UNIVERSITY OF CALIFORNIA,
IRVINE

Designing Environment-Oriented Pricing and Traffic Rationing Schemes for Travel Demand Management

DISSERTATION

submitted in partial satisfaction of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

in Civil Engineering

by

Daniel Rodríguez Román

Dissertation Committee:
Professor Stephen G. Ritchie, Chair
Professor R. Jayakrishnan
Professor Michael G. McNally

2015

DEDICATION

To my parents and sister

Table of Contents

List of Figures

List of Tables

Acknowledgements

I would like to acknowledge the guidance of my advisor, Professor Stephen G. Ritchie. I appreciate the confidence that he placed in me when, not long after my arrival at UCI, he asked me to join the CSFFM project. He was also very supportive as I explored possible research directions. I am also grateful for the terrific instruction of Professors Will Recker, R. Jayakrisnan, Michael McNally, and Wenlong Jin. I am indebted to my friend, Dr. Joseph Chow, for he has always been available to discuss research ideas, offer suggestions on my work, and provide encouragement. To my friends at ITS and the rest of UCI, thank you for making my time in Irvine so memorable.

I would also like to acknowledge the guidance provided during my undergraduate education by Professors Alberto M. Figueroa and Benjamin Colucci, and Ms. Gisela Gonzalez. Ms. Gonzalez also provided helpful advice at important junctures of my graduate life. In this regard, I also acknowledge the help of Professor Ismael Pagan.

I am grateful for the economic support provided by the University of Puerto Rico, the University of California, UCCONNECT, and Caltrans. Finally, I thank my parents, sister, and my uncle Ángel Emilio for their support.

Curriculum Vitae

Education

2015	Ph.D. in Civil Engineering (Transportation Systems Engineering) University of California, Irvine
2010	M.S. in Civil Engineering (Transportation Engineering) University of California, Berkeley
2009	B.S. in Civil Engineering University of Puerto Rico, Mayaguez

Research Experience

2011 – 2013	California Statewide Freight Forecasting Model University of California, Irvine
2009 – 2010	Survey of Self-Declaration Systems and Implementations University of California, Berkeley

Publications

Ranaiefar F, Chow JYJ, Rodríguez Román D, Camargo PV, Ritchie SG. 2013. Geographic scalability and supply chain elasticity of a structural commodity generation model using public data. Transportation Research Record, 2378:73–83.

Rodríguez Román D, Masoud N, Kyungsoo J, Ritchie SG. 2014. Goal-programming approach to allocate Freight Analysis Framework mode flow data. Transportation Research Record 2411:82–89.

Conference Proceedings

Masoud N, Ranaiefar F, McNally M, Rodríguez Román D, and Ritchie SG. 2014. An alternative method to estimate balancing factors for the disaggregation of OD matrices. 93rd Annual Meeting of the Transportation Research Board, Washington, D.C.

Rodríguez Román D, Ritchie SG. 2015. Accounting for population exposure to pollutants in the toll design problem. 94th Annual Meeting of the Transportation Research Board, Washington, D.C.

Abstract of the Dissertation

Designing Environment-Oriented Pricing and Traffic Rationing Schemes for Travel Demand Management

By

Daniel Rodriguez Roman

Doctor of Philosophy in Civil Engineering

University of California, Irvine, 2015

Professor Stephen G. Ritchie, Chair

Optimization-based approaches are presented for the design of environment-oriented road pricing and traffic rationing schemes, particularly with the objective of curbing human exposure to motor vehicle generated air pollutants. In addition, surrogate-based solution algorithms are developed to accelerate the search of good solutions for the problems considered.

A toll design problem is proposed for selecting tolling locations and levels that minimize environmental inequality and human exposure to pollutants, subject to budget constraints and pollutant concentration constraints at receptor points. A mixed-integer variant of the metric stochastic response surface algorithm and a hybrid genetic algorithm-metric stochastic heuristic are presented to solve the mixed integer toll design problem. Numerical tests suggest that the proposed algorithms are promising solution methods for transportation network design problems.

In addition, an optimization problem is presented for the design of cordon and area-based road pricing schemes subject to environmental constraints. Flexible problem formulations are considered which can be easily utilized with state-of-the-practice transportation planning models. A surrogate-based solution algorithm that utilizes a geometric representation of the charging area boundary is proposed to solve cordon and area pricing problems.

Lastly, a bi-objective traffic rationing problem is considered where the planner attempts to maximize auto usage while minimizing pollutant exposure inequality, subject to constraints on the levels of greenhouse gas emissions and pollutant concentration levels. A surrogate-assisted differential evolution algorithm for multiobjective continuous optimization problems with constraints is proposed.

Chapter 1

Introduction

Considerable advances have been made in the past decades in the reduction of vehicle generated pollutants. Yet, despite these improvements, cutting motor vehicle emissions and their related negative effects remains a priority for societies around the world. The reason for this concern is that, even in countries with stringent environmental regulations and the resources to benefit from the latest vehicle technologies, motor vehicles continue to be a major source of pollutants that contribute to anthropogenic climate change and air quality degradation. For example, in the US road vehicles generate 23 percent of greenhouse gas (GHG) emissions (EPA, 2014), and are estimated to be the largest contributors to ozone and particulate matter-related premature mortalities (Caiazzo et al., 2013). Worldwide, road emissions are estimated to account for 10 percent of GHG emissions (OECD/ITF, 2010), and to cause the deaths of more than 150,000 people each year (World Bank, 2014). In light of the threat to public health and the environment posed by vehicle emissions, several vehicle emissions control strategies have been proposed and implemented. Among these strategies is the use of road pricing and quantity control schemes to manage travel demand.

In the travel demand management (TDM) context, pricing refers to the use of price signals, such are tolls or taxes, to influence the transportation decisions of individuals or firms. These decisions could include when and where to enter a transportation network, what mode to use, or what type of vehicle to purchase. London's Low Emission Zone (LEZ) program and British Columbia's fuel carbon tax are examples of real-world implementation of emission pricing schemes. In the LEZ, the price signal is a daily charge levied based on a vehicle's emission classification, while in British Columbia it is taxes levied on fuels based on their carbon content.

TDM quantity control instruments set (or cap) the quantity of a polluting activity, as opposed to setting the price for the activity. TDM control schemes are somewhat analogous to cap-and-trade schemes. A famous example of an implemented quantity control scheme is Mexico City's "no driving day program"

(Hoy No Circula, HNC) (Goddard, 1997). The HNC program restricts the days private vehicles can circulate in the city. More recently, German cities have implemented Low Emissions Zones where only low emitting PM10 vehicles are allowed access to city centers (Wolf and Perry, 2010). Cities like Paris, Milan, and Beijing have also imposed strict traffic rationing schemes to combat air pollution.

The international experience with pricing and traffic rationing schemes shows that these strategies can be effective in reducing traffic congestion and pollution. Their political acceptance, however, constitutes a significant challenge. This resistance partly stems from the considerable uncertainties surrounding the strategies' economic, social, and environmental effects. Hence, the importance of the planning and design process of pricing and rationing schemes. In this regard, optimization-based analysis methods can be a useful tool in the TDM planning and design process, as this analysis approach can provide guidance on what courses of action could be most effective in achieving a schemes' intended objectives. In particular, discrete transportation network design problems, and related solution algorithms, have been proposed to aid planners in answering a variety of network management questions. Among these questions is how to manage travel demand in order to mitigate the environmental impacts of vehicle emissions. Existing models can be used to determine, for example, toll levels on roads that minimize network emissions (e.g., Yin and Lawphongpanich 2006), tolls levels that ensure link emission compliance with environmental constraints (e.g., Ferrari 1995), and the distribution of travel permits among network users to achieve a series of suitability related goals (e.g., Nargurney 2000a).

In this dissertation, optimization problems are proposed that extend the scope of previous environment-oriented transportation network design problems. The proposed problems can be used to determine the location and level of tolls according to pollutant exposure and environmental equity objectives, to design area-based charging schemes that help curb pollutant concentration in cities, and to select traffic rationing levels that both maximize auto trips in a city and minimized air pollutant intake inequality in a population. In addition, new surrogate-based solution algorithms are proposed to solve the presented problems. Network design problems are notoriously difficult to solve, in part, because of the time consuming nature of the models utilized to evaluate candidate solutions. Surrogate based-solution algorithms attempt to

address this computational problem by utilizing computationally inexpensive approximations of the computationally expensive problems to guide the search for good solutions. The proposed algorithms are derivative-free, so they can easily be utilized to solve other optimization problems with similar characteristics to the ones presented here.

A motivating example is presented in the next section. This section is intended to illustrate why an optimization-based perspective is potentially useful in the design of environment-oriented pricing and traffic rationing schemes. The example also highlights a few of the primary objectives that are considered in the design of TDM strategies intended to improve environmental sustainability. This is followed by a brief introduction to the surrogate-based optimization literature in transportation. The chapter's last section provides an overview of the structure and content of this dissertation.

1.1. Motivating Example: A Pollutant Exposure Pseudo-Paradox

Environment-oriented road pricing research is partly motivated by the observation that minimizing traffic congestion does not necessarily reduce emissions; in fact, the opposite could be the case. To illustrate this point, Nagurney (2000b) presented three emission paradoxes that demonstrate how strategies aimed at improving network performance, such as adding a road or decreasing travel demand, may result in an increase in total system emissions. The emission paradoxes highlighted the importance of incorporating network topology, cost structures, and levels of travel demand in the design of policies aimed at reducing traffic emissions. In this section an additional example is presented to illustrate the importance of explicitly accounting for population information and meteorological uncertainty when designing demand management policies aimed at mitigating the adverse effects of vehicle emissions.

Consider the network presented in Figure 1-1. Traffic flows from node 1 to node 2 along two 10-mile links: link a and link b. Travel times on the links are estimated using the Bureau of Public Roads (BPR) link performance function, as shown in the figure. It is assumed that the free flow travel time in both links is 9.2 minutes, the capacities of links a and b are 5000 vehicles per hour (vph) and 3000 vph, respectively, and there is a fixed travel demand of 8000 vph.

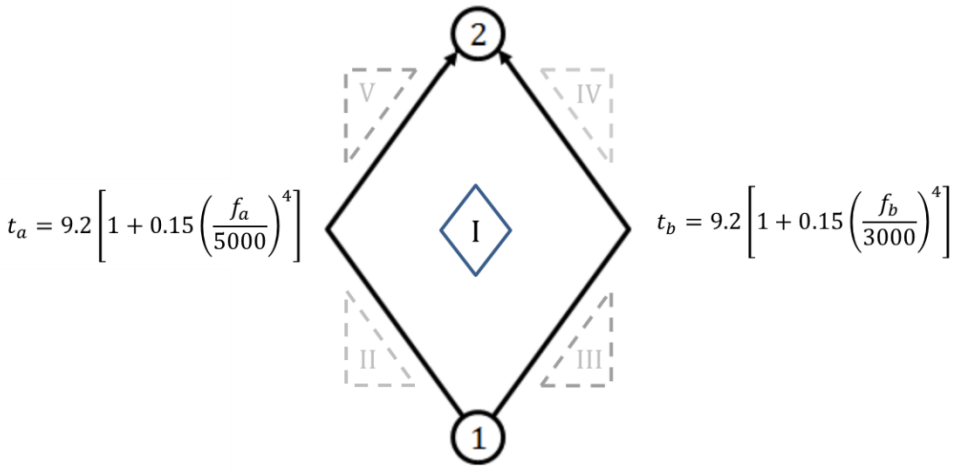


Figure 1-1 Network topology for toy example

Link emissions (e, in grams per vehicle) of carbon monoxide (CO) are computed using the following speed-based model (Sugawara and Niemeier 2002).

$$e = \exp[1.8 - 0.05627y + 2.1(10)^{-3}y^2 - 7.6(10)^{-5}y^3 + 1.2(10)^{-6}y^4] \times l \quad (1.1)$$

where y is equal to the link velocity minus 16, and l is the link's length. Eastward winds transport link a's emissions to population zone I, the only populated zone in the regions (zones II through V are uninhabited). A transfer coefficient of 0.0001 is multiplied to link a's emissions to compute the traffic-related pollutant concentration in zone I. Link b's emissions do not affect zone I given the wind direction; if westward winds were assumed the situation would be reverse (i.e., only link b's emissions would affect zone I).

Consider three network performance indicators: total system travel time (TT), total emissions (TE), and total population exposure to pollutants (EP). Figure 1-2 presents the normalized TT, TE, and EP curves given different link loadings. The abscissa shows the fraction of the total demand assigned to link a. If deterministic user-optimized routing behavior is assumed, the network is in equilibrium when five eighths of the total demand travels on link a. The deterministic user equilibrium (DUE) solution equals the system optimal solution, so a manager interested exclusively in reducing total travel time could not improve the network's performance. Now consider two additional myopic managers: the TE manager, who is only

concerned with minimizing total emissions, and the EP manager, who is only concerned with minimizing population exposure to air pollutants. The optimal network performance from the TE manager's perspective would occur if all the demand was assigned to link a. Relative to DUE flows, assigning all the demand to link a would result in a 13.2 percent decrease in system emissions. However, assigning all the demand to link a would also result in a 38.8 percent increase in zone I's pollutant concentration and a 72.4 percent increase in total travel time. Conversely, from the EP manager's perspective all the demand should be assigned to link b since this would completely eliminate the carbon monoxide concentration in zone I. Yet, assigning all the demand to link b would increase total emissions by 210 percent and total travel time by 647 percent.

Note that the EP manager's objective function value, and the manager's resulting optimal intervention (e.g., a toll), is highly dependent on meteorological conditions. For example, the TE and EP managers' optimal strategies would align if the wind blew from east to west. Given that meteorological conditions vary from day to day, even from second to second, an EP manager should be interested in incorporating the variability of wind speed, wind direction, and atmospheric stability, among other factors, as part of a robust policy evaluation process. Furthermore, the location, density, and characteristics of the populations adjacent to the network would affect the design of the EP manager's interventions. Even each zone's population activity patterns and the building characteristic (e.g., ventilation) would also be relevant information in the process of reducing human exposure to air pollutants.

If zones II through V were also populated, and each zone had different population groups and density levels, the EP manager might be interested in considering a comprehensive population exposure measure, in addition to air pollutant concentration, in the design of the travel demand management strategy. The planner also faces an environmental equity problem, as shifting traffic flow from one link to the other would transfer pollutant concentrations between the different zones. Assuming eastward winds and that the population in zone I is considerably larger than the combined populations in zones II and V, it might still be optimal for the EP manager to shift most of the traffic flow to link b, but that would undoubtingly result

in an increase in the concentrations levels of zones II and V. The question then becomes if the redistribution in emissions and pollutant concentrations is fair.

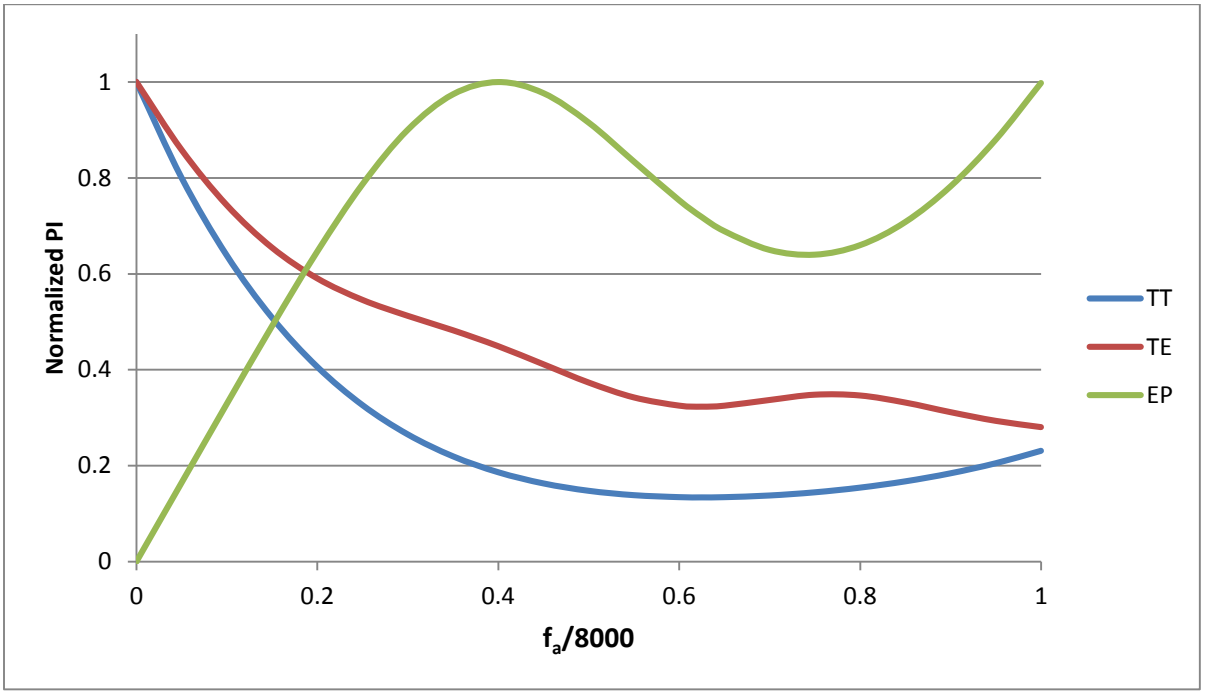


Figure 1-2 Normalized performance indicators as a function of f_a

Obviously, network managers are not myopically concerned with a single objective. What this toy example illustrates, however, is that environment-oriented demand management interventions could have unintended consequences, as there might be tradeoffs in the reduction of congestion, emissions, pollutant concentrations, and environmental equality. Particularly in regards to the latter two issues, the problems proposed in this document extend previous road pricing and traffic rationing models, providing planners additional methods for designing road pricing and rationing schemes that explicitly account for population exposure to pollutants.

1.2. Applications of Surrogate-Based Solution Algorithms to Transportation Optimization Problems

Computer-based models are commonly part of contemporary engineering design optimization problems. Given the time consuming nature of these models, considerable computational cost is incurred in finding good candidate designs. This computational challenge is commonly encountered in transportation network design problems (NDPs), which usually employ multiple types of interconnected computer models to simulate network users' behaviors and the resulting traffic conditions in a road network. In traditional NDPs intended for congestion mitigation, the computational cost of finding optimal solutions can be attributed primarily to the run times of traffic assignment models. In the case of environment-oriented NDPs, these computational costs are compounded by the introduction of additional models to compute, for example, link emissions and pollutant concentrations.

Surrogate-based optimization procedures have been developed in response to the computationally intensive nature of computer-based optimization problems. In the context of optimization algorithms, surrogate models (also known as metamodels or response surface models) are models that provide computationally inexpensive approximations to computationally expensive models. Commonly used models for constructing surrogates include polynomial regression models, kriging models, radial basis functions, and support vector machines, among others (Queipo et al. 2005, Forrester and Keane 2009). Surrogate models could be used, for example, to quickly screen for promising candidate solutions before actually evaluating any designs with the time consuming traffic assignment and related models. This in turn reduces the number of model runs necessary in the search for good solutions. In some methodologies, the surrogate completely substitutes the computer-based model in the optimization process (Queipo et al. 2005).

Particularly in the last five years, there is growing interest in the field of transportation engineering in using surrogate-based optimization methods to solve network design problems. In Table 1-1 a list of surrogate model applications for transportation optimization problems is presented. Previous studies have developed surrogate-assisted solution algorithms for continuous (e.g., Chow et al. 2010, Chen et al. 2013)

and discrete (e.g., Xiong and Schneider 1994, Wismans 2012) NDPs. Only one algorithm has been proposed for a mixed integer problem, in a problem where there is no dependence between the integer and continuous variables. Also, the algorithm proposed by Chow and Regan (2014) appears to be the only transportation application with a multiobjective problem.

Table 1-1 Previous applications of surrogate-based optimization in transportation NDPs

Authors	Year	Variable	Problem
Xiong and Schneider	1994	Integer	Network expansion
Chow et al.	2010	Continuous	Capacity expansion
Osorio and Bierlaire	2010	Continuous	Signal control
Fiske	2011	Integer	Dynamic traffic management
Wismans	2012	Integer	Dynamic traffic management
Chow and Regan	2014	Continuous	Road pricing
Lamotte	2014	Continuous	Signal control
Chen et al.	2014a	Continuous	Road pricing
Chen et al.	2014b	Continuous	Toll and transit fare optimization
Chen et al.	2015	Mixed	Network and capacity expansion

Besides the proposed environment-oriented NDPs, a second contribution of this dissertation are four surrogate-based solution algorithms that can be used to find good solutions for mixed integer optimization problems, like the toll design problem, for cordon and area based pricing problems, or for constrained multiobjective continuous problems, as the proposed traffic rationing problem. The general strategy of the proposed algorithms is presented in Figure 1-3. The initial step of the algorithms is to generate an initial set of candidate solutions that are then evaluated with the computationally expensive models. Using the model outputs as independent variables and the candidate design solutions as the dependent variables, a surrogate model is constructed. The surrogate model is used to identify the most promising solutions that are evaluated in each iteration. The model is updated as new information is acquired from the model evaluation of candidates. This process continues until a convergence criterion is met, which is defined as a maximum number of iterations in all the application presented.

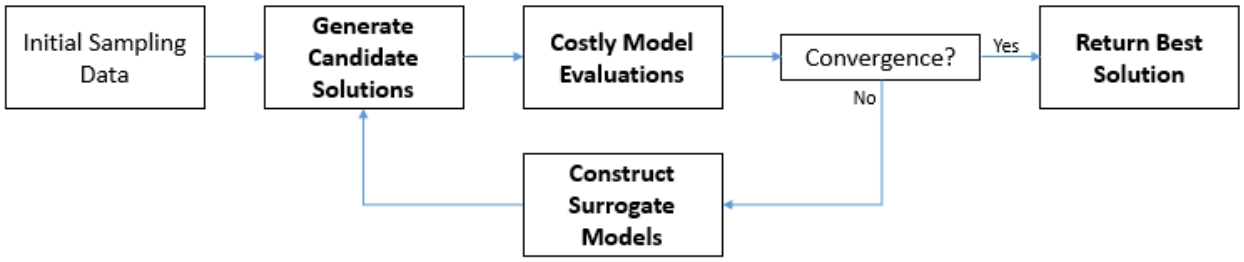


Figure 1-3 General structure of proposed surrogate-based solution algorithms

1.3. Research Objectives and General Outline of Dissertation

The general objective of this work is to propose new optimization problems that can be used by planners to design road pricing and traffic rationing schemes with the objective of reducing human exposure to vehicle generated pollutants. The pollutant exposure minimization objective is defined either explicitly (in the form an objective function) or implicitly (in the form of constraints). In addition, since the proposed optimization problems would undoubtedly be computationally challenging to utilize in any real-world applications, surrogate-based solution algorithms are presented that could help analysts find good solutions in relatively short amount of time. The following three chapters have the same general structure. The chapters open with an introduction to the problem of interests, followed by a review of relevant literature, the formulation of the optimization problem, a discussion of the surrogate-based solution algorithm proposed for the problem, and, lastly, numerical tests with applications of the problem and its solution algorithm.

In Chapter 2, a toll design problem is proposed that accounts for population exposure to pollutants and environmental equity objectives subject to chance constraints on the pollutant concentration levels at receptor points. A mixed-integer variant of the metric stochastic methodology (Regis and Shoemaker 2007) and a hybrid genetic algorithm-metric stochastic heuristic are presented to solve the mixed integer toll design problem.

An area-based pricing problem under environmental constraints is presented in Chapter 3. Three flexible objective function formulations (consumer surplus maximization, minimization of auto travel demand deviations from the status quo, and revenue maximization) are proposed which can be easily utilized along with state-of-the-practice transportation planning models to design cordon and area pricing

schemes. A surrogate-based solution algorithm is proposed which utilizes a geometric representation (simple polygons) of the charging area boundary to estimate the response surface models.

A bi-objective traffic rationing problem is proposed in Chapter 4. In this problem, a traffic rationing scheme is sought which maximizes auto travel demand served while minimizing population pollutant exposure inequality, this subject to constraints on the levels of greenhouse gas emissions and pollutant concentration levels. A surrogate-assisted differential evolution algorithm for multiobjective continuous optimization problems with constraints is proposed. In contrast to the algorithms presented in Chapters 2 and 3, in the proposed differential evolution variant separate surrogates are estimated to produce surrogate ensemble-based predictions for the computationally expensive objective function (the pollutant exposure objective function) and the constraints.

Chapter 2

Accounting for population exposure to vehicle-generated pollutants and environmental equity in the toll design problem

2.1 Introduction

Extensive research has been conducted on the use of tolls or fees to control the negative externalities of vehicular traffic (e.g., see Yang and Huang, 2005). In the last two decades several studies have proposed transportation network design problems (NDPs) that incorporate environmental considerations in the design of road pricing schemes. NDPs provide transportation planners an optimization-based approach for determining what is the best network intervention given a series of objectives and constraints. Coupled with judgment-based considerations, NDPs could be useful tools in the planning process of travel demand management schemes. Generally speaking, environment-oriented NDPs, such as the toll design problem (TDP), have been proposed from the perspective of planners interested in either reducing total network emissions, ensuring that link emissions do not exceed predetermined emission constraints, or guaranteeing that pollutant concentration standards in areas adjacent to the road network are met. In this chapter, an extension to the TDP is presented that explicitly accounts for population exposure to air pollutants and environmental equity, which is defined here as the equitable distribution of environmental risk, as indicated by proxy pollutant exposure measures, among spatial locations and population groups.

In the TDP the planner's problem is to select toll locations and charging levels such that the expected human exposure to vehicle emitted pollutants and environmental inequality are minimized, subject to a budget constraint and probabilistic constraints on pollutant concentration levels. Both pollutant exposure and environmental inequality are quantified using estimated pollutant concentration levels at receptor points. These two objectives are common goals of environmental protections agencies. For instance, estimates of a population's pollutant exposure are an integral part of cost and benefits calculations in environmental impacts studies (Rabl et al. 2014). And, while the primary policy, research developments,

and public advocacy on environmental equity has occurred in the US, reducing pollutant exposure disparities is an emerging issue in various countries (e.g., see Buzzelli 2008, Laurent, 2011).

In addition to an extension to the TDP, new solution algorithms are presented which can be applied to mixed integer NDPs. Transportation NDPs have been the subject of a considerable amount of research, in part, because of the computational challenges associated with solving this type of problem (Yang and Bell, 1997). For environment-oriented NDPs these practical challenges are compounded by need to connect additional models to the ones used to simulate the network users' response to the planner's interventions. For example, in the problem considered in this chapter an analyst would need to connect at least three distinct models to compute link flows, link emissions, and air pollutant dispersion. The resulting combined model might require considerable time to evaluate the impact of a single candidate set of tolling locations and levels. Furthermore, it might not be possible to represent the model with a closed form analytical expression. This computational challenge motivates the second contribution of this chapter: two new derivative-free surrogate-based solution algorithms (SBSA) for mixed integer TDPs.

In the next section a literature review is presented on previous environment-oriented discrete network pricing studies (excluding noise-related studies), on equity and road pricing models, and on the algorithms that have been proposed to solve mixed integer TDPs. The problem formulation is presented in section 2.4, followed by an explanation of the proposed surrogate-assisted solution algorithms. The results of a numerical examples are also discussed. In the last section conclusions and possible future research directions are presented.

2.2 Literature Review

Next, previous discrete network studies that account for environmental considerations in the design of road pricing schemes are reviewed, followed by a brief discussion of equity considerations in road pricing NDPs. Lastly, an overview of solutions algorithms for mixed integer TDPs is offered.

2.2.1 Discrete network toll setting problems with environmental considerations

Environmental air quality regulations usually establish maximum air pollutant concentrations thresholds that can be violated only a predetermined number of times. Thus, it is not surprising that environmental considerations are often introduced in discrete network-based road pricing models in the form of environmental constraints. An example of such constraints are link-based environmental capacity constraints (Ferrari 1995). A roadway's environmental capacity is a traffic flow limit defined by a regulatory agency based on environmental concerns. This concept has been used in bi-level optimization road pricing problems with elastic demand (Yang and Bell 1997), uncertain demand (Li et al. 2012), and in dynamic pricing models (Zhong et al. 2012). A related type of environmental constraint is the link-based emission constraint, which directly limits total emissions on a link. Emission constraints have been incorporated in road pricing models with fixed and elastic demand (Nagurney 2000), land-use considerations (Li et al. 2007), multi-criteria routing behavior (Jaber and O'Mahony 2009), and in models that have considered corridor-level pricing (Mishra and Welch 2012). Dhanda et al. (1999) proposed the concept of link-based environmental target constraints in road pricing models; the targets are met by the implementation of link subsidies or tolls. Constraints on total system emissions have also been proposed, especially in models concerned with minimizing GHG emissions (Sharma and Mishra 2013).

Discrete network optimization problems with environment-oriented objective functions are an alternative approach for designing road pricing schemes. Yin and Lu (2000) proposed a bi-level optimization problem to determine tolls that minimize total network emissions. Yin and Lawphongpanich (2006) extended this work, in part, by presenting a bi-objective problem where total travel time and total emissions are simultaneously minimized. Other studies have considered the design of pricing schemes with emission minimization objectives in the presence of public transit (Si et al. 2012; Sharma and Mishra 2013), area-based emission budget constraints (Yang et al. 2013), and joint toll and rebate schemes (Chen and Yang 2012), or in conjunction with dynamic assignment models (Friesz et al. 2013). The model proposed by Wang et al. (2014) is perhaps the most relevant work to this study. Wang et al. proposed a multiobjective

toll setting problem where the pollutant intake of travelers is minimized. Pollutant intake is defined in terms of the median or maximum pollutant intake experienced by road travelers in different routes. An alternative to emission minimization objective functions is the minimax objective function formulated to minimize the maximum deviations between observed link emissions (Hizir 2006) or concentrations (Kolak et al. 2013) relative to an emission or concentration limit, respectively.

Besides the work of Wang et al. (2014), human exposure to pollutants has been previously considered in the context of two other types of transportation network design models. Gouge et al. (2013) proposed a transit network design problem that considers the minimization of an intake fraction measure. In the context of robust signal timing optimization, Zhang et al. (2013) proposed the use of a mean excess exposure measure, where pollutant exposure was defined as a function of pollutant concentrations and population density.

In the reviewed road pricing studies, models were proposed mainly with the objectives of minimizing emissions or ensuring that emission or pollutant concentration levels do not exceed a preselected threshold. The minimization of human exposure to vehicle-generated air pollutants has only been considered in the work of Wang et al. (2014), but in terms of pollutant exposure of road travelers. Furthermore, in these studies the toll locations are predefined; the planner's decision variables are only the toll levels in the selected links. The TDP extension presented in the section 2.4 can be used to select both toll locations and their corresponding levels according to human exposure and environmental equity considerations and in compliance with pollutant concentration constraints at receptor points. As it will be illustrated in the example presented in the section 1.1, this is a useful extension given that neither human exposure to air pollutants nor environmental inequality are necessarily minimized in models aimed at minimizing emissions or congestion.

2.2.2 Equity considerations and discrete network road pricing problems

The equity impacts of road pricing schemes are a major concern for transportation planners (Levinson 2009), and accordingly equity-based constraints and objectives have been incorporated to toll setting

models, particularly in terms of the distribution of travel benefits and costs between different socioeconomic groups and geographical locations. For example, Yang and Zhang (2002) proposed a multi-class toll setting problem with spatial and social equity constraints that control the allowable relative change in travel cost with respect to the equilibrium costs. In other models constraints are introduced that ensure that no user is worse off after the pricing intervention (e.g., Song et al. 2009, Lawphongpanich and Yin 2010). Yin and Yang (2004) proposed two road pricing optimization problems with the objectives of minimizing the difference in travel disutility among user classes and origin-destination pairs. While Wu et al. (2012) proposed a model for designing congestion pricing schemes that maximize social benefit and equity. The Gini coefficient was used by Wu et al. to measure disparities in the distribution of incomes due to the pricing intervention.

However, although environmental justice is recognized as an important equity dimension in the transportation planning process (Forkenbrock and Schweitzer 1999, Feitelson 2002), it appears that it has not been considered in transportation NDPs, including toll pricing models. Nevertheless, the use of road charges to achieve environmental equity objectives has been discussed in the context of traffic assignment models (Rilett and Benedek 1994), and there are simulation studies that support the contention that road pricing could be a useful mechanism for reducing disparities in the distribution of pollutant concentrations (Mitchell 2005).

Environmental equity considerations are incorporated in the TDP presented in section 2.4 as part of the planners objective. It is posited that the planner can quantify environmental inequity utilizing an inequality index. Levy et al. (2006) examined different inequality indicators in the context of health benefit analysis for pollution control policies, and they concluded that the Atkinson index offers is the most appropriate indicator, partly because it satisfies the Pigou-Dalton transfer principle and it is subgroup decomposable. The Atkinson index has been used in the assessment of pollutant exposure and mortality risk inequalities caused by mobile sources (e.g., Levy et al. 2009, Marshall et al. 2014), and it is selected as the inequality index in the numerical examples presented in section 2.5.

2.2.3 Solution algorithms for mixed-integer TDPs

Several optimization algorithms have been proposed for different types of TDP. Bai (2004) developed a dynamic slope scaling procedure for the minimum toll booth problem. Verhoef (2002) proposed the use of welfare gain indicators in solution algorithms for the social welfare maximization TDP with elastic demand, an approach later extended by Shepherd and Sumalee (2004) and Ekström et al. (2009). Yang and Zhang (2002) presented a simulated annealing-genetic algorithm (SA-GA) heuristic to solve the social welfare maximization TDP and the travel cost minimization TDP. Mixed integer linear approximations (Ekström et al. 2012) and GA (Fan and Gurmu 2014) are two other solution approaches developed for the travel cost minimization TDP. For the case of the cordon-based TDP, Zhang and Yang (2004) and Sumalee (2004) proposed GA-based approaches, while Zhang and Sun (2013) proposed a dual-based heuristic algorithm.

As previously mentioned, finding optimal solutions to NDPs is generally computationally expensive, partly given the time required to run traffic simulation models. In the case of environment-oriented NDPs, each candidate solution's evaluation is further burdened by the need to compute, for example, vehicle emissions and pollutant concentrations. With the increasing complexity of the models used to simulate travel behavior and assess the environmental impact of traffic it becomes even more important to develop efficient solution algorithms if optimization based approaches are to application to practical engineering problem. Hence, there is a growing interest in developing SBSAs for NDPs. In the context of optimization algorithms, surrogate models (also known as metamodels or response surface models) are models that provide computationally inexpensive approximations to computationally expensive models. Surrogate models could be used, for instance, to quickly screen for promising candidate solutions before actually evaluating any designs with the time expensive models. This in turn reduces the number of model runs necessary in the search for good solutions. Previous studies have developed SBSAs for primarily for continuous (e.g., Chow et al. 2010; Osorio and Bierlair 2013; Lamotte 2014*)* and discrete (Xiong and Schneider 1992; Fiske 2011; Wismans, 2012) NDPs. Of particular interest are the SBSAs proposed by Chow and Regan (2014) and Chen et al. (2014), which were developed for continuous toll pricing problems. In this chapter, a SBSA is developed for the mixed integer TDP. Concurrently, Chen et al. (2015) developed

a SBSA for the link addition and capacity expansion mixed integer NDP. To the authors' knowledge, this is the only other study that has considered the use of SBSA for mixed integer NDP. The proposed algorithm differs from previous SBSA in that it is applicable to mixed integer problems where binary variables indicate which continuous variables can assume non-zero values, which is the case of TDPs. Given that it is a derivative-free algorithm, the proposed SBSA can be applied to any other optimization problem (including other NDPs) that features a similar relationship between binary integer and continuous variables.

2.3 Model Formulation

This section opens with the introduction of a TDP intended for pollutant exposure minimization, with environmental equity considerations incorporated in the pollutant concentration constraints. This is followed by a discussion of two indicators that can be used to quantified population exposure to pollutants. Lastly, a TDP objective function is presented that explicitly considers environmental inequality. To set the problem, first consider a planner that is interested in determining toll locations $\boldsymbol{y}$ and associated toll levels $\boldsymbol{\tau}$ that minimize the expected population exposure to vehicle-generated air pollutants. Given are a set of preselected candidate links $\bar{A}$ for tolling, the costs h_a incurred for charging a toll on each link a ($a \in \bar{A}$), a budget H, and the requirement that pollutant concentrations levels C_r at each receptor point r ($r \in R$) have to be less than or equal to a threshold $C_{r,max}$ with probability θ_r or higher (for notational simplicity, a single pollutant is considered). As in previous TDPs, this problem can be formulated as a bi-level optimization problem, where the upper level contains the planner's objectives and constraints, and the lower level models the response $\boldsymbol{v}(\boldsymbol{y}, \boldsymbol{\tau})$ of the network users to the planner's intervention (i.e., the tolling scheme).

2.3.1 Upper level problem

Human exposure to pollutants is a function of multiple factors, including total emissions, the location, density, and characteristics of the population, and meteorological conditions (Levy et al., 2010). Let I be an indicator of human exposure to vehicle-generated air pollutants that accounts for population characteristics, and assume that this indicator varies with changing meteorological conditions (as pollutant

concentrations vary with meteorological conditions). Let k represent a meteorological scenario and π^k represent its probability of occurrence. Additionally, let I^k represent the region's aggregate exposure to a pollutant of interest under scenario s. Utilizing a sample average approximation approach, the upper level is formulated as follows:

$$\min_{\boldsymbol{\tau},\boldsymbol{y}} \varphi(\boldsymbol{\tau},\boldsymbol{y}) = \sum_k \pi^k I\big(\boldsymbol{v}(\boldsymbol{y},\boldsymbol{\tau})\big)^k \tag{2.1}$$

subject to

$$P\big(C_r(\boldsymbol{v}(\boldsymbol{y},\boldsymbol{\tau}),\boldsymbol{\gamma}) \le C_{r,max}\big) \ge \theta_r \qquad \forall\, r \in R \tag{2.1.1}$$

$$\sum_{a\in\bar{A}} h_a y_a \le H \tag{2.1.2}$$

$$0 \le \tau_a \le y_a \tau_{max} \qquad \forall a \in \bar{A} \tag{2.1.3}$$

$$y_a \in \{1,0\} \qquad \forall a \in \bar{A} \tag{2.1.4}$$

In the upper level problem the expected pollutant exposure is minimized subject to probabilistic constraints on pollutant concentration (2.1.1), budget constraint (2.1.2), bounds constraints for the toll levels (2.1.3), and integer constraints for the toll locations variables (2.1,4). For each receptor, constraints 2.1.1 ensure that the probability $P\big(C_r(\boldsymbol{v}(\boldsymbol{\tau}),\boldsymbol{\gamma}) \le C_{r,max}\big)$ is greater than or equal to θ_r. A simple approach to operationalize constraint set 2.1.1 is to restate it as follows:

$$\sum_k \pi^k \Gamma_r^k\big(\boldsymbol{v}(\boldsymbol{\tau}),\boldsymbol{\gamma^k}\big) \ge \theta_r \qquad \forall\, r \tag{2.2}$$

where $\Gamma_r^s(\boldsymbol{v}(\boldsymbol{\tau}),\boldsymbol{\gamma^s})$ is an indicator variable defined as:

$$\Gamma_r^k(\boldsymbol{v}(\boldsymbol{\tau}),\boldsymbol{\gamma^s}) = \begin{cases} 1 & ,if\ C_r\big(\boldsymbol{v}(\boldsymbol{\tau}),\boldsymbol{\gamma^k}\big) \le C_{r,max} \\ 0 & ,if\ C_r\big(\boldsymbol{v}(\boldsymbol{\tau}),\boldsymbol{\gamma^k}\big) > C_{r,max} \end{cases} \tag{2.3}$$

Two reasons for incorporating the probabilistic concentration constraints 2.1.1 are the planner's interest of meeting certain air quality regulatory standards or reducing the level of environmental equity inequality in the distribution of pollutant concentrations. Minimizing total expected pollution intake does not guarantee that air quality standards are met. In fact, it is possible that a toll location and level configuration that reduces the overall human exposure to pollutants could result in worse pollutant concentration levels

in some communities, thus raising the second issue of environmental equity. Environmental equity considerations could be reflected in the determination of $C_{r,max}$. For example, let $C_{r,base}$ represent concentration levels in receptor r previous to the road pricing scheme. Relative to $C_{r,base}$, define a maximum allowable increase (δ_r) in pollutant concentrations resulting from the road pricing scheme. $C_{r,max}$ could then be computed as $\min(\delta_r C_{r,base}, C_{reg})$, where C_{reg} represents an air quality standard.

2.3.1.1 Population-level indicators of pollutant exposure

I is assumed to be a population-level exposure measure, that is, a measure of human exposure computed using aggregate population information (including encountered pollutant concentrations) at the grid cell or receptor level. This is a common approach used in the assessment of human exposure to air pollutants (e.g., see BENMAP, Abt Associates 2012) as it can be applied with readily available data (e.g., using census tract information). Here two specifications of I are considered: I_{intake} and I_{health}. I_{intake} is defined as an air pollution intake measure (Evans et al 2002), which is a proxy measure for the public health impact of emitted pollutants. Assume that a population ρ can be divided according to their proximity to discrete receptor points r ($r = \{1, \dots, R\}$) and demographic classification g (g= $\{1, \dots, G\}$) characterized by a breathing rate B_g. Each group ρ_{gr} ($\rho = \sum_r \sum_g \rho_{gr}$) is exposed to a pollutant concentration C_{gr}. Then, I_{intake} is formulated as:

$$I_{intake} = \sum_g \sum_r \rho_{gr} C_{gr} B_g \tag{2.4}$$

C_{gr} depends on the pollutant contribution c_a of each link, which is a function of the spatial location of the link relative to the location r (d_{ar}), the link emissions Q_a, meteorological conditions $\boldsymbol{\gamma}$, terrain or urban form effects $\boldsymbol{\varepsilon}$, and background concentrations C_{rb}. Furthermore, given that direct exposure to outdoor levels of ambient pollutant concentrations varies among the population, ϕ_{rg} is used to factor the outdoor pollutant concentrations to a value that is more representative of the concentrations that the average group g person associated with receptor r is exposed to. C_{gr} is defined as:

$$C_{gr} = \phi_{gr} \sum_{a} c_a(d_{ar}, Q_a, \boldsymbol{\gamma}, \boldsymbol{\varepsilon}) + C_{rb} \tag{2.5}$$

The concentrations levels to which each group is exposed to can have different health outcomes, so it might be of interest for a planner to convert the pollutant exposure estimates into a measure that directly accounts for health effects, which is the motivation for the I's second specification, I_{health}. I_{health} can be specified in terms of the risks of an adverse health outcome o (e.g., chronic bronchitis) by multiplying each population groups' pollutant exposure I_g by a corresponding unit risk factor μ_{og} (Lai et al. 2000). Assuming that I_g appropriately characterizes the long-term pollutant exposure of each population group, the health impact performance indicator can be formulated as:

$$I_{health} = \sum_{g} \mu_{og} I_g \tag{2.6}$$

In the numerical examples section an application of the proposed TDP is presented with I_{health} specified in terms of years of life lost (YOLL) due to exposure to particulate matter ($PM_{2.5}$). Note that although I_{health} provides only a rough estimate for assessing health outcomes (given the aggregate nature of the analysis), it is a useful metric as it provides an objective way for weighting the health implications of each group's pollutant exposure levels. Naturally, other non-linear specifications of I_{health} are possible (e.g., depending on the pollutant, using nonlinear exposure response functions) that given the flexible nature of the bilevel formulation and the derivative-free solution algorithms proposed in the sections 2.4 and 2.6 could be used to specify I.

2.3.1.2 Minimizing expected pollutant exposure and environmental inequality

Minimizing a total population exposure to pollutants would result in obvious aggregate benefits, but this objective does not directly address environmental equity concerns. Instead of reflecting environmental equity concerns only in the specification of the concentration constraints, an alternative approach is to explicitly incorporate an environmental inequality metric as part of the planner's objective. In the current context, an inequality index simply indicates the level of pollutant exposure disparities that exists among

different locations and population groups. Since minimizing pollutant exposure inequality does not necessarily lead to a minimization in the region-wide impacts of vehicle emission, the environmental equity objective can be viewed as a complementary objective that is weighed against the objective of minimizing total pollutant exposure. These two objectives can be combined into a single equation using linear scalarization. Let $\tilde{\varphi}$ represent the normalized expected pollutant exposure function in equation (2.1), $\tilde{A}$ represent a normalized pollutant exposure inequality index, and w_I and w_{EQ} represent the respective weight of each objective ($w_{EQ} = 1 - w_I$). Then, the combined objective function is formulated as:

$$\min_{\boldsymbol{\tau},\boldsymbol{y}} \varphi(\boldsymbol{\tau},\boldsymbol{y}) = w_I\, \tilde{\varphi}(\boldsymbol{\tau},\boldsymbol{y}) + (1 - w_I)\tilde{A}(\boldsymbol{\tau},\boldsymbol{y}) \quad (2.7)$$

subject to

$constraints\ (2.1.1)\ to\ (2.1.4)$

In the numerical examples, φ is normalized by using the computed pollutant exposure indicator value for the untolled network as the scaling factor. Similarly, the inequality index is normalized by the pollutant concentration inequality levels estimated for the untolled network. The Atkinson index is utilized as the inequality measure, per the analysis of Levy et al. (2006). The Atkinson index varies between zero (complete equality) and one (complete inequality).

2.3.2 Lower level problem

The lower level problem is formulated as a fixed-demand DUE problem, a common assumption used in NDP studies. The fixed demand between each origin-destination (OD) pair w ($w \in W$) there is represented by q_w. f_k^w is a path k flow between for each OD pair w. δ_{ak}^w is an indicator variable with value of one when link *a* is in path *k*, and zero otherwise. Homogeneous drivers are assumed with a single value of time (β) is used. The link flow vector $\boldsymbol{v}$ is determined by solving problem:

$$\min_{\boldsymbol{v}} \sum_{a \in A \setminus \bar{A}} \int_0^{v_a} t_a(\omega) d\omega + \sum_{a \in \bar{A}} \int_0^{v_a} \left(t_a(\omega) + \frac{\tau_a}{\beta} \right) d\omega \tag{2.8}$$

$$subject\ to$$

$$v_a = \sum_w \sum_k f_k^w \delta_{ak}^w \qquad \forall\, a \in A \tag{2.8.1}$$

$$\sum_k f_k^w = q_w \qquad \forall w \in W \tag{2.8.2}$$

$$f_k^w \geq 0 \qquad \forall w \in W, k \in K \tag{2.8.3}$$

Equations 2.8.1 are definitional constraints, equations 2.8.2 are the flow conservation constraints, and constraints 2.8.3 specify the non-negativity condition. Note that the intent of this section is to present a flexible bi-level optimization framework for the design of tolling schemes that minimize human exposure to vehicle-generate pollutants. Nevertheless, for completeness and purposes of the numerical tests a specific lower level problem assumption (i.e., DUE) is made. Next, a new derivative-free solution algorithm for mixed-integer problems like the TDP is presented.

2.4 Solution Algorithm: Mixed Integer Variant of the LMSRS Algorithm

In practice, it is likely that finding good solutions to the proposed TDP would present a significant challenge given the computational cost and likely black-box nature of the multiple computer software applications necessary to evaluate candidate tolling schemes. Therefore, a new derivative-free SBSA algorithm is proposed to tackle these problems. The algorithm presented next is a variation of the local metric stochastic response surface (LMSRS) methodology (Regis and Shoemaker 2007; Muller et al. 2014) that is applicable to mixed integer problems where binary integer variables indicate which of the continuous variables can assume non-negative values.

The problems shown in (2.1) and (2.7) have two constraints: the budget constraint and the concentration constraints. The budget constraint is easily evaluated as it is only a function of the costs of charging tolls on the candidate links. Therefore, constraint (2.1.2) is computed before evaluating any candidate solution; if the constraint is violated a new link combination is generated that complies with the budget constraint. In contrast, the concentration constraints require outputs from the traffic assignment, emission, and air

dispersion models, and consequently it is computationally expensive to determine if a candidate solution meets these constraints. To circumvent this problem, the concentration constraints are handled using a parameter-free penalty approach (Deb 2000). Let φ_{max} be the objective function value of the worst known feasible solution, and κ be a penalty factor. The upper level problem's objective function (2.1) is re-formulated (arguments omitted) as:

$$\Phi = \begin{cases} \varphi & if\ P(C_r \le C_{r,max}) \ge \theta_r \quad \forall r \\ \varphi_{max} - \kappa \sum_r \min\left(P(C_r \le C_{r,max}) - \theta_r, 0\right) & otherwise \end{cases} \tag{2.9}$$

φ_{max} is updated as worst feasible solutions are discovered. If there is no known feasible solution, a placeholder value (e.g., zero) is assumed for φ_{max}. The same adjustment is made for function (2.7).

2.4.1 Preliminary concepts of the LMSRS algorithm

The LMSRS method was previously applied by Chow et al. (2010) to the capacity expansion NDP, and extended for the multiobjective toll pricing problem by Chow and Regan (2014). Muller et al. (2013) presented a LMSRS methodology for general mixed integer problems. Here the LMSRS strategy is adapted for the mixed integer TDP, where the integer variables are binary and determine where the continuous variables (i.e., toll levels) can assume positive values. In the algorithm to be presented, the surrogate models are estimated using the toll level vectors and their associated objective function values Φ.

The basic idea behind the LMSRS methodology is to use a surrogate model to iteratively select the most promising solutions among a set of randomly generated candidate solutions, which are then evaluated with the computationally expensive models to determine if there is a candidate that improves on the best known solution. The most promising solution is the candidate with the minimum weighted score $W = w^{RS}V^{RS} + w^{D}V^{D}$, where V^{RS} is the response surface criterion score, V^{D} is the distance criterion score, and w^{RS} and w^{D} are their corresponding weights (where $w^{RS} + w^{D} = 1$). Assume that $\boldsymbol{\tau}^m$ $(m = 1, \dots, M)$ is a candidate toll vector and $\boldsymbol{y}^m$ is its associated toll location vector, both with dimension $1 \times |\bar{A}|$. Furthermore, let $s(\boldsymbol{\tau}^m)$ represent the objective function value predicted by the surrogate model. Then,

$$V^{RS}(\boldsymbol{\tau}^m) = \frac{s(\boldsymbol{\tau}^m) - s_{min}}{s_{max} - s_{min}} \tag{2.10}$$

where s_{min} and s_{max} are the minimum and maximum predicted objective function values, respectively, among the candidate solutions. If s_{min} and s_{max} are equal, then $V^{RS}(\boldsymbol{\tau}^m) = 1$. For the distance criterion score, let $\Delta(\boldsymbol{\tau}^m)$ represent the minimum distance (e.g., Euclidean) between candidate $\boldsymbol{\tau}^m$ and the previously n evaluated points. Additionally, let $\Delta_{min} = \min_{1 \le j \le n} \{\Delta(\boldsymbol{\tau}^m)\}$ and $\Delta_{max} = \max_{1 \le j \le n} \{\Delta(\boldsymbol{\tau}^m)\}$. The distance criterion is computed as:

$$V^{D}(\boldsymbol{\tau}^m) = \frac{\Delta_{max} - \Delta(\boldsymbol{\tau}^m)}{\Delta_{max} - \Delta_{min}} \tag{2.11}$$

If Δ_{min} equals Δ_{max}, then $V_n^{dist}(\boldsymbol{\tau}^m) = 1$. For the weight w^{RS}, an ordered set $\Upsilon = \langle v_1, \dots, v_\kappa \rangle$ ($0 \le v_1 \le \cdots \le v_\kappa \le 1$) is defined. In each iteration n the value of w^{RS} is adjusted by sequentially cycling through the set elements. By varying the weights' values the algorithm gradually oscillates between an exploitative and an explorative search focus.

Another important element of the LMSRS method is the set of strategies used for generating candidate solutions. Let $\boldsymbol{\tau}^{best}$ be the toll level vector with lowest known Φ (Φ^{best}), and $\boldsymbol{y}^{best}$ be its associated location vector. Three types of candidate points are generated in each iteration of the proposed algorithm. In the first group of candidates (group I), M^I candidates are generated by varying the toll levels in the locations indicated in $\boldsymbol{y}^{best}$. First, define the probability p^I for perturbing a coordinate of $\boldsymbol{\tau}^{best}$. Let ω be a random draw from a standard uniform distribution. For each tolled coordinate i of $\boldsymbol{\tau}^{best}$, if random draw $\omega_i > p^I$, then $\tau_i^m = \tau_i^{best}$; otherwise, the tolls are perturbed according to $\tau_i^m = \tau_i^{best} + z(0, \sigma^I)$, where z is a random draw from a normal distribution with mean zero and variance σ^I. This procedure is repeated M^I times. The variance σ^I is initialized with value σ_{init}^I, but it is modified depending on the number of success or failures in finding a better solution than $\boldsymbol{\tau}^{best}$.

The second group of candidates (group II) are generated by creating $\boldsymbol{y}^m$ vectors in the neighborhood of $\boldsymbol{y}^{best}$. First, M^{II} location vectors are generated by changing each coordinate of y^{best} with probability

p^{II}. $\boldsymbol{y}^m$ is said to be in the neighborhood of $\boldsymbol{y}^{best}$ if $0 < \sum_i |y_i^m - y_i^{best}| \leq \eta^{II}$, where η^{II} is a parameter that defines the maximum distance that can exists between $\boldsymbol{y}^m$and $\boldsymbol{y}^{best}$. If candidate $\boldsymbol{y}^m$ is determined to be in the neighborhood of $\boldsymbol{y}^{best}$, the budget constraint is checked; candidate toll location vectors that violate the budged constraint are adjusted until the constraint is satisfied. For each neighbor $\boldsymbol{y}^m$ a total of λ^{II} candidate toll levels are generated (naturally, in the toll locations indicated by $\boldsymbol{y}^m$). Toll locations shared between $\boldsymbol{y}^m$and $\boldsymbol{y}^{best}$ are perturbed with probability p^{II}. If $\omega_i > p^{II}$, then $\tau_i^m = \tau_i^{best}$; otherwise, the tolls are perturbed according to $\tau_i^m = \tau_i^{best} + z(0, \sigma^{II})$, where z is a random draw from a normal distribution with mean zero and variance σ^{II}; σ^{II} is assumed fixed. Toll levels for the toll locations not shared by $\boldsymbol{y}^m$ and $\boldsymbol{y}^{best}$are generated by randomly drawing from the interval $[0, \tau_{max}]$.

In the last group (group III) a candidate that is not in the neighborhood of $\boldsymbol{y}^{best}$ is generated. A two-stage procedure is followed. First, M^{III} candidate toll locations are generated and adjusted according to the budget constraint. A candidate $\boldsymbol{y}^m$ with $\sum_i |y_i^m - y_i^{best}| > \eta^{III}$ is randomly selected. η^{III} is the minimum distance for which $\boldsymbol{y}^m$ is said to be outside the neighborhood of $\boldsymbol{y}^{best}$. Next, λ^{III} toll level vectors are generated in the positions indicated in $\boldsymbol{y}^m$ by randomly drawing from the interval $[0, \tau_{max}]$; the λ^{III} toll vectors constitute group III. Although a total (M) of $M^I + \lambda^{II} M^{II} + \lambda^{III}$ candidate points are generated, in each iteration only three points (one from each group) are evaluated with the computationally expensive models. Next the steps of the mixed integer LMSRS algorithm is presented.

2.4.2 LMSRS algorithm for the TDP

The algorithm steps and the notation used in this description are given below.

Decision variables

$\boldsymbol{\tau}$: toll level vector

$\boldsymbol{y}$: toll location vector

Indices

m : indices for candidate solution vectors generated according to the rules of groups I, II, and III ($m = 1, \dots, M$)

u : indices for the three candidate solutions selected for evaluation ($u = \mathrm{I}, \mathrm{II}, \mathrm{III}$)

j : indices for solutions that were evaluated with the models ($j = 1, \dots, n$)

Counters

n : counter for the number model runs

G_{succ} : counter for number of successes in improving upon best known solution

G_{fail} : counter for number of failures in improving upon best known solution

Functions

$\varphi(\cdot)$: objective function

$\Phi(\cdot)$: penalty-adjusted objective function

$o(\cdot)$: feasibility indicator equal to $\sum_r \min\left(P\left(C_r \leq C_{r,max}\right) - \theta_r, 0\right)$

$s(\cdot)$: surrogate model

σ^I : variance used in the generation of candidates for group I

$W(\boldsymbol{\tau}^m)$: weighted candidate score of $\boldsymbol{\tau}^m$

χ : function that indicates if a feasible solution was found

μ : function that indicates if φ_{max} was updated

Parameters

n_0 : initial number of evaluated solutions

n_{max} : maximum value for n

E_{succ}, E_{fail} : maximum counter values for G_{succ} and G_{fail}, respectively

σ^I_{init}, σ^I_{min} : initial and minimum value of σ^I_n, respectively

Sets

T : set for $\boldsymbol{\tau}^j$ values

Λ : set for $\varphi(\boldsymbol{\tau}^j)$ values

Ψ : set for $o(\boldsymbol{\tau}^j)$ values

Ω : set for $\Phi(\boldsymbol{\tau}^j)$ values

Algorithm Steps

1. Initialization:

1.1. Set $n = n_0$, $G_{fail} = 0$, $G_{succ} = 0$, $\sigma^I = \sigma^I_{init}$, and $\mu = 0$.

1.2. Generate n $\boldsymbol{y}^j$ vectors, and adjust those vectors that do not satisfy the budget constraint.

1.3. Based on the $\boldsymbol{y}^j$ vectors, generate $\boldsymbol{\tau}^j$ vectors.

1.4. Add vectors $\boldsymbol{\tau}^j$ to T, and initiate Λ , Ψ, and Ω as empty sets.

2. Initial point evaluation and selection of best solution:

2.1. For each $\boldsymbol{\tau}^j$, compute $\varphi(\boldsymbol{\tau}^j)$ and $o(\boldsymbol{\tau}^j)$, and add values to Λ and Ψ.

2.2. If there are feasible solutions according to the $o(\boldsymbol{\tau}^j)$ values, the $\boldsymbol{\tau}^j$ and $\boldsymbol{y}^j$ vectors for the candidate with lowest $\varphi(\boldsymbol{\tau}^j)$ are labeled $\boldsymbol{\tau}^{best}$ and $\boldsymbol{y}^{best}$, respectively. Additionally, the highest feasible $\varphi(\boldsymbol{\tau}^j)$ is labeled φ_{max}, and $\chi = 1$. Otherwise, if there are no feasible solutions, $\boldsymbol{\tau}^{best}$ and $\boldsymbol{y}^{best}$ are assigned the values of the candidate with the lowest $o(\boldsymbol{\tau}^j)$ value, φ_{max} is assigned a placeholder value, and $\chi = 0$.

2.3. For each initial point, compute $\Phi(\boldsymbol{\tau}^j)$ and add values to Ω.

2.4. Determine Φ^{best}.

3. Fit surrogate model s using information in T and Ω.

4. Candidate point generation and selection:

4.1. Generate candidates according to the specifications of groups I, II, and III.

4.2. For each group and each candidate point $\boldsymbol{\tau}^m$, compute $W(\boldsymbol{\tau}^m)$.

4.3. For each candidate point group, select the candidate solutions with minimum $W(\boldsymbol{\tau}^m)$.

5. Candidate evaluation and updates of parameters and archives:

5.1. Evaluate the three selected candidates u with the computationally expensive models to determine $\varphi(\boldsymbol{\tau}^u)$ and $o(\boldsymbol{\tau}^u)$.

5.2. Update φ_{max} if possible: If $\chi = 1$ and there is a feasible candidate u with $\varphi(\boldsymbol{\tau}^u) > \varphi_{max}$, then $\varphi_{max} = \varphi(\boldsymbol{\tau}^u)$. Else, if $\chi = 0$ and there are one or more of the three candidate points that are feasible, update φ_{max} with the worst feasible $\varphi(\boldsymbol{\tau}^u)$ and set $\chi = 1$. If φ_{max} is updated, $\mu = 1$.

5.3. For each candidate point, compute $\Phi(\boldsymbol{\tau}^{\boldsymbol{u}})$.

5.4. If $\mu = 1$, use the new φ_{max} to update values in Ω, update Φ^{best}, and then set $\mu = 0$.

5.5. For the group I candidate, if $\Phi(\boldsymbol{\tau}^I) < \Phi^{best}$, then $\Phi^{best} = \Phi(\boldsymbol{\tau}^I)$, $\boldsymbol{\tau}^{best} = \boldsymbol{\tau}^I$, $\boldsymbol{y}^{best} = \boldsymbol{y}^I$, $G_{succ} = G_{succ} + 1$ and $G_{fail} = 0$. Otherwise, only adjust counters: $G_{succ} = 0$ and $G_{fail} = G_{fail} + 1$. Similarly, sequentially check if $\Phi(\boldsymbol{\tau}^{II})$ or $\Phi(\boldsymbol{\tau}^{III})$ are less than Φ^{best}. If so, update Φ^{best}, $\boldsymbol{\tau}^{best}$, and $\boldsymbol{y}^{best}$ accordingly, but in both cases set $G_{succ} = 0$ and $G_{fail} = 0$.

5.6. If $G_{succ} = E_{succ}$, then reset $\sigma^I = 2\sigma^I$ and set $G_{succ} = 0$. If $G_{fail} = E_{fail}$, then reset $\sigma^I = \max\left(\sigma^I_{min}, \sigma^I/2\right)$ and set $G_{fail} = 0$.

5.7. Increase counter $n = n + 3$, and add information from evaluated candidate points to Λ, Ψ, T, and Ω. If $n \leq n_{max}$ return to step 3; otherwise, continue to step 6.

6. Return Φ^{best}, $\boldsymbol{\tau}^{best}$, and $\boldsymbol{y}^{best}$.

2.5 Numerical Examples for Toll Problem Extension and the LMSRS Algorithm

Numerical tests were performed to provide a sample application of the proposed TDP and to examine the performance of the proposed LMSRS algorithm. Four fixed-demand TDPs were considered in the performance tests: a total system travel cost minimization (TT) problem (in time units), a total system emission minimization (TE) problem, a pollutant intake minimization (PIK) problem using I_{intake}, and a health impact minimization (HI) problem using I_{health}. The algorithm's performance was compared against

the SA-GA heuristic (Yang and Zhang 2002) and a GA-based approach (FG-GA; Fan and Gurmu 2014); which to the authors' knowledge are the only derivative-free heuristics that have been proposed for TDPs. A series of numerical test were also conducted with the combined PIK and inequality minimization (PIK-EQ) problem in equation (2.7).

In the HI problem, chronic mortality (measured in YOLL) due to $PM_{2.5}$ exposure (Rabl et al. 2014) was used as a proxy to the adverse health impacts caused by vehicle emissions. Linear exposure-response functions (ERF) were assumed for different populations group. Let the slope of the group g's ERF be represented by S_g (units: YOLL/(person×year× $\mu g_{PM2.5}/m^3$)). Additionally, let VL represent the value of a life year, set equal across all groups in this example. With the given assumptions, the health impact objective function for the TDP test was specified as:

$$I_{health} = VL \sum_{g} S_g \left(\sum_{r} \rho_{gr} C_{gr} \right) \tag{2.12}$$

In this example, C_{gr} represents daily concentrations, which were modeled as an average of the vehicle-generated $PM_{2.5}$ concentrations during two time periods, with the second time period's OD trip matrix being simply the transpose of the first period's OD trip matrix (i.e., return trips). Traffic conditions are assumed to be invariant throughout the year. Furthermore, it was assumed that tolls are charged only during the first period (and therefore the concentrations of the second period were assumed fixed).

As previously mentioned, the Atkinson index was used to quantify the level of environmental inequality, which in the examples is defined in terms of the differences in average CO pollutants concentrations at the receptor points ($\bar{C}_r$). So for this problem the Atkinson index is defined as:

$$A = 1 - \left[\frac{1}{R} \sum\nolimits_{r=1}^{R} \left(\frac{\bar{C}_r}{\bar{C}} \right)^{1-\varepsilon} \right]^{\frac{1}{1-\varepsilon}} \tag{2.13}$$

$\bar{C}$ is the average concentration over all receptors, while ε is a parameter that indicates the level of aversion to inequality (assumed here not to be equal to one).

2.5.1 Test networks

Two test networks were used in the algorithm performance tests: the Sioux Falls network and a grid network. The Sioux Falls network data was obtained from Bar-Gera's transportation network test problem database (Bar-Gera 2014). This network is composed of 76 links, 24 nodes, and 24 OD zones and has a total demand of 360,000 trips. The given node coordinates were used to define the spatial location of the links, assuming shortest connections between nodes. The unit length in the Sioux Falls' coordinate system was set to be equivalent to 0.023 meters. In Figure 2-1(a) the network is presented and the 10 candidate links selected for the test problems are identified (dashed arrows).

The grid network is composed of 1444 links, 400 nodes, and 38 OD zones, and the total travel demand was set to 104,694. An 800 meter spacing is specified between nodes. The square grid network is shown in Figure 2-1(b); 32 candidate links are considered for tolling (dashed lines). The red lines in Figure 3(b) represent links with capacities of 7200 vph and free flow travel times of 0.5 minutes. All other links were randomly assigned capacities and free flow travel times form the intervals [4000, 6000] and [0.7, 1], respectively. For both networks a budget of 10 cost units was assumed, along with a value of time β of \$20 per hour and a \$5 maximum toll (τ_{max}) for all candidate link locations. Furthermore, the Bureau of Public Road (BPR) function was used to compute average link travel time in both networks.

2.5.2 Models and associated inputs

In all problems traffic flow was modeled using the fixed-demand DUE problem, which was solved using the Frank-Wolfe algorithm. Carbon monoxide (CO) was selected as the pollutant of interest in the PIK problem, the TE problem, and the combined PIK and inequality minimization problem. CO emissions were computed using the speed-emission function utilized in the motivating example presented in section 1.1 (Sugawara and Niemeier 2002). In the health impact problem, $PM_{2.5}$ emissions were computed using average speed-based emissions factors obtained from the Motor Vehicle Emission Simulator (MOVES; EPA 2014). Pollutant concentrations in the pollutant intake and health impact problems were computed

using the Gaussian plume approach proposed by Zhang et al. (2010). A hundred meteorological scenarios were created with different wind speeds, wind directions, and atmospheric stability classes, and equal probability of occurrence. The receptor locations were spaced 200 meters from the links' centerlines. $C_{r,max}$ for all receptors was assumed to be twice the concentration resulting from the base UE flows (background concentrations were set to zero), and ϕ_{rg} was set to two thirds for all receptors and population groups. Random number generation procedures were used to define the populations associated with each receptor. For the pollutant intake problem a homogeneous population was assumed with breathing rate of 12.2 cubic meters per day (m^3/d). In the health cost impact problem, two population groups were assumed. Population Group 1 was characterized by a loss of $S_1 = 6.8E\text{-}4$ YOLL/(person×year× $\mu g_{PM2.5}/m^3$) and Group 2 was characterized by a loss of $S_2 = 6.2E\text{-}4$ YOLL/(person×year× $\mu g_{PM2.5}/m^3$). For both groups VL was set to \$50,000. The inequality aversion parameter ε is set to 0.75 (Levy et al. 2006, Marshall et al. 2014) in the PIK-EQ problem.

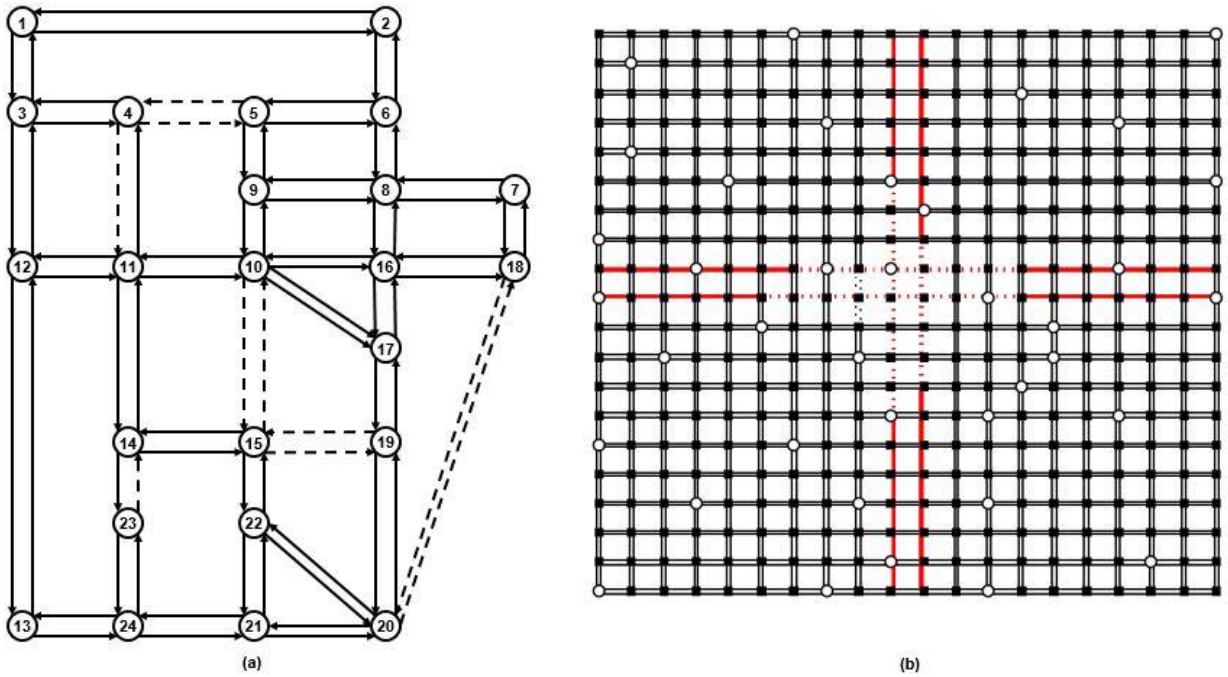


Figure 2-1 The Sioux Falls (a) and grid (b) networks.

2.5.3 Parameters for solution algorithms

The SA-GA heuristic as presented by Yang and Zhang (2002) was implemented, including the Hooke-Jeeves (HJ) local search procedure. GA is used to explore the tolling locations, while the HJ-assisted SA algorithm determines the toll levels for each tolling location vector. The initial step size, termination parameter, and acceleration factor of the HJ algorithm with discrete steps (Bazaraa and Shetty 1979) were set to 0.25, 0.001, and 1, respectively. The SA parameters were specified as follows: $\delta = 0.1$, $\chi_0 = 0.9$, $t = 0.7$, $L = 1$, $T_s = 10^{-3}$, and $\mathrm{m}_0 = 30$ (initial population) (Dekkers and Aarts 1991). A crossover rate of 0.7, a mutation rate of 0.05, and a population size of 64 was used for the GA (Fan and Gurmu 2014). The tournament selection size for GA was set to three, with the best solution in each generation always preserved. The same GA parameters were used for FG-GA. In addition, FG-GA was initialized by setting the initial number of toll locations equal to the best number of locations found in preliminary MI-LMSRBF runs (this information was not utilized in the SA-GA tests). This additional information is required because FG-GA is designed to determine the best toll configuration incrementally, starting by finding the best tolling scheme for a single toll location, and continuing until the maximum number of possible toll locations is reached. This strategy means that the algorithm might spent considerable computational time on tolling configurations (e.g., schemes with only one toll location) that are suboptimal. Therefore, for FG-GA it is assumed that the analysts has a priori knowledge on what is the total number of tolling locations that should be considered.

The following parameters were used for the LMSRS algorithm: $M^I = 1000$, $M^{II} = 2000$, $M^{III} = 1000$, $\lambda^{II} = 10$, $\lambda^{III} = 1000$, $\sigma^I_{min} = 0.1$, $p^I = p^{II} = 0.8$, $p^{II} = 0.1$, $E_{succ} = 8$, $E_{fail} = 12$, $\Upsilon^I = \{0.15, 0.40, 0.75, 0.95, 0.97, 1\}$, $\Upsilon^{II} = \{0.6, 0.95, 0.99, 1\}$, and $\Upsilon^{III} = \{1\}$. Additionally, $\eta^{II} = \eta^{III} = 3$ and $n_0 = 30$ was set for the Sioux Falls tests, and $\eta^{II} = \eta^{III} = 5$ and $n_0 = 50$ for the grid network tests. The radial basis function (RBF) model with cubic functional form (Gutmann 2001) was used as the surrogate model for the mixed integer LMSRS (henceforth, MI-LMSRBF). The previous algorithm parameter values were obtained from the references cited in this section or based on the authors' judgment.

2.5.4 Test results

A total of 30 model runs were performed for each problem, network, and solution algorithm, each run being stopped once 1,500 or more candidate solutions were evaluated. Figures 2-2 and 2-3 present the progression of the average best objective function values for the tests conducted using the Sioux Falls and grid networks, respectively. Tables 2-1 and 2-2 present the best solutions (after 1,500 model evaluations) for the Sioux Falls and grid networks, respectively. For all the problems the best solutions were found by the MI-LMSRBF algorithm. Interestingly, on average SA-GA found better solutions than MI-LMSRBF for short periods during the TT and TE tests in the grid network. However, past 150 model evaluations the proposed solution algorithm found better solutions than the GA solutions for all problems, which suggests that MI-LMSRBF is a promising solution approach for TDP problems.

Relative to the DUE concentrations, the population weighted average pollutant concentration at receptor points decreased in the solutions for both the PIK and the HI problems, and in both networks. For example, the best solution found for the grid network's PIK problem results in a 6 percent decrease in the population weighted mean concentration CO (averaging across all receptor concentrations). In the same network but for the HI problem, the best solution results in a 12 percent reduction in of $PM_{2.5}$ concentrations (11 percent reduction for Group 1 and 13 percent reduction for Group 2). However, partly because the tolls are simply rerouting traffic flows (given that travel demand is fixed), some receptors may actually experience an increase in pollutant concentrations, a previously stated possibility. Again taking the grid network's results as examples, in 66 and 60 percent of the receptors the average pollutant concentrations increased for the PIK and HI problems' solutions, respectively, although the overall regional pollutant intake and health impact decreased by around 6 percent. This underscores the importance of considering the environmental equity impacts of tolling decisions. Fortunately in the grid examples, the Atkinson inequality index improved in both solutions; relative to the UE concentration levels, the inequality index was reduced by 4.5 percent in the PIK problem results and by 7 percent in the results of the HI problem. Naturally, if minimizing environmental inequality is incorporated in the planner's objective function, as in

the PIK-EQ problem, then more significant inequality index reductions can be attained. Table 2-3 shows solutions for the PIK-EQ problems for different weights w_I. As expected there is a trade-off between minimizing an aggregate pollutant exposure objective and minimizing disparities in pollutant concentrations; what is the optimal weighting of the objectives will obviously depend on the circumstances and goals of each community. Other noteworthy observations from the numerical examples are the tradeoffs that occur with the TT and TE objectives. For example, in the grid network's PIK problem the best solution results in increases in travel cost (2.9 percent) and CO emissions (3.3 percent). In the Sioux Fall's PIK problem the best solution results in an increase in travel cost (1 percent) and a decrease in CO emissions (1 percent). These results emphasize the general observations made in section 1.1.

2.5.5 Sensitivity tests of the neighborhood distance η

Varying the neighborhood distance η has an impact on the speed in which good solutions are found. In Figure 2-4 and Figure 2-5 the results of tests with different values of η are presented. All tests used the TT objective function (the least time consuming problem). The results presented in Figure 2-4 were obtained from tests performed using the Sioux Falls, while Figure 2-5 presents the results obtained with the grid network. Both sets of test results suggest that low values of η produce good solutions in the initial 250 model evaluations; a $\eta = 2$ resulted in the best objective functions values. However, the results in Figure 2-5 suggest that in the long run η around half the number of variables yields better solutions. Therefore, it could be preliminary stated that the selection of the neighbor distance depends on available simulation time. Further sensitivity tests with additional networks and objective functions are needed to determine how general these observations are.

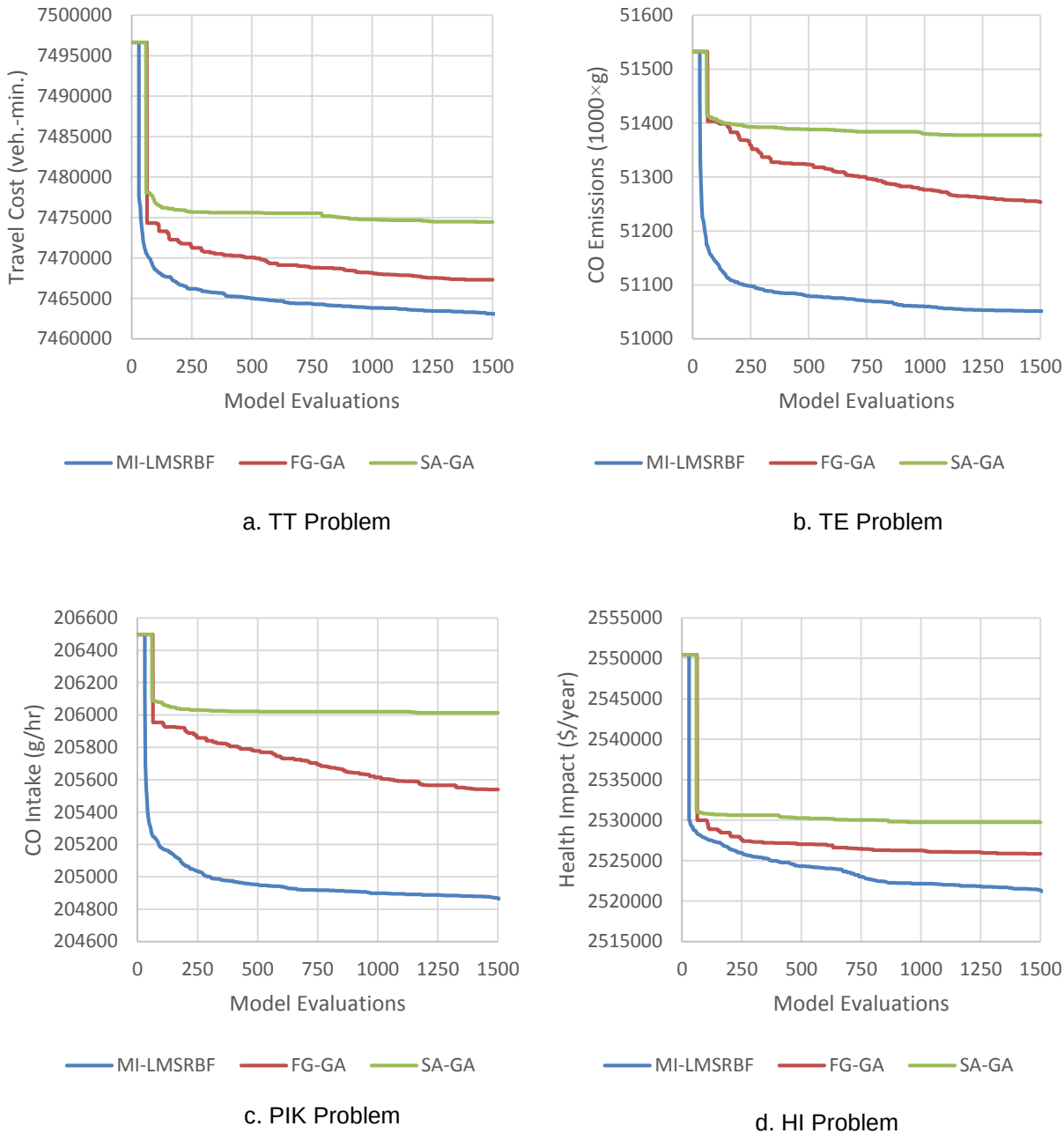


Figure 2-2 Progression of average best known objective function values for the Sioux Falls network tests

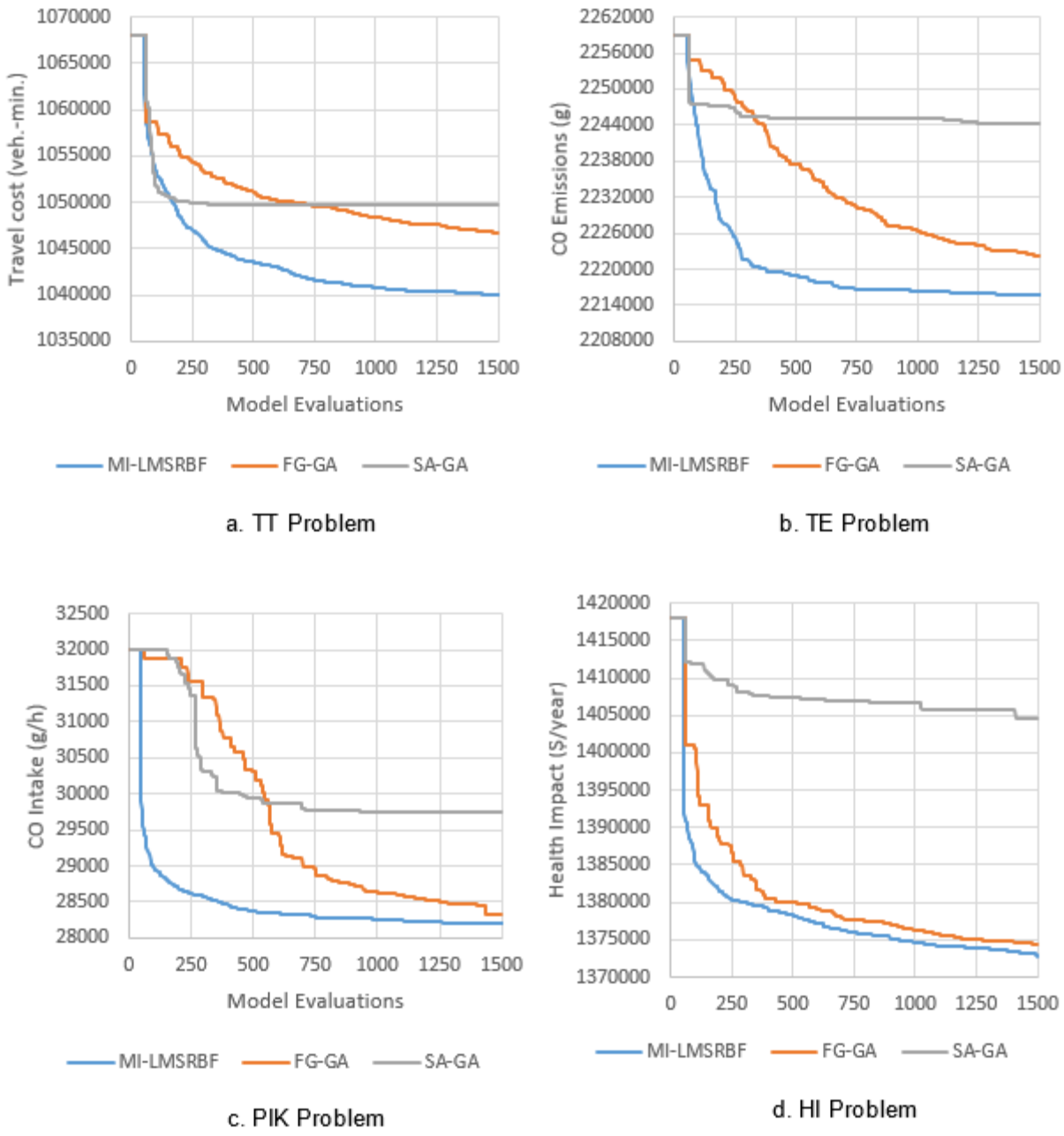


Figure 2-3 Progression of average best known objective function values for the grid network tests

Table 2-1 Best solutions obtained for the Sioux Falls tests problems

Problem	Tolling locations as indicated by y^{best}	τ^{best}	Φ^{best} (units)	Δ
TT	{2, 6, 8,10}	{4.44, 3.49, 3.08, 1.23}	7,461,412 (veh.-min.)	-0.25
TE	{2, 5, 6, 7, 8, 9, 10}	{4.68, 4.68, 4.93, 4.88, 5.00, 0.08, 5.00}	51,027,841 (g)	-1.14
PIK	{2, 5, 6, 7, 8, 10}	{4.54, 5.00, 3.12, 4.90, 4.73, 4.98}	204,680 (g/hr)	-1.00
HI	{6, 7, 8, 9}	{4.95, 4.14, 3.02, 2.28}	2,517,382 ($/year)	-0.63

$Note$: Δ refers to the percent difference between the best solution objective function value and the objective function value obtained with the DEU flows.

Table 2-2 Best solutions obtained for the grid network problems

Problem	Tolling locations as indicated by y^{best}	τ^{best}	Φ^{best} (units)	Δ
TT	{1, 7, 9, 13, 18, 19, 24, 26, 32}	{1.21, 0.87, 0.68, 0.24, 0.58, 1.13, 0.76, 1.01, 0.39}	1,035,727 (veh.-min.)	-2.43
TE	{8, 9, 14, 22, 23, 27}	{1.84, 0.35, 3.31, 0.49, 0.24, 0.22}	2,213,057 (g)	-0.83
PIK	{1, 2, 3, 4, 6, 7, 8, 9, 10, 11, 14, 15, 21, 26, 27, 29, 30}	{3.57, 2.88, 3.27, 4.57, 1.39, 0.67, 4.22, 3.56, 0.75, 1.33, 1.99, 3.20, 1.84, 2.68, 4.67, 3.90, 1.83}	27,863 (g/hr)	-6.43
HI	{1, 2, 5, 7, 9, 11, 12, 15, 16, 18, 19, 20, 25, 26, 27, 28, 29, 30, 32}	{4.62, 2.97, 1.36, 4.30, 4.67, 3.95, 0.51, 1.28, 1.02, 4.52, 0.27, 4.57, 1.17, 2.93, 2.06, 2.70, 1.57, 3.05, 0.15}	1,366,651 ($/year)	-6.14

$Note$: Δ refers to the percent difference between the best solution objective function value and the objective function value obtained with the DEU flows.

Table 2-3 Solutions obtained for the joint intake-environmental equity problem

w_I	Intake (g/hr)	Δ Intake	Atkinson Index	Δ Atkinson Index
1	27,863	-6.43	0.1870	-4.45
0.75	27,964	-6.09	0.1854	-5.26
0.25	28,642	-3.81	0.1849	-5.51
0	29,230	-1.84	0.1844	-5.78

Note: Δ represents the relative change with respect to the no toll condition.

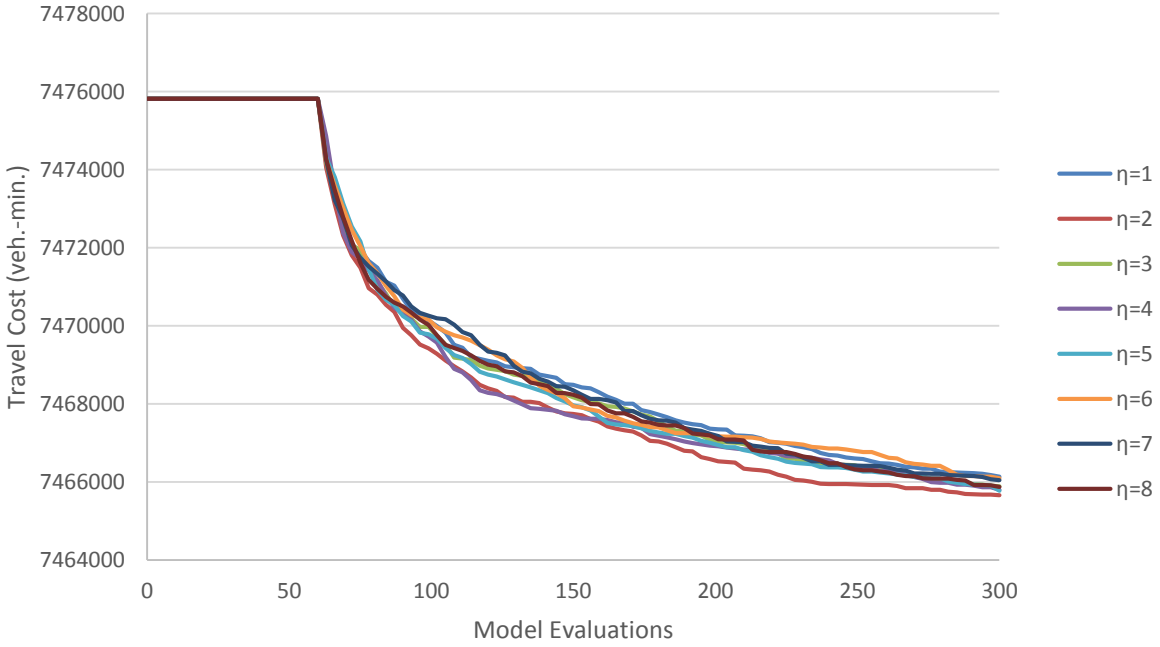


Figure 2-4 Sensitivity tests conducted using Sioux Falls network

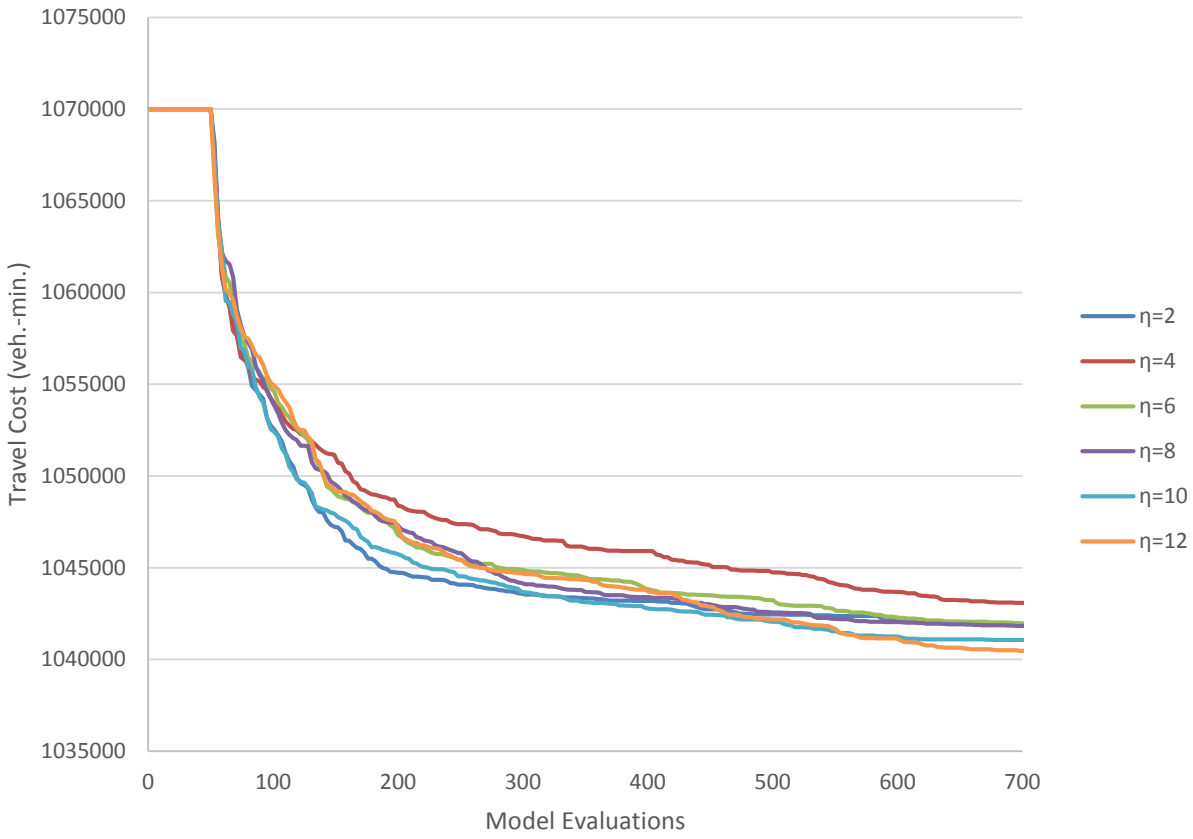


Figure 2-5 Sensitivity tests conducted using grid network

2.6 Solution Algorithm: GA-LMS approach for Mixed Integer TDPs

The SA-GA heuristic had the worst performance in the numerical tests. This could be explained in part by the nature of the algorithms used to search for the optimal tolls of each toll location vector (or parent). In particular, the nature of the HJ algorithm results in the evaluation of a significant number of candidate solutions that differ very little among themselves. Despite this issue, for a short period of the grid network model runs the SA-GA heuristic obtained better results, on average, for the TT and TE problems. For this reason in this section the basic idea of the SA-GA heuristic, namely the use of GA to explore toll locations and a continuous optimization heuristic to explore toll values, is modified by substituting the SA and HJ algorithms by a surrogate based heuristic, namely, the LMSRS algorithm. Next the steps of the new algorithm, GA-LMS, are presented. The steps are presented for an unconstrained TDPs, but they can be easily adjusted by using the formulation presented in equation (11).

2.6.1 GA-LMS heuristic for the TDP

In this algorithm surrogate models are used to explore and preselect the candidate toll levels for each parent toll location vector (i.e., the time consuming models). In addition to selecting the most promising toll level for each parent using a surrogate model, for each toll location parent a surrogate model is used to select which toll location vector mutation is evaluated in each iteration. The notation and steps of the algorithm are as follows:

Notation

n	:	generation counter
n_{max}	:	maximum value for n
n_0	:	parent population size
n_{TS}	:	maximum number of iterations for LMSRS algorithm
n_{MC}	:	number of children toll location vectors created $\boldsymbol{yc^u}$ for each parent $\boldsymbol{y}^j$
n_{TC}	:	number of children toll level vectors created for each $\boldsymbol{yc^u}$

Algorithm Steps

1. Initialization

1.1 Set $n = 1$.

1.2 Generate n_0 $\boldsymbol{y}^j$ parent vectors, and adjust those vectors that do not satisfy the budget constraint.

1.3 For each parent toll location vector $\boldsymbol{y}^j$, generate two toll level vectors $\boldsymbol{\tau}^j$ vectors (two is used here, but naturally more than two initial toll level vectors per location vector can be generated).

1.4 Add vectors $\boldsymbol{\tau}^j$ to T, and initiate Λ as an empty set.

2. Initial point evaluation and selection of best solution:

2.1 For each $\boldsymbol{\tau}^j$, compute $\varphi(\boldsymbol{\tau}^j)$, and add values to Λ.

2.2 For each parent $\boldsymbol{y}^j$, select the best known toll level vector $\boldsymbol{\tau}^j$

3. For each parent, improve upon its best known toll vector.

3.1 Apply LMSRS algorithm n_{TS} times in order to find a better toll vector for each parent toll location vector.

3.2 In each iteration of the LMSRS algorithm, update the T and Λ sets.

4. For each parent $\boldsymbol{y}^j$, generate children toll location vectors $\boldsymbol{yc^u}$ by applying GA mutation and crossover operators, and by using the surrogate models to screen for best children.

4.1 For each parent $\boldsymbol{y}^j$, apply mutation and crossover operations n_{MC} time to generate n_{MC} toll location vector children $\boldsymbol{yc^u}$, and for each $\boldsymbol{yc^u}$ child generate n_{TC} toll level vectors (therefore, for each parent toll location vectors there are $n_{MC} \times n_{TC}$ candidate children).

4.2 For each $\boldsymbol{yc^u}$ vector, score each toll level child based on its surrogate-based predicted objective function value and its distance to previous solutions using the same procedure as the one discussed in section 2.4.2.

4.3 Score each child $\boldsymbol{yc^u}$ using the n_{TC} scores obtained from its child toll level vectors. Among many possibilities, the score of each $\boldsymbol{yc^u}$ vector could be either the mean of the children toll level vector scores or the minimum score found.

4.4 Accept child location vector with the lowest (best) score ($\boldsymbol{yc^{best}}$).

4.5 Evaluate the best predicted toll level vector found for $\boldsymbol{yc^{best}}$, and update the T and Λ sets.

5. Apply LMSRS algorithm to find better toll level values for each of the children toll location vectors (same procedure as step 3) , while updating the T and Λ sets.

6. If $n = n_{max}$, return the best solution found in the archieve sets T and Λ sets. Else, $n = n + 1$ and continue to next steps

7. Given parent and children toll location vector objective function values, conduct tournament selection procedure (Golberg and Deb 2001) to select next parents and go to step 4.

2.6.2 Testing GA-LMS heuristic

This algorithm was tested for the TT problem using the Sioux Falls network. The results were compared to the MI-LMSRBF and SA-GA heuristics. The initial population n_0 was set to 60 for all the algorithms. For the GA-LMS heuristic n_{TS}, n_{MC}, and n_{TC} were set to 5, 10, and 20, respectively. For the LMSRS algorithm, a weight of 0.8 was used for w^{RS}. A total of 2,500 candidate toll values were generated using a 0.7 variance in each iteration of the LMSRS algorithm. In the exploration of the best child toll location vectors, a weight of 0.9 was assigned for the response surface criterion score. For MI-LMSRBF and SA-GA the parameters presented in section 2.5.3 were used.

In Figure 2-6 the results of the tests are shown. Like in previous tests, 30 runs were performed for each algorithm, and 1,500 model evaluations were performed for each test. Again, the MI-LMSRBF algorithm found better solutions in fewer model evaluations. The GA-LMS heuristic obtained better results than SA-GA, but only after 225 model evaluations were performed, which practically speaking might not be too useful. Further improvements are necessary for the GA-LMS heuristic in order to have confidence that it is a better alternative that SA-GA heuristic.

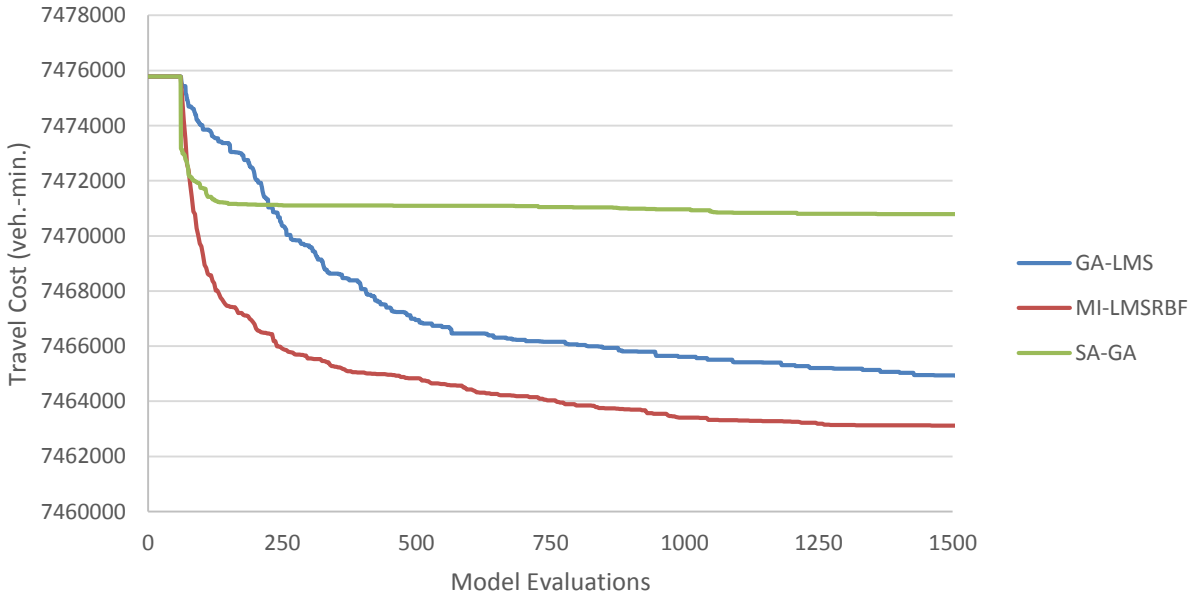


Figure 2-6 Numerical tests for GA-LMS heuristic

2.7 Closing Remarks

A TDP was presented that accounts for human exposure to air pollutants and environmental inequality given a budget constraint and chance constraints on pollutant concentration levels. Two possible indicators were suggested: a population intake measure and a health impact measure. Several extension to the proposed model can be considered, including incorporating models that account for mode and destination selection and differential pricing based on vehicle class emission intensity. Furthermore, additional studies are required to account for in-vehicle exposure to pollutants in TDPs. Future research could approach the in-vehicle pollutant exposure problem from the personal exposure perspective, rather than the aggregate population approach employed here. A personal exposure methodology would specify I using personal exposure measures (I_{micro}) based on individual agents' exposures. Agent-level pollutant exposure levels could be estimated based on a simulation of each agent's activity selection and scheduling, and a simulation of the concentration levels encountered in the different microenvironment visited during the course of the day. This microsimulation approach is far more challenging both in terms of data requirement (e.g., requires activity pattern data) and computational resources, so in all likelihood a SBSA approach would be useful. Even though this agent-based approach is not incorporated in the lower level problem used in this chapter, the proposed upper level formulation and derivative-free solution algorithms presented can be utilized by planners interested in applying a personal exposure microsimulation approach.

Numerical examples were used to illustrate the application of the presented TDP and to study the performance of the proposed LMSRS-based algorithms. The results suggest that the MI-LMSRBF algorithm generally outperforms the evaluated GA methods when the number of candidate solution evaluations is restricted. The test results add further evidence that surrogate-based optimization methods are a useful solution tools for NDPs, particularly when the computational cost of the models utilized is a concern. Future research could explore modifications to the GA-LMS heuristic. In particular, the number of evaluations performed by the LMSRS algorithm for each parent toll location vector (n_{TS}) could be adaptively varied based on the improvement observed as new candidates are evaluated. This would prevent

the use of simulation time on parents that are not too promising. Additionally, several modifications could be made to the proposed MI-LMSRBF algorithm, including the use of adaptive η^{II} and η^{III} parameters and adaptive selection of surrogate models. The use of surrogate model ensembles for NDPs is another potentially useful research direction. Surrogate ensembles could circumvent the need for testing and selecting a single type of surrogate model, and they could prove to be more robust solution search tools. Furthermore, algorithmic developments are necessary for problems where the models utilized have non-deterministic outputs, such as the aforementioned probabilistic personal-level pollutant exposure models.

Chapter 3

Designing area pricing schemes to reduce vehicle-generated pollutant concentrations

3.1 Introduction

Since Singapore's Area Licensing Scheme in 1975, cordon and area-based congestion pricing have proven effective tools in reducing congestion and pollutant emissions. For example, London's congestion charge zone reduced delays by more than 20 percent and particulate matter emissions by approximately 10 percent (Beevers and Carslaw 2005, Tonne et al. 2011). Stockholm's congestion charging system resulted in traffic and air pollutant reductions of around 15 and 10 percent, respectively (Eliasson et al. 2009). And Milan's Ecopass produced comparatively positive results both in terms of emissions and congestion reduction. (Rotaris et al. 2010). Ecopass is of particular interest because, as its name suggests, it was implemented specifically to curb pollution. Prior to the start of the program, Milan was experiencing particulate matter concentrations in exceedance of European Union environmental regulations. In response, the city planners implemented this area-based scheme, in which vehicles were charged according to their emission standard classification. Starting in 2012, Ecopass was converted into a standard congestion pricing scheme (i.e., no price differentiation according to vehicle emission intensity).

International experience demonstrates that cordon and area-based road pricing can be effective in mitigating negative externalities produced by motor vehicles. It also shows that there is interest by planners to design and implement schemes for the achievement of environmental objectives. As it will be discuss in the next section, discrete network optimization-based approaches for congestion mitigation have been proposed for the design of cordon and area-based road pricing schemes, but explicit consideration of environmental objectives has not been accounted for. In this chapter, an optimization-based approach is presented for the design of area pricing schemes intended to reduce pollutant concentrations in a city. The problem is formulated taking into consideration the fact that most transportation planning agencies have

models that do not incorporate elastic demand traffic assignment procedures, the modeling approach used in previous studies. In addition, in this chapter a new surrogate-based solution algorithm is proposed for area and cordon pricing problems.

The next section contains a review of discrete network methodologies for the design of cordon and area congestion charging schemes. In section 3.3, the problem setting and formulation are discussed, followed by the presentation of the surrogate-based solution algorithm. A numerical example is presented in section 3.6. The chapter closes with the chapter's summary and future research ideas.

3.2 Literature Review: Designing Cordon Pricing Schemes

In this section an overview of cordon pricing design studies is presented. Two analysis approaches are considered: the continuum modeling approach and the discrete network optimization approach.

3.2.1 Continuum modeling studies on the design of optimal cordon pricing schemes

Continuum modeling studies are particularly useful for gaining general theoretical insights into the impacts of cordon pricing schemes, independent of the particular configuration of a transportation network. For example, in the model developed by Mun et al. (2003), a continuous space monocentric city is considered in which a planner is interested in determining the cordon location and tolling level that maximizes social welfare. The cordon location variable is defined in terms of the cordon boundary from the city center. Their analysis suggests that cordon pricing results in welfare levels that are close to those of the first-best pricing optimum (a toll equal to marginal cost). A similar conclusion was reached by Verhoef (2005), who extended Mun et al.'s model by assuming endogenous urban density. Of particular interest is the model by Li et al. (2013), who consider the design of a cordon pricing schemes that account for vehicle-generated air pollutant costs. The model indicates that the optimal cordon scheme (in terms of social welfare) depends on traffic levels and related emissions, and, surprisingly, that compact cities are less efficient in terms of total traffic emissions.

3.2.2 Discrete network optimization approach for the design of cordon pricing schemes

Static discrete network optimization problems offer a practical approach for the design of cordon and area pricing schemes for specific networks. Like in the continuous approximation studies, in these models the optimal location and tolling level of the cordon are sought. Given the computational cost associated with solving this type of problem, usually a single toll level is determined for the cordon. Zhang and Yang (2004) proposed using a GA-based approach to solve the cordon-based congestion pricing problem. The problem was formulated as a bi-level optimization problem where the planner's objective is to maximize the social welfare of network users, who are assumed to have elastic demands. A cordon's feasibility was defined based on the graph theory concept of a cutset. Sumalee (2004) noted that Zhang and Yang's algorithm could result in infeasible offspring after the GA's crossover and mutation procedures where applied to feasible parent solutions. Sumalee developed a branch-tree encoding of the GA solution and related crossover and mutation operations that prevents the generation of infeasible solutions. Additionally, in this version of the problem the cordon pricing implementation cost was introduced to the upper level problem's social welfare function. Zhang and Sun (2013) formulated the cordon problem as a mathematical program with complementarity constraints, and proposed a dual-heuristic solution algorithm. The applicability of the solution algorithm depends on the problem formulation (i.e., it is not a derivative free algorithm). Maruyama et al. (2014) also used a GA-based approach to solve the cordon pricing problem. In their methodology the cordon location was determined using a computational geometry approach, instead of a direct graph-based approach. In contrast to the GA methods proposed by Zhang and Yang (2004) and Sumalee (2004), this GA heuristic is not anchored to a central position; it optimizes the central location of the charging cordon while using convex region constraints to control the shape of the charging boundary. Of interest are also the operations performed to generate new candidate solutions. These operations include: expanding the cordon boundary, contracting the boundary, and mutations via rotation of the cordon boundary geometry. Hult (2006) documents similar efforts in the UK Department of Transport (DoT) to use a GA-based computational geometry approach for the cordon pricing problem. The DoT method

mapped the network nodes into a Delaunay triangulation, and the cordon was expanded from a randomly selected triangle.

3.3 Problem Formulation: An Area-based Pricing Problem with Environmental Constraints

The studies discussed in the previous section used a static equilibrium assignment model with elastic demand, and, given this assumption, proposed a social welfare maximization objective. In this section, a departure is made from the elastic demand formulation, instead choosing a more flexible framework that can be applied along with models commonly used by most transportation planning agencies (e.g., sequential four-step models). Like in previous studies, a bi-level simulation framework is adopted. In the upper level the planner's objective is optimized subject to pollutant concentration constraints. The network users response to a candidate changing scheme is simulated using transportation planning models, which generally do not lend themselves to be succinctly expressed in mathematical forms Hereafter, the references will be made only to area pricing schemes, but it should be evident that the discussion can be easily extended to cordon pricing schemes.

In the area pricing problem considered, the planner attempts is define a charging zone area and its toll level, assumed fixed for in all time periods, that maximizes an objective F subject to concentration constraints on selected receptor points. Define $\bar{C}_r(\tau, \boldsymbol{\varepsilon}, \boldsymbol{v}(\tau, \boldsymbol{\varepsilon}))$, as the average pollutant concentration levels at receptor r, which is partially a function of the toll τ over the charging area represented by a collection of tolled links $\boldsymbol{\varepsilon}$**,** which results in link flows $\boldsymbol{v}(\tau, \boldsymbol{\varepsilon})$ (other factors like meteorological conditions are omitted for notational simplicity). $C_{r,max}$ is defined as the maximum allowable value of $\bar{C}_r(\tau, \varepsilon)$. ε_a is a variable that indicates if a link a leads to the charging zone area. Let Θ be the set of all possible closed charging areas, so $\boldsymbol{\varepsilon} \in \Theta$. The toll τ belongs to a finite set, where τ_{min} and τ_{max} are its minimum and maximum elements, respectively. Given this notation, in general the upper level problem can be expressed as:

$$\max_{\tau,\boldsymbol{\varepsilon}} F\big(\tau, \boldsymbol{\varepsilon}, \boldsymbol{v}(\tau, \boldsymbol{\varepsilon})\big) \tag{3.1}$$

$$subject\ to$$

$$\bar{C}_r(\tau, \boldsymbol{\varepsilon}, \boldsymbol{v}(\tau, \boldsymbol{\varepsilon})) \leq C_{r,max} \qquad \forall\, r \tag{3.1.2}$$

$$\tau \in \{\tau_{min}, \dots \tau_i, \dots, \tau_{max}\} \tag{3.1.3}$$

$$\varepsilon_a \in \{0,1\} \qquad \forall a \tag{3.1.4}$$

$$\boldsymbol{\varepsilon} \in \Theta \tag{3.1.5}$$

$\boldsymbol{v}(\tau, \boldsymbol{\varepsilon})$ is the simulation output obtained from the transportation planning models used. The next subsections present three objectives that could be easily used in conjunction with most transportation planning models utilized by planning agencies.

3.3.1 Maximizing consumer surplus

Consumer surplus is an important policy evaluation measure. Provided that a logit choice model exists to simulate the response of consumers to a policy, consumer surplus can be easily computed using the model's logsum. Fortunately, logit models are widely used in transportation planning (Cambridge Systematics et al. 2012), so utilizing a logsum-based objective function in the design of an area pricing scheme would likely be feasible for most planning agencies.

Assume that the analyst has at her disposal a multinomial logit model as part of a larger transportation planning model. These MNL model could simulate, for example, mode choice, joint destination and mode choice, or joint destination, departure time, and mode choice. The planning model has N representative groups, each group with population size ρ_n distinguished by separate zone membership and marginal utility of income α_n (assumed constant with respect to income). A group n obtains a deterministic utility V_{nj} for alternative j previous to the introduction of the charging scheme, and utility $V_{nj}(\boldsymbol{\varepsilon}, \tau)$ with the candidate scheme $(\boldsymbol{\varepsilon}, \tau)$. With this information, the design objective F_{CS} can be expressed as:

$$\max_{\boldsymbol{\tau},\boldsymbol{\varepsilon}} F_{CS} = \sum_t \sum_n \frac{\rho_{nt}}{\alpha_n} \left(\ln\left(\sum_j e^{V_{ntj}(\boldsymbol{\varepsilon},\tau)} \right) - \ln\left(\sum_j e^{V_{ntj}} \right) \right) \tag{3.2}$$

F_{CS} represents the expected change in consumer surplus, a common metric of consumer benefits (de Jong et al. 2007). In this expression, the index t represents the different model periods in the planning model.

3.3.2 Minimize deviations from status quo

Another sensible objective is to minimize deviations from the base auto trip demand conditions. Let q_{wt} represent the auto trip demand for origin-destination pair w and time period t, and $q_{tw}(\tau, \boldsymbol{\varepsilon})$ represent the demand under charging scheme $(\tau, \boldsymbol{\varepsilon})$. The objective of minimizing deviation from the status quo can be expressed as:

$$\max_{\boldsymbol{\tau},\boldsymbol{\varepsilon}} F_{SQ} = -\sum_{t}\sum_{w} |q_{wt}(\tau, \boldsymbol{\varepsilon}) - q_{wt}|^{\xi} \tag{3.3}$$

F_{SQ} is an objective that attempts to maintain the status quo given the pollutant concentration constraints. ξ ($\xi \geq 1$) is a parameter that can be used to reflect how acceptable large deviations from the original demands are.

3.3.3 Maximizing revenue generation

Like with any pricing scheme, area based pricing can be used to generate revenue. Letting v_{ta} represent traffic flow on like a for time period t, the objective maximizing revenue can be formulated as:

$$\max_{\boldsymbol{\tau},\boldsymbol{\varepsilon}} F_{\text{Rev}} = \tau \sum_{t}\sum_{a} \varepsilon_a v_{at} \tag{3.4}$$

If the auto demand is considered to be fixed, it might also be meaningful to consider the objective of minimizing total revenue collected.

3.4 Solution Algorithm

There are at least four GA-based solution algorithms that could be applied to solve the proposed area pricing problem: the cutset-based heuristic (Zhang and Yang 2004), the branch-tree method (Sumalee 2004), the

Delaunay triangulation approach (Hult 2006), and the shape constrained algorithm (Maruyama et al. 2014). In this section a surrogate-based solution algorithm is proposed to solve cordon and cordon pricing problems. Surrogate-based optimization approaches have proven to be more effective in finding good solutions in fewer iterations than traditional evolutionary heuristics (e.g., see Chapter 2), an ability that is particularly useful when the evaluation of candidate solution takes too much time. This computational cost problem is encountered in the design of area pricing schemes with environmental constraints, so a new surrogate-based solution algorithm is presented for this type of problem.

Next, the surrogate-based cordon/area pricing solution algorithm (SB-CAPSA) is explained. In this algorithm the concentration constraints are handled using a parameter-free penalty approach (Deb 2000). From the maximization perspective, let F_{min} be the objective function value of the worst known feasible solution, and κ be a penalty factor. Objective function (3.1) is restated (arguments omitted) as:

$$\tilde{F} = \begin{cases} F & if\ \bar{C}_r \le C_{max} \quad \forall r \\ F_{min} + \kappa \sum_r \min\left(C_{r,max} - \bar{C}_r, 0\right) & otherwise \end{cases} \tag{3.5}$$

F_{min} is updated as worst feasible solutions are discovered. If there is no known feasible solution, a placeholder value is assumed for F_{min} .

3.4.1 Preliminaries of SB-CAPSA

Next, a description is presented of how the charging area is represented in SB-CAPSA, how candidate schemes are generated, and the algorithms steps are detailed.

3.4.1.1 Representing the charging area

A fundamental feature in area-based pricing solution algorithms is the manner in which the charging area is represented. In SB-CAPSA, the charging boundary representation was motivated in part by the cutset-based heuristic proposed by Zhang and Yang. In this algorithm, preselected network nodes are classified as either inside or outside the charging area. The charging area itself is represented by a binary vector (chromosome) $\boldsymbol{\eta}$ of length equal to the number of preselected nodes, where each element of the vector is a

binary variable indicating if a node is in or outside the charging area. $\boldsymbol{\eta}$ is then used to define which network links have to be tolled. Although it is technically possible to use the representation $\boldsymbol{\eta}$, along with the toll information $\boldsymbol{\tau}$, to estimate a surrogate $s(\boldsymbol{\eta}, \boldsymbol{\tau})$ of the objective function $F(\boldsymbol{\eta}, \boldsymbol{\tau})$, generally it would impractical. A surrogate model is estimated using as data points a set of candidate solutions and their corresponding objective function values (e.g., $\{(\boldsymbol{\eta}, \boldsymbol{\tau}, \boldsymbol{F}(\boldsymbol{\eta}, \boldsymbol{\tau}))\}$, with $\boldsymbol{\eta}$ and τ as the dependent variables). For the surrogate model to be useful, or even capable of being estimated, the number of data points used in its estimation would have to be least greater than the number of unknowns in the model. With the $\boldsymbol{\eta}$ representation, the number of unknowns is $d = |\boldsymbol{\eta}| + 1$ (plus one for the toll level), where $|\boldsymbol{\eta}|$ could easily be greater than a 100 nodes. As an example, if a cubic radial basis function (RBF) model is used as the surrogate for $F(\boldsymbol{\eta}, \boldsymbol{\tau})$, the minimum number of initial candidate solutions needed for its estimation is $d + 1$, and, in practice, usually more than $d + 1$ data points are used in order to estimate a model with sufficient predictive power. So, representing the charging zone using the $\boldsymbol{\eta}$ approach would require considerable computational time.

For a surrogate-based solution algorithm, the charging area representation used would, ideally, contain all the relevant information of the charging scheme, impose minimal data requirement for the estimation of the surrogate model, and, yet, be useful in the search for solutions. A geometric representation of the charging area's boundary is used in SB-CAPSA because it can be used to represent complex shapes with relatively minimal information. One simple and general approach is to model the boundary's shape as a piecewise polynomial function that is defined in part by a collection of points ("knots"). Specifically, in this chapter the charging area boundaries are represented using simple polygons, which are defined by an ordered collection of p points or vertices $\boldsymbol{P} = \{(r_i, \theta_i): i = 1, \dots, p\}$. r_i is the point's radial distance from an artificial area center O, and θ_i is its angle relative to the horizontal direction from point O. A piecewise linear representation is used, as opposed to a more complex shapes (e.g., cubic B-splines), because of its simplicity; the polygon is completely specified by the vertices without the need for additional parameters.

Associated to each candidate polygon boundary are $2p + 1$ unknowns (the radii and angles of the p vertices and the toll).

In the proposed algorithm a charging area boundary is created by generating vertices $\boldsymbol{P}$ for a simple polygon. Network nodes that are inside the polygon area are inside the charging area, and, naturally, nodes outside the polygon are outside the charging area. This leads to the secondary $\boldsymbol{\eta}$ representation proposed by Zhang and Yang. So $\boldsymbol{P}$ maps to $\boldsymbol{\eta}$, and $\boldsymbol{\eta}$ maps to the network as tolling indicators for the links. The objective function value F, as computed with the computationally expensive models, is calculated by using $\boldsymbol{\eta}$ and toll τ. An objective function estimate $\hat{F}^m$ of charging area candidate m can be obtained from a surrogate model s estimated using a collection of n previously evaluated candidates, represented by n data points $\{(\boldsymbol{P_j}, \tau_j, F_j)\}$. Given a candidate boundary $\boldsymbol{P}^m$ and toll τ^m, the surrogate is used to predict $\hat{F}^m = s(\boldsymbol{P}^m, \tau^m)$. Next, how the surrogate model is used to search for good solutions to the area pricing problem is discussed.

3.4.1.2 Overview of surrogate-based search procedure

SB-CAPSA follows the general logic of the metric stochastic response surface algorithms proposed by Regis and Shoemaker (2007). An initial population of n polygons and tolls is generated and evaluated with the computationally expensive models. The resulting data points $\{(\boldsymbol{P_j}, \tau_j, F_j)\}$ are used to estimate surrogate model s. In each iteration, five groups of candidate points are generated according to different set of rules (explained in the next section), and, among thousands of candidates, a candidate is selected for each group. This selection is partly made based on the surrogate model's prediction of each candidate's objective function value. Surrogate model s is updated after each iteration as new information (i.e., new data points $(\boldsymbol{P_j}, \tau_j, F_j)$) is learned.

During the candidate generation procedures, each candidate is given a score ($W = w^{RS}U^{RS} + w^{D}U^{D}$) according to its predicted objective function value (surrogate score U^{RS}), a measure of each candidate's distance to previously evaluated points (distance score U^{D}), and weights w^{RS} and w^{D} corresponding to

these two criterions ($w^{RS} + w^{D} = 1$). For a candidate scheme $(\boldsymbol{P}^m, \tau^m)$, the score $U^{RS,m}(\boldsymbol{P}^m, \tau^m)$ is computed using:

$$U^{RS,m}(\boldsymbol{P}^m, \tau^m) = \frac{s(\boldsymbol{P}^m, \tau^m) - s_{min}}{s_{max} - s_{min}} \tag{3.6}$$

where s_{min} and s_{max} are the minimum and maximum predicted objective function values, respectively, among the candidate solutions of each group. If s_{min} and s_{max} are equal, then $U^{RS}(\boldsymbol{P}^m, \tau^m) = 1$.

The purpose of the distance criterion score U^D is to measure how different a candidate solution is from previously evaluated solutions. Given that multiple polygons can result in the same charging zone (in terms of which nodes are within the charging area), the distance criterion is computed using the $\boldsymbol{\eta}$ mapping of each polygon. The minimum distance $\Delta(\boldsymbol{\eta^m}, \tau^m)$ between candidate $(\boldsymbol{\eta^m}, \tau^m)$ and the previously n evaluated schemes is computed along with the related minimum $\left(\Delta_{min} = \min_{1 \le j \le n} \{\Delta(\boldsymbol{P}^m, \tau^m)\}\right)$ and maximum $\left(\Delta_{max} = \max_{1 \le j \le n} \{\Delta(\boldsymbol{\tau}^m)\}\right)$ distances. Score $U^{D,m}(\boldsymbol{P}^m, \tau^m)$ is calculated using:

$$V^{D}(\boldsymbol{\tau}^m) = \frac{\Delta(\boldsymbol{\tau}^m) - \Delta_{min}}{\Delta_{max} - \Delta_{min}} \tag{3.7}$$

Again, if Δ_{min} equals Δ_{max}, then $V_n^{dist}(\boldsymbol{\tau}^m) = 1$. The weight of each criterion (w^{RS} and w^{D}) is cyclically adjusted in each iteration. For this purpose, an ordered set $\Upsilon = \langle v_1, \dots, v_\kappa \rangle$ $(0 \le v_1 \le \cdots \le v_\kappa \le 1)$ is defined, and with each changing iteration w^{RS} is sequentially assigned a value from Υ. The purpose of adjusting the weights in this manner is to alternate between an exploitative and an explorative search focus.

After each candidate m is given a score W, in each candidate group the scheme $(\boldsymbol{P}^m, \tau^m)$ with the highest score is selected and evaluated with the computationally expensive models. The information of the evaluated candidates is stored in data archives which are used to re-estimate the surrogate model at the start of each iteration, as previously mentioned. An additional data archive is used to store the best known charging area scheme (i.e., the feasible solution with highest objective function value). This best known solution is used in the generation of candidates. SB-CAPSA stops when a convergence criterion is met,

which given the computationally expensive nature of the problem, is specified as a maximum number of iterations.

3.4.1.3 Procedure to generate candidate solutions

Of the five types of candidates generated in each iteration of SB-CAPSA, the first four candidate groups are created using information from the best known charging area boundary $\boldsymbol{P^{best}} = \{\boldsymbol{r^{best}}, \boldsymbol{\theta^{best}}\}$. The first group of candidates (type I) share the boundary $\boldsymbol{P^{best}}$, but differ in their toll level τ^m. The number of type I candidates, M^I, depends on how many of $\boldsymbol{P^{best}}$ possible toll values have been evaluated. In the unlikely case that all tolls are evaluated for $\boldsymbol{P^{best}}$, the analyst could decide to simply stop producing type I candidates or switch the $\boldsymbol{P^{best}}$ information with data from another feasible solution (e.g., the second best known solution). In numerical tests this situation was not encountered, as the best known solution changes before all tolls have been considered for a particular $\boldsymbol{P}$.

In the second group of candidates (type II), M^{II} candidate charging area boundaries are generated by enlarging the $\boldsymbol{P^{best}}$ boundary. The radii of $\boldsymbol{P^{m,II}}$ candidates is created by perturbing the $\boldsymbol{r^{best}}$ elements according to the formula $r_j^{m,II} = r_j^{best}(1+\omega)$, where ω is a randomly generated value in the interval [0,1]. ω can be generated, for example, by making random draws from a beta distribution. A similar approach is used for generating type III candidates, but using the formula $r_j^{m,III} = r_j^{best}(1-\omega)$; the $\boldsymbol{P^{best}}$ boundary is contracted in the M^{III} type III candidates. For the fourth group of candidates (type IV), both expansion and contraction operations are performed on the $\boldsymbol{P^{best}}$ boundary. The radii of the M^{IV} candidates are computed using $r^{m,IV} = r_j^{best}(1+b\omega)$, where b is a random variable that can assume the values of -1 and 1 with equal probability. The angles of candidates for groups II, III, and IV are calculated with the formula $\theta_j^m = \theta_j^{best} + z(0,\sigma)$, where z is a random draw from a normal distribution with mean zero and variance σ. Lastly, for the last group of candidates (type V), M^{V} boundaries are generated randomly, without using $\boldsymbol{P^{best}}$ as a reference.

After generating the boundary locations, the associated toll levels are generated for each candidate m. Here it is assumed that the planner is interested in $|\boldsymbol{\tau}|$ number of tolls from a finite set $\tau = \{\tau_{min}, \dots, \tau_{max}\}$. For each candidate charging boundary $(\boldsymbol{P^m}, \tau^m)$, all possible tolls are considered; $|\boldsymbol{\tau}|$ candidate schemes are created with the coupling of boundary $(\boldsymbol{P^m}, \tau^m)$ with all possible tolls. So from groups II through V a total of $M = |\boldsymbol{\tau}| \times (M^{II} + M^{III} + M^{IV} + M^{V})$ schemes are generated.

M candidates are generated, but a W score is not generated for the M candidates. A candidate m must pass a series of checks before being considered for evaluation. The polygon $\boldsymbol{P^m}$ has to be a simple polygon (i.e., composed of non-intersecting line segments), as intersecting polygons generally have no practical meaning. If the polygon is simple, it is then checked that its transformation to $\boldsymbol{\eta}^m$, along with its toll τ^m, have not been previously evaluated. Finally, here we impose the cutset requirement used by Zhang and Yang, so the only polygons accepted are those that reduce the rank of the road network's graph incidence matrix by one when the columns representing the tolled links leading to the charging area are eliminated (see Zhang and Yang 2004).

3.4.2 Algorithm steps

The algorithm steps and the notation used in the description of SB-CAPSA are given below.

Decision variables

τ	**:**	charging zone toll level
$\boldsymbol{P}$	**:**	vertices defining changing zone boundary

Indices

m	**:**	indices for candidate solution vectors generated according to the rules of groups I through V
u	**:**	indices for the five candidate solutions selected for evaluation ($u = \mathrm{I}, \mathrm{II}, \mathrm{III}, \mathrm{IV}, \mathrm{V}$)
j	**:**	indices for solutions that were evaluated with the models ($j = 1, \dots, n$)

Counters

n : counter for the number model runs

Functions

$F(\cdot)$: objective function

$\tilde{F}(\cdot)$: penalty-adjusted objective function

$o(\cdot)$: feasibility indicator equal to $\sum_r \min\left(C_{r,max} - \bar{C}_r, 0\right)$

$s(\cdot)$: surrogate model

$W(\cdot)$: weighted candidate score

χ : function that indicates if a feasible solution was found

μ : function that indicates if F_{min} was updated

Parameters

n_0 : initial number of evaluated solutions

n_{max} : maximum value for n

Sets

T : set for τ values

P : set for $\boldsymbol{P}$ values

Λ : set for $F(\cdot)$ values

Ψ : set for $o(\cdot)$ values

Ω : set for $\tilde{F}(\cdot)$ values

Algorithm Steps

1. Initialization:

1.1. Set $n = n_0$.

1.2. Generate n initial candidates $(\boldsymbol{P}_j, \tau_j)$.

1.3. Add vectors $\boldsymbol{P}_j$ and τ_j to P and T, respectively, and initiate Λ , Ψ, and Ω as empty sets.

2. Evaluate initial candidates and select initial best solution:

2.1. For each candidate $(\boldsymbol{P}_j, \tau_j)$, compute $F(\boldsymbol{P}_j, \tau_j)$ and $o(\boldsymbol{P}_j, \tau_j)$, and add values to Λ and Ψ.

2.2. If there are feasible solutions according to the $o(\boldsymbol{P}_j, \tau_j)$ values, the $\boldsymbol{P}_j$ and τ_j vectors for the candidate with the highest $F(\boldsymbol{P}_j, \tau_j)$ are labeled $\boldsymbol{P}^{\boldsymbol{best}}$ and τ^{best}, respectively. Additionally, the lowest feasible $F(\boldsymbol{P}_j, \tau_j)$ is labeled F_{min}, and $\chi = 1$. Otherwise, if there are no feasible solutions, $\boldsymbol{P}^{\boldsymbol{best}}$ and τ^{best} are assigned the values of the candidate with the highest $o(\boldsymbol{P}_j, \tau_j)$ value, F_{min} is assigned a placeholder value, and $\chi = 0$.

2.3. For each initial point, compute $\tilde{F}\left(\boldsymbol{P}_{\boldsymbol{j}}, \tau_j\right)$ and add values to Ω.

2.4. Determine $\tilde{F}^{best}$.

3. Fit surrogate model s using information in P, T and Ω.

4. Candidate point generation and selection:

4.1. Generate candidates according to the specifications of groups I through V.

4.2. For each group and each candidate point m, compute $W(\boldsymbol{P}^{\boldsymbol{m}}, \tau^m)$.

4.3. For each candidate group, select for model evaluation the candidate solutions with maximum $W(\boldsymbol{P}^{\boldsymbol{m}}, \tau^m)$.

5. Candidate evaluation and updates of parameters and archives:

5.1. Evaluate the five selected candidates u with the computationally expensive models to determine $F(\boldsymbol{P}^{\boldsymbol{u}}, \tau^u)$ and $o(\boldsymbol{P}^{\boldsymbol{u}}, \tau^u)$.

5.2. Update F_{min} if possible: If $\chi = 1$ and there is a feasible candidate u with $F(\boldsymbol{P}^{\boldsymbol{u}}, \boldsymbol{\tau}^u) < F_{min}$, then $F_{min} = F(\boldsymbol{P}^{\boldsymbol{u}}, \boldsymbol{\tau}^u)$. Else, if $\chi = 0$ and there are one or more of the five candidate points that are feasible, update F_{min} with the worst feasible $F(\boldsymbol{P}^{\boldsymbol{u}}, \boldsymbol{\tau}^u)$ and set $\chi = 1$. If F_{min} is updated, $\mu = 1$.

5.3. If $\mu = 1$, use the new F_{min} to update values in Ω, update $\tilde{F}^{\text{best}}$, and then set $\mu = 0$.

5.4. If any feasible candidate $\tilde{F}(\boldsymbol{P^u}, \tau^u)$ is greater than $\tilde{F}^{best}$, set $\tilde{F}^{best} = \tilde{F}(\boldsymbol{P^u}, \tau^u)$, $\boldsymbol{P^{best}} = \boldsymbol{P^u}$, and $\tau^{best} = \tau^u$.

5.4 Increase counter $n = n + 5$, and add information from evaluated candidate points to Λ, Ψ, P, T, and Ω. If $n \leq n_{max}$ return to step 3; otherwise, continue to step 6

6. Return $\boldsymbol{P^{best}}$ and τ^{best}

3.5 Numerical Tests

The tests presented in this section explore the accuracy of the surrogate models estimated with data obtained from the proposed geometric representation of charging area boundaries. Additionally, a sample application of the algorithm is presented. All tests assume a sequential three-step planning model composed of trip distribution, mode split, and traffic assignment procedures, which are applied for four time periods in a day (morning, midday, afternoon, night). The output of the planning model is used to compute link emissions and, subsequently, pollutant concentration at receptor points. In Figure 3-1 the overall model structure used to compute the upper level objective function and constraints is shown. A range of $1 to $40 dollars was set for $\boldsymbol{\tau}$, with $1 increments from $1 to $40. In this example, it is assumed that the planner is interested in reducing the average daily pollutant concentration experienced in the city of Chicago.

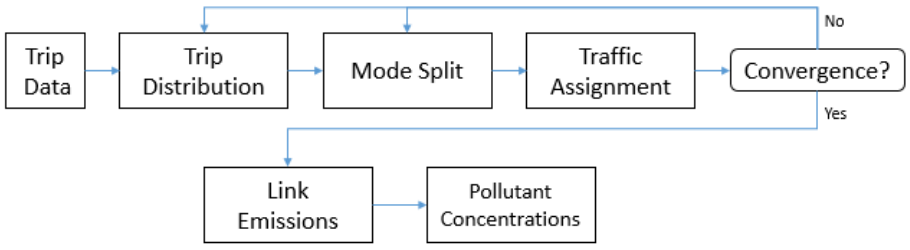


Figure 3-1 Structure of integrated model used in numerical tests

3.5.1 Models and related parameters

The network users behavior was simulated using a standard travel demand planning model structure, without the trip generation step. A portion of the travel demand distribution was assumed fixed, meaning

that the charging area did not have any effect on the users' choice of destinations. In the short term, this would be expected for people whose work requires them to be present at a particular location. The distribution of a relatively small portion of the travel demand is allowed to vary. The destination choice of trips originating in a particular zone was modeled using a multinomial logit formulation, where each destination's utility was specified as a function of the mode choice logsum and an activity level variable, which in this case was computed based on the employment level of each zone. Only auto and transit modes were consider in the binary logit mode split model. Mode choice was specified as a function of travel time and cost. For auto the cost was computed using a fixed per mile cost (60 cents per mile) and the corresponding tolls, while for transit only fixed per mile costs (2 cents per mile) were considered. Also, transit was assumed to have fixed travel times (i.e., changes in road network congestion have no effect on the transit travel time). The final component of the model structure is the traffic assignment model. User equilibrium behavior was assumed, and the resulting problem was solved using the gradient projection method (Jayakrishnan et al. 1994). A value of time of $30 per hour was used to convert tolls to link specific costs. The three components are applied to find links flows for four model time periods. A feedback loop based on the link flows was implemented at the trip distribution and mode split steps using the method of successive averages (Boyce et al. 2008). Only three loops were employed in this example, because of time constraints. Tables 3-1 and 3-2 present the parameters assumed for the destination and mode choice models.

Table 3-1 Parameters for destination choice model

Variable	Parameter value
Attraction	10
Impedance (logsum)	0.5

Table 3-2 Parameter for mode split model

Variable	Parameter value
Car constant	4
Travel time (minutes)	-0.1
Travel cost (cents)	-0.002

Particulate matter ($PM_{2.5}$) is the pollutant of interest in this example. Link emissions were estimated using speed-based factors extracted from EPA's MOVES model. Once the link emissions were computed, concentration at the receptor points were estimated using a Gaussian plume model and the cell-based methodology utilized by Zhang et al. (2010). Four different meteorological conditions (with distinct wind speeds, directions, and atmospheric stability classes) for each time period were employed to compute the average pollutant concentrations at the receptor points. The blue region in Figure 3-2 represents the area with the 14,057 receptor points of interest for which concentrations were computed. The receptors were spaced 656 feet apart. The circle in the figure represents the artificial point O from which the charging boundary polygon vertices are referenced. The farthest receptor of interest is approximately 50000 feet from point O. In the sample application, the area pricing scheme is feasible if concentrations at all receptors points are reduced by 5 percent relative to the base situation (before the pricing scheme).

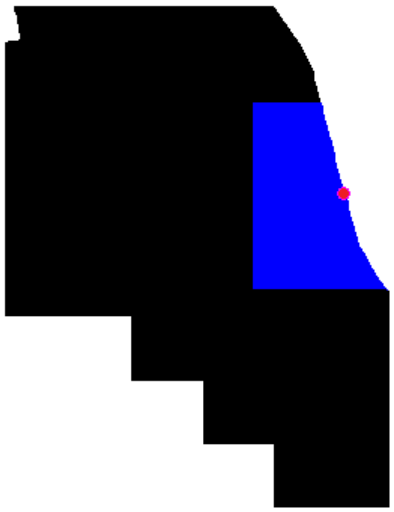


Figure 3-2 Receptor region

3.5.2 Network and demand data

A section of the Chicago Sketch Network, corresponding to the Cook and DuPage County road networks, was used in the numerical tests. Figure 3-3 presents the network. This network has 1330 links and 421 nodes, of which 160 are origin-destination zones. The network data and its related trip distribution were obtained from an online data repository (Bar-Gera 2014). The provided trip data was augmented by a factor

of 1.5, and assumed as the fixed travel demand distribution for the morning period. For the remaining model periods, the fixed travel demand distribution was assumed to be the given trip data times factors 0.3, 1.05, and 0.15, respectively. Additionally, variable travel demand distributions were created for each time periods by multiplying the computed fixed travel demand by factors 0.2, 0.1, 0.2, and 0.5, respectively.

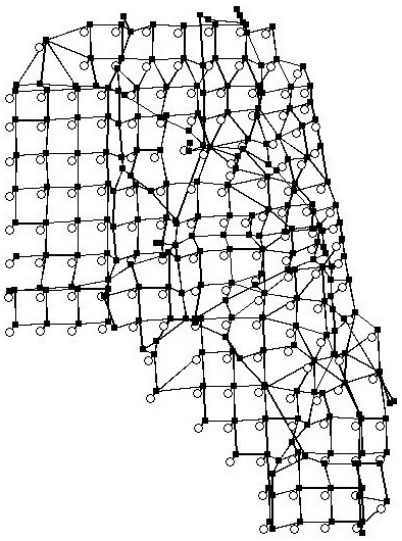


Figure 3-3 Section of the Chicago Sketch Network

3.5.3 Testing accuracy of surrogate models

A geometric representation of a charging area's boundary is arguably intuitively reasonable. However, it is not immediately clear if such a representation would be useful in a surrogate-based solution algorithm for cordon and area pricing problems. For this reason a series of tests were conducted to explore the predictive accuracy of surrogate models estimated using the proposed simple polygon representation of charging boundaries. The tests were conducted using a sample of 400 charging schemes with randomly generated polygons and tolls.

Each scheme's polygon boundary was generated using 10 vertices. The radii and angles of these vertices were measured from a reference point set at (710070, 1931400) (in feet, see node coordinates in Bar Gera 2014), in the vicinity of Chicago's central business district. Radii were allowed to assume values in the interval [16500, 75500] (feet), while the angles were bounded to [1.8, 5] (radians). For each candidate

polygon, two vertices had their angles fixed to 1.8 and 5 radians (i.e., fixed at the angle bounds). For each pricing scheme the total toll revenue was computed using the output of the transportation planning model previously discussed. Radial basis functions (RBF) were used as surrogate models.

Four RBF functional forms were tested: linear, cubic, thin plate spline, and multiquadratic (with parameter 0.001) (Gutmann 2001). For each functional form a series of k-fold cross validation tests were performed. In *k*-fold cross validation the sample is divided into *k* subsamples. A partition *k* is used once as the validation data set, while the remaining partitions are used to fit the surrogate model; this is repeated for all partitions. Folds of 2, 5, and 10 were used in the tests, with sampling for each fold performed 30 times. In addition to the k-fold cross validation tests, cross validation was performed by randomly selecting 100 data points (candidate schemes) as the training data set, while the remaining 300 data points were used for validation. Again, these tests were repeated 30 times.

With the validation data a series of measures of fit were computed. The correlation between the surrogate model predictions and the planning model values was computed using the Pearson correlation coefficient (PC). Mean average percent error (MAPE), root mean squared error (RMSE), and normalized root mean square error (NRMSE) were used as error metrics. The test results are presented in Table 3-3. On average, a correlation of 0.76 was observed between the surrogate predicted values and the planning model values. The highest correlations were observed with the multiquadratic and linear functional forms. Average MAPE, RMSE, and NRMSE for all tests were 15%, 161, and 3%. Again, based on these metrics the multi-quadratic functional form had the best performance. There are several other RBF functional forms, other types of surrogate models, and other geometric approaches for representing charging boundaries that should be examined. But, these preliminary results, particularly the level of correlation between predicted and modeled values, suggest that the use of the proposed surrogate model concept to predict charging scheme's objective functions is an idea worth exploring.

Table 3-3 Validation results from cross-validation tests

Form	Cross-validation	mean PC	MAPE (%)	RMSE	NRMSE (%)
linear	10-fold	0.795	13.4	146	3.02
	5-fold	0.799	13.5	150	2.97
	2-fold	0.795	14.2	157	2.95
	100/300 split	0.774	15.5	165	3.05
cubic	10-fold	0.734	14.9	159	3.34
	5-fold	0.733	15.5	168	3.29
	2-fold	0.708	16.6	179	3.38
	100/300 split	0.661	18.8	194	3.63
thin-plate	10-fold	0.766	13.7	149	3.01
	5-fold	0.777	13.8	153	3.01
	2-fold	0.771	14.4	159	3.00
	100/300 split	0.729	16.5	173	3.21
multi-quadratic	10-fold	0.792	13.4	144	3.00
	5-fold	0.800	13.5	150	2.95
	2-fold	0.791	14.2	157	2.97
	100/300 split	0.780	15.9	168	3.13
	All tests	0.763	14.9	161	3.12

3.5.4 Sample application of SB-CAPSA

The objective functions discussed in section 3.3 were used in the applications of the proposed algorithm. For the consumer surplus objective function F_{CS}, all population groups were assumed to have the same marginal utility of income, so this term was dropped (equivalently, it was assumed to be 1). Additionally, the logsum of the base condition ($\ln\left(\sum_j e^{V_{ntj}}\right)$), given that it is a constant, was removed from the formulation. For the auto trip objective function, ξ was set to one, and the q_{wt} terms were removed from the function, again, because these terms are constants.

An initial population n_0 of 100 polygons, with randomly assigned tolls, was generated for all tests. The maximum number of model evaluations, n_{max}, was set to 250. The polygons were produced using the same information used for the accuracy tests in section 3.5.3. M^{II}, M^{III}, M^{IV}, and M^{V} where set to 5,000. The ω parameter used in the radius perturbation was drawn from a beta distribution with shape parameters set to $Beta(2{,}20)$. The θ_j^{best} angles were perturb using a variance σ of 0.3. The set Υ was specified as $\{0.75, 0.85, 0.90, 0.95, 0.97, 1\}$. A parameter free cubic radial basis function (RBF) model was employed as the surrogate model (Gutmann 2001).

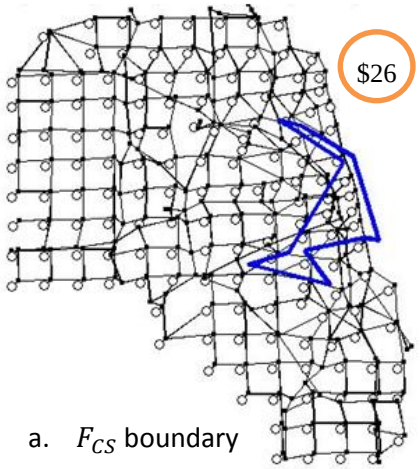
In Figure 3-4, the best cordon boundaries found after 250 candidate evaluations, along with their corresponding tolls, are presented. The change in the best known solution with each model evaluation is given in Figure 3-5. Figure 3-6 presents the mode shares and trip distributions for the based condition and the three solutions. As it would be expected, different objectives result in different cordons. The status quo objective resulted in an aggregate mode share most similar to the base situation. The revenue objective resulted in the highest possible toll. No general insights can be derived, partly because the presented solutions are most likely not the optimal solutions; just the best solutions found after 250 model function evaluations.

3.6 Closing Remarks

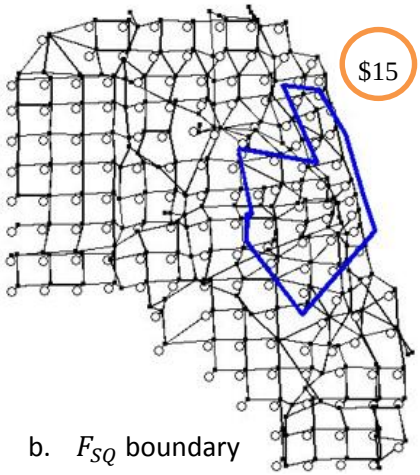
A variant of the cordon and area-based pricing problems was proposed which incorporates pollutant concentration constraints. Departing from the formulation of previous problems, a series of general objective functions were proposed that can be utilized with state-of-the-practice travel demand planning models. Additionally, a new surrogate-based solution algorithm, SB-CAPSA, was proposed for cordon and area-based discrete network optimization problems. As part of the new algorithm, a geometric-based method for representing charging boundaries and estimating surrogate models was proposed. Numerical tests suggest that arguably good predictive accuracy can be obtained by applying the proposed surrogate model methodology with radial basis functions. Lastly, an example was provided to illustrate the application of SB-CAPSA.

The formulated problem can be extended to account for several practical considerations that are part of the public and political debate surrounding the design of cordon and charging area schemes, including discounts for citizens that live within the charging boundary, time-of-day toll variability, and toll exceptions for particular users or vehicles. Further research is also needed to explore SB-CAPSA's multiple components. For instance, more sophisticated approaches could be utilized to create boundary mutations. Currently, random perturbations are independently applied to each polygon vertex. An alternative approach would be

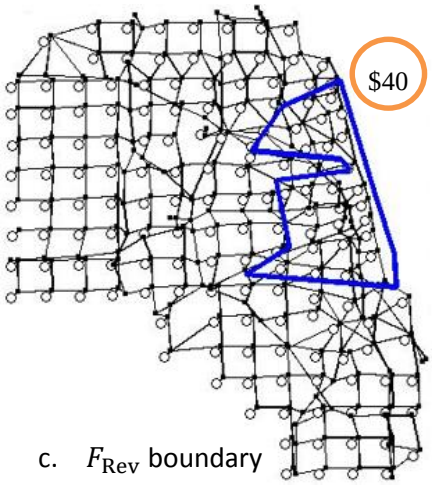
to coordinate the perturbations, perhaps producing smoother shape transformation and improving the success rate in finding better solutions. Also, alternative geometric representations (e.g., using cubic B-splines) should be explored. Even with the simple polygon representation, additional tests are required to explore what is the impact of the number of vertices utilized.



a. F_{CS} boundary

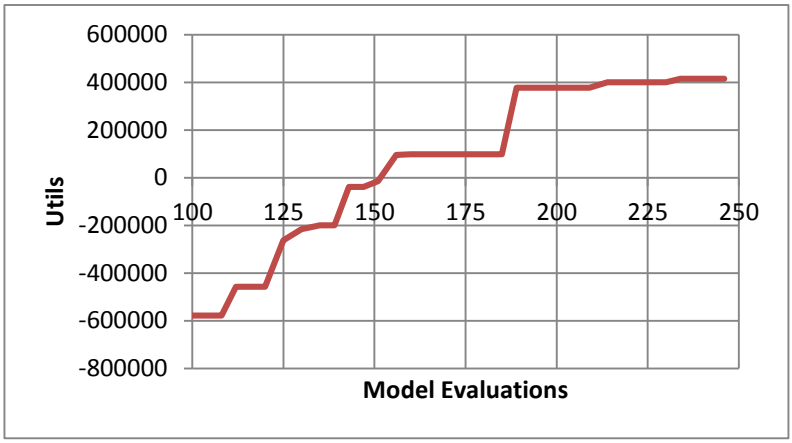


b. F_{SQ} boundary

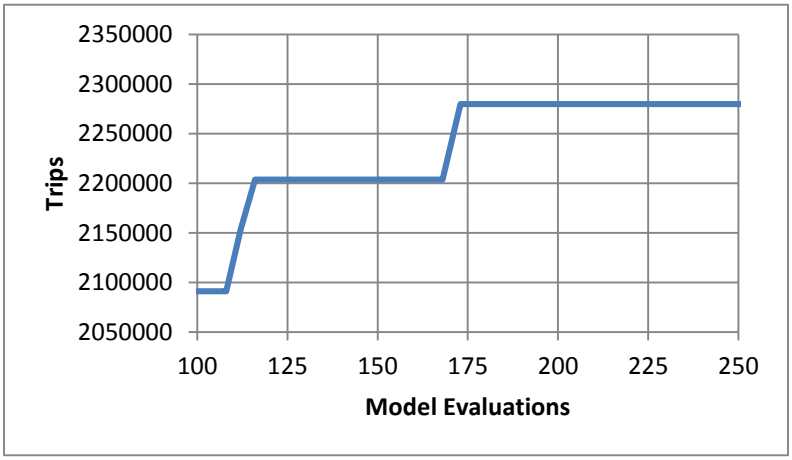


c. F_{Rev} boundary

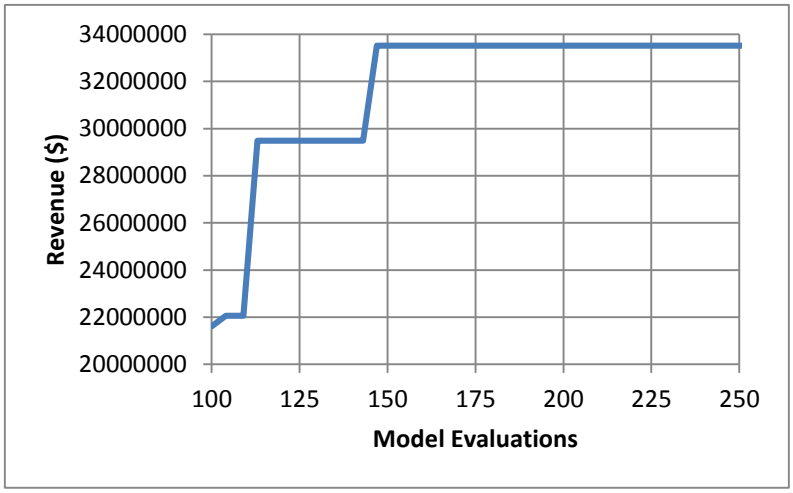
Figure 3-4 Charging boundaries for Chicago Sketch Network



a. F_{CS} objective

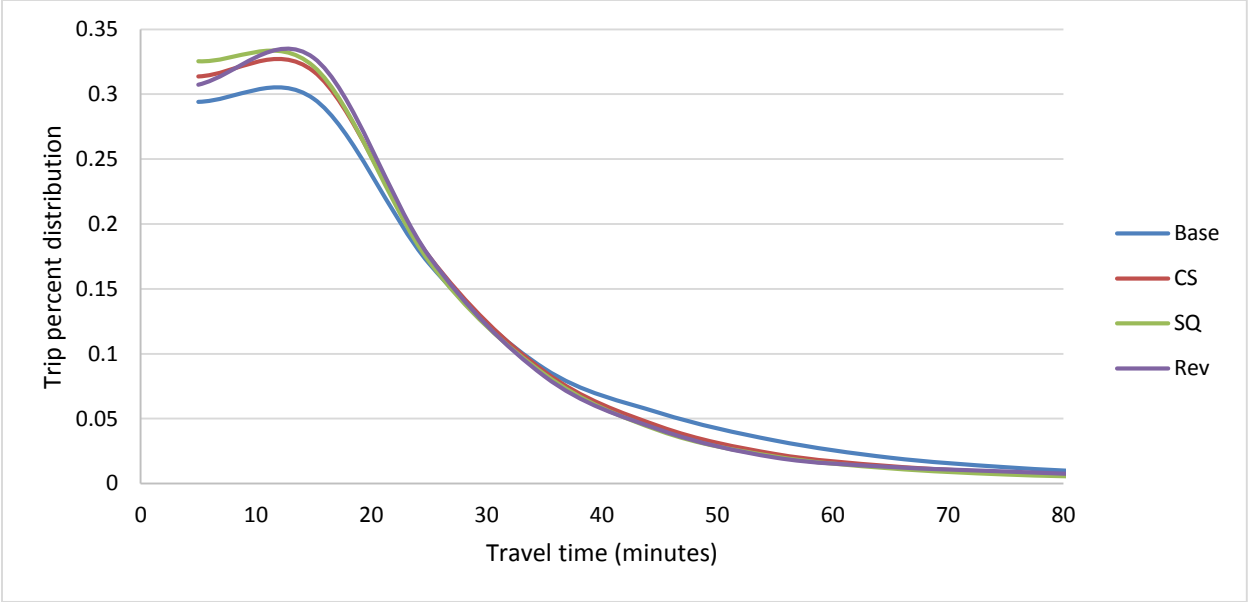


b. F_{SQ} objective

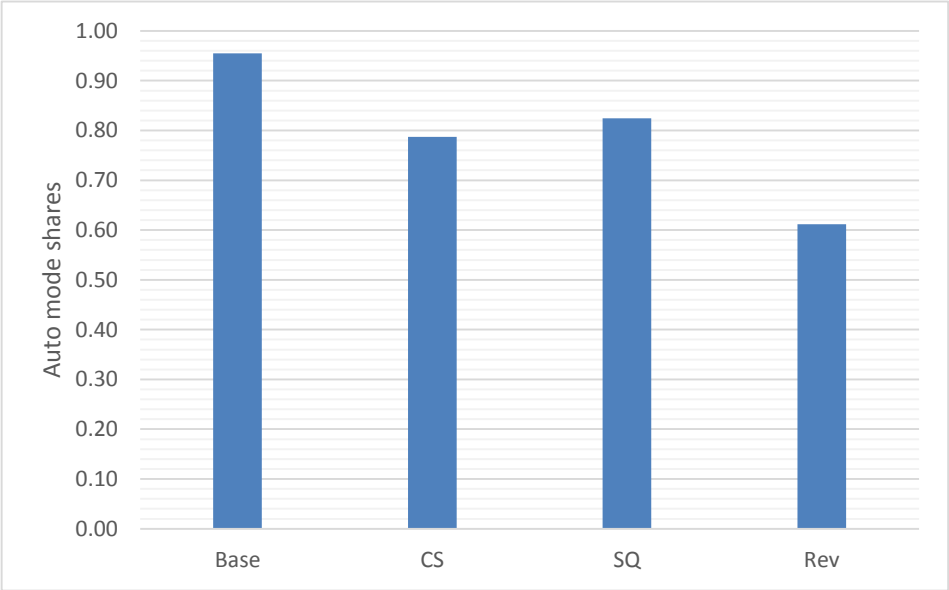


c. F_{Rev} objective

Figure 3-5 Change in best known objective function value



a. Trip length distribution



b. Auto mode share

Figure 3-6 Differences in trip length distribution and mode share due to charging schemes

Chapter 4

Designing traffic rationing schemes to reduce air pollution and environmental inequality

4.1 Introduction

The political intractability of road pricing and its possible negative equity impacts has in recent years revived interest in the study of non-pricing based travel demand management strategies. One of such strategies is what in this chapter will be called traffic rationing, that is, schemes that restrict the number of vehicles on a road network using non-pricing control mechanism, such as permit systems or license plate based programs. In contrast to road pricing schemes, there are multiple examples of major cities implementing traffic rationing schemes with the explicit objective of reducing critical levels of air pollution (e.g., Mexico City, Paris, Milan, and Beijing).

In this chapter a bi-objective simulation-based optimization problem is proposed to determine traffic restriction levels that maximize total road traffic and minimize pollutant exposure inequalities while reducing emissions and pollutant concentration levels below given maximum values. A probabilistic pollutant exposure approach is used to simulate pollutant exposure at the person level. Like the road pricing problems discussed in previous chapters, the problem considered in this chapter poses a significant computational challenge. For this reason, a new surrogate assisted multi-objective differential evolution algorithm (SAMDE) is proposed to solve the traffic rationing problem.

This chapter is organized as follows. In the next section, literature related to the design of traffic control strategies is reviewed, both for the cases of congestion and pollution control. Additionally, research on probabilistic modeling of personal pollutant exposure is discussed. This is followed by the formulation of the bi-objective traffic rationing problem. Next, the steps of the SAMDE algorithm are detailed and the heuristic is applied to a sample problem. Closing remarks and future research opportunities are presented in the last section of the chapter.

4.2 Literature Review

This section presents a literature review of models proposed to design non-pricing traffic control strategies, and of models used to estimate the exposure of individuals to air pollutants.

4.2.1 Quantity control schemes based on environmental considerations

Quantity control measures have been considered as possible vehicle emission reduction policies since at least the 1960s (Howitt and Altshuler 1999), and despite their apparent political impracticability these strategies have been implemented in a few cities. In practice, the type of program that have been implemented can be describes as "command-and-control" programs in the sense that they prescribe a goal that must be achieved without a direct financial incentive. More succinctly, Raux (2004) described this type of scheme as "cap but no trade" programs. Traffic rationing systems, such as Mexico's HNC program, are examples of this type of strategy. The design of rationing systems with no trade mechanisms has been studied from a congestion reduction perspective (e.g., Daganzo 1995; Han et al. 2010), and their effects on total emissions and pollutant concentration have been documented (e.g., Wang et al. 2007; Chen et al. 2008).

Another quantity control strategy are tradable permit systems. This type of system is analogous to a cap-and-trade scheme. Verhoef et al. (1997) discussed possible road transportation applications of tradable permits, including driver-oriented applications such as vehicle ownership permits, road usage permits, and parking permits. These user-oriented programs are known as downstream trading system, as opposed to the upstream trading systems that target industries in the fuel and vehicle manufacturing sectors. The following five factors are identified as important design considerations of tradable permit systems: type of policy target, geographic domain, initial distribution of permits, enforcement and monitoring of system, and degree of differentiation in the permit market given spatial and temporal considerations. Associated to the first factor is the important distinction between emission permit systems and ambient permit systems, the latter systems are aimed at controlling the amount of emissions while the former are systems designed to meet

ambient standards. Expanding on Verhoef et al., Raux and Mariot (2005) and Raux (2010) proposed fuel consumption permit systems for autos and trucks, a system that would fall under Verhoef et al.'s road usage permit system classification.

Though the use of tradable pollution permit programs to control the emission of firms has received considerable attention since the work of Montgomery (1972), few studies have considered this strategy for mobile emission sources. In light of Mexico City's experience with HNC, Goddard (1997) presented one of the first models to consider the use of travel demand restrictions, coupled with tradable permits, to control vehicle generated air pollution. Goddard's model, developed from an ecological economics perspective, is formulated as a cost minimization problem (or equivalently, a welfare maximization problem) where the decision maker's objective is to "minimize the total cost of emissions reduction subject to a constraint on allowable emissions". Bulteau (2012) developed a microeconomic theoretic model where consumers trade permits that regulate the number of kilometers that they can travel. In this model the decision maker defines the total amount of traffic emission in the region and freely distributes the permits among the consumers. Zhu et al. (2013) also developed a theoretical model for rationing, but in terms on vehicle ownership and usage and with emphasis on congestion mitigation. Feng et al. (2010) and Feng and Timmermans (2014) propose practical, bi-level optimization methods for determining vehicle ownerships rates and usage given environmental (emission) constraints. The model presented by Feng and Timmermans includes the objective of maximizing spatial equity.

In the field of transportation networks, research on tradable pollution permit systems has been conducted primarily by Nagurney and her colleagues. Like in previous environmental pricing studies, their analysis assumes that maximum allowable emission amounts can be defined on network components (e.g., links, paths, or origin and destination (OD) pairs), and these amounts can serve as an air quality standard constraint on the maximum emissions generated by the network users. The allowable amount of emission is represented by a number of permits. The traffic network equilibrium problems are solved subject to a permit constraint. Using a variational inequality approach, these researchers developed models for single and multimodal networks; for the cases of user compliance or noncompliance with the system (Nagurney

et al. 1998); for link, path, or origin and destination based systems (Nagurney 2000), and to consider different dynamic aspects surrounding permit systems (Nagurney and Zhang 2001).

4.2.2 Quantity control instrument for congestion mitigation

Until this point only environment-oriented quantity control studies have been discussed. A related body of literature concerned with controlling the amount of traffic from a congestion mitigation perspective might be helpful in conceptualizing alternative environment-oriented travel demand management strategies or considering alternative environmental goals.

Temporal considerations are a neglected component in the previously discussed quantity control models. The programs control where the emissions occur, but when the emissions occur during the day is not explicitly considered. In contrast, several models for quantity control travel demand management strategies have been proposed which consider the temporal effects of traffic on congestion. Teodorovic and Edara (2005) and Edara and Teodorovic (2008) proposed a highway booking system in which users send to the transportation manager requests to enter the highway at a given time for a specific trip (defined in terms of the trip's date and time, the access and egress ramps, the tariff paid, and the type of vehicle), and the transportation manager decides to accept or reject the request. The allocation of highway space for different types of trips is determined by solving an integer program with the objective of maximizing total travel passenger miles while not exceeding the links capacity. Similarly, Ma et al. (2010) developed a methodology to determine the optimal time slot allocation of demand that minimizes travel time along a highway. In this model the demand is allocated to time slots by determining a departure distribution that minimized the cost of time-slot-shifting (relative to the base time slot demand) given the constraint that the resulting flow on each link has to be less than its capacity. Akamatsu (2007) proposed a tradable bottleneck permit system where the decision maker issues time-dependent permits for the bottlenecks in the network with the objective of minimizing social transportation costs. Two types of permits distribution schemes are proposed: free distribution and auction based distribution. An alternative permit distribution method is proposed by Nie and Yin (2013). In their model drivers arriving at a bottleneck during the peak period must

pay either a relatively high toll or a number of mobility credits, while drivers arriving during the off-peak period are awarded mobility credits that they can either save to travel during a future peak period or sell to other driver.

Another type of quantity control strategy for congestion mitigation are area-based control systems. Teodorovic et al. (2008) developed a model for an auction-based downtown time slot program. In this system the city manager auctions time slots and users bid for the slots corresponding to their desired visit duration. The objective of the model is to maximize revenue subject to a maximum number of allowable vehicles in the downtown area, which is defined in terms of the number of parking spaces and the maximum accepted delay in links. Similarly, Zhao et al. (2010) proposed a downtown space reservation system which is analogous to the highway space reservation system, but different in that the allocation of trips is constrained by the downtown's network capacity, as opposed the links capacity of a highway, and the city manager is assumed to maximize two objectives: revenue and system throughput.

Static road network equilibrium studies have also explored the use of quantity control mechanisms to minimize congestion. Yang and Wang (2011) proposed the use of tradable credit to manage network congestion, and they studied the properties of the network equilibrium resulting from this type of scheme. Several extensions have been presented for the tradable credit models. For example, Wang et al. (2012) relaxed the homogenous user assumption by assuming users with different value of time. Nie (2012) considered the effects of transaction costs in the credit market. Shirmohammadi et al. (2013) studied the effects of demand and supply uncertainty in the analysis and design of tradable credit schemes. Wu et al. (2012) studied the design of tradable credit schemes in a multimodal networks considering different user income levels.

4.2.3 Probabilistic simulation of human exposure to air pollutants

A common strategy in state-of-the-practice models, like BenMAP, is to estimate population intake of pollutants by assuming that a person's pollutant exposure depends only on the pollutant concentration levels at the person's home location. This assumption is usually made, in part, because it allows for a model that

is not too data or computationally expensive. An alternative to this approach are probabilistic personal exposure models that attempt to simulate a person's pollutant exposure based on activity patterns. Naturally, this modelling approach requires considerable more data and computational power than the home-based approach. Early model developments utilized data from activity pattern diaries to estimate probability distributions that were used to probabilistically simulate, for example, starting time of work trips, trip duration, and mode choice, among other important factors (Ott et al. 1988). This basic idea has been extended by incorporating an even greater number of factors that determine the degree of pollutant exposure, such as the simulation of type of activities performed by individuals and the breathing rates associated with each activity (Zidek et al. 2005).

Examples of integrated transportation and personal exposure models can be found in the literature. For example, Hatzopoulou et al. (2011) developed an activity-based model to evaluate the pollutant exposure impacts of a series of policy interventions in Toronto, Canada. In this study an activity-based model (TASHA) was integrated with a multi-agent transportation simulation (MATSIM), a vehicle emissions model (MOBILE), and a pollutant dispersion model (CALPUFF). Similarly, Dhondt et al (2012) utilized an integrated activity-based model to simulate population exposure to pollutants in Flanders and Brussels, Belgium. In this study the activity-patterns were aggregated into OD matrices which were then assigned to the road transportation network using equilibrium traffic assignment models. This hybrid micro-macro approach is used in the numerical example presented in section 4.5.

4.3 Traffic Rationing Problem Formulation

Nagurney (2000a) considers the problem of determining what is the maximum achievable travel demand given a network-level environmental quality standard. Here a related problem is proposed. Consider a region where pollutant concentration levels and total greenhouse gas emissions exceed given thresholds. Additionally, the region's authorities are concerned about the level pollutant exposure inequality among the population. The planners decide to cap car usage in different zones in the region using a permit system (or any other control mechanism). The region is divided into Z zones, and for each zone a number of car

usage permits are randomly distributed among the population of car users. The number of car users in each zones is assumed fixed; not receiving a permit is the only reason for a car user to shift to other modes. Furthermore, it is assumed that transit services can accommodate the additional demand diverted from auto.

4.3.1 Objective functions

The planner has two objectives: maximizing total auto usage and minimizing inequalities in pollutant intake. Define q_z as the total demand for car usage in zone z ($z = 1,2, \dots, Z$). For each zone, define the number of distributed permit (p_z) as:

$$p_z = \lfloor (1 - x_z) q_z \rfloor \tag{4.1}$$

x_z is the percent level reduction of the total demand for car usage demand in zone z. x_z is bounded by $[x_{z,min}, x_{z,max}]$. The first objective is defined as follows:

$$\max_{\boldsymbol{x}} P = \sum_z p_z = \sum_z \lfloor (1 - x_z) q_z \rfloor \tag{4.2}$$

Alternatively, objectives could be formulated to consider disparities in travel costs between rationing districts as a result of the rationing scheme. Minimizing deviations between rationed and original travel demand offers a way of reflecting both the objective of maximizing auto usage and the objective of reducing rationing disparities among zones. This objective can be expressed as:

$$\min_{\boldsymbol{x}} P = \sum_z \left((1 - x_z) q_z - q_z \right)^2 = \sum_z (x_z q_z)^2 \tag{4.3}$$

For the second objective, the Atkinson index A is used to represent the level of pollutant intake inequality among the population. Let I_n represent an individual person n's pollutant intake. Then, the second objective is defined as:

$$\min_{\alpha} A = 1 - \left[\frac{1}{N} \sum\nolimits_{n=1}^{N} \left(\frac{I_n}{\bar{I}} \right)^{1-\varepsilon} \right]^{\frac{1}{1-\varepsilon}} \tag{4.4}$$

where N is the total population in the region, $\bar{I}$ is the average pollutant intake in the region, and ε is a parameter that indicates the level of aversion to inequality (here assumed not to be equal to one). Next, the method used here to compute personal level pollutant intake is discussed.

4.3.1.1 Person-level modeling of pollutant intake

An individual n's air pollutant intake I_n during a day is a function of the sequence of activities undertaken in the day, the location and duration of each activity, and the pollutant concentration encountered in each location, among other factors. Intake I_n can be decomposed into $I_{n,S}$, the intake of the individual while the person is at a particular location (e.g., home, office) and $I_{n,M}$, the person's pollutant intake while moving from one location to another (e.g., going from home to work). I_n is computed using:

$$I_n = I_{n,S} + I_{n,M} \tag{4.5}$$

Let meS be a microenvironment visited by person n for a duration δ_{meS}. Assuming no indoor pollutant sources, the concentration at meS can be computed using $pF_{meS} \times C_{meS,t}$, where pF_{meS} is a infiltration factor that converts the outdoor concentration $C_{meS,t}$ at time period t to the microenvironment's pollutant concentration. Also, for simplicity a single breathing rate B_n is assumed for all time periods of the day. Then, summing over all microenvironments where the person remains relatively stationary during a period of time, $I_{n,S}$ can be defined as:

$$I_{n,S} = B_n \sum_{m \in S} \delta_{meS} \times pF_{meS} \times C_{meS,t} \tag{4.6}$$

To compute the intake resulting from moving from an origin to a destination, the selected path pollutant concentration profile is needed. Assume that the path $\varkappa$ can be deconstructed into a series of links a, each link with average pollutant concentration $C_{a,t}$ at time period t and average travel time d_a/v_a, with d_a representing the length of the link and v_a representing the average velocity. Then, the pollutant intake estimate for path $\varkappa$ can be formulated as:

$$I_{n,\varkappa} = B_n \sum_a \frac{d_a}{v_a} \times pF_{mode,\varkappa} \times C_{a,t} \tag{4.7}$$

$pF_{mode,\varkappa}$ is the inflation factor for the mode used in a particular movement. Finally, $I_{n,M}$ can be computed using:

$$I_{n,M} = \sum_{\varkappa} I_{n,\varkappa} \tag{4.8}$$

4.3.2 Constraints

As previously mentioned, it is assumed that the planner is concerned with the levels of pollutant concentrations and GHG emissions. The pollutant concentrations $C_{r,t}$ of interest are measured at selected locations r ($R = \{1, \dots, r\}$) and time periods t. The planner is interested in maintaining the average pollutant concentrations level $\bar{C}_r$ at every receptor point below a threshold C_{max}. Let C_L be equal to $C_L = \max_{l \in r} \bar{C}_l$. The pollutant concentration constraint is stated as:

$$\bar{C}_L < C_{max} \tag{4.9}$$

In the case of the emission constraints, the interest is in the level of GHG emissions are the network level (not at link level). The total network emissions Q_{GHG} in the network have to be less than Q_{max}. For completeness:

$$Q_{GHG} < Q_{max} \tag{4.10}$$

Lastly, the decision variables are constrained in the interval $[x_{z,min}, x_{z,max}]$.

4.4 Solution Algorithm: A Surrogate Assisted Multiobjective Differential Evolution Algorithm

As its name suggests, the differential evolution algorithm is an evolutionary algorithm, meaning that it follows the general procedure of generating a parent population of solutions and iteratively applying mutation, recombination, and selection procedures to create an improved population until a convergence criterion is met (Storn and Price, 1997). During the past two decades, evolutionary algorithms, including

DE, have been combined with surrogate models to improve their efficiency. Interest in surrogate-assisted evolutionary algorithms (SEAs) has grown as the simulation models embedded in optimization problems have increased in complexity (Jin 2011). In the context of the problem presented in this chapter, an analysts could be interested in combining models for traffic assignment, emission generation, atmospheric chemistry, air pollutant emission dispersion, and population pollution exposure. Given the high computational cost that such combined model would entail, the use of SEAs is a promising solution approach for practical design problems.

A common strategy in SEAs is the use of surrogate models to prescreen multiple solutions before being evaluated by the costly models. For example, Zhang and Sanderson (2009) proposed a radial basis function (RBF) assisted DE. The surrogate model was used to evaluate multiple mutations for a single parent, and then select the most promising candidate. The algorithm proposed in this section incorporates this surrogate strategy to an adaptive multiobjective differential evolution algorithm proposed by Zhang and Sanderson (2009) (here referred to as JADE-MO). Additional algorithm modifications are introduced to handle computationally expensive constraints. The next sections discuss the basics of JADE-MO and the proposed modifications for the traffic rationing problem.

4.4.1 Preliminaries of JADE-MO

Like other DE algorithms, in JADE-MO mutations are performed based on the difference of vectors of candidate solutions. What distinguishes JADE-MO from other algorithms is the adaptive nature of its mutation and crossover parameters, and the rules used to select the parents for the mutation procedure. In this section the mutation, crossover, and selection procedures of JADE-MO are explained.

Assume that there are NP parents in the population, and for each parent a child solution is produced. The child $\boldsymbol{v}_{i,g}$ (i for parent relation and g for iteration) is generated by combining the parent $\boldsymbol{x}_{i,g}$ with three other randomly selected candidate solutions: a member $\boldsymbol{x}_{best,g}$ from a collection the best solutions in the current population, another member $\boldsymbol{x}_{r1,g}$ from the parent population, and a solution $\boldsymbol{x}_{r2,g}$ from an archive

that contains all the evaluated candidate solutions until generation g. The following mutation function is used to generate $v_{i,g}$:

$$\boldsymbol{v}_{i,g} = \boldsymbol{x}_{i,g} + F_i\left(\boldsymbol{x}_{best,g} - \boldsymbol{x}_{i,g}\right) + F_i\left(\boldsymbol{x}_{r1,g} - \boldsymbol{x}_{r2,g}\right) \tag{4.11}$$

$\boldsymbol{x}_{best,g}$ refers to the n_{best} solutions with the best crowding density values. Crowding density is a measure of the distance between the objective function values of a candidate solution to the objective function values of other candidate solutions (see Zhang and Sanderson 2009). A solutions is said to be more crowded than others if its distance to other solutions is small, relatively speaking. F_i is an adaptive mutation factor, and it is randomly generated from a Cauchy distribution with location parameter μ_F and scale parameter 0.1. The location parameter μ_F is adaptive; it is updated after each iteration using the current μ_F parameter, the Lehmer mean S_F of all successful F_i of the current iteration (successful in the sense that child i was better than parent i), and a recombination parameter c. The parameter μ_F is updated using the next equation:

$$\mu_{F,new} = (1 - c)\mu_F + cS_F \tag{4.12}$$

The child $\boldsymbol{v}_{i,g}$ is then modified in the crossover step. Let $v_{i,j,g}$ and $x_{i,j,g}$ represent the element j of vectors $\boldsymbol{v}_{i,g}$ and $\boldsymbol{x}_{i,g}$, respectively. In the crossover step the final version of the child, $\boldsymbol{u}_{i,g}$, is produced as follows:

$$u_{i,j,g} = \begin{cases} v_{i,j,g} & if\ \omega < CR_i \\ x_{i,j,g} & otherwise \end{cases} \tag{4.13}$$

CR_i is the crossover parameter and, like F_i, it is randomly generated using a distribution, in this case, a normal distribution with mean μ_{CR} and variance 0.1. After each iteration the parameter is updated using the current μ_{CR}, the arithmetic mean S_{CR} of all successful CR_i of the current iteration, and a recombination parameter c. The parameter μ_{CR} is computed using the next equation:

$$\mu_{CR,new} = (1 - c)\mu_{CR} + cS_{CR} \tag{4.14}$$

After the NP children are generated the next generation of parents has to be selected. The process begins by pooling the parents and offspring, thus creating a set of solutions of size $2NP$. The selection process ends once NP solutions are selected from the pool. The selection process consists of three selection rules or steps. In the first step the Pareto rank of the vectors in the pool of solution is determined. If a child has a better Pareto rank than its parent, the parent is removed from the pool, and vice versa. Note that the child and parent can have the same Pareto rank, so after this step there might still exist more than NP solutions in the pool. In the second stage, the remaining solutions are pruned further based on their Pareto rank. Solutions are saved from pruning in increasing order of their Pareto rank until at least NP solutions are selected. For example, assume that there is a pool with 18 solutions, and $NP = 10$ (i.e., the pool has to be reduced to 10 solutions). There are five solutions with Pareto rank 1, four solutions with Pareto rank 2, six solutions with Pareto rank 3, and three solution with Pareto rank 4. So first the solutions with rank 1 are accepted, followed by the solutions with rank 2, followed by the solutions with rank 3, at which points there are more than NP candidate selected, so the selection procedure is stopped. If more than NP solutions remain (like in the previous example), in the third step vectors with the worst rank are eliminated based on their crowding density; the least crowded solutions are preserved. In the example, after the second step there are 15 solutions, of which six belong to the worst ranked group. So the five more crowded solutions are rejected and the least crowded solution is preserved to form part, along with the rank 1 and rank 2 solutions, of the parent population for the next iteration. Next the modifications to JADE-MO are discussed.

4.4.2 SAMDE: A surrogate-based variant of JADE-MO

SAMDE can be considered an extension of JADE-MO given that it follows the general strategy of the algorithm proposed by Zhang and Sanderson (2009). It defers from JADE-MO in that it is applicable to constrained optimization problems and it incorporates surrogate-based strategies to speed the search of good solutions to computationally expensive problems. In the next paragraphs the offspring generation (OG), offspring selection (OS), and parent selection (PS) strategies are explained.

4.4.2.1 OG strategy

Usually in evolutionary algorithms like DE, each parent is subjected to mutation and recombination operations to generate a single child. In surrogate based OG strategies, multiple children are generated for a single parent by applying to it the mutation and recombination operators multiple times. Here, the previously described JADE-MO mutation and recombination strategies are applied $K \times W$ times to create $K \times W$ candidate offspring, of which one is selected as most promising according to the surrogate models predictions. This is accomplished by selecting K different $\boldsymbol{x}_{best,g}$, $\boldsymbol{x}_{r1,g}$, and $\boldsymbol{x}_{r2,g}$ for each parent $\boldsymbol{x}_{i,g}$, and, for each of those K sets of vectors, W mutation and recombination parameters are generated. For each, of the $K \times W$ sets of vectors and parameters the previous mutation and crossover equations are applied to generate the offspring.

4.4.2.2 OS strategy

The next step is to select the most promising offspring from the $K \times W$ children using the surrogate models. This follows a three-step process: select the feasible solutions, then select solutions with the best Pareto rank, then select the solution with the lowest crowding density. Surrogate models are estimated for each constraint and objective function. In the traffic rationing problem, only the inequality objective function makes use of the computationally expensive model outputs, so no surrogate model is need for the car usage objective function. Based on the constraints' surrogate models, the children that are predicted to be feasible are selected, and the rest discarded. If none are predicted to be feasible, then solutions with the minimum number of constraint violations are selected. If all solutions have the same number of constraint violations, then the solutions with the minimum magnitude of constraint violations are selected. Naturally, this last condition is problem specific. In the case of the traffic rationing problem, let sE represent the surrogate model for estimating the level of GHG emissions, and sC be the surrogate model to estimate the maximum pollutant concentration. Additionally, denote the maximum and minimum predictions for each surrogate model as sE_{min}, sE_{max}, sC_{min}, and sC_{max}, respectively. Then, the constraint violation score vs_i of a candidate $u_{i,g}$ is calculated as follows:

$$vs_i = w_E \frac{sE(u_{i,g}) - sE_{min}}{sE_{max} - sE_{min}} + w_C \frac{sC(u_{i,g}) - sC_{min}}{sC_{max} - sC_{min}} \quad (4.15)$$

where w_E and w_C ($w_E + w_C = 1$) are weights that reflect the relative difficulty in complying with each constraint. Having used the constraints' surrogate models to select the feasible (or least infeasible) candidates, the objective function surrogate model is then used to predict each selected candidate's intake inequality objective function value. Based on their predicted intake inequality objective function values and the computed traffic restriction objective function value, each candidate is assigned a Pareto rank. Solutions with rank 1 are selected. Of the solutions with Pareto rank 1, then the solution with the lowest crowding density is finally picked as the most promising offspring of parent $\boldsymbol{x}_{i,g}$

4.4.2.3 PS strategy

The parent selection strategy for SAMDE is identical to the JADE-MO strategy, but for one difference: after each solution is given a Pareto rank, the rank of infeasible solutions is adjusted to be equal to the highest rank found in the pool of candidates, plus one, thus ensuring that infeasible solutions are always removed when compared to feasible solutions.

4.4.2.4 Surrogate models and ensembles

The traffic rationing problem has one computationally expensive objective functions and two computationally expensive constraints. Therefore, three different surrogate models are estimated and used to screen for the most promising candidate solutions. Instead of conducting expensive tests to select the best surrogate model type (i.e., radial basis functions, artificial neural networks, kriging) for each function, one could use an ensemble of surrogate models.

A surrogate model ensemble refers to the weighted linear combination of multiple surrogates. The expectation is that by combining surrogate models errors will cancel out thus improving overall prediction accuracy (Viana et al. 2010). As the optimization progresses the weights for each model change. Zhou et al. (2011) grouped ensemble weighting methods into three categories: weighting based on prediction error

variance, on cross-validation errors minimization, or on mean square error minimization. In addition, Zhou et al. proposed a recursive simple averaging method to compute ensemble weights. A common feature of these methods is that the weights are constrained to sum to one. A method that does not appear to have been explored in the surrogate optimization literature is the Granger-Ramanathan averaging method (Granger and Ramanathan 1984). This method estimates unconstrained ordinary least squares weights (i.e., weights do not have to sum to one) based on the surrogate models predictions (explanatory variables) and the true function values (dependent variable). In the context of forecasting model accuracy, Diks and Vrugt (2010) state that the Granger-Ramanathan averaging method exploits the possible presence of covariance structure in the prediction errors, and they show that this simple method performs better than more sophisticated ensemble building techniques. Here the Granger-Ramanathan averaging method is used because of its simplicity and its good performance in other applications. It is anticipated that the averaging could have a smoothing effect on the surrogate model predictions. The smoothing effect is of interest given that the simulation models used in the sample application (Section 4.5) have non-deterministic outputs.

Let $\boldsymbol{\Phi}$ represent a vector of model outputs (either for an objective function or a constraint), and $\widehat{\boldsymbol{\Phi}}$ be a matrix of predicted values of the elements in $\boldsymbol{\Phi}$ (i.e., each row $\widehat{\boldsymbol{\Phi}}_{\boldsymbol{j}}$ contains different predictions that were made previous to the determination of $\boldsymbol{\Phi}_{\mathbf{j}}$). Then, the Granger-Ramanathan weights $\boldsymbol{w}_{GR,g}$ are estimated by:

$$\boldsymbol{w}_{GR,g} = \left(\widehat{\boldsymbol{\Phi}}_{\boldsymbol{g}}^{\boldsymbol{T}}\widehat{\boldsymbol{\Phi}}_{\boldsymbol{g}}\right)^{-1}\widehat{\boldsymbol{\Phi}}_{\boldsymbol{g}}^{\boldsymbol{T}}\boldsymbol{\Phi}_{\boldsymbol{g}} \tag{4.16}$$

In each iteration the combined predictions $\widehat{\boldsymbol{\Phi}}_{\boldsymbol{com}}$ are equal to $\widehat{\boldsymbol{\Phi}}_{\boldsymbol{g+1}}\boldsymbol{w}_{GR}$. The weights can be initialized in several ways. For example, given the initial population and its evaluated outputs, the best surrogate model type can be determined for each function (i.e., using a cross-validation approach) and assigned a weight of one, while the other models are assigned a zero weight. A simpler method is to use equal weight averaging, where all the models are assigned the same weight; this is the approach used in the example presented in section 4.5. After each iteration of the algorithm, the surrogate model predictions for each candidate

solution are stored in a matrix $\hat{\boldsymbol{\Phi}}$, and after each candidate is evaluated with the actual model and the $\boldsymbol{\Phi}$ vectors are obtained, equation 4.16 is used to update the $\boldsymbol{w}_{GR}$ weights for the next iteration.

4.4.2.5 SAMDE steps

The steps and related notation of the algorithm are detailed next.

Decision variable

$\boldsymbol{x}$	:	vector of the percent level reduction of the total demand for car usage in rationing zones

Functions

OF_D	:	total car usage objective function
OF_{EQ}	:	exposure inequality objective function
Z_E	:	estimate of total GHG emissions
Z_C	:	estimate of maximum pollutant concentration

Counters

g	:	generation counter

Parameters

M	:	number of different surrogate models

Sets

X	:	set of evaluated solution vectors
Λ_{EQ}	:	set of computed OF_{EQ} values
O_E	:	set of computed Z_E values
O_C	:	set of computed Z_C values
Γ_{EQ}	:	set of surrogate model OF_{EQ} predictions for parents' offspring
H_E	:	set of surrogate model Z_E predictions for parents' offspring
H_C	:	set of surrogate model Z_C predictions for parents' offspring

Steps

1. Initialization.

1.1 Set generation counter $g = 1$.

1.2 Generate initial set of points (parent population) $X_g = \{x_{1,g}, \ldots, x_{i,g}, \ldots, x_{NP,g}\}$ within predefined bounds.

1.3 Set initial model $\boldsymbol{w}_{GR}$ weights equal to $1/M$ for all objective functions and constraint RBF models.

1.4 Add solutions in X_g to X.

2. Initial parent evaluation.

2.1 For each point in X_g, evaluate objective function OF_D and OF_{EQ} and constraints Z_E and Z_C.

2.2 Add data to archives Λ_{EQ}, O_E, and O_C.

3. Generate surrogate models.

3.1 Given X_g and the data in archives Λ_{EQ}, O_E, and O_C, fit surrogate models for the OF_{EQ} objective function and constraints Z_E and Z_C.

3.2 Using $\boldsymbol{w}_{GR}$ weights, create surrogate ensemble.

4. Create offspring using the OG strategy.

5. Select most promising offspring Y_g.

5.1 Using surrogate models, screen for the most promising offspring using the OS strategy.

5.2 For each selected offspring, save the predicted objective function and constraint values in archives Γ_{EQ}, H_E, and H_C.

6. Evaluate offspring Y_g.

6.1 Evaluate the most promising offspring with the computationally expensive models.

6.2 Update the X, Λ_D, Λ_{EQ}, O_E, and O_C sets with the new information.

7. Apply PS strategy to choose the parents of the next generation.

8. If $n = n_{max}$, stop and return the current parent population solutions. Else, reset $n = n + 1$ and continue to Step 9.
9. Update ensemble weights and algorithm parameters.

9.1 Using costly model outputs and the data in achieves Γ_{EQ}, H_E, and H_C, update the objective function and constraints surrogate model weights utilizing the Granger-Ramanathan averaging method.

9.2 Clear data from sets Γ_{EQ}, H_E, and H_C.

9.3 Update mutation and crossover parameters.

9.4 Return to step 3.

4.5 Numerical Example

For illustrative purposes, a numerical example is presented in this section. In this example the planner seeks a rationing schedule that maximizes auto trips, maximizes pollutant exposure equality, and reduces GHG emissions and the maximum pollutant concentration by 20 percent relative to conditions prior to the introduction of the scheme. The proposed problem is applied in a hypothetical version of Sioux Falls. Figure 4-1 presents the network. Zones are aggregated to 10 rationing districts. The planner's decision is the degree to which auto drivers in each zone will be forced to take the bus, the only alternative mode besides walking.

4.5.1 Network and population data

The Sioux Falls network is composed is composed of 76 links, 24 nodes, and 24 OD zones. Figure 4-1 presents the network. The link attributes were obtained from an online database (Bar-Gera 2014). A population of 600,000 is assumed to live in the 24 zones. Each individual agent is characterized by an age group class, an income class, a mode class, and an activity profile class. There are five activity profiles: full-time worker, part-time worker during the day, part-time worker during the night, student, and unemployed. Tied to each activity profile is also a series of mandatory activities that the agent always performs. Each agent is also characterized by a start and end time for each of the mandatory activities. Regarding the mode class classification, it is assumed that every agent only makes binary choices involving

their assigned mode class and walking. For example, if a person is an auto driver, at a particular destination and mode decision point (to be explained in the next section), the agent only considers walking as an alternative to using the car. Similarly, an agent classified as a bus user only considers using the bus or walking. For all rationing districts, the rationing fraction x_z was bounded in the interval [0.5, 1].

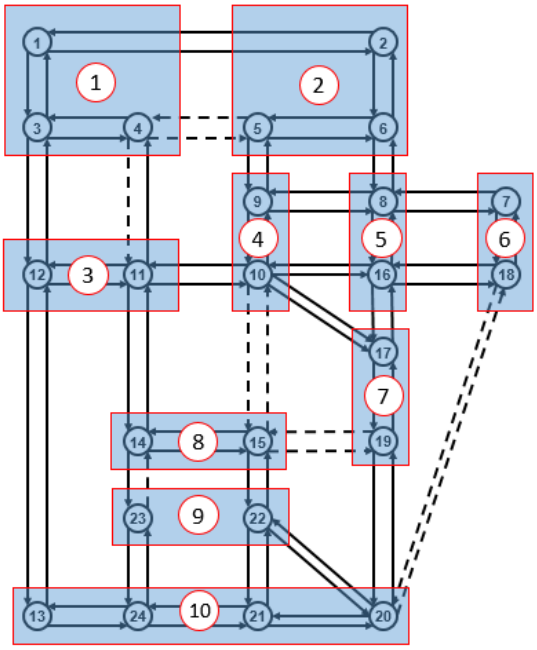


Figure 4-1 Sioux Falls Network and its rationing zones

4.5.2 Simulation model structure

The structure of the simulation model is presented in Figure 4-2. The model is composed of an activity scheduling model, a traffic assignment model, a vehicle emissions model, a pollutant dispersion model, and a pollutant exposure model.

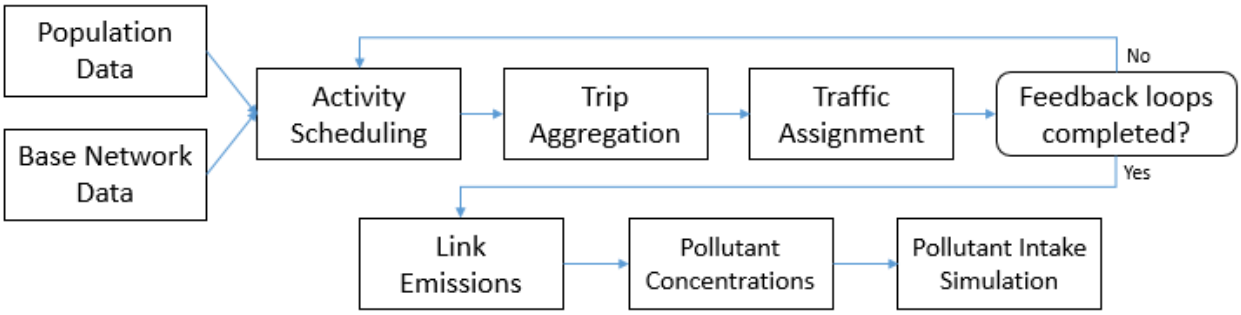


Figure 4-2 Components of integrated model used to compute agents' pollutant intake

A simple behavioral model structure was utilized to simulate individual agent's activity schedules, meaning, a list of each agent's activities, the location of the activities, the time period when they occur, and, in the case of transportation activities, the mode utilized. Activity participation was modeled using a combination of rules and sequential logit models. Each agent has an assigned activity type profile that determines the types of mandatory activities (work and school) that the agent has to engage in, along with the number of non-mandatory activities that the person can participate in and when each non-mandatory activity can occur. These activity anchors constituted the rule-based model component of the activity scheduling model. For example, a full time worker has at most two non-mandatory activities: one in the middle of the work period and one after work. In the middle of the work period the worker decides to either perform an activity outside of the workplace or stay in the work location. After work the person decides to either go home or perform an activity. If the agent decided to return to their home, then it again has to decide whether to perform an activity. If at any stage the agent is simulated to perform an activity, a joint decision is made on where the activity will be performed and which mode will be used. The activity participation and joint destination-mode choice decisions were modeled sequentially using a logit formulation. The binary logit model was used to compute the probability of deciding to participate in an activity or not, and a multinomial logit model was used to compute the probability of selecting a destination-mode alternative. Based on these probabilities, a Monte Carlo simulation procedure was used to determine the agent's choices. The rules and specification of the logit models are presented in Appendix A.

The activity scheduling model results in individual activity lists that contain information on people's origin and destination, modes, and the time period when the trip occurs. Initially, these lists were generated, in part, using free flow road network travel times. The travel times were updated based on the output of the traffic assignment model. After the activity schedules were determined, the lists' trips were aggregated to mode and time specific OD matrices. The auto OD matrices were assigned to the road network by the traffic assignment model, while bus OD demand was assumed to travel on the shortest path and to have no effect on the auto travel time. The gradient projection method (Jayakrishnan et al. 1994) was employed to find the user equilibrium traffic flows on the network for the different model time periods. The output of the traffic model were auto link flows, OD path sets, and their corresponding path flows. With this information the auto OD travel time were updated by combining predecessor link flow information with updated link flow information using the method of successive averages (Boyce et al. 2008), which results in an updated set of link travel times, and, consequently, new auto OD travel times. The new auto OD travel times were looped back to the activity scheduling model, and a new simulation of the agent's activity schedules was performed. This feedback procedure was repeated only three times, given time constraints.

In this example, only autos emit pollutants. Link GHG and $PM_{2.5}$ emissions were estimated using speed-based emission factors obtained from MOVES. Following methodology proposed by Zhang et al. (2010), a Gaussian plume model was utilized to compute average $PM_{2.5}$ pollutant concentrations over a 14,641 receptor grid (with 200 meter spacing) that covers the Sioux Falls region. Only four meteorological scenarios (with distinct wind speeds, directions, and atmospheric stability classes) were utilized.

Given the activity schedules, the path sets and flows, and the time specific $PM_{2.5}$ concentration at all receptor points, each agent's daily $PM_{2.5}$ intake was estimated using the pollutant exposure model. As discussed in section 4.2.3, pollutant intake was modeled here as the sum of the agent's intake while at particular locations and the intake while moving between locations. The pollutant intake model determined the time spent in each microenvironment based on each agent's schedule, along with the concentration of the microenvironment. For activities at a particular location, the microenvironment's concentration $C_{me,S}$

were computed with the formula $pF_{me,S} \times C_{r,mes}$, where $C_{r,mes}$ is the concentration of the closest receptor to where the agent is located. The infiltration factor $pF_{me,S}$ was generated by taking random draws from a normal distribution with mean 0.66 and variance 0.04. In the case of concentrations encountered while moving via auto or bus, the movement's concentration exposure profile was constructed based on the movement's path links, the average $PM_{2.5}$ concentrations on these links (as indicated by the receptors in the proximity of each link), and the time it takes to traverse each link. The infiltration factors for autos and buses were generated in a similar fashion, but with a normal distribution with mean 0.9 and variance 0.01. A random walk algorithm was used to simulate the path that walkers take from their origin to their destination, and the intake was computed based on concentrations encountered in the random path; walking speed was fixed at 3 miles per hour. Daily breathing rates were also randomly generated using a truncated normal distribution with mean of 12 m^3/d and variance of 2 $(m^3/d)^2$. Additional details on the pollutant intake model can be found in Appendix A.

4.5.3 Solution algorithm parameters

The parent population NP was set to 40. A maximum of 10 iterations where performed, so in total 400 candidate solutions were evaluated. The mutation factor μ_F and cross-over factor μ_{CR} were initiated at 0.5 and 0.8, respectively. A value of 0.3 was used for the recombination factor c. Only eight solutions were used for n_{best}. The factors K and W were set to 10 and 100, so 1000 candidate offspring were prescreened for each parent. Radial basis function (RBF) models were used as surrogate models. The ensembles were constructed using three parameter free RBF functional forms: linear, cubic, and thin-plate spline.

4.5.4 Results

In Figure 4-3, the Pareto frontier for the numerical example is presented. The feasible solutions have total traffic restrictions ranging from 23 percent to 48 percent, and Atkinson index ranging from 0.90 to 0.93. Without the rationing scheme the Atkinson index is 0.95. High levels of inequalities are found, in part, because very similar wind directions were used, the network is relatively sparse, and agents are allowed to

perform activities at any point within the receptor grid. So, agents performing activities in the relatively small number of locations that are downwind and close to links inhale significantly more pollutants than their counterparts. This result, however, highlights the possibility that improvements in pollutant exposure inequality could be hard to achieve if there are structural reasons (e.g., meteorology, population distribution, urban form) for the observed inequality that are unaffected by traffic management schemes. Here small reductions in the Atkinson index are achieved even though there are significant reductions in traffic. In Figure 4-4, a box plot of the reduction level variables is presented. For each variable, the whiskers represent the minimum and maximum values, the edges of the box represent the 25 and 75 percentile, and the box's central mark is the variable's median. As can be seen, there is great variation in the values that the rationing variable can assume. The rationing variable for district 6 is the least disperse and the variable with the highest mean value. District 5 has the lowest median reduction. Note that most of the variables are skewed to the left, with over 25 reduction in car usage.

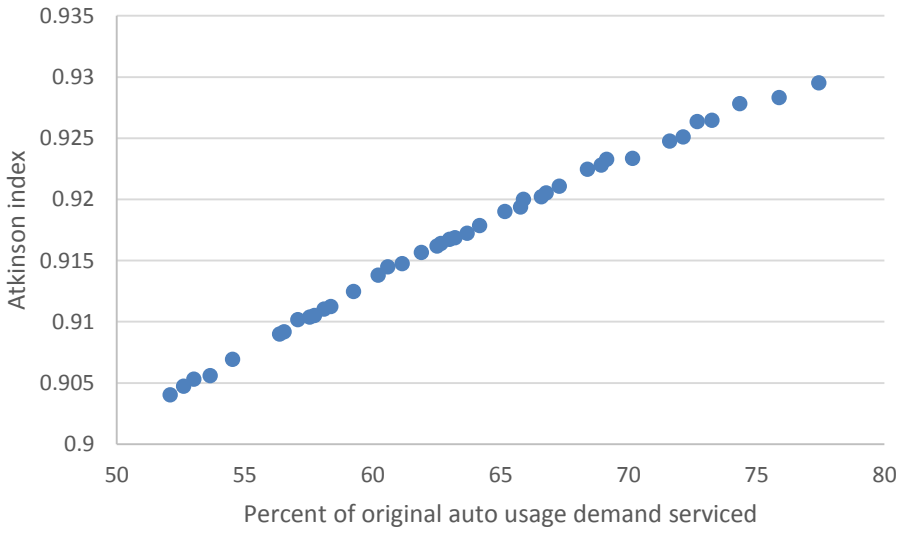


Figure 4-3 Pareto frontier for numerical example

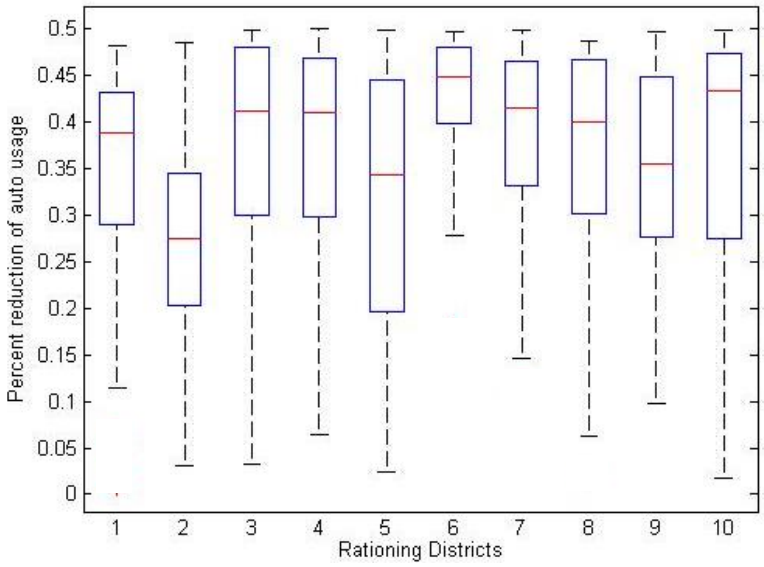


Figure 4-4 Box plot of rationing variables

4.6 Closing Remarks and Future Research

A traffic rationing problem was introduced that attempts to balance the planner's objectives of improving air quality conditions, reducing GHG emissions, and equitably distributing environmental pollution burdens and benefits, with the objective of allowing citizens the opportunity to use their desired mode of transportation. To find good solutions for this problem, a multiobjective differential evolution algorithm was proposed that incorporates surrogate model ensembles to predict which candidate solutions are feasible and more likely to be better than their parent solutions.

Objective function 4.1 reflects the planner's desire to reduce the disutility of the control scheme by maximizing the demand served. Since x_z will not be one for all ODs, the planner is simultaneously faced with a second problem: what to do with the rationed demand $x_z q_z$. A possible response is to improve or introduce a transit system to service the capped demand. Therefore, the proposed problem can be extended by integrating objectives related to transit network design, where decisions on transit network service (e.g., in terms of number of buses and routes) would be made considering demand $x_z q_z$, budget constraints, and possibly equity measures. This second objective can serve as a way to discriminate among the multiple possible solutions the proposed problem. The traffic rationing problem could also be modified by

substituting objective function 4.1 with a function that accounts for the differences in disutility between capped demand given the variations of travel costs between OD pairs (i.e., accounting for inequalities in travel costs).

In the context stochastic optimization problems like the one presented here, SAMDE could potentially be improved by substituting the surrogate ensemble approach used with the radial basis functions proposed by Jakobsson et al. (2010), which are intended to be used with noisy black box functions. Additionally, the adjustments implemented on JADE-MO could also be implemented in the LMSRS-inspired multiobjective optimization algorithm proposed by Chow and Regan (2014), and tests could be conducted to determine if the resulting algorithm performs better than SAMDE.

Chapter 5

Summary and Future Research

A summary of research contributions and related future research is presented next.

5.1. Summary of Contributions

The general objective of the presented research was to provide planners with optimization-based approaches to design pricing and traffic rationing schemes that minimize human exposure to motor vehicle generated pollutants. To this end, network design problems and related surrogate-based solution algorithms were proposed.

In Chapter 2, the TDP was extended to account for population exposure to pollutants and environmental inequalities between zones. The planner's upper level problem in the TDP is to find the set of toll locations and levels that minimize population intake to pollutants and environmental inequality subject to a budget constraints and pollutant concentration chance constraints. In this problem, environmental inequality was defined in terms of zonal differences in pollutant concentrations at receptor points. To solve this problem a mixed integer variant of the LMSRS algorithm was proposed. Numerical results suggest that the proposed algorithm performs better than previous GA methods. In addition, the GA-LMS algorithm was presented; the performance of this algorithm was promising, but not better than the mixed integer variant of the LMSRS algorithm. The two proposed algorithms are derivative-free, so they can be applied to similar network design problems, like network capacity expansion problems where the planner must select which links to improve (integer variable), and by how much the capacity will be expanded (continuous variable).

The research presented in Chapter 3 can be utilized by transportation planners to design cordon and area pricing schemes subject to environmental constraints. Like in previous work, the area pricing problem is structured as a bi-level optimization problem. But, departing from previous elastic demand formulations of the problem, objective functions were proposed that can be easily applied with most state-of-the-practice planning models. A surrogate-based solution algorithm was proposed for the area pricing problem, which

is also applicable to cordon pricing problems. Preliminary numerical tests suggest that RBF models estimated using a geometric representation of a charging boundary result in predictions of objective function values that are relatively good, with correlations between model and surrogate values greater than 0.7.

A bi-objective traffic rationing problem with environmental constraints was considered in Chapter 4. In this problem, restrictions on zonal trips and inequalities in pollutant intake were minimized, subject to constraints on the levels of GHG emissions and $PM_{2.5}$ concentrations. In contrast to the population intake methodology utilized in Chapter 1, in this chapter an individual exposure approach is used, with the assumption that the planner has access to models that can simulate agent-level activity scheduling. A surrogate-assisted differential evolution algorithm was presented that can be used for multiobjective continuous optimization problems with linear and non-linear constraints.

5.2. Future Research

Future research directions are broadly discussed in terms of additional optimization problems that can be considered in light of the problems proposed in previous chapters, and in terms of further research on the application of surrogate models.

5.2.1. Planning models

The problems proposed in this work dealt exclusively with personal transportation travel demand, with no consideration of urban freight movements. Urban freight movements by trucks are a major source of GHG emissions and other pollutants (Piera et al. 2011). For this reason, multiple strategies have been proposed to control truck generated pollution, including the introduction of programs aimed at shifting truck traffic to periods of the day that experience less traffic congestion (e.g., Holguin Versa et al. 2011). Although it is argued that this shift would reduce truck emissions, studies have shown that shifting truck traffic to off-peak hours could increase pollutant concentrations in urban areas due to the relatively stable atmospheric conditions in those periods (Sathaye et al. 2010). There appears to be a trade-off between

minimizing emissions and minimizing pollutant concentrations when deciding how many trucks to shift to off-peak hours. Additionally, depending on the industry, there could be trade-offs between reducing the environmental impact of truck traffic and the economic profitability of firms. In light of these multiple competing objectives, research could be conducted on the development of an optimization-based framework to aid decision makers design truck traffic management programs.

In addition, although transit choice was considered in Chapters 3 and 4, it was treated in the abstract, with no simulation of how increase patronage would affect the operations of the transit system. Given the budget constraints that surround the expansion of any transportation system, and the likely need to expand such systems when pricing and traffic rationing programs are implemented, it would be useful to couple the proposed problems with transit network design problems.

5.2.2. Surrogate-based solution algorithms and other applications

Comparative studies of algorithm performance are necessary, particularly in the case of the proposed algorithm for cordon and area-based pricing. Like the tests conducted in Chapter 2, these studies will involve multiple model formulations, including the elastic demand formulations used in previous discrete network cordon pricing studies. Additional tests are also required to explore alternative surrogate-based methods for handling stochastic simulations, which is particularly important in the case of models that simulate agent-level behavior in non-deterministic manner.

An open question is which surrogate model works best in finding good solutions for the transportation network design problems found in the literature. Accuracy-based criteria is commonly used in the selection of surrogate models. However, Viana et al. (2010) noted that using the most accurate surrogate models does not necessarily lead to the best solutions. In fact, a surrogate model could have low approximation accuracy and, yet, it could be more useful in the search of good design solutions than a more accurate model (Jin 2011). Extensive research remains to be done to determine which surrogate models are best suited for the different transportation optimization problems that can be tackled with these surrogate-based techniques.

Another potentially useful applications of surrogate models is as substitutes or complements of particularly time consuming components of joint transportation-air pollution models. For example, instead of applying a computationally expensive air dispersion model over a large area to compute concentrations at thousands of grid receptors, the model could initially be applied to compute the concentrations over only a fraction of the receptor points. Then a surrogate model, fitted with the modeled concentration at the selected receptors, could be used to interpolate the concentration in the remaining receptors not considered with the computationally expensive models. This could potentially reduce the computational burden that is incurred by embedding an air dispersion model in an optimization problem. Naturally, in this case the accuracy of the surrogate model is critical. Preliminary tests conducted with a simple Gaussian plume model suggests that this is a promising line of research. The tests were performed using the Sioux Falls grid receptor network utilized in Chapter 4, the Sioux Falls' user equilibrium traffic flows provided by Bar-Guera (2007), MOVES' $PM_{2.5}$ emission factors, the Gaussian plume model, and a Gaussian RBF function as the surrogate model. The surrogate model was estimated with 53 percent of all the receptor concentration information, and used to interpolate the concentrations at the remaining locations. In Figure 5-1 the Gaussian model estimate are plotted versus the RBF predictions. The correlation between the Gaussian model values and the RBF interpolations is of 0.99. Future tests will be conducted with more complex air dispersion models, different receptor grip arrangements, and other types of surrogate models.

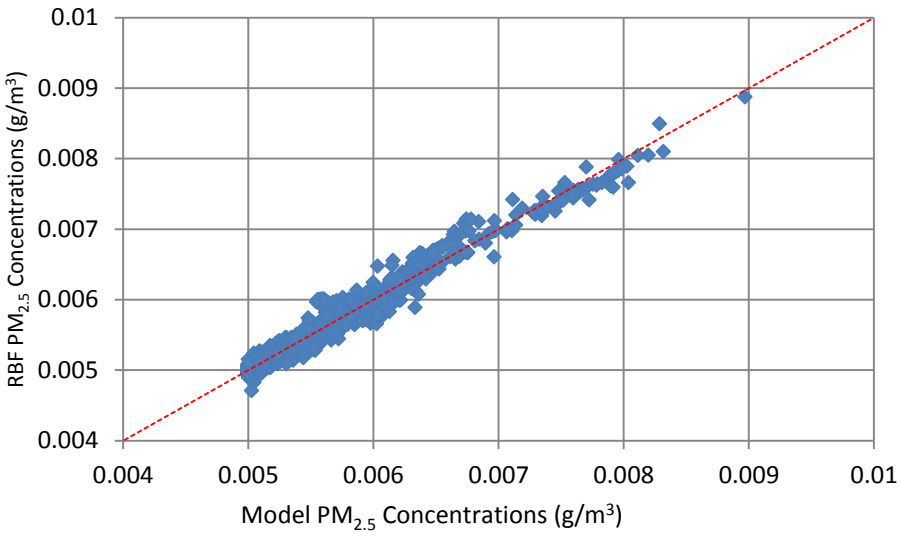


Figure 5-1 Gaussian RBF pollutant concentration interpolations versus Gaussian plume model estimates

Appendix A

Description of Activity Scheduling Model and Pollutant Intake Model

This appendix discusses the structure and assumptions used in the development of the activity scheduling model and the pollutant intake model. Both are simple models used simply for the numerical example in Chapter 4.

A-1 Description of Activity Scheduling Model

The activity scheduling model is used to simulate activity participation and joint destination-mode choices of individual agents. Each agent in the model was characterized by a class type, and based on the class the agent was assigned a series of mandatory (fixed) activities and a series of potential non-mandatory activities. In Table A-1 the class types are listed, along with the number of mandatory activities that each class performs and the maximum number of possible non-mandatory activities. Each agent was also characterized by an age group class, an income class, a travel mode class, a home location, the location of any mandatory activity, and the time each activity can be performed. Based on these attributes, an agent's participation in non-mandatory activities was simulated using a binary logit model (parameters shown in Table A-2). If it was simulated that an activity was performed, then a multinomial logit model was used to select the mode and destination for the activity (parameters shown in Table A-3). So, a sequential logit model structure was used to simulate agent choices.

The full-time worker has two potential non-mandatory activities at two different time periods: in the middle of the work period and after work. In the mid-work period, the agent has the choice of staying at the work location or performing an activity outside the work location. After work, the agent has a choice of returning home or performing an activity. If the agent goes home, a choice is then made on whether to stay home or perform an activity. The part-time daytime workers can perform an activity after work. Similar to the full-time worker, the after-work choice is to perform an activity or go home, and if home is picked,

another activity choice simulation is perform for the agent while at home. Part-time nighttime workers and students do not have non-mandatory activity choices. The choice of the unemployed person is either to stay home or perform an activity.

For all classes, once it was determined that the agent would perform an activity, the next choice was to select the destination and mode. Only two modes were consider: the travel model class assigned to the person (auto, bus, walk) and walk. The joint destination-mode choice was a function of an attraction factor for each destination and the modes' travel times and out-of-pocket costs.

Table A-1 Agent classes

Class	Description	# mandatory activities	# non-mandatory activities
1	Full-time worker	1	2
2	Part-time daytime worker	1	1
3	Part-time nighttime worker	1	0
4	Student	1	0
5	Unemployed	0	1

Table A-2 Parameters for activity participation binary logit model

Variable	Parameter value
Alternative specific (do activity) constant	1.00
Age group 1	-0.10
Age group 2	-0.08
Age group 3	-0.01
Age group 4	0.01
Age group 5	1.00
Income group 1	0.01
Income group 2	0.00
Income group 3	-0.10
Income group 4	-0.15
Income group 5	-0.20

Table A-3 Parameters for activity participation binary logit model

Variable	Parameter value
Auto constant	2.00
Bus constant	-0.50
Travel time (minutes)	-0.02
Travel cost (cents)	-0.008
Destination attraction factor	0.05

A-2 Description of Pollutant Intake Model

The pollutant intake model was used to simulates the time (t_{me}) spent in each of the agent's activities, the concentration of the microenvironment (me) during different periods, and the pollutant intake (I_{me}) resulting from these factors. As part of this simulation each agent was given a breathing rate B. The basic intake I_{me} formula is:

$$I_{me} = B \times t_{me} \times C_{me} \quad \text{(A1)}$$

The time spent on each microenvironment was specified as a function of the activity being performed. The time spent in the home location before a mandatory activity was the only activity that is a function of the start time of another activity. The time when an agent departs from home for a mandatory activity was a function of the activity's start time, the average travel time to the mandatory activity location, and a randomly generated factor that determines how early or late the agent arrives to the mandatory activity location. This factor was generated using a normally distributed random number generator with mean of 5 minutes and variance of 0.5 squared minutes. The duration of other activities were generated by taking random draws from a normal distribution with mean μ and variance σ. Table A-4 presents the normal distribution parameters used for different activities. The work activity parameters refer to a single block of work activity. For full-time workers, two draws were made from the work duration normal distribution for the two blocks of mandatory activity.

Pollutant intake from movements by auto and bus was a function of the links in the movement path, the time spent on each link, and the infiltration factor for the transporting vehicle. For each origin-destination (OD) pair, bus users were assumed to take the shortest path between each OD. In the case of auto users,

multiple paths can exists for each OD. For an OD with multiple paths, an agent was assigned to a single path with a probability equal to the path's traffic flow divided by the total flows between the OD pair (the sum of all path flows for that OD pair). This path information was obtained from the output of gradient projection algorithm.

Table A-4 Parameters for the normal distributions used in the simulation of activity duration

Activity	$\boldsymbol{\mu}$	$\boldsymbol{\sigma}$
Waiting at bus stop	5	0.5
Work	240	30
Non-mandatory activity	60	8
School	240	30

References

Abt Associates. 2012. BenMAP's user manual. U.S. Environmental Protection Agency.

Akamatsu T. 2007. Tradable network permits: A new scheme for the efficient use of network capacity. Working Paper, Tohoku University, Japan. <http://www.plan.civil.tohoku.ac.jp/~akamatsu/Publications/PDF/TS-TNP-P1%28070311%29.pdf > (accessed May 12, 2013)

Bai L. 2004. Computational methods for toll pricing models. PhD Diss., University of Florida, US.

Bar-Gera H. Transportation network test problems. http://www.bgu.ac.il/~bargera/tntp/. Accessed January 9, 2014.

Bazaraa MS, Shetty CM. 1979. Nonlinear programming: Theory and algorithms. New York: Wiley.

Beevers SD, Carslaw DC. 2005. The impact of congestion charging on vehicle emissions in London. Atmospheric Environment 39(1):1-5.

Boyce D, O'Neill CR, Scherr W. 2008. Solving the sequential travel forecasting procedure with feedback. Transportation Research Record 2077:129-135.

Bulteau J. 2012. Tradable emission permit system for urban motorists: The Neo-classical standard model revisited. Research in Transportation Economics 36:101-109.

Buzzelli M. 2008. Environmental justice in Canada. Canadian Policy Research Networks. http://www.cprn.org/documents/50875_EN.pdf. Accessed April 7, 2015.

Caiazzo F, Ashok A, Waitz IA, Yim SHL, Barrett SRH. 2013. Air pollution and early deaths in the United States. Part I: Quantifying the impact of major sectors in 2005. Atmospheric Environment 79:198-208.

Cambridge Systematics, Vanasse Hangen Brustlin, Gallop Corporation, Bhat C, Shapiro Transportation Consulting, Martin/Alexiou/Bryson. 2002. Travel demand forecasting: Parameters and techniques. NCHRP Report 716.

Cheng YF, Heintzenberg J, Wehner B, Wu ZJ, Su H, Hu M, Mao JT. 2008. Traffic restrictions in Beijing during the Sino-African Summit 2006: Aerosol size distribution and visibility compared to long-term in situ observations. Atmospheric Chemistry and Physics 8:7583-7594.

Chen L, H Yang. 2012. Managing congestion and emissions in road networks with tolls and rebates. Transportation Research Part B 46:933-948.

Chen X, Zhang L, He X, Xiong C. 2014. Surrogate-based optimization of expensive-to-evaluate objective for optimal highway toll charges in transportation network. Computer-Aided Civil and Infrastructure Engineering 29:359-381.

Chen X, Yin M, Song M., Zhang L, Li M. 2014. Social welfare maximization of multimodal urban transportation systems using metamodels: Theory, model, and application to the Tianjin Eco-City in China. In Proceedings of the 93rd Transportation Research Board Annual Meeting.

Chen X, Zhu Z, He X, Zhang L. 2015. Surrogate-based optimization for solving mixed integer network design problem. In Proceedings of the 94th Transportation Research Board Annual Meeting.

Chow JYJ, Regan AC. 2014. A surrogate-based multiobjective metaheuristic and network degradation simulation model for robust toll pricing. Optimization and Engineering 15(1):137-165.

Chow JYJ, Regan AC, Arkhipov DI. 2010. Faster converging global heuristic for continuous network design using radial basis functions. Transportation Research Record 2196:102-110.

Daganzo CF. 1995. A Pareto optimum congestion reduction scheme. Transportation Research Part B 29 (2):303-311.

Deb K. 2000. An efficient constraint handling method for genetic algorithms. Computer Methods in Applied Mechanics and Engineering 186:311-338.

de Jong G, Daly A, Pieters M, van der Hoorn T. 2007. The logsum as an evaluation measure: Review of the literature and new results. Transportation Research Part A 41:874-889

Dekkers A, Aarts E. 1991. Global optimization and simulated annealing. Mathematical Programming 50:367-393.

Dhanda KK, Nagurney A, Ramanujam P. 1999. Environmental networks: A framework for economic decision-making and policy analysis. Cheltenham: Edwards Elgar.

Dhondt S, Beckx C, Degraeuwe B, Lefebvre W, Kochan B, Bellemans T, Panis L, Macharis C, Putman K. 2012. Health impact assessment of air pollution using a dynamic exposure profile: Implications for exposure and health impact estimates. Environmental Impact Assessment Review 36:42-51.

Diks CHG, Vrugt JA. 2010. Comparison of point forecast accuracy of model averaging methods in hydrologic applications. Stochastic Environmental Research and Risk Assessment 24:809-820.

Edara P, Teodorovic D. 2008. Model of an advance-booking system for highway trips. Transportation Research Part C 16:36-53.

Ekström J, Engelson L, Rydergren C. 2009. Heuristic algorithms for a second-best congestion pricing problem. Netnomics 10:85-109.

Ekström J, Sumalee A, Lo HK. 2012. Optimizing toll locations and levels using a mixed integer linear approximation approach. Transportation Research Part B 46:834-854.

Eliasson J, Hultkrantz L, Nerhagen L, Smidfelt Rosqvist L. 2009. The Stockholm congestion-charging trial 2006: Overview of effects. Transportation Research Part A 43(3):240-250

EPA. 2014. Motor Vehicle Emission Simulator (MOVES): User guide for MOVES2014. US Environmental Protection Agency, EPA-420-B-14-055.

Evans JS, Wolff SK, Phonboon K, Levy JI, Smith KR. 2002. Exposure efficiency: An idea whose time has come? Chemosphere 49:1075-1091.

Fan W, Gurmu Z. 2014. Combined decision making of congestion pricing and capacity expansion: Genetic algorithm approach. Journal of Transportation Engineering 140 (8):1-9.

Feitelson E. 2002. Introducing environmental equity dimension into the sustainable transport discourse: Issues and pitfalls. Transportation Research Part D 7:99-118.

Feng T, Zhang J, Fujiwara A, Timmermans HJP. 2010. An integrated model system and policy evaluation tool for maximizing mobility under environmental capacity constraints: A case study in Dalian City, China. Transportation Research Part D 15:263-274

Feng T, Timmermans HJP. 2014. Trade-offs between mobility and equity maximization under environmental capacity constraints: A case study of an integrated multi-objective model. Transportation Research Part C 43:267-279.

Ferrari P. 1995. Road pricing and network equilibrium. Transportation Research Part B, 29 (5): 357-372.

Fiske K. 2011. Accelerating the search for optimal dynamic traffic management. Master's Thesis, University of Twente, The Netherlands.

Forrester AIJ, Keane AJ. 2003. Recent advances in surrogate-based optimization. Progress in Aerospace Sciences 45:50-79

Friesz TL, Han K, Liu H, Yao T. 2013. Dynamic congestion and tolls with mobile source emission. Procedia - Social and Behavioral Sciences 80:818-836.

Global Road Safety Facility, The World Bank, Institute for Health Metrics and Evaluation. 2014. Transport for health: The global burden of disease from motorized road transport. Institute for Health Metrics and Evaluation and World Bank.

Goddard HC. 1997. Using tradable permits to achieve sustainability in the world's large cities: Policy design issues and efficiency conditions for controlling vehicle emissions, congestion and urban decentralization with an application to Mexico City. Environmental and Resource Economics 10:63–99.

Golberg DE, Deb K. 1991. A comparative analysis of selection schemes used in genetic algorithms. In *Foundations of Genetic Algorithms*, Ed. Gregory JRR. Morgan Kauffman Publishes, US.

Gouge B, Dowlatabadi H, Ries FJ. 2013. Minimizing the health and climate impacts of emissions from heavy-duty public transportation bus fleets through operational optimization. Environmental Science and Technology 47:3734-3742.

Granger CWL, Ramanathan, R. 1984. Improved methods of combining forecasts. Journal of Forecasting 3:197-207.

Gutmann HM. 2001. A radial basis function method for global optimization. Journal of Global Optimization 19:201-227.

Han D, Yang H, Wang X. 2010. Efficiency of the plate-number-based traffic rationing in general networks. Transportation Research Part E 26:1095-1110.

Hatzopoulou M, Hao JY, Miller EJ. 2011. Simulating the impacts of household travel on greenhouse gas emissions, urban air quality, and population exposure. Transportation 38:871–887.

Hizir AE. 2006. Using emission functions in mathematical programming models for sustainable urban transportation: An application in bi-level optimization. Master's Thesis, Sabanci University, Turkey.

Howitt AM, Altshuler A. 1999. The politics of controlling auto air pollution. Eds. Gomez-Ibañez J, Tye WB, Winston C. Essays in Transportation Economics and Policy.

Hult EE. 2006. Closed charging cordon design problem. University of Cambridge, UK.

International Transportation Forum. 2010. Reducing transport greenhouse gas emissions: Trends and data 2010. OECD and International Transportation Forum.

Jaber X, O'Mahony M. 2009. Mixed stochastic user equilibrium behavior under traveler information provision services with heterogeneous multiclass, multicriteria decision making. Journal of Intelligent Transportation Systems: Technology, Planning, and Operations 13 (4):188-198.

Jayakrishnan R, Tsai WK, Prashker JNM, Rajadhyaksha S. 1994. A faster path-based algorithm for traffic assignment. Transportation Research Record 1443:75-83

Jakobsson S, Patriksson M, Rudholm J, Wojciechowski A. 2010. A method for simulation based optimization using radial basis functions. Optimization and Engineering 11: 501–532.

Jin Y. 2011. Surrogate-assisted evolutionary computation: Recent advances and future challenges. Swarm and Evolutionary Computation 1:61-70.

Kolak OI, Feyzioglu O, Birbil I, Noyan N, Yalcindag S. 2013. Using emission functions in modeling environmentally sustainable traffic assignment policies. Journal of Industrial and Management Optimization 9(2): 341-363.

Lai ACK, Thatcher TL, Nazaroff WW. 2000. Inhalation transfer factors for air pollution health risk assessment. Journal of the Air and Waste Management Association 50(9):1688-1699.

Lamotte RAF. 2014. A fast multi-objective optimization approach to solve the continuous network design problem with microscopic simulation. Master's Thesis, Concordia University, Canada.

Laurent E. 2011. Issues in environmental justice within the European Union. Ecological Economics 70: 1846-1853.

Lawphongpanich S, Yin Y. 2010. Solving the Pareto-improving toll problem via manifold suboptimization. Transportation Research Part C 18:234-246.

Levinson D. 2010. Equity effects of road pricing: A review. Transport Reviews 30(1):33-57

Levy JI, Chemerynski SM, Tuchmann JL. 2006. Incorporating concepts of inequality and inequity into health benefits analysis. International Journal for Equity in Health 5(1), 2.

Levy JI, Greco SL, Melly SJ, Mukhi N. 2009. Evaluating efficiency-equality tradeoffs for mobile source control strategies in an urban area. Risk Analysis 29(1):34-47.

Levy JI, Buonocore JJ, von Stackelberg K. 2010. Evaluation of the public health impacts of traffic congestion: A health risk assessment. Environmental Health 9(65):1-12.

Li XQ, Szeto WY, O'Mahony M. 2007. Incorporating land use, transport and environmental considerations into time-dependent tolling strategies. Journal of Eastern Asia Society of Transportation Studies 7:360-375.

Li ZC, Lam WHK, Wong SC, Sumalee A. 2012. Environmentally sustainable toll design for congested road networks with uncertain demand. International Journal of Sustainable Transportation 6:127-155.

Li ZC, Wang YD, Lam WHK, Sumalee A, Choi K. 2013. Design of sustainable cordon toll pricing schemes in a monocentric city. Networks and Spatial Economics DOI: 10.1007/s11067-013-9209-3

Ma Y, Zuylen HJ, Chen Y, Dalen J. 2010. Allocating departure time slots to optimize dynamic network capacity. Transportation Research Record 2197:98-106.

Maruyama T, Takaki R, Mizokami S. 2014. Incorporating computational geometry into second-best congestion pricing design problem: Algorithm development and applications. In the 93rd Transportation Research Board Annual Meeting.

Marshall JD, Swor KR, Nguyen NP. 2014. Prioritizing environmental justice and equality: Diesel emissions in Southern California. Environmental Science & Technology 48(7): 4063-4068.

Mishra S, Welch TF. 2012. Joint travel demand and environmental model to incorporate emission pricing for large transportation networks. Transportation Research Record 2302:29-41.

Mitchell G. 2005. Forecasting environmental equity: Air quality responses to road user charging in Leeds, UK. Journal of Environmental Management 77:212-226.

Montgomery WD. 1972. Markets in licenses and efficient pollution control programs. Journal of Economic Theory 5(3):395-418.

Muller J, Shoemaker CA, Piché R. 2013. SO-MI: A surrogate model algorithm for computationally expensive nonlinear mixed-integer black-box global optimization problems. Computers and Operations Research 40:1383-1400.

Mun S, Konishi K, Yoshikawa K. 2003. Optimal cordon pricing. Journal of Urban Economics 54:21-38

Nagurney A, Ramanujam P, and Dhanda KK. 1998. A multimodal traffic network equilibrium model with emission pollution permits: Compliance vs. noncompliance. Transportation Research Part D 3(5):349-374.

Nagurney A. 2000a. Sustainable transportation networks. Cheltenham, UK: Edward Elgar

Nagurney A. 2000b. Congested urban transportation networks and emission paradoxes. Transportation Research Part D 5(2): 145–151.

Nagurney A 2000c. Alternative pollution permit systems for transportation networks based on origin/destination pairs and paths. Transportation Research Part D 5:37-58.

Nagurney A, Zhang D. 2001. Dynamics of a transportation pollution permit system with stability analysis and computations. Transportation Research Part D 6:243-268.

Nie Y. 2012. Transaction costs and tradable mobility credits. Transportation Research Part B 46:189-203.

Nie Y, Yin Y. 2013. Managing rush hour travel choices with tradable credit scheme. Transportation Research Part B 50:1-19.

Osorio C, Bierlaire M. 2013. A simulation-based optimization framework for urban transportation problems. Operations Research 61(6):1333-1345.

Ott W, Thomas J, Mage D, Wallace M. 1988. Validation of the simulation of human activity and pollutant exposure (SHAPE) model using paired days from the Denver, CO, carbon monoxide field study. Atmospheric Environment. 22(10):2101-2113.

Queipo NV, Haftka RT, Shyy W, Goel T, Vaidyanathan R, Tucker PK. 2005. Surrogate-based analysis and optimization. Progress in Aerospace Sciences 41:1-28.

Rabl A, Spadaro JV, Holland M. 2014. How much is clean air worth? Calculating the benefits of pollution control. Cambridge, UK: Cambridge University Press.

Raux C. 2004. The use of transferable permits in transport policy. Transportation Research Part D 9(3): 185-197.

Raux C, Mariot G. 2005. A system of tradable CO2 permits applied to fuel consumption by motorists. Transport Policy 12(3):255–265.

Raux C. 2010. The potential for CO2 emissions trading in transport: The case of personal vehicles and freight. Energy Efficiency 3:133-148.

Regis RG, Shoemaker CA. 2007. A stochastic radial basis function method for the global optimization of expensive functions. INFORMS Journal on Computing 19(4):497-509.

Rilett LR, Benedek CM. 1994. Traffic assignment under environmental and equity objective. Transportation Research Record 1443:92-99.

Rotaris L, Danielis R, Marcucci E, Massiani J. 2010. The urban road pricing scheme to curb pollution in Milan, Italy: Description, impacts and preliminary cost–benefit analysis assessment. Transportation Research Part A 44:359-375.

Sharma S, Mishra S. 2013. Intelligent transportation systems-enabled optimal emission pricing models for reducing carbon footprints in a bimodal network. Journal of Intelligent Transportation Systems: Technology, Planning, and Operations 17 (1):54-64.

Shepherd S, Sumalee A. 2004. A genetic algorithm based approach to optimal toll level and location problems. Networks and Spatial Economics 4:161-179.

Shirmohammadi N, Zangui M, Yin Y, Nie M. 2013. Analysis and design of tradable credit schemes under uncertainty. Transportation Research Record 2333:27-36.

Si B, Zhong M, Yang X, Gao Z. 2012. Modelling the congestion cost and vehicle emission within multimodal traffic network under the condition of equilibrium. Journal of System Science and System Engineering 21 (4):385-402.

Song Z, Yin Y, Lawphongpanich S. 2009. Nonnegative Pareto-improving tolls with multiclass network equilibria. Transportation Research Record 2091(1):70-78.

Storn R, Price K. 1997. Differential evolution: A simple and efficient heuristic for global optimization over continuous spaces. Journal of Global Optimization, 11:341-359.

Sugawara S, Niemeier DA. 2002. How much can vehicle emissions be reduced? Exploratory analysis of an upper boundary using an emissions-optimized trip assignment. Transportation Research Record 2260:29-37.

Sumalee A. 2004. Optimal road user charging cordon design: A heuristic optimization approach. Computer-Aided Civil and Infrastructure Engineering 19:377-392.

Teodorovic D, Edara P. 2005. Highway space inventory control system. In Proceedings of the ISTTT 16: Transportation and Traffic Theory: Flow, Dynamics and Human Interaction, Elsevier Publishers, Maryland, 43–62.

Teodorovic D, Triantis K, Edara P, Zhao Y, Mladenovic S. 2008. Auction-based congestion pricing. Transportation Planning and Technology 31(4):399-416.

Tonne C, Beevers S, Armstrong B, Kelly F, Wilkinson P. 2008. Air pollution and mortality benefits of the London Congestion Charge: Spatial and socioeconomic inequalities. Occupational and Environmental Medicine 65(9): 620-627.

U.S. Environmental Protection Agency. 2014. Inventory of US greenhouse gas emissions and sinks: 1990-2012. EPA 430-R-14-003, US Environmental Protection Agency, Washington DC.

Verhoef ET, Nijkamp P, Rietveld P. 1997. Tradable permits: Their potential in the regulation of road transport externalities. Environment and Planning B 24 (4):527-548.

Verhoef ET. 2002. Second-best congestion pricing in general networks: Heuristic algorithms for finding second-best optimal toll levels and toll points. Transportation Research Part B 36:707-729.

Verhoef ET. 2005. Second-best congestion pricing schemes in the monocentric city. Journal of Urban Economics 58:367-388

Viana FA, Gogu C, Haftka RT. 2010. Making the most out of surrogate models: Tricks of the trade. In Proceedings of the ASME 2010 International Design Engineering Technical Conferences and Computers and Information in Engineering Conference.

Wang Y, McElroy MB, Boersma KF, Eskes HJ, Veefkind JP. 2007. Traffic restrictions associated with the Sino-African summit: Reduction of NOx detected from space. Geophysical Research Letters 34(8), L08814, doi:10.1029/2007GL029326.

Wang X, Yang H, Zhu D, Li C. 2012. Tradable travel credits for congestion management with heterogeneous users. Transportation Research Part B 48:426-437.

Wang JYT, Ehrgott M, Dirks KN, Gupta A. 2014. A bi-level multi-objective road pricing model for economic, environmental and health sustainability. Transportation Research Procedia 3: 393-402.

Wismans LJJ. 2012. Toward sustainable dynamic traffic management. PhD Diss., University of Twente, The Netherlands.

Wolf H, Perry L. 2010. Trends in clean air legislation in Europe: Particulate matter and Low Emission Zones. Review of Environmental Economics and Policy 4(2):293-308.

Wu D, Yin Y, Lawphongpanich S, Yang H. 2012. Design of more equitable congestion pricing and tradable credit schemes for multimodal transportation networks. Transportation Research Part B 46(9): 1273-1287.

Xiong Y, Schneider JB. 1992. Transportation network design using a cumulative genetic algorithm and neural network. Transportation Research Record 1364:37-44.

Yang H., Bell MGH. 1997. Traffic restraint, road pricing, and network equilibrium. Transportation Research Part B 31 (4):303-314.

Yang H, Zhang X. 2002. Optimal toll design in second-best link-based congestion pricing. Transportation Research Record 1857:85-92.

Yang H, Zhang X. 2002. Multiclass network toll design problem with social and spatial equity constraints. Journal of Transportation Engineering 128 (5):420-428.

Yang H, Huang HJ. 2005. Mathematical and economic theory of road pricing. Oxford, UK: Elsevier.

Yang H, Wang X. 2011. Managing network mobility with tradable credits. Transportation Research Part B 45:580-594.

Yang Y, Yin Y, Lu H. 2013. Designing emission charging schemes for transportation conformity. Journal of Advanced Transportation DOI: 10.1002/atr.1226

Yin Y, Lu H. 2000. Traffic equilibrium problems with environmental concerns. Journal of Eastern Asia Society for Transportation Studies 3:195-206.

Yin Y, Yang H. 2004. Optimal tolls with a multiclass, bicriteria traffic network equilibrium. Transportation Research Record 1882: 45–52.

Yin Y, Lawphongpanich S. 2006. Internalizing emission externality on road networks. Transportation Research Part D 11:292-301.

Zhang J, Anderson AC. 2009. Adaptive differential evolution: A robust approach to multimodal problem optimization. Springer-Verlag,

Zhang L, Sun J. 2013. Dual-based heuristic for optimal cordon pricing design. Journal of Transportation Engineering 139 (11):1105-1116.

Zhang L, Yin Y, Chen S. 2013. Robust signal timing optimization with environmental concerns. Transportation Research Part C 29:55-71.

Zhang X, Yang H. 2004. The optimal cordon-based network congestion pricing problem. Transportation Research Part B 38:517-537.

Zhang Y, Lv J, Ying Q. 2010. Traffic assignment considering air quality. Transportation Research Part D 15:497-502.

Zhao Y, Triantis K, Teodorovic D, Edara P. 2010. A travel demand management strategy: The downtown space reservation system. European Journal of Operational Research 205:584-594.

Zhong RX, Sumalee A, Maruyama T. 2012. Dynamic marginal cost, access control, and pollution charge: A comparison of bottleneck and whole link models. Journal of Advanced Transportation 46:191-221.

Zhou XJ, Ma YZ, Li XF. Ensemble of surrogates with recursive arithmetic average. Structural and Multidisciplinary Optimization 44:651-671.

Zhu S, Du L, Zhang L. 2013. Rationing and pricing strategies for congestion mitigation: Behavioral theory, econometric model, and application in Beijing. Transportation Research Part B 57:210-224.

Zidek JV, Shaddick G, Meloche J, Chatfield C. 2005. Using a probabilistic model (pCNEM) to estimate personal exposure to air pollution. Environmetrics 16:481-493.