a^0
1	7.0	800	0.5
2	9.0	400	0.1
3	9.0	200	0.2
4	12.0	800	0.3
5	3.0	350	0.2
6	9.0	400	0.5
7	5.0	800	0.2
8	13.0	250	0.2
9	5.0	250	0.1
10	9.0	300	0.4
11	9.0	550	0.1
12	10.0	550	0.5
13	9.0	600	0.3
14	6.0	700	0.5
15	9.0	500	0.1
16	8.0	300	0.4
17	7.0	200	0.2
18	14.0	400	0.1
19	11.0	600	0.1

TABLE 3-2 O-D Travel Demand for the Example Network

O/D	2	3
1	400	800
4	600	200

In the example, the FSP impact function was assumed to be $s'_a(z) = \varepsilon_a^0 \cdot e^{-0.5 * z_i}$, where ε_a^0 is given in Table 3-1. The setup of beats is also illustrated in Figure 3-2, which is: Beat 1 = {links 1, 6}; Beat 2 = {links 3, 5, 7}; Beat 3 = {links 17, 8}; Beat 4 = {links 12, 14}; Beat 5 = {links 4, 13} and Beat 6 = {links 10, 16}.

A SQP subroutine with finite-differencing derivatives in Matlab was used to solve the master problem (3-17)-(18) as well as one self-programmed subroutine to solve the bi-level programming model (3-19)-(23).

We first examined the convergence of the iterative algorithm. Figure 3-3 plots the value of the objective function against outer iteration number of the SQP procedure, where the size of FSP fleet was 10. It can be observed that the outer iteration of the algorithm had a fast convergence; convergence was achieved in about 15 iterations. However, total computation was quite demanding due to the use of finite-differencing derivatives, which suggests that the iterative algorithm may not be applicable for a large-scale network. The resultant fleet allocation is 3.1, 1.5, 0, 0.7, 1.9 and 2.8 for Beats 1 to 6 respectively.

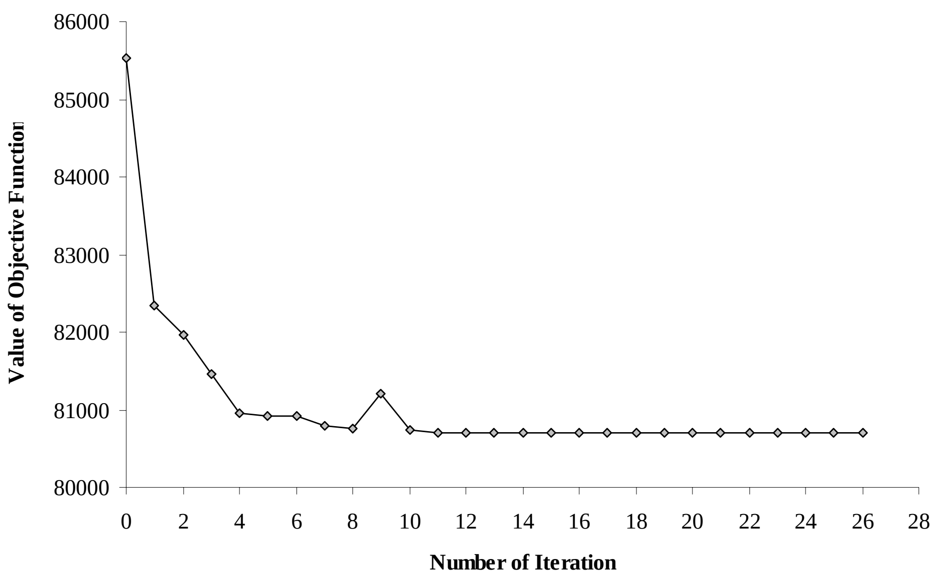


FIGURE 3-3 Convergence of the Iterative Algorithm for the Min-Max Model

To validate the effectiveness of the proposed model, we compared system performance that optimal and uniform allocation may result in. Here the total fleet was set as 6 and uniform allocation means one assigned truck on each beat. We computed differences of total travel times that uniform and optimal allocation could achieve under their corresponding worst cases. In order to examine the impacts of capacity uncertainty and network congestion, we varied the level of capacity uncertainty by multiplying ε_a^0 listed in Table 3-1 by an amplifier, changing from 0 to 1.6, and considered two demand levels: 100% and 70% of the demand given in Table 3-2.

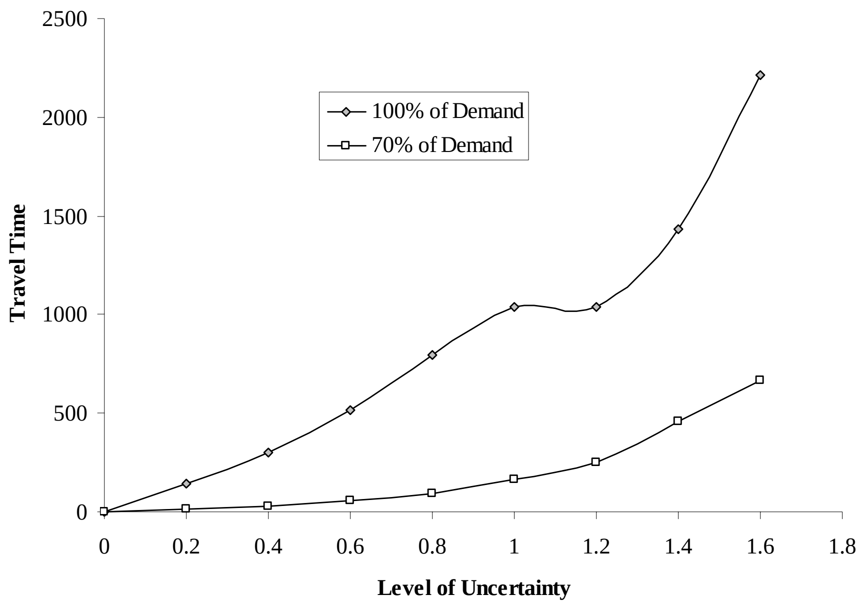


FIGURE 3-4 System Performance Differences between Uniform and Optimal Allocation (Total Travel Time of Uniform minus that of Optimal)

Figure 3-4 depicts performance differences of optimal and uniform allocation (the worst-case system travel time of uniform allocation minus that of optimal allocation) against varied values of the amplifier for two demand levels. The results are within expectation and intuitively realistic. We have two observations from the figure:

- The performance difference is always positive, suggesting that the optimal fleet allocation always performs better against the worst-case scenarios than the uniform allocation.

- The performance difference becomes more significant with higher levels of capacity uncertainty or network congestion, suggesting that the model may contribute more in the situations of high frequencies of incidents or high levels of network congestion.

4. CONCLUSIONS

This report has applied robust optimization methodology to determine robust strategies to improve the performances of highway systems, confronted with fluctuations in traffic demands or irregular incidents.

The first-half of the report has addressed network design with demand uncertainty. Three alternate models, sensitivity-based, scenario-based and min-max respectively have been presented to determine robust optimal improvement schemes for road networks under demand uncertainty. The usefulness and validity of these three models have been demonstrated through numerical examples and simulation tests. It is suggested that if decision makers aim to achieve a mean-variance tradeoff, the sensitivity-based and scenario-based models should be used with particular caution placed on the minimization of the sensitivity of total travel time with respect to demand perturbations or minimization of the variance of total travel times under various demand scenarios. If fluctuations of travel demand are believed to be non-significant, the sensitivity-based model is more appropriate to use, because it is simpler and requires much less computation efforts. Otherwise, the scenario-based model should be used, and additional efforts are needed to generate the demand scenarios and determine the corresponding probabilities of occurrence. On the other hand, if decision makers are more concerned with the worst-case scenarios, the min-max model should be applied.

It should be pointed out that although only demand uncertainty was of concern in the three models, the proposed modeling framework is quite general and is applicable to accommodate other types of uncertainty, such as incident-induced travel time uncertainty. As aforementioned, non-recurrent events, such as incidents will reduce the effective road capacity and deteriorate the network performance rapidly. Transportation network design may be improved significantly if these impacts can be predicted. To address incident-induced travel time uncertainty (or more specifically, link capacity uncertainty), we can add another additional component to the objective function of the sensitivity-based model to minimize the sensitivity of total travel time to the link capacity, generate another set of scenarios to approximate the link capacity uncertainty and then incorporate them into the scenario-based model, or add another additional optimization layer to identify the worst-case capacity scenarios into the min-max model.

The second-half of the report has investigated how to make use of FSP systems to improve the robustness of the highway systems. We have presented a min-max bi-level programming model to determine the optimal fleet allocation strategy for FSP services. A heuristic solution algorithm has been proposed to solve the model. Both the model formulation and the solution algorithm have been validated and demonstrated through a numerical example. It has been observed that the robust optimal fleet allocation always performs better against the worst-case scenarios. Moreover, the performance difference becomes more significant with higher levels of capacity uncertainty or network congestion, suggesting that the model may contribute more in the situations of high frequencies of incidents or high levels of network congestion.

In summary, this report has delivered a proof of concept that optimal planning and management strategies can be formulated through applying robust optimization methodology such that limited resources could be allocated more rationally, and reliability of a highway network improved more efficiently.

Further research would be applying the robust optimization concept to help Metropolitan Planning Organizations to refine their regional transportation planning to further improve the robustness or reliability of road systems at the regions or to allow Departments of Transportation to schedule maintenance activities in networks in a manner that is "robust" to demand uncertainties or incidents. The concept can also be adopted to fine-tune the existing FSP systems at specific locations, such as the San Francisco Bay Area or Los Angeles or to determine optimal deployment strategies of new FSP systems. Several tasks will be needed when applying the concept: 1) validation of the assumptions of the model formulations and calibration of parameters by using actual data; 2) development of more efficient algorithms for the models; 3) extension of the FSP fleet allocation model to simultaneously determine the setup of beats and fleet allocation; 4) data collection and evaluation.

REFERENCE

Bell, M.G.H. (2000) A Game Theory Approach to Measuring the Performance Reliability of Transport Networks. *Transportation Research*, Vol.34B, pp.533-545.

Beckmann, M.J., McGuire, C.B. and Winsten, C.B. (1956) *Studies in the Economics of Transportation*. Yale University Press, New Haven, Conn.

Ben-Tal, A. and Nemirovski, A. (2002) Robust Optimization – Methodology and Applications. *Mathematical Programming*, Series B 92, pp. 453-380.

Borcherds, P.H. (2000) Importance Sampling: An Illustrative Introduction. *European Journal of Physics*, Vol.21, pp.405-411.

Braess, D. (1968) *Über ein paradox der verkerhsplannung*, Unternehmenstorchung, 12, 258-268.

Cantarella G.E. (1997) A General Fixed-Point Approach to Multimode Multi-user Equilibrium Assignment with Elastic Demand. *Transportation Science*, Vol.31, No.2, pp.107-128.

Cascetta, E., Nuzzolo, A., Russo, F.and Vitetta, A. (1996) A Modified Logit Route Choice Model Overcoming Path Overlapping Problems: Specification and Some Calibration Results for Interurban Networks. *Proceedings of 13th International Symposium on Transportation and Traffic Theory*, eds J.B. Lesort, Lyon, France, July 24-26, Pergamon, pp.697-711.

Chen, A., Yang, H., Lo, H.K. and Tang, W.H. (2002) Capacity Reliability of A Road Network: An Assessment Methodology and Numerical Results. *Transportation Research*, Vol.36B, pp.225-252.

Chen, A., Subprasom, K. and Ji, Z. (2003) Mean-Variance Model for the Build-Operate-Transfer Scheme under Demand Uncertainty. *Transportation Research Record*, 1857, National Research Council, Washington, D.C., pp.93-101.

Chiou, C.W. (2005) Bilevel Programming for the Continuous Transport Network Design problem. *Transportation Research*, Vol.39B, 361-383.

Clark, S. and Watling, D. (2004) Modeling Network Travel Time Reliability under Stochastic Demand. *Transportation Research*, Vol.39B, pp.119-140.

Daganzo C.F. (1983) Stochastic Network Equilibrium with Multiple Vehicle Types and Asymmetric, Indefinite Link Cost Jacobians. *Transportation Science*, Vol.17, No.3, pp.282-300.

Davies, L., Mauch, M. and Skabardonis, A. (2004) Freeway Service Patrol Ppredictor Model: Methodology and Implementation. *83rd Annual Meeting Compendium of Papers CD-ROM*, Transportation Research Board, Washington, D.C.

Davis, G.A. (1994) Exact Local Solution of the Continuous Network Design Problem via Stochastic User Equilibrium Assignment. *Transportation Research*, Vol.28B, pp.61-75.

Dowling, R., Skabardonis, A., Carroll, M. and Wang, Z. (2004) Methodology for measuring recurrent and non-recurrent traffic congestion. *Transportation Research Record*, 1867, 60-68.

El Ghaoui, L. (2003) *A Tutorial on Robust Optimization*, Presentation at the Institute of Mathematics and Its Applications, March 11, 2003. Available at http://robotics.eecs.berkeley.edu/~elghaoui/

Federal Highway Administration (2003) *Status of the Nation's Highways, Bridges, and Transit:2002 Conditions and Performance Report - Chapter 10: Sensitivity Analysis.*

Fisk, C. (1979) More Paradoxes in the Equilibrium Assignment Problem. *Transportation Research*, Vol.13B, pp. 305-309.

Fenno, D. and Odgen, M. (1998) Freeway Service Patrols: A State of Practice. *Transportation Research Record* 1643, Transportation Research Board, Washington, D.C., 28-38.

Han, S. P. (1976) Superlinearly Convergent Variable Metric Algorithms for General Nonlinear Programming Problems. *Mathematical Programming*, 11, 263-282.

Khattak, A, Rouphail, N, Monast, K and Havel, J (2004) Method for Priority-Ranking and Expanding Freeway Service Patrols, *Transportation Research Record*, 1867, 1-10.

Koppelman, F.S. and Wen, C.H. (2000) The Paired Combinatorial Logit Model: Properties, Estimation and Application. *Transportation Research*, 34B, 75-89.

Kouvelis, P. and Yu, G (1997) *Robust Discrete Optimization and its Applications*. Kluwer Academic Publisher, Norwell, MA.

Laguna, M. (1998) Applying Robust Optimization to Capacity Expansion of One Location in Telecommunications with Demand Uncertainty. *Management Science*, Vol.44, No.11, pp.101-110.

Levinson, D., Gillen, D. and Parthasarathi, P. (2003) The Insurance Value of Freeway Service Patrols: A Stated Preference Analysis. *82nd Annual Meeting Compendium of Papers CD-ROM*, Transportation Research Board, Washington, D.C.

Linderoth, J., Shapiro, A. and Wright, S. (2004) The Empirical Behavior of Sampling Methods for Stochastic Programming. *Optimization Technical Report 02-01*, Computer Sciences Department, University of Wisconsin-Madison.

List, G.F., Wood, B., Nozick, L.K., Turnquist, M.A., Jones, D.A., Kjeldgaard, E.A. and Lawton, C.R. (2003) Robust Optimization for Fleet Planning under Uncertainty. *Transportation Research*, Vol.39E, pp.209-227.

Meng, Q., Lee, D.H., Yang, H. and Huang, H.J. (2004) Transportation Network Optimization Problems with Stochastic User Equilibrium Constraints. *Transportation Research Record*, 1882, National Research Council, Washington, D.C., pp.113-119.

Moore II, J., Giuliano, G. and Cho, S. (2004) Secondary Accident Rates on Los Angeles Freeways. *Journal of Transportation Engineering, ASCE*, Vol. 130, No.3, 280-285.

Morris, M. and Lee, W. (1994) Survey of Efforts to Evaluate Freeway Service Patrols. *Transportation Research Record 1446*, Transportation Research Board, Washington, D.C., 77-85.

Mulvey, J.M., Vanderbei, R.J. and Zenios, S.A. (1995) Robust Optimization of Large-Scale Systems. *Operation Research*, Vol.43, No.2, pp.264-281.

Nelder, J. A. and Mead, R. (1965) A Simplex Method for Function Minimization. *Computer Journal*, 7, 308-313.

Nguyen, S. and Dupuis, C. (1984) An Efficient Method for Computing Traffic Equilibria in Networks with Asymmetric Transportation Costs. *Transportation Science*, 18, pp. 185–202.

Ordonez, F. and Zhao, J. (2004) Robust Capacity Expansion of Transit Networks. USC-ISE Working Paper #2004-01, University of Southern California.

Ozbay, K. and Bartin, B. (2003) Incident Management Simulation. *Simulation*, Vol.79, No.2, 69-82.

Ozbay, K., Xiao, W. and Iyigun, C. (2004) Probabilistic Programming Models for Response Vehicle Dispatching and Resource Allocation in Traffic Incident Management. *83rd Annual Meeting Compendium of Papers CD-ROM*, Transportation Research Board, Washington, D.C.

Pal, R. and Sinha, K. (2002) Simulation Model for Evaluating and Improving Effectiveness of Freeway Service Patrol Programs. *Journal of Transportation Engineering, ASCE*, Vol.128, No.4, 355-365.

PB Farradyne (2000) *Traffic Incident Management Handbook*. Report USDOT-13286, US DOT, 2000.

Petty, K.F. (1997) Incidents on the Freeway: Detection and Management. *Doctoral Dissertation*, Department of Electrical Engineering and Computer Science, University of California, Berkeley.

Prashker, J.N. and Bekhor, S. (1998) Investigation of Stochastic Network Loading Procedures. *Transportation Research Record*, 1645, 94-102.

Sherali H. and Subramanian, S. (1999) Opportunity Cost-Based Models for Traffic Incident Response Problems. *Journal of Transportation Engineering*, Vol.125, 176-185.

Skabardonis, A., Petty, K., Varaiya, P. and Bertini, R. (1998) Evaluation of Freeway Service Patrol at Los Angeles Freeway Site. *PATH Research Report, Report UCB-ITS-PRR-98-13*, University of California, Berkeley.

Taylor, M.A.P. (1999) Dense Network Traffic Models, Travel Time Reliability and Traffic Management. II: Application to Reliability. *Journal of Advanced Transportation*, Vol.33, No.2, pp. 235-244.

Tobin, R.L. and Friesz, T.L. (1988) Sensitivity Analysis for Equilibrium Network Flow. *Transportation Science*, 22, No.4, 242-250.

Waller, S.T., Schofer, J.L. and Ziliaskopoulos, A.K. (2001) Evaluation with Traffic Assignment under Demand Uncertainty. *Transportation Research Record*, 1771, National Research Council, Washington, D.C., 2001, pp.69-74.

Waller, S.T. and Ziliaskopoulos, A.K. (2001) Stochastic Dynamic Network Design Problem. *Transportation Research Record*, 1771, National Research Council, Washington, D.C., 2001, pp.106-113.

Yang, H. and Bell, M.G.H. (1998) Models and Algorithms for Road Network Design: A Review and Some New Developments. *Transport Review*, 18, No.3, pp. 257-278.

Yang, H., Yagar, S., Iida, Y. and Asakura Y. (1994) An Algorithm for the Inflow Control Problem on Urban Freeway Networks with User-Optimal Flows. *Transportation Research*, 28B, pp.123-139.

Yin, Y. (2000) Genetic-Algorithm-Based Approach for Bi-level Programming Models. *Journal of Transportation Engineering, ASCE*, Vol. 126, No.2, pp.115-120.

Yin, Y. and Ieda, H. (2002) Optimal Improvement Scheme for Network Reliability. *Transportation Research Record*, 1783, National Research Council, Washington, D.C., pp.1-6.

Ying, J.Q. and Miyagi, T. (2001) Sensitivity Analysis for Stochastic User Equilibrium Network Flows – A Dual Approach. *Transportation Science*, Vol.35, pp.124-133.