Université de Montréal

City decision-making: Optimization of the location and design of urban green spaces

par

Caroline Leboeuf

Département d'informatique et de recherche opérationnelle
Faculté des arts et des sciences

Mémoire présenté en vue de l'obtention du grade de
Maître ès sciences (M.Sc.)
en Informatique

April 30th, 2022

Université de Montréal

Faculté des arts et des sciences

Ce mémoire intitulé

City decision-making: Optimization of the location and design of urban green spaces

présenté par

Caroline Leboeuf

a été évalué par un jury composé des personnes suivantes :

Emma Frejinger

(présidente-rapporteuse)

Margarida Carvalho

(directrice de recherche)

Michel Gendreau

(membre du jury)

Résumé

Le besoin grandissant pour une planification urbaine plus durable et pour des interventions publiques visant à l'amélioration du bien-être collectif, ont grandement contribué à un engouement pour les espaces verts. Les parcs sont reconnus pour leur impact positif en zone urbaine dense, et nous sommes intéressés par l'application des concepts théoriques du domaine de la recherche opérationnelle pour assister les décideurs publics afin d'améliorer l'accessibilité, la distribution et la conception des parcs. Étant donné le contexte, nous sommes particulièrement motivés par le concept d'équité, et étudions le comportement des usagers des parcs à l'aide d'un modèle d'interaction spatiale, tel qu'appliqué dans les problèmes d'emplacement d'installations dans un marché compétitif. Dans cette recherche, nous présentons un modèle d'emplacement d'installations à deux étapes pouvant être adapté pour assister les décideurs publics à l'échelle de la ville. Nous étudions spécifiquement l'application aux espaces verts urbains, mais soulignons que des extensions du modèle peuvent permettre d'aborder d'autres problèmes d'emplacements d'installations sujets à des enjeux d'équité. La première étape de notre problème d'optimisation a pour but d'évaluer l'allocation la plus équitable du budget de la ville aux arrondissements, basé sur une somme du budget pondérée par des facteurs d'équité. Dans la deuxième étape du modèle, nous cherchons l'emplacement et la conception optimale des parcs, et l'objectif consiste à maximiser la probabilité totale que les individus visitent les parcs. Étant donné la non-linéarité de la fonction objective, nous appliquons une méthode de linéarisation et obtenons un modèle de programmation linéaire mixte en nombres entiers, pouvant être résolu avec des solveurs standards. Nous introduisons aussi une méthode de regroupement pour réduire la taille du problème, et ainsi trouver des solutions quasi optimales dans un délai raisonnable. Le modèle est testé à l'aide de l'étude de cas de la ville de Montréal, Canada, et nous présentons une analyse comparative des résultats afin de justifier la performance de notre modèle.

Mots Clés : Emplacement des installations; programmation en nombres entiers; modèles d'interaction spatiale; espaces verts urbains; prise de décision de la ville; équité.

Abstract

The recent promotion of sustainable urban planning combined with a growing need for public interventions to improve well-being and health in dense urban areas have led to an increased collective interest for green spaces. Parks have proven a wide range of benefits in urban areas, and we are interested in the application of theoretical concepts from the field of Operations Research to assist decision-makers to improve parks' accessibility, distribution and design. Given the context of public decision-making, we are particularly concerned with the concept of fairness, and are focused on an advanced assessment of users' behavior using a spatial interaction model (SIM) as in competitive facility locations' frameworks. In this research, we present a two-stage fair facility location and design (2SFFLD) model, which serves as a template model to assist public decision-makers at the city-level for the urban green spaces (UGSs) planning. We study the application of the 2SFFLD model to UGSs, but emphasize the potential extension to other applications to location problems concerned with fairness and equity. The first-stage of the optimization problem is about the optimal budget allocation based on a total fair-weighted budget formula. The second-stage seeks the optimal location and design of parks, and the objective consists of maximizing the total expected probability of individuals visiting parks. Given the non-linearity of the objective function, we apply a "Method-based Linearization" and obtain a mixed-integer linear program that can be solved with standard solvers. We further introduce a clustering method to reduce the size of the problem and determine a close to optimal solution within reasonable time constraints. The model is tested using the case study of the city of Montreal, Canada, and comparative results are discussed in detail to justify the performance of the model.

Keywords : Facility location problem; mixed-integer programming; spacial interaction models; urban green spaces; city decision-making; fairness.

Contents

List of Tables

List of Figures

List of Acronyms & Abbreviations

2SFFLD	Two-stage fair facility location and design
ACM	Attraction Choice Model
Can-ALE	Canadian Active Living Environment
CFLDP	Competitive Facility Location and Design Problem
DCM	Discrete Choice Model
FLP	Facility Location Problem
FSA	Forward Sortation Area
MCI	Multiplicative Competitive Interaction
MIP	Mixed-Integer Program
MILP	Mixed-Integer Linear Program
MINLP	Mixed-Integer Non-Linear Program

NDVI	Normalized Difference Vegetation Index
SIM	Spatial Interaction Model
UGS	Urban Green Space

Acknowledgements

I would like to express my sincere gratitude to my research supervisor, Dr. Margarida Carvalho, who has shown an exceptional support towards the completion of this thesis. I had the privilege to work with one of the few women in the department of Computer Science and Operations Research, and she has shown enormous generosity with her knowledge, time, and sense of guidance. I am incredibly thankful for this collaboration during the last two years, especially in a context that was severely impacted by the pandemic we have faced.

I also wish to extend my special thanks to Dr. Nurit Oliker, who has given us substantial inputs and considerations through the conception of our model. Working with Dr. Margarida Carvalho and Dr. Nurit Oliker was a great pleasure, and never would I have hoped for a better team to lead me in the completion of this research project.

A special mention is given to the SPHERElab team, including Benoit Thierry and Yan Kestens, who have provided us with data of exceptional quality and have underlined key considerations in the implementation of this project to a concrete application of great interest.

And finally, I would personally like to thank my best friends, who have reminded me while I was in the work industry of the value and importance of academia, and who have helped me find the courage to return to school.

Introduction

Context

The need to improve green spaces coverage and their accessibility in urban areas has been extensively acknowledged in different fields of studies, whether it is to respond to environmental or health challenges. In the 2017 WHO[1] report "Urban Green Space Interventions and Health: A review of impacts and effectiveness" [WHO, 2017], it was made evident that increased attention should be given by policymakers towards urban greening investments. In this detailed paper, we are provided with recommended methods for implementing green space interventions in urban settings. These recommendations are built with the intent of improving equity and health, in all of its forms. This is defended by suggesting, for example, to make use of existing collected datasets about green spaces usage and deprivation levels, or to gain a better understanding of the local demographics and parks' users. Although few studies have measured with precision the magnitude of the impact of green space installations on health, well-being, social equity or pollution, the authors also emphasize that few other public infrastructures can also address these targeted benefits with as much potential effect as green spaces. In respond to this report, it appears that a collaboration between researchers and practitioners in the field of health and urban planning would lead to interesting solutions and initiatives.

In addition, the current context of the Covid-19 pandemic brings to light the perceived value of society towards parks and green spaces. Indeed, a surprising stream of studies emerged with the intent to assess the distribution inequalities and the increase of perceived importance of urban green spaces in times of crisis. Studies and surveys conducted in Belgium [da Schio et al., 2021], Berlin [Collins et al., 2022], China [Zhang et al., 2022], Madrid [Maury-Mora et al., 2022], Mexico City [Mayen Huerta and Utomo, 2021], New-York City [Lopez et al., 2021] and Poland [Noszczyk et al., 2022] unanimously suggest that policymakers invest in improved accessibility to green spaces.

We should also mention that such interventions are considered as sustainable urban development initiatives, which is of particular interest when intense urban densification is observed

[1] World Health Organization

[Haaland and van den Bosch, 2015]. Furthermore, the increasing collective enthusiasm of the last years towards the field of data science provoked a surge of research using large scale data to improve predictions and better model real-world problems.

In this context, we are faced with the question of how can the field of Operations Research and computer science contribute to the concrete implementation of urban greening actions, and prove how the collaboration between practitioners and researchers can lead to realistic solutions for social good. For this purpose, we are interested in a model formulation which encompasses common urban planning considerations, while making use of reliable data as an input to recommend a method for improving the distribution and accessibility of parks. As suggested in WHO [2017], interventions include either the installation of new urban green spaces, or improvement of existing ones, with design options deemed with an important role in this decision process. In the following sections, we give a brief introduction of our work that was motived by this context. We start by stating the research problem that we attempt to answer with our work, followed by our contribution to the field, and conclude with the outline of this thesis.

Problem Statement

The question we are interested in is: how can we leverage existent data and promote fairness in city-decisions to improve the accessibility and distribution of urban green spaces? Based on this premise and on our expertise, we want to build an optimization model with sufficient flexibility to serve as a template, which can be modified to account for the attributes that are specific to the green spaces of a city. We are also concerned in proving that such optimization problem can be solved in a reasonable time or good quality solutions can be found efficiently. The model should further account for realistic considerations from the policymakers and the decision-making structure in practice while using accurate datasets. In this manner, resulting decisions with respect to the model's solution should have reduced bias, potential for discrimination and subjective motivations. We also focus on the definition of green spaces as parks, excluding public open spaces. The distinction can be found in WHO [2017]. Following this research question, we present a model, formulated as a two-stage decision process. In the first-stage, the city allocates a budget among the boroughs while seeking fairness. In the second-stage, independently, each borough optimizes the usage of urban green spaces by deciding both about their location and design, while restricted by the budget assigned by the city.

Contributions

Our contributions can be divided in four different axes: *(i)* modeling, *(ii)* a practical case-study, *(iii)* methodological, and *(iv)* experimental.

Our first contribution is the modeling of the two-stage decision-making process in cities for the planning of urban green spaces. To start, we identify the baseline procedure through which cities distribute budgets among neighborhoods (or boroughs). Then, we propose a redistribution of the budgets driven by fairness. Afterward, we model the problem faced by each neighborhood. The latter is an application of a special case of the commonly known Operations Research' facility location problem, with the use of a spatial interaction model to account for individuals' usage patterns of green spaces. More precisely, we modify the standard competitive facility location problem formulation to our context of public facilities provision, and argue how this can improve results compared to traditional distance minimization objectives.

Our two-stage model finds its strength in its flexibility to adapt to different contexts with different model parametrizations, and can easily be extended to other types of applications than urban green spaces, should minor changes to the model formulation be made. In this line, our contributions are twofold. First, we concretely define the first-stage problem for the city of Montreal, namely, we propose the fair budget allocation to be driven in terms of known statistical indexes for inequalities. Second, we generate second-stage instances based on existing datasets for the city of Montreal. This enables us to test our model on instances reflecting real-world problem topologies.

On one hand, the accuracy of our second-stage problem is highly related to the modeling of the demand (i.e., usage of a park by a category of users). On the other hand, a granular modeling of the demand results into large second-stage optimization problems. Thus, our methodological contribution comes in the form of aggregating similar demand points, resulting in smaller problems. Concretely, we propose a clustering technique for the demand points, allowing us to reduce the time to solve otherwise large instances, at the cost of a slight reduction of accuracy. We provide empirical evidence for the value of our approach to reduce the size of the second-stage problems.

Finally, we solve our two-stage model for the city of Montreal. This allows to discuss the effects of budget allocation, the performance of our second-stage size reduction approach, as well as the importance of properly estimating the model parameters for practical use.

To the best of our knowledge, this is the first work that uses a competitive facility location problem in a context of public planning, and that explicitly accounts for fairness through a two-stage optimization process. Some challenges of this application include the lack of available data for modeling the green space users accurately using statistical techniques. However, we expect our contributions to be valuable in assisting public decision-makers.

Outline

Here, we provide a clear outline of the thesis structure. We start with the literature review in Chapter 1, and introduce key researches that have motivated and inspired our current work. We first outline the different methods used for assessing accessibility to urban green spaces and the resulting discussion regarding inequalities using multiple case studies. Then, we detail how varying applications of the facility location problem compare to our specific problem, with a special consideration for the use of choice modeling. In Chapter 2, a formal background is given on facility location problems and choice modeling, with the corresponding notation that is used in the work that follows. Chapter 3 provides the detailed problem formulation of the two-stage fair facility location and design model. Moreover, it details the linearization methodology used to obtain a mixed-integer linear program for the second-stage problems. We then detail the case study of the city of Montreal in Chapter 4, and give a background about the data source and data manipulations to use it as an input in our model. Then, we provide an extensive discussion of the results in the computational experiments of Chapter 5 along with the methodology to reduce the size of the second-stage problems. We conclude in Chapter 6 with a brief review of our work, including our contributions and the shortcomings, and provide a discussion on the potential future work extensions.

Chapter 1

Literature Review

In this work, we are interested in the problem of optimal location and (re)design of existing and new urban green spaces (UGSs), and more specifically, of parks. We model this problem as a particular case of the discrete competitive facility location problem (CFLP) with design options and a spatial interaction model (SIM) to encompass the users' preferences. Given that the purpose is to assist public decision-making, we are taking the stance of policymakers, for whom the objective is to cover as much of the population's demand as possible with attractive and accessible UGSs. In other words, the objective is to maximize the probability of citizens visiting parks. In this way, the benefits of UGSs are expected to spread throughout the population, promoting benefits such as physical activity, mental health and social interaction. This underscores that policymakers must also be concerned with equity, or more broadly, fairness. Fairness is defined in terms of an equitable accessibility to the benefits of green spaces, an issue also explored in our work.

We remark that the term "competitive" is used in a broad sense, as in Aros-Vera et al. [2013], because there is no actual market competition in the context of public facilities' location problem. Instead, policymakers, also referred to as decision-makers, compete against the no-choice alternative (Bechler et al. [2021]) of not choosing any park due to insufficient attractiveness or accessibility of the available choices. We also define a choice model to input the parks' users behavior in the same way CFLPs account for customers preferences to increase their market share.

In this chapter, we provide a literature review motivating the studied problem as well as the positioning of our contributions within the existent work. In Section 1.1, we introduce studies that have used advanced techniques to assess UGSs accessibility and inequity, with methods to improve it. Then, in Section 1.2, we present papers on the subject of facility location problem (FLP) with similar characteristics to our model formulation, including choice models, competitive and non-competitive variants, design options and consideration of fairness.

1.1. Urban Green Spaces

Our research project finds its motivation from the numerous studies underlining the importance of better distribution of UGSs. In Section 1.1.1, we present several works from the fields of geography and urban planning which study accessibility to UGSs and the resulting inequity from the current spatial distribution using different approaches. To address accessibility issues underlined in previous studies, in Section 1.1.2, we review spatial optimization models that were developed with the aim to assist policymakers in the decision-making process, specifically for UGSs.

1.1.1. Accessibility and Inequity

One of the most popular approaches in assessing UGSs accessibility is to use Geographic Information Systems (GISs). In Oh and Jeong [2007], the distribution of parks in the city of Seoul is addressed due to inefficient accessibility. The authors debate that standard statistical indices on the parks' serviceability do not accurately convey whether a park location is in a central or outer area, neither how citizens benefit from them. For instance, previous studies about urban parks accessibility use the linear distance to parks, instead of accounting for the population real path choices and travel time, and thus, failing to quantify the population patronizing each park. In addressing this issue, the authors demonstrate the importance of considering such factors as land use and population density around a park. In Comber et al. [2008], Kabisch and Haase [2014] and Hoffimann et al. [2017], GISs were also used to assess inequalities of green space accessibility among diverse socioeconomic groups in Leicester, England, in Porto, Portugal, and in Berlin, Germany, respectively. In Coombes et al. [2010], using a GIS database of neighborhood and green spaces characteristics in Bristol, England, the authors investigate the relationship between accessibility to green spaces, physical activity and overweight. In Chang and Liao [2011], the authors are interested in an equitable public facility distribution in urban development using GIS and spatial analysis models in the city of Tainan, Taiwan. In all of these studies, proximity is deemed necessary when planning the location of UGSs. It should be underlined that in Hoffimann et al. [2017], the authors suggest that conclusions made in a specific location are geographically biased and should not be generalized to other cities.

Alternative methods were suggested to assess accessibility to UGSs. For example, in Ngom et al. [2016], the authors are interested in individuals' access to UGSs and suggest refining distance metrics with the use of travel costs and SIMs. An ANOVA regression model is developed to explain the distance to green spaces and their total coverage area using significant explanatory variables. Two case studies are used for this purpose, including databases from the cities of Quebec and Montreal, Canada. The results of the regression model show that Montreal displays less favorable access to green spaces in poorer areas, while wealthier areas have increased access to green spaces. Boone et al. [2009] recommend a novel approach based on Thiessen polygons to define each park and on an asymmetric reapportioning of census data, allowing to measure the crowding of a park. They apply the methodology in Baltimore, Maryland, and conclude unfair park access for African Americans. In Dai [2011], we are introduced to a Gaussian-based *two-step floating catchment area model* to assess spatial accessibility, and to an *ordinary least squares* model to assess socioeconomic disparities in the UGSs of Atlanta, Georgia. Using geospatial analysis and equity mapping exercises, Wolch et al. [2013] show inequalities between sociodemographic groups in Los Angeles, United States. Ibes [2015] uses a multistep statistical analysis to classify urban parks according to their specific physical and geographic dimensions, as well as their features. This is then mapped to a base layer of social variables in the city of Shanghai, China. In Ye et al. [2018], the two-step floating catchment area method is used to display changes of UGSs access between 2010 and 2015 in the city of Macau, China.

1.1.2. Spatial Optimization

Spatial optimization is known as a reliable approach in the field of geography and urban planning to help urban planners make decisions about the location of public facilities. For a detailed review of spatial optimization concepts, we refer the reader to Ligmann-Zielinska [2016]. Neema and Ohgai [2010] consider the siting of urban parks and open spaces through the case study of the city of Dhaka, Bangladesh. In this paper, the authors formulate a multi-objective model to account for the population's density, the air and noise pollution, and the areas with lack of accessibility to urban parks and open spaces. A genetic algorithm with dynamic weighting is used to solve the optimization problem, successfully generating Pareto optimal solutions. The results suggest that air pollution has the greatest impact on the locations. In Vallejo et al. [2017], two heuristics are proposed to optimize the allocation of green spaces over time. The authors explore an "offline" and an "online" approach using an evolutionary algorithm in a sequential decision-making problem. In Yoon et al. [2019], the authors study the optimal location and type of green spaces in a planning model given their greening benefits. The problem defines a multi-objective formulation, maximizing cooling and connectivity, and minimizing costs. The problem is solved using a non-dominated sorting genetic algorithm. Yu et al. [2020] formulate a multi-objective function to maximize simultaneously the economic, ecological and social value of green spaces and solve the problem using a genetic algorithm. In this paper, the social value acts as the input to model the users' demand for UGSs. More recently, Li and Ma [2022] aim to support the decision-making process of UGSs planning using an optimization method that minimizes the land conversion cost of newly added UGS parcels. This solution aims to lower construction costs and improve utility of UGSs. The authors highlight that previous spatial optimization solutions can hardly be used in real-life context, failing to meet the actual constraints of equity and costs.

1.2. Facility Location Problem

The problem of location and design of UGSs inherits from concepts of multiple variants of the well-known FLP. As mentioned in Section 1.1.1, estimating the number of individuals who patronize each UGS (facility) is extremely relevant to a proper assessment of their accessibility. Therefore, this metric is essential to anticipate the added value of a new UGS or an enhanced UGS. Thus, we focus our literature review on FLPs that make use of choice models to estimate users' demand for each facility within a competitive framework. An extensive stream of studies has made use of the random utility model (RUM), first introduced by McFadden et al. [1973], which we review in Section 1.2.1. Then, in Section 1.2.2, we introduce a stream of literature which makes use of a proportional (probabilistic) choice rule with deterministic utilities, also known as gravity models or SIMs. The complexity of the CFLP with both types of choice models resides in the non-linearity of the objective function, with scholars recommending appropriate algorithmic solutions depending on the specified problem and context. In Section 1.2.3, we conclude with studies that introduce a fairness component into their problem formulation, a key consideration when locating public facilities.

1.2.1. Random Utility Model

Many FLPs studies have made use of the RUM, and authors have provided numerous solutions to overcome the non-linearity of the objective function and computational performance challenges. We note the work of Aros-Vera et al. [2013], in which a mixed linear programming formulation is proposed to model the problem of optimizing the location of park and ride facilities to maximize commuters usage, given a demand that follows a logit model in a non-competitive framework. The authors suggest a linearization method to solve small instances, and a modification of the heuristic concentration integer procedure to tackle large instances. Motivated by the problem of school network planing, Haase and Müller [2013] propose a method for solving a non-competitive discrete location model with endogenous demand, in which students' preferences are modeled with a RUM. The target objective is defined as the maximization of the total expected utility over all students, weighted by the number of students at each location and selected school. Utilities are then simulated given that choice probabilities based on the mixed-multinomial logit model do not display a closed-form formula, and the authors prove that this model is a better and more realistic alternative to the traditional multinomial logit model. In Haase and Müller [2014], a firm's objective is to maximize the frequency of visits, or alternatively, the probability that customers will choose their facility over their competitor's. The objective function is formulated as a mixed-integer non-linear program (MINLP). Then, the authors describe three linear reformulations from the literature with a unified notation. They further compare each formulation's performance using artificial data. Limitations relating to the constant substitution assumption make these formulations unrealistic and difficult to apply to real-life problems. In Müller and Haase [2014], the same authors stretch the importance of customers segmentation to create homogeneous subgroups with shared characteristics at each demand point. The reasoning behind this suggestion is to reduce the bias from the unrealistic assumption of independence of irrelevant alternatives in the multinomial logit model. Ljubić and Moreno [2018] propose the first branch-and-cut algorithm for the maximum capture FLP when customers demand is modeled through an multinomial logit model, and show its computational effectiveness. Mai and Lodi [2020] address the same problem and recommend a modified version of the multi-cut outer-approximation algorithm that uses a cutting plane approach as opposed to the state-of-art branch-and-cut method by Ljubić and Moreno [2018]. Their method is shown to be more robust and more efficient, especially on large scale instances.

1.2.2. Spatial Interaction Model

The main alternative to RUMs in choice modeling within the framework of CFLPs is based on Luce's choice axiom (refer to Luce [1959] for more details), in which the choice probability is defined as a constant ratio of deterministic utilities. Luce's axiom leads to the proportional choice rule, in which customers' probability of choosing a facility is proportional to deterministic utilities [Lin and Tian, 2021]. The proportional choice rule we are interested in is the SIM, also referred to as the gravity model or Huff-model (refer to Chapter 2 for more details). As underlined in Berman and Krass [1998], the proportional choice rule is deemed more appropriate for real-life application than the alternative deterministic rule, also referred to as the binary-rule, the all-or-nothing approach or the full capture model [Aboolian et al., 2007b]. For this reason, we focus on SIMs. Moreover, we should mention that CFLPs with customers demand modeled using SIMs have many applications. Concrete examples include health-care facility location [Ammari et al., 2000] and electric vehicle charging stations placement [Anjos et al., 2020], to name a few. Below, we review key articles and their corresponding methodologies to tackle the objective that is being optimized.

A stream of studies using CFLPs focus on the maximization of market share generated from newly installed facilities. In Aboolian et al. [2007b], the authors introduce the competitive facility location and design problem (CFLDP), where each facility has specific design options in a competitive framework. Here, the goal is to simultaneously optimize a facility's location and its design components. The customer's utility is defined as proportional to the facility's attractiveness and to the inverse of the distance. The problem is formulated as an integer program with non-linear objective function and constraints, and focuses on a specific CFLDP where only a specific number of design options are available to ensure solvability. Two distinct algorithmic approaches are discussed for solving medium and large size instances, showing accurate results with reasonable computation time for both methods. In Aboolian et al. [2021], the same authors introduce the generalized facility location and design problem, a generalized version of the previously introduced CFLDP. This model enables accounting for two distinct types of demand allocation, including the gravity type and all-or nothing models. In addition, it captures demand expansion, as well as the cannibalization effect. Due to the non-linearity, integrality constraints and high dimensionality, the problem is deemed very hard to solve. It is further simplified using dimensionality reduction and linearization to obtain a solvable mixed-integer programming (MIP) model. In Drezner et al. [2018], the authors extend the gravity model by assuming that facilities' attractiveness is random, which is argued as a more realistic approach. The authors suggest two solutions, one concerns discretization of the attraction levels, the second uses the concept of effective attraction.

A more extensive line of studies approached the CFLP with the objective to maximize profit, defined as the total expected sum of revenues minus costs and expenses. We mention Kucukaydin et al. [2011], in which a firm's objective is to locate new facilities and define their attractiveness level to maximize their profit. The authors use a MINLP formulation to model the problem and propose three solutions. Gila Arrondo et al. [2014] consider a competitive framework in which the demand depends on the market conditions. Although more realistic, this problem is deemed much more complex to solve and makes use of global optimization techniques, an evolutionary algorithm and a parallel version of it to improve solvability. Redondo et al. [2015] formulate a bi-objective function for both the franchisor and the franchisee to maximize profit, which in this specific case is equivalent to maximizing the market share. An evolutionary algorithm is suggested for obtaining the corresponding Pareto-front. In Fernández et al. [2017a], the authors introduce the multi-deterministic choice rule, in which customers patronize only one facility for each firm, and distribute their demand proportionally to each facility's attractiveness. Here, the continuous location problem is studied. More recently, Lin and Tian [2021] introduced a generalized version of the CFLP and proposed a branch-and-cut algorithm based on Benders decomposition, which is proven to outperform the state-of-art exact approaches.

We end this section with additional references for the interested reader. A survey on models estimating demand based on SIMs is provided in Eiselt and Laporte [1989]. The tutorial by Berman et al. [2009] and the recent book chapter by Drezner [2019] survey and detail various aspects and approaches to CFLPs using SIMs.

1.2.3. Fairness

Little attention has been given to the question of including fairness as a component of the FLP's objective function, or as an alternative objective to the classic distance minimization. Marsh and Schilling [1994] made the first detailed literature review of the use of fairness and equity in FLPs, and introduced a unified notation for all seven measures recognized at the time. They include the center (as used in the p-center problem), the variance, the mean absolute deviation, the sum of absolute deviations, the Gini coefficient, the range, and modified versions of the sum of absolute deviations. More recent studies in location problems attempted to introduce a fairness component in the problem formulation. Below, we introduce two papers that consider fairness in the objective function. In both papers, the concept of demand is not used in the same way as in the literature covering competitive frameworks. Indeed, these work are not interested in market share or revenue maximization, but rather on the promotion of equitable solutions where customers receive a similarly "satisfactory" service by the facility that they patronize.

In Jung et al. [2019], the authors underline the limitations of minimizing traditional measures such as k-center, k-means and k-medians to achieve a fair solution to the classical FLP. Here, the demand is not factored in the problem. An alternative fair k-center measure is introduced, which accounts for the population density through the neighborhood radius. The α-fairness measure is further defined as the maximum ratio of the distance to the nearest facility over the neighborhood radius. Upper and lower bounds on the α-fairness value are derived, and the authors prove that the fair k-center measure performs best in achieving minimal α-fairness when compared with traditional measures.

In Filippi et al. [2021], the authors study the fair single-source capacitated facility location problem, a modified version of the classic SSCFL problem which presumes that customers will access the provided facilities at an incurred cost. The first objective function is defined as the minimization of the costs for both the planner (installation cost) and the customers (assignment cost). The second objective component of the bi-objective formulation includes the fairness element, with the minimization of the $\beta\%$-worst assignment costs. The concept of demand in this paper is referred to as the amount of demand that is to be satisfied by the assigned facility. Small instances are solved using a weighted sum method, while larger instances are solved using Benders decomposition. Some authors have also studied fairness through the measure of envy [Filippi et al., 2021].

Chapter 2

Background

As introduced in the previous sections, we are interested in the problem of discrete FLP applied to the case of UGSs, and more specifically of neighborhood parks. The terms UGSs and parks will be used interchangeably in this thesis. Given that the set of existing or potentially new UGSs are usually assigned to predefined geographic coordinates, our work only refers to discrete versions of the problem, unless specified otherwise. Our goal is to suggest a flexible model to assist decision-makers in the process of locating and selecting the design of existing and new UGSs. Therefore, this research is focused on the applicability of such a model, and a consideration of computation complexity and model formulation should reflect this intent. The literature has taught us that few researches have studied the location of public facilities using a competitive framework with users' behavior modeling. Indeed, on one hand, as underlined in Gorji [2015], the existing models for the problem of locating public facilities rather use a simplistic choice rule such as patronizing the closest facility, which is far from the observed behavior (refer to Section 1.1 for a detailed justification). On the other hand, most CFLPs are applied in a context of market competition, with the intention to either maximize a firm's market share or revenues. In this work, we extend the concept of market share maximization as a proxy for maximizing the probability of visiting UGSs. Furthermore, advanced methods to model customers' behavior (or demand) have been developed for CFLPs, and can be extended to UGSs users to improve the assessment of the distribution and design of parks, and therefore increase their usage.

We start the background review in Section 2.1 with a preliminary formulation of a generic discrete FLP. We then introduce the CFLP in Section 2.2, and present its main variations, while justifying how they relate to our application case. Then, Section 2.3 details the chronological background of choice models used in location models, considered as one of the most important and complex modeling elements in a CFLP.

2.1. Preliminaries

In this section, we define a unified notation for discrete FLPs, which will be extended later to the specific case of CFLPs. Let I be the set of demand points in the location problem, which are generally defined as a geographic zone (e.g., neighborhood, forward sortation area (FSA) or postal codes). In discrete planning, a geographical location is often given as the centroid of the associated zone. We also define J as the set of candidate locations for facilities installation. The decision variable to optimize is x_j, which is equal to 1 if a facility is installed in location j, and 0 otherwise. Below is a generic formulation of the discrete FLP:

$$\max_{x} \sum_{j \in J} x_j \left(\sum_{i \in I} f_{ij}(x) \right) \tag{2.1.1a}$$

$$s.t. \quad x \in X \tag{2.1.1b}$$

$$x_j \in \{0,1\}, \forall\, j \in J, \tag{2.1.1c}$$

where each real-valued function $f_{ij}(x)$ in the Objective (2.1.1a) relates the demand point $i \in I$ with the facility $j \in J$ and the decision x. These functions model the element to be optimized according to the context. For example, $f_{ij}(x)$ can correspond to the negative of the average distance between i and j, to the market share of demand point i captured by the facility j, or to the expected revenue obtained from i visiting j. In Problem (2.1.1), X is a set of constraints related with the installment of the facilities (e.g., budget, coverage). Following this formulation, discrete FLPs have broadly been defined as the problem of selecting a subset of locations from a candidate set J, given the demand of users in the set of demand points I [Laporte et al., 2015]. Traditional FLPs have focused on optimization problems like the minimization of the demand-weighted average distance in the p-Median problem, the minimization of the maximum distance in the p-Center problem, or the maximization of coverage given a budget constraint. Studies have extended the classic FLP concept to more complex models, often to address realistic applications. For a detailed overview of location problems and their application, the reader is referred to Laporte et al. [2015].

2.2. Competitive Facility Location Problem

As emphasized earlier, the CFLP is the version of the discrete FLP that best serves the purpose of our research project. This is because it enables the consideration of a "competitive" alternative for the UGS users, which will be designated by the *no-choice alternative* and denoted by 0. This makes the set of available choices for I to $J \cup \{0\}$. Of course, in general competitive settings, the alternative choice set is not necessarily a singleton.

The first paper that was recognized for introducing the competitive framework into location problems is Hotelling [1929]. Here, two facilities are competing on a one-dimensional straight line, and identical customers patronize the facility providing the lowest incurred cost.

Several extensions of the CFLP have inherited from the concepts introduced in Hotelling [1929], either in continuous or discrete solution spaces. Variations of the original model have considered the number of facilities to locate, the type of competition, the type of demand, the inclusion of design options and the choice rule [Gorji, 2015], which are reviewed in this section. We also introduce less common characteristics of the CFLP related to previous work, which are particularly useful in the UGSs location and design problem.

Number of facilities. In earlier studies, Drezner [1994a] and Drezner [1994b] tackled the location of a single facility in a continuous plane. Most subsequent studies have focused on the location of multiple facilities for more realistic applications like Berman and Krass [1998] and Drezner [1998]. In the case of location of UGSs, it seems evident that our problem considers multiple facilities.

Type of competition. Another characteristic of CFLP that is worth mentioning, is the type of competition that is being modeled. Indeed, all papers mentioned in Chapter 1 use a *static* competitive framework. However, there exists a stream of studies relating to dynamic (sequential) competitive location decisions. First introduced in Hay [1976] and Prescott and Visscher [1977], this model is also referred to as the Stackelberg game in Game Theory, or the leader-follower problem in location science [Laporte et al., 2015]. In this type of problem, we consider a bi-level optimization formulation in which a leader first locates one or multiple facilities, and then, a follower reacts to this location decision and defines the location of its facilities accordingly. Given that we are concerned with the maximization of UGSs usage by minimizing the alternative of not visiting any park, the bi-level modeling scheme seems inappropriate. We refer the reader to Fernández et al. [2017b] for a detailed review of the leader-follower model formulations. In this manner, we will presume a static competitive framework for the rest of this thesis.

Demand. Recent studies have also focused on the integration of various types of demand in a competitive framework. A type of demand is defined here as either elastic or inelastic. Inelastic demand is considered appropriate for essential goods, while elastic demand is generally associated to non-essential goods [Gorji, 2015]. Elastic demand is expected to expand as the diversification of the services and the resulting total utility is increased in a competitive market [Aboolian et al., 2007b]. This effect is known as the “market expansion” or the “cannibalization effect” [Aboolian et al., 2007a]. First introduced in Berman and Krass [2002], and later used in Aboolian et al. [2007a], most works have not accounted for demand elasticity given the consequent increased complexity of the objective function. In our context, demand is presumed as inelastic, given that UGSs are deemed as public facilities, and are often defended as an essential service. Although it could be argued that demand might vary according to the amount of UGSs available, we presume that the demand elasticity is negligible for our purpose.

We now mention additional CFLP characteristics that are specific to our problem formulation and to the field of application. First, as suggested in Haase and Müller [2014], the definition of the set of demand points can be refined by segmenting each location in subgroups to create more homogeneous categories of individuals with shared characteristics, such as sociodemographic attributes. Indeed, in the context of UGSs location, we make the assumption that age has a significant impact on the frequency of usage, sensitivity to distance and preferred design options. For this reason, we introduce the set of segments S. Each pair $(i,s) \in I \times S$ corresponds to the UGS users from the demand point i in the age segment s.

Design options. Our location problem is further refined by allowing the model to define the design option of existing and new UGSs. We distinguish the concept of design options from the one defined in Aboolian et al. [2007b], in which they are a set of variables acting as attributes of facilities that need to be optimized for market share maximization. The reason for not modeling attributes as a decision variable is that the design of UGSs is usually undertaken by specialists that are given a predefined budget, and therefore, we are not interested in the specific design attributes. Instead, we refer to Section 4 of Aboolian et al. [2007b] which presents discrete design scenarios and which have the same definition as the design options we refer to in our problem formulation. In this section, the conventional location decision variable x_j is redefined as x_{jr} and equals 1 if the facility with design scenario r is selected, and 0 otherwise. In the next chapter (Chapter 3), we will use the same convention introduced in the discrete design scenarios formulation for the design option and associate a cost c_{jr} to each design option.

Choice rule. Another distinction in our model is the consideration of the no-choice option as an alternative in our "competitive" framework. This was mentioned in Bechler et al. [2021], who emphasized that the total utility resulting from a market should account for the no-choice alternative. Therefore, our set of competitors in the UGSs context is an empty set, but the total set of choices includes the choice of not visiting a park, as well as visiting parks that are not subject to optimization (this is discussed in detail in Chapter 3).

2.3. Choice Modeling

In this section, we detail the different choice rules used in CFLPs through a chronological background with relevant literature references. We first introduce the traditional deterministic rules, then follow with the more realistic probabilistic approaches and how they relate to our problem. We conclude this section by discussing the solution methods for location problems using the surveyed choice rules.

Classic deterministic rules. The simplistic choice rule by Hotelling [1929] introduced in Section 2.1 presumes an all-or-nothing approach, also known as the full-capture, the binary or the deterministic approach. Other influential papers have used this proposition, such as Revelle [1986], Hakimi [1983] and Drezner [1994a]. They presume that customers patronize only one facility to satisfy their demand, either according to distance minimization or attractiveness (utility) maximization. Gorji [2015] reinforced that such a framework would be appropriate in some cases, such as in central planning (e.g., polling location).

Probabilistic (deterministic) rules. Other choice rules have gained attention to respond to the criticisms and evident shortcomings of the all-or-nothing approach in competitive location models. The gravity-based approach, also known as the SIM, is widely used given its reliability in estimating market share and the simplicity of its formulation. The first introduction of the gravity model in location models was made in Reilly [1931], where customers account for facility's attractiveness and not only distances for making their choice. Following this important work, Huff [1964] then introduced the well-known Huff-model. It defines the probability of selecting a facility as a ratio of its total space area to the power of the distance, as below

$$p_{ij} = \frac{A_j/d_{ij}^{\beta}}{\sum_{k\in J} A_k/d_{ik}^{\beta}} \ , \tag{2.3.1}$$

where i is a demand point in the set I, j is a facility location point in the set J, A_j corresponds to the space area of the facility j (the letter A refers to the attraction), d_{ij} is the distance between demand point i and facility location j, and β is the distance sensitivity for the type of facility considered in the problem. The reasoning is that a customer patronizes a facility according to its attractiveness, exemplified in Huff's model by the space area, and it is negatively impacted by its distance. His work also inherited of Luce's choice axiom in Luce [1959], which stipulates a probabilistic choice rule with deterministic utilities as the ratio of a choice's utility divided by the sum of all available alternatives' corresponding utilities. The ratio can also be seen as an estimation of a facility's market share MS_{ij}, such that

$$p_{ij} = MS_{ij} = \frac{u_{ij}}{\sum_{k\in J} u_{ik}} \ , \tag{2.3.2}$$

where u_{ij} is the utility perceived by the user i for the choice j. Luce's axiom also defines the assumption of independence of irrelevant alternatives resulting from the probabilistic choice rule. This assumption stipulates that two options will display constant ratio of choice probabilities, and is independent of the other alternatives in the choice set [Haase and Müller, 2013]. Nakanishi and Cooper [1974] extended Huff's model and proposed the multiplicative competitive interaction (MCI) model by generalizing the space area with the multiplication of a facility's attribute's attractiveness:

$$p_{ij} = \frac{\prod_{l\in Q} x_{lij}^{\beta_l}}{\sum_{k\in J}\prod_{l\in Q} x_{lij}^{\beta_l}} \ , \tag{2.3.3}$$

where Q is the set of attributes, x_{lij} is the l-th attribute of location j for the user i, and β_l is the sensitivity to this l-th attribute. Parameters estimation based on statistical methods were developed in the same paper. Other variations of the Huff model surged and redefined the attraction function or the distance decay function, such as in Aboolian et al. [2007b] which extended the concept of the MCI model. In this paper, the design components of a facility are considered, and the authors define the choice rule using Luce's axiom for the market share formula, while utility is defined as

$$u_{ij} = \frac{A_j}{(1+d_{ij})^{\beta}} \ , \tag{2.3.4}$$

where u_{ij} is the deterministic utility of facility j from the perspective of customer i, A_j is the general attractiveness of facility j, and $1/(1+d_{ij})^{\beta}$ is the distance decay function. Justification for the use of this function over Huff's version $1/d_{ij}^{\beta}$ is given in Aboolian et al. [2007a] who recommend its use, particularly, given very small distances d_{ij}. In such cases, facilities' attractiveness should be the defining component for a customer's choice. The attractiveness definition further inherits from Nakanishi and Cooper [1974]'s model, such that

$$A_j = x_j \alpha_j \prod_{l\in Q} (1+x_{lj})^{\beta_l} \tag{2.3.5}$$

where x_j keeps its meaning as a binary variable equal to 1 if the facility j is chosen and 0 otherwise, and α_j is the base attractiveness of location j. Parameter β_l has the same definition as in the original MCI model, while variable $(1+x_{lj})$ is the replacement for the original variable x_{lij}. In the context of optimization and CFLPs, we remind the reader that the gravity-based rule is a probabilistic choice rule, with deterministic utilities. As noted in Huff [2021], rigorous parameter estimation in the Huff-model is often neglected and lacks statistical methods. Indeed, given that the choice rule must be linearized in order to apply statistical regression models to estimate the sensitivity parameters β, many authors have ignored the details related to parameters estimation. In this thesis, we will also focus on the formulation of the choice rule and the location problem without developing on the parameters' estimation due to the lack of data. We refer the interested reader to Nakanishi and Cooper [1974] and Huff and McCallum [2008] for recommended model calibrations.

Random utilities. An alternative probabilistic rule was introduced in McFadden et al. [1973], assuming random (stochastic) utilities. This model is referred to as the RUM, or alternatively, discrete choice models (DCMs). In this context, the customer will choose the facility that maximizes its utility. The utility is therefore defined as the sum of a deterministic component v_{ij} and a random component ϵ_{ij}, such that

$$u_{ij} = v_{ij} + \epsilon_{ij} \ . \tag{2.3.6}$$

The distribution of ϵ_{ij} is what defines the type of RUM, either the multinomial logit model, the mixed-multinomial logit model, the *generalized extreme value* models or *probits*. Other models are available, and the reader is referred to Train [2009] for a detailed review of DCMs. Drezner et al. [2011] suggested another approach to the classic deterministic (proximity or attraction approach) or probabilistic (SIM and RUM) choice rules introduced previously, namely the cover-based approach. In this model formulation, each facility displays a sphere of "influence" defined by a radius, and customers will patronize their demand equally across all the facility spheres they are in. Nevertheless, we must remark that most studies have focused on the application of SIMs more importantly, and RUMs.

Solution approaches to optimization problems. The main challenge with the use of choice model within CFLPs is the resulting non-linearity of the objective function. For example, $f_{ij}(x)$ in the Objective (2.1.1a) could be equal to p_{ij} in Equation (2.3.1), i.e.,

$$f_{ij}(x) = \frac{u_{ij}}{\sum_{k \in J} u_{ik} \cdot x_k},$$

clearly making the objective function non-linear. Thus, linearization methods are required in order to obtain a mixed-integer linear problem (MILP) that can be solved with standard optimization software (solvers). Linearization methods have been documented in Haase and Müller [2013] and Bechler et al. [2021], and were applied in many subsequent studies. Unfortunately, the final problem is known to be NP-hard (Benati [1999] and Bechler et al. [2021]), and the additional variables introduced with the linearization yield very large problems, often leading to long-running times or even out-of-memory issues. For this reason, authors have studied alternative algorithmic approaches to achieve optimal solutions, such as in Ljubić and Moreno [2018] and Mai and Lodi [2020] for RUMs, and Aboolian et al. [2021] for SIMs. Both probabilistic choice rules have been extensively studied, providing variant advantages depending on the field of application. Given the recognized applicability of the SIM and the absence of market competition with UGSs, we develop a simple choice rule based on the SIM. In Chapter 6, we discuss potential model extensions to consider RUM as a choice rule.

Chapter 3

Problem Formulation

In this chapter, we formally present the notation and the model for locating and (re)designing UGSs as a two-stage fair facility location and design (2SFFLD) model. See Figure 3.1 for an overview of our model.

To emphasize on the context of public decision-making, we consider the concept of fairness in the first stage of the optimization problem. In this step, the budget allocation of a city is distributed among neighborhoods (typically, city administrative subdivisions) such that inequalities of accessibility to UGSs are accounted for. This step is presented in detail in Section 3.1. Later, in Chapter 4, we will describe the actual budget allocation for our case study, the city of Montreal and, in Chapter 5, we will compare it to the one we propose here.

The second-stage problem is about the location and (re)design of UGSs for each neighborhood, given its predefined budget derived in the first-stage process. Here, we do not incorporate fairness in the usual sense of the term, i.e., we do not directly penalize planning solutions that favor a majority of the population. Thus, we also provide a discussion on the problematic use of standard fairness concepts, motivating the objective function of our model. This second step of decision-making is described in Section 3.2.

To our knowledge, this is the first research project simultaneously *(i)* considering a city decision-making process in sequence, intrinsic to their administrative subdivision structure, *(ii)* including the concept of fairness through the city's budget-allocation to neighborhoods (subdivisions), and *(iii)* modeling the population access to facilities administrated by each neighborhood using SIMs. As this enumeration of model characteristics highlights, although we are particularly focused on the application of our model to UGSs to assist public decision-makers, we will present a formulation that can be easily extended to other fields of application.

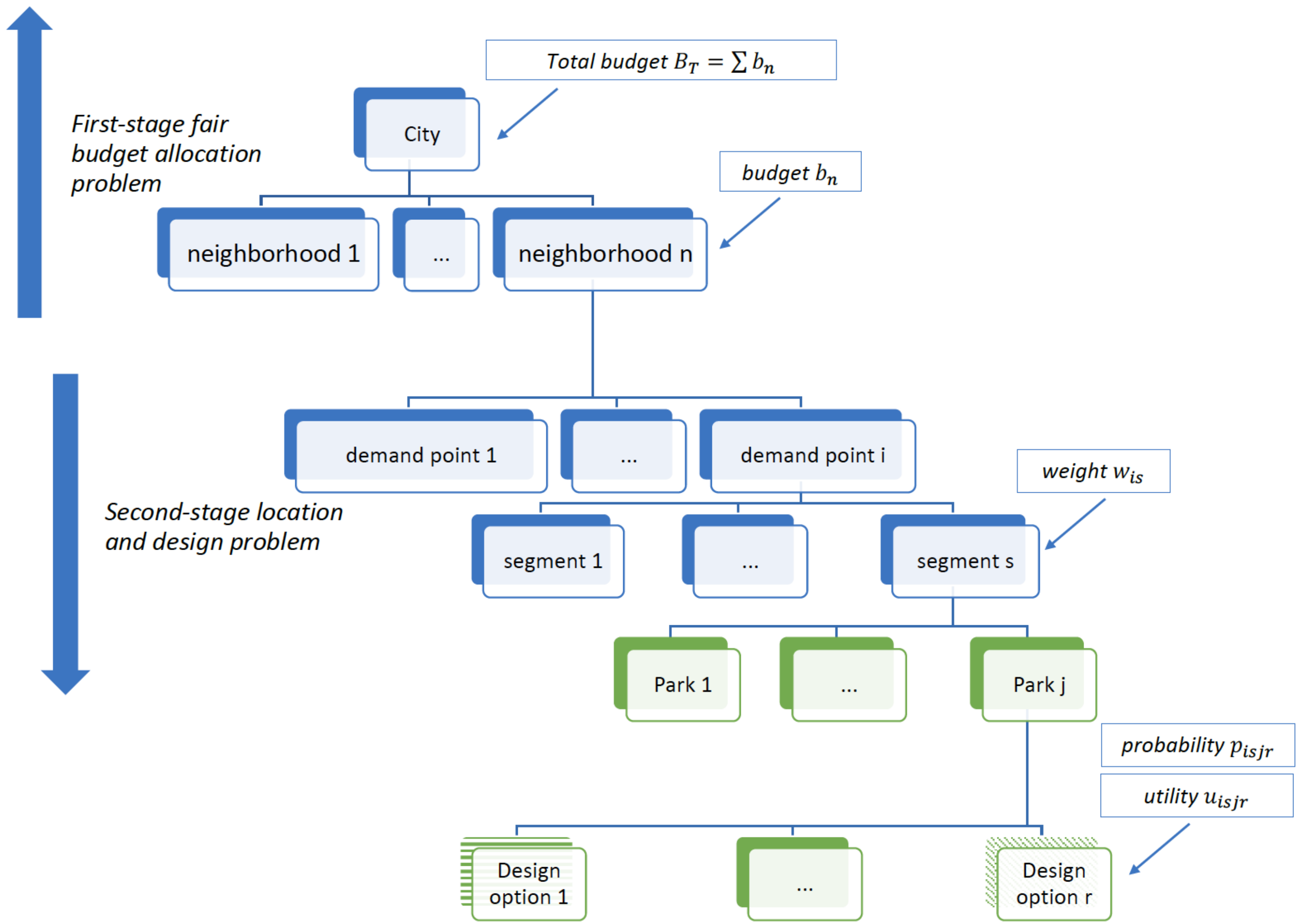


Figure 3.1. Two-stage fair facility location and design model

3.1. First-Stage: Fair Budget Allocation Model

The first-stage fair budget allocation step requires distributing a total city budget B_T fairly among neighborhoods for UGS planning. Defining concretely, fairness depends heavily on the decision-maker's preferred outcome, which varies from one application to another. There is an extensive literature investigating the axiomatic definition of fair resource allocations, such as proportional fairness and max-min fairness; see Bertsimas et al. [2011] for a general overview of the most well-known and fundamental fairness concepts. One particular area of research in fair allocations is fair division. The latter focuses on allocations satisfying properties such us envy-freeness and Pareto-efficiency, to name a few; see the survey by Bouveret et al. [2016]. In our practical context, the allocation of budget to neighborhoods by cities can be strongly restricted due to maintenance costs, previous allocation record, etc. Thus, due to the imminent limited freedom to deviate from a pre-determined budget allocation, we will present a simple division of the budget guided by measures of each neighborhood disadvantages.

We denote the set of neighborhoods by N. For each neighborhood $n \in N$, the allocated budget is given by the decision variable b_n. We further define a fixed baseline budget $\bar{b}_n$ for each neighborhood $n \in N$. In our context, this is the budget that follows estimation of budget per capita for UGSs investment based on city recommendation, without consideration of other fairness elements. The total city-budget B_T is therefore formulated as

$$B_T = \sum_n \bar{b}_n \ . \tag{3.1.1}$$

The minimum budget per neighborhood $n \in N$ required to cover the maintenance cost is given by $\underline{b}_n$.

The optimization problem consists of maximizing the *fair-weighted city budget* (3.1.2):

$$\max_{b \in \mathbb{R}^N_{\geq 0}} \sum_{n \in N} b_n \rho_n \tag{3.1.2a}$$

$$\text{s. t.} \quad \sum_{n \in N} b_n \leq B_T \tag{3.1.2b}$$

$$|\frac{b_n - \bar{b}_n}{\bar{b}_n}| \leq \delta\% \quad \forall n \in N \tag{3.1.2c}$$

$$b_n \geq \underline{b}_n \qquad \forall n \in N \ . \tag{3.1.2d}$$

The objective function (3.1.2a) is guided by the weighting parameters $\rho_n \ \forall \ n \in N$, which act as multiplicative factors, favoring disadvantaged neighborhoods according to predefined attributes. Given the context of UGSs, these fairness attributes include population size, social and material deprivation index, and pollution. These attributes, as well as all the remaining parameters presented here, will be computed for our case study in Chapter 4. Constraint (3.1.2b) ensures that the total budget of the city is not exceeded. Constraints (3.1.2c) enforce the percentage threshold δ on the maximum neighborhood budget deviation from the baseline budget $\bar{b}_n$. In this way, we prevent an excessive deviation from the baseline budget. Constraints (3.1.2d) guarantee that maintenance costs are at the very least covered.

We remark that model (3.1.2) could also be applied for a provincial, territorial or state level budget to fairly allocate it to its municipalities, regions or counties. Although, we presume a city-decision making process with fair budget distribution to its neighborhoods, a direct application to other types of regional division is straightforward.

3.2. Second-Stage: Green Space Location and Design Model

In the second-stage of the decision process, we take the (independent) perspective of each neighborhood $n \in N$. The goal of the neighborhood is to maximize the overall expected proportion of the neighborhood population visiting its parks. To this end, the neighborhood seeks the optimal (re)design options of existing parks and the optimal location and design of new parks. In the following notation description, for the sake of simplicity, we omit the reference to a neighborhood n and presume that the optimization occurs at the neighborhood level.

Next, we adapt the general notation of Section 2.2 to the current context. We define I as the set of demand points whose location corresponds to the centroid of the geographic region defining the demand zone. The demand zone of $i \in I$ can correspond to a postal code, FSA, neighborhood, etc. The set of segments S allows creating homogeneous clusters of individuals with similar usage behavior according to sociodemographic characteristics. This methodology follows the work of Haase and Müller [2013]. In our application, S will define age group segments. For each pair $(i,s) \in I \times S$, we define the weight w_{is} representing the population size percentage located in the demand zone i and in segment s such that $\sum_{i\in I}\sum_{s\in S} w_{is} = 1$.

The set of locations J contains the existing and potential new park locations, which are respectively referred to as $\bar{J}$ and $\tilde{J}$. As for the demand points, facility locations are associated to the centroid of the geographic region defined by an UGS. The distance from demand point $i \in I$ to park $j \in J$, d_{ij}, reflects the Euclidean distance between centroid coordinates, and is adjusted to account for approximate walking distance; this will be detailed in Chapter 4. For each location $j \in J$, the set of design options $R(j)$ is specific to it, meaning that the model encompasses a varying number of design options (referred to as "scenarios" in Aboolian et al. [2007b]) for existing and new parks. Existing parks' baseline design relates to the scenario in which no improvement is made, and maintenance is the only expense to consider. For simplicity, the set of design options equals an ordered sequence of integers, for example $R(j) = [1,\ 2,\ 3]$ signifies that three design options are available at location j, where option 1 is the baseline, option 2 is an improved and more costly option than the baseline, and option 3 is the most expensive option with the most improvement. Park improvement examples (for our case study) include the addition of park installations, sports fields or children playground, and tree planting [Ville de Montréal, 2021c]. For an existing park $j \in \bar{J}$, the associated cost of improving the park with design option $r \in R(j)$ is denoted by c_{jr}. On the other hand, for new locations $j \in \tilde{J}$, the notation c_{jr} represents the associated cost of installing a park in the location j with design option $r \in R(j)$. The optimization problem is constrained to a total budget B set by the first stage allocation.

As discussed in Chapter 2, we will follow the framework of the SIM of Huff [1964] and Nakanishi and Cooper [1974] to model the demand of parks' users. We use a slightly modified version of the general formulation of the utility function introduced in Aboolian et al. [2007b], such that

$$u_{isjr} = \frac{f(A_{sjr})}{h(d_{ij})} \ . \tag{3.2.1}$$

The difference with Aboolian et al. [2007b] is that we use parameter A_{sjr} instead of A_j, to illustrate the specific preference of the segment $s \in S$ of the population for the facility j with design option r. Consideration of the demand points' specificities in the attraction parameter is also accounted for in Nakanishi and Cooper [1974], but not in Huff [1964] which only considers a firm's facility total space. In our specific case, it seems appropriate to account for age groups, given their varying behavior towards UGSs. Indeed, in Åsa Ode Sang et al. [2016], the authors suggest that children, women and elderlies value more the importance of park given that they spend more time in environments near their home. We now introduce the exact utility function that will be used in our model formulation:

$$u_{isjr} = \frac{A_{sjr}}{(1+d_{ij})^{\beta_s}} \ , \tag{3.2.2}$$

$$A_{sjr} = \alpha_j \cdot (1+\theta_{sjr}) \ , \tag{3.2.3}$$

where $\alpha_j > 0$ is a fix parameter of the baseline attractiveness of facility location j (e.g. walking score), and θ_{sjr} can be understood as a percentage increase in the attractiveness of demand point's segment s for the design option r. The distance decay function $\frac{1}{(1+d_{ij})^{\beta_s}}$ inherits from the framework introduced by Aboolian et al. [2007b] as justified in Chapter 2, where β_s is the distance sensitivity parameter for age group s. We also introduce the parameter u^0_{is}, the utility of demand point i in segment s for the no-choice option. In the context of UGSs, we define de following function for quantifying this value:

$$u^0_{is} = \frac{\frac{1}{|J|}\sum_{j \in J} \alpha_j}{(1+d_{\text{large}})^{\beta_s}} \ , \tag{3.2.4}$$

where $\frac{1}{|J|}\sum_{j \in J} \alpha_j$ is the average baseline attractiveness of all park locations, and d_{large} is a minimum distance threshold value that is deemed too large for someone to want to visit a park (e.g., 1 km). In this way, for parks with low utility, the no-choice option is more attractive. This motivates the optimal UGS planning to install parks and improve parks' design in order to capture more demand that otherwise chooses the no-choice. It is this modeling aspect that justified the consideration of a competitive facility location formulation. Here, the competition is with the no-choice option. If we do not consider it, then 100% of the demand is covered by any feasible planning, which would be unrealistic.

We now present the model formulation using the notation and parameters introduced above:

$$\max \sum_{i\in I}\sum_{s\in S} w_{is}\cdot\left(\sum_{j\in J}\sum_{r\in R(j)} p_{isjr}\right) \tag{3.2.5a}$$

$$\text{s. t.} \sum_{j\in J}\sum_{r\in R(j)} x_{jr}\cdot c_{jr} \leq B \tag{3.2.5b}$$

$$\sum_{r\in R(j)} x_{jr} \leq 1 \quad \forall\, j\in J \tag{3.2.5c}$$

$$\sum_{r\in R(j)} x_{jr} = 1 \quad \forall\, j\in \bar{J} \tag{3.2.5d}$$

$$x_{jr}\in\{0{,}1\} \qquad \forall\, j\in J,\ \forall\, r\in R(j), \tag{3.2.5e}$$

where

$$p_{isjr} = \frac{u_{isjr}\cdot x_{jr}}{u^0_{is}+\sum_{k\in J}\sum_{t\in R(k)} u_{iskt}x_{kt}}\ ,$$

and x_{jr} is the location and design decision variable which equals to 1 if design option r is selected for park location j, and 0 otherwise. The objective function (3.2.5a) corresponds to the total park visits' frequency or, equivalently, to the total market share of the parks controlled by the neighborhood. Constraint (3.2.5b) enforces the budget restriction. Constraints (3.2.5c) imply that at most one design option is associated to a park j, while Constraints (3.2.5d) guarantee that exactly one design option is selected for an existent park $j\in\bar{J}$.

Problem (3.2.5) is an MINLP. Given that the objective (3.2.5a) is non-linear, linearization is required to ensure solvability with existent (powerful) mixed-integer solvers. Using the "Method-Based Linearization" technique reviewed in Bechler et al. [2021], we introduce the following non-negative auxiliary variables

$$v_{is} = \frac{1}{u^0_{is}+\sum_{j\in J}\sum_{r\in R(j)} u_{isjr}x_{jr}} \quad \forall i\in I, \forall s\in S\ ,$$

with v_{is} taking values in the interval

$$\left[\frac{1}{u^0_{is}+\sum_{j\in J}\sum_{r\in R(j)} u_{isjr}},\ \frac{1}{u^0_{is}}\right]\ .$$

The objective function becomes

$$\sum_{i\in I}\sum_{s\in S} w_{is}\left(\sum_{j\in J}\sum_{r\in R(j)} u_{isjr}x_j v_{is}\right) \ , \tag{3.2.6}$$

and we add the constraint

$$p_{is}^0 + \sum_{j\in J}\sum_{r\in R(j)} x_{jr}p_{isjr} = 1 \tag{3.2.7}$$

$$\iff u_{is}^0 v_{is} + \sum_{j\in J}\sum_{r\in R(j)} u_{isjr}x_{jr}v_{is} = 1 \ ,$$

where

$$p_{is}^0 = \frac{u_{is}^0}{u_{is}^0 + \sum_{k\in J}\sum_{t\in R(k)} u_{iskt}x_{kt}} \ .$$

In this formulation, we still remain with variable multiplication (bilinear terms) $x_{jr} \cdot v_{is}$. Proceeding with the following re-expression of variable multiplications,

$$z_{isjr} = x_{jr} \cdot v_{is} \quad \forall i \in I, \forall s \in S, \forall j \in J, \forall r \in R(j) \ ,$$

we obtain the resulting MILP problem:

$$\max \ \sum_{i\in I}\sum_{s\in S} w_{is}\left(\sum_{j\in J}\sum_{r\in R(j)} u_{isjr}z_{isjr}\right) \tag{3.2.8a}$$

$$\text{s. t.} \ \ (3.2.5\text{b}) - (3.2.5\text{e})$$

$$u_{is}^0 v_{is} + \sum_{j\in J} u_{isjr}z_{isjr} = 1 \qquad \forall j \in J \tag{3.2.8b}$$

$$z_{isjr} \geq 0 \qquad \forall i \in I, \forall s \in S, \forall j \in J, \forall r \in R(j) \tag{3.2.8c}$$

$$z_{isjr} \leq v_{is} \qquad \forall i \in I, \forall s \in S, \forall j \in J, \forall r \in R(j) \tag{3.2.8d}$$

$$z_{isjr} \leq K_{isjr(1)}x_{jr} \qquad \forall i \in I, \forall s \in S, \forall j \in J, \forall r \in R(j) \tag{3.2.8e}$$

$$z_{isjr} \geq v_{is} + K_{isjr(2)}(x_{jr} - 1) \ \forall i \in I, \forall s \in S, \forall j \in J, \forall r \in R(j) \ , \tag{3.2.8f}$$

where $K_{isjr(1)}$ and $K_{isjr(2)}$ are sufficiently large numbers. In this case, we will set it to the maximum upper bound of $z_{isjr} = x_{jr} \cdot v_{is} \leq v_{is} \leq \frac{1}{u_{is}^0} = K_{isjr(1)} = K_{isjr(2)}$.

We conclude this section with a brief discussion of the objective function of the second stage of the decision-making process. Given the consideration of fairness within the first stage of the problem, one could also ask about its incorporation in the second stage. Indeed, in the beginning of this research problem, we had considered two fairness schemes.

- The L_1-fairness. For the sake of simplicity, we explain this concept assuming that each demand point patronizes exactly one location or the no-choice. The goal of the L_1-fairness objective is to find an UGS plan that minimizes the sum of the absolute difference between the average traveled distance to a park (of all demand points) and the park patronized by each pair (i,s) $\in I \times S$. Alternatively, one could adapt the concept to use utilities instead of distances. We did not proceed with this objective because we observed in our preliminary tests that the model could provide a *bad* UGS plan: if the average is a large value and the distance to the park patronized by each pair (i,s) is also large, then the L_1-fairness can be close to zero. A possible solution could be to use the L_2-fairness where instead of the absolute difference to the average, we consider the L_2-norm. However, this would add more non-linearity to the model. Thus, in our experimental results, we will simply analyze the L_2-norm metric to evaluate our solutions in terms of the L_2-fairness.

- The min-max fairness[1]. As above, for simplification, we describe this concept considering that each demand point patronizes exactly one location or the no-choice. Here, the aim is to find an UGS plan that minimizes the traveled distance to visit the patronized park by the demand point (or pair) traveling the greatest distance to the selected park. Analogously, this concept could be defined in terms of the utilities. In this case, the UGS plan may have to significantly sacrifice all demand points at the expense of optimizing the least favored one. Moreover, since the proportions w_{is}, for $(i,s) \in I \times S$, can present a significant variation and that they are not considered in the min-max metric, one may question its fairness. Therefore, we do not present results optimizing this metric, but we use it to evaluate the obtained solutions in the experimental part of this work.

[1]The max-min fairness term is the standard designation. This term is defined according to the demand point benefits. Therefore, when the context analyzes instead non-benefits, such as distance, one should keep in mind that we are in fact considering a min-max.

Chapter 4

Case Study

Given the research scope to assist decision-makers in UGSs planning, we show how to apply the model formulation given in Chapter 3 with a concrete case study of the city of Montreal, Canada. To motivate this choice, we refer to Ngom et al. [2016], where the authors discuss how Montreal and Quebec City, Canada, compare with respect to UGSs accessibility using statistical learning. They conclude that Montreal displays more disparity between socioeconomic groups and argue for an improved decision process. With easy access to the city's database relating to green spaces and sociodemographic census, combined with growing discussions on how to improve sustainable public installations and public health, Montreal appears as an ideal location to test the suggested model. As underlined in "The 2021 Canadian city parks report" [Stark et al., 2021], the Covid-19 pandemic is another reason for putting forward discussions relating to parks' accessibility, and Montreal is especially concerned with important investments for its green spaces. Data preparation for the needs of this research was completed by the SPHERElab [1] team from the "Centre de Recherche en Santé Publique", a research team which specializes in projects that aim to improve populations' health with the use of urban environment interventions.

In Section 4.1, we list the original data sources, and provide an overview of data manipulations and the resulting data analysis. In Section 4.2, we detail the process of instance generation with the use of the datasets and using assumptions about UGSs usage in the city of Montreal to test the applicability of our model.

4.1. Data

We start this section with a data description including a list of the datasets provided by the SPHERElab team, followed by an overview of the data manipulation needed to obtain a final dataset with the required format and fields in Section 4.1.1. The extensive list of fields

[1] https://www.spherelab.org/

of each dataset can be found in the Appendix A. We then provide a brief data analysis in Section 4.1.2 to help the reader understand the scope and the scale of the datasets.

4.1.1. Data Description

Our case study focuses on the application of the 2SFFLD model to the planning of UGSs in Montreal. We specify here that we will exclude in the following work all the independent suburbs of the city of Montreal that have distinct budget planning, and will only consider its 19 boroughs listed in Ville de Montréal [2022a], as well as the parks managed by its boroughs. Therefore, the large parks of Montreal managed at the city-level are excluded of the optimization problem. The list of large parks can be found in Ville de Montréal [2022b]. We justify this choice by underlining that large parks and neighborhood parks have different objectives, and different usage patterns. Indeed, neighborhood parks are meant to be accessible at a short distance and for daily usage, while large parks can sometimes be accessed with cars and require longer travel time. Next, we list *(i)* the datasets that allow us to locate the demand points, either the set of postal codes, the FSA or a cluster of postal codes, and *(ii)* the parks, acting as the set of facilities in our location problem. Additional datasets resulting from data manipulations are presented for the needs of our model.

Neighborhood dataset. This dataset consists of the list of all neighborhoods in Montreal, including the suburbs. It was retrieved from the Open Data page of Montreal City's website and can be accessed in Ville de Montréal [2020]. It contains the list of IDs and name of each neighborhood with their corresponding geographic coordinates and centroid points in GEOjson format, the full land area in square meters (m^2), the 2016 population, and the population density. It has a total of 33 rows, including the 19 boroughs and the additional 14 independent suburbs.

FSA dataset. Equivalent to the neighborhood dataset, the FSA dataset contains each corresponding geographic coordinates and centroid points in GEOjson format, the zone area in m^2, the 2016 population, and the population density. Given that the FSA is one of the options to define the set of demand points in our problem formulation, we requested the age group distribution based on Statistic Canada's 2016 census. The reader is referred to Statistics Canada [2016] for the full dataset. This distribution allows us to create the demand point's segments and evaluate its corresponding weight (see Section 3.2 of Chapter 3 for details). With the intention to gather information regarding fairness, we add the Quebec's Material and Social Index at the FSA level, also referred to as Pampalon's deprivation index. Details relating to this measure can be found in INSPQ [2019]. This measure easily illustrates social and economic inequalities. Finally, the smoke pollution measure is added and equals the PM2.5 metric under the Canadian Optimized Statistical Smoke Model, and was aggregated at the FSA level by the SPHERElab team using ArcGIS tools. PM2.5 metrics, indexed to DMTI Spatial Inc postal codes, were provided by CANUE[2]. This measure is intended to approximate the pollution level of an area according to the measured concentration of fine particle in the air, and is used to establish the weighting parameters in the fair budget allocation step (see Section 3.1 of Chapter 3 for details). The full dataset consists of a total of 97 rows for all of Montreal's FSAs. The final version of the FSA dataset was achieved by adding the corresponding neighborhood for each FSA. For the specific FSAs which belong to more than one neighborhood, we select the neighborhood that covers the largest area of the FSA for simplicity.

Montreal postal code dataset. Given our intention to test multiple aggregation level for the set of demand points, we gathered equivalent information of the FSA dataset, but at the postal code level. It simply consists of a list of Montreal's 44,117 postal codes and the centroid geographic coordinates, and is queried from Service Objects' public "ZIP and Postal Code Database with GeoCoordinates for US and Canada" in ServiceObjects [2020]. As a step of validation, we replace any outlier value from this database using Python's "gmaps" package. In the final version of the postal code database, we add for each postal code the FSA age group distribution and assume that the distribution is equivalent for the FSA as for the postal code aggregation level.

Parks dataset. The data is queried from the list of the 1,368 large parks, neighborhood parks and public spaces dataset from the city of Montreal found in Ville de Montréal [2022d]. It consists of the list of park IDs and names, GEOjson format geographic coordinates and centroid points, the green spaces' area in m^2, the type of park (e.g. neighborhood park, green island, public space, pass, urban park) and the management level, such as neighborhood, private, municipal and provincial, to name a few. Additional fields are added to the parks' dataset, including the number and type of installations in each park from Montreal's list of recreational, sports and cultural outdoor installations, found in Ville de Montréal [2022c]. For the scope of our project, we also requested the Landsat Normalized Difference Vegetation Index (NDVI) and the Canadian Active Living Environment (Can-ALE) score. The NDVI is a spectral index that quantifies the level of greenness of a geographic zone according to measures retrieved through satellite images. For more details on how this index is calculated, the reader is referred to the Landsat handbook [Landsat Missions, 2022]. The Can-ALE is equivalent to the commonly known walking score, and is computed using GIS information collected through Canada. Canadian Active Living Environments Index (Can-ALE), indexed to DMTI Spatial Inc postal codes, were provided by CANUE. Both the Landsat NDVI and Can-ALE measures are added with the intention to better understand the attractiveness of the existing parks. Figure 4.1 displays a map of the city

[2]Canadian Urban Environmental Health Research Consortium

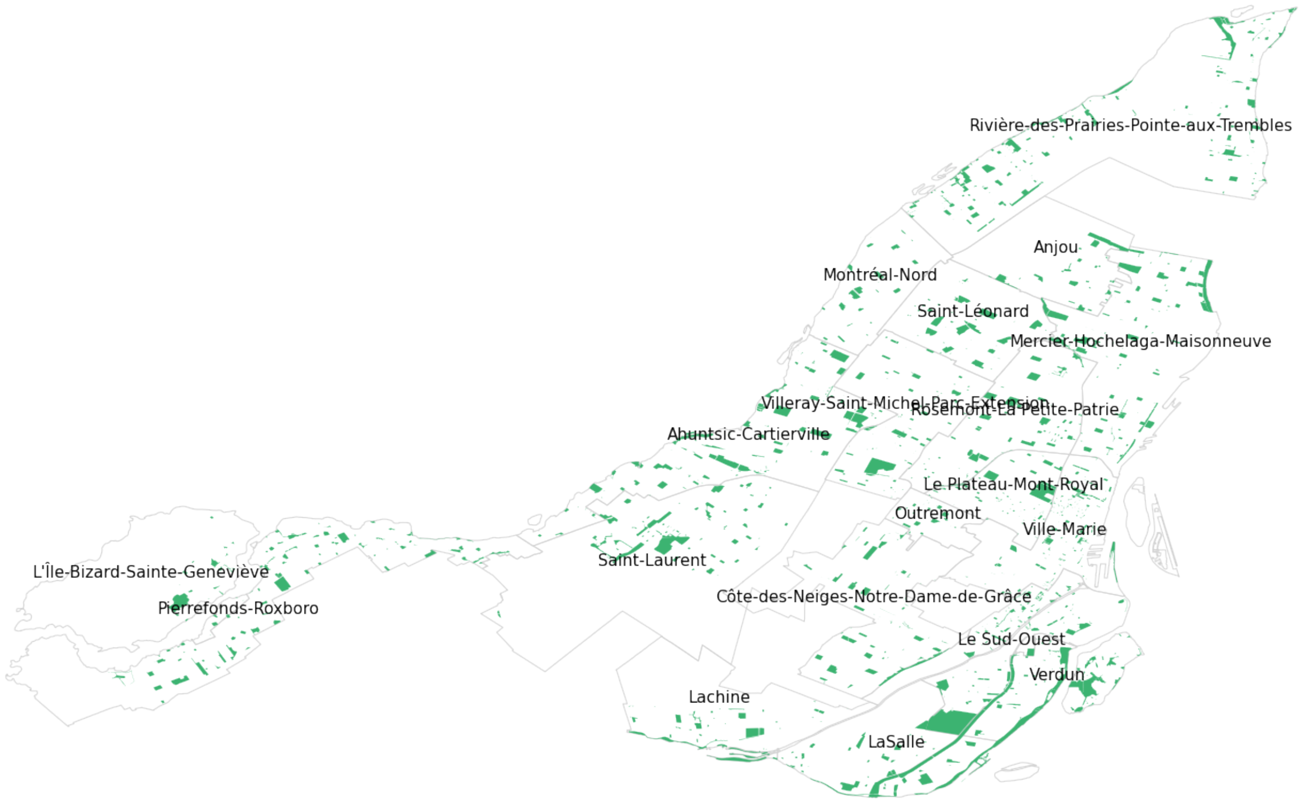


Figure 4.1. Montreal's boroughs and green spaces

of Montreal with its 19 boroughs, and the borough parks are illustrated with the green zones.

Clusters dataset. The *clusters dataset* is generated to reduce the excessive computation time resulting from using the postal codes as the aggregation level for the set of demand points. Recall the second-stage problem from Section 3.2: the size of its mathematical programming formulation depends on $|I|$; moreover, the linearization process, i.e., the introduction of the z_{isjr} variables, increases the model size, and it depends also on $|I|$. This dataset consists of a set of postal codes for each neighborhood with a predefined size resulting from the *k-means* clustering method using the latitude and longitude coordinates. This algorithm allows creating clusters of postal codes using an iterative process where each postal code is associated to the cluster with the nearest average geographic coordinates. For each neighborhood, the number of predefined clusters is given in Appendix B.

Parks & FSA dataset. This dataset consists of a merge between all possible combinations of parks and FSAs within a neighborhood. The goal of this step is to compute the Euclidean distance between parks and FSA centroids to estimate the travel time to reach a park when the FSA is used as a basis for the set of demand points. All combinations result in a dataset of 6,183 rows.

Parks & Postal Code dataset. As for the Parks & FSA dataset, the goal here is to list all the combinations of parks and postal codes within a neighborhood, and compute the Euclidean distance for each row. This yields an extremely large dataset of 2,752,628 rows.

4.1.2. Data Analysis

In this section, we present a brief data analysis for the reader to gain a better understanding of the case study of Montreal city and the scale of the dataset.

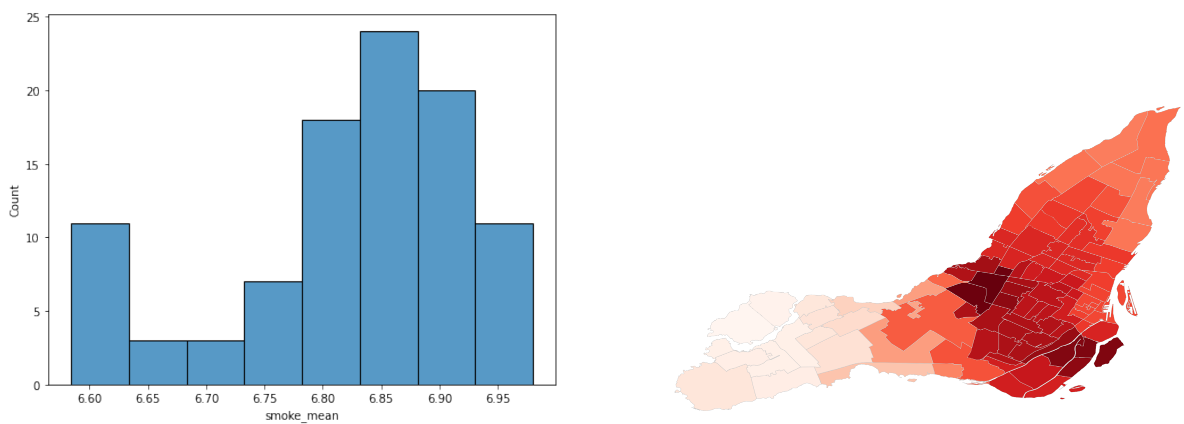
Table 4.1 lists the 19 neighborhoods of the city of Montreal that are subject to the distribution of the city budget and the optimization problem described in Section 3.2. For each neighborhood, we list the corresponding total land area in square kilometers, the population from Statistic Canada's 2016 census, the population's density per square kilometers, and the average income per household. These measures show how urban neighborhoods can vary from each other, and should be considered with the objective of making fair public decision-making.

Table 4.2 shows how the park's distribution varies according to the neighborhoods. The first column consists of the percentage of the neighborhoods' coverage with neighborhood parks, while the second column consists of the ratio of park area (m^2) per capita. One aspect to bear in mind here is that only neighborhood parks are considered, and not large parks or any other urban green spaces not managed at the neighborhood level. For this reason, some special considerations should be made in areas where very large natural parks exist, such as L'Île-Bizard-Sainte-Geneviève, Pierrefonds-Roxboro and Rivière-des-Prairies-Pointe-aux-Trembles. Such considerations will be made a posteriori when analyzing the empirical results.

We now analyze how different index and measures introduced in Section 4.1.1 vary in the city of Montreal. Figure 4.2 displays an analysis of the distribution of the smoke pollution measure, on the left side with a histogram of FSAs' index, and on the right side, a heatmap of the normalized values of the index per FSA, where darker values indicate a higher concentration of polluting particles. The histogram suggests a range of values between 6.60 and 6.95, with a peak at 6.85. The heatmap indicates that the denser and more industrial areas of Montreal have a higher risk of pollution exposure. Figures 4.3 and 4.4 illustrate the distribution of the Social and Material deprivation index at the FSA level, where higher deprivation is related to darker colors. The heatmap shows that the west side of Montreal displays more favorable values with lower deprivation indexes, while zones with high population density have a higher risk of social deprivation like in subareas of Le Plateau-Mont-Royal neighborhood. The material deprivation heatmap in its case confirms that neighborhoods with higher

Neighborhood	Area	Population	Density	Average Income
Ahuntsic-Cartierville	24.3	135,000	5,600	70,000
Anjou	13.9	42,810	3,000	69,000
Côte-des-Neiges-Notre-Dame-de-Grâce	21.5	166,000	7,700	68,000
L'Île-Bizard-Sainte-Geneviève	23.7	18,000	800	115,000
Lachine	17.9	44,600	2,500	70,000
LaSalle	16.4	77,000	4,700	64,000
Le Plateau-Mont-Royal	8.1	104,000	12,800	67,000
Le Sud-Ouest	15.8	79,000	5,000	66,000
Mercier-Hochelaga-Maisonneuve	25.5	136,000	5,300	60,000
Montréal-Nord	11.0	83,000	7,600	52,000
Outremont	3.8	24,000	6,300	175,000
Pierrefonds-Roxboro	27.2	69,000	2,500	87,000
Rivière-des-Prairies-Pointe-aux-Trembles	42.5	108,000	2,500	75,000
Rosemont-La Petite-Patrie	15.9	140,000	8,800	64,000
Saint-Laurent	43.1	98,000	2,200	81,000
Saint-Léonard	13.6	79,000	5,800	64,000
Verdun	9.7	69,000	7,100	85,000
Ville-Marie	16.0	88,000	5,500	73,000
Villeray-Saint-Michel-Parc-Extension	16.5	143,000	8,700	54,000

Table 4.1. Neighborhood data analysis: Area (km^2), population, population density per area (km^2), and average income per household ($).



(a) Histogram

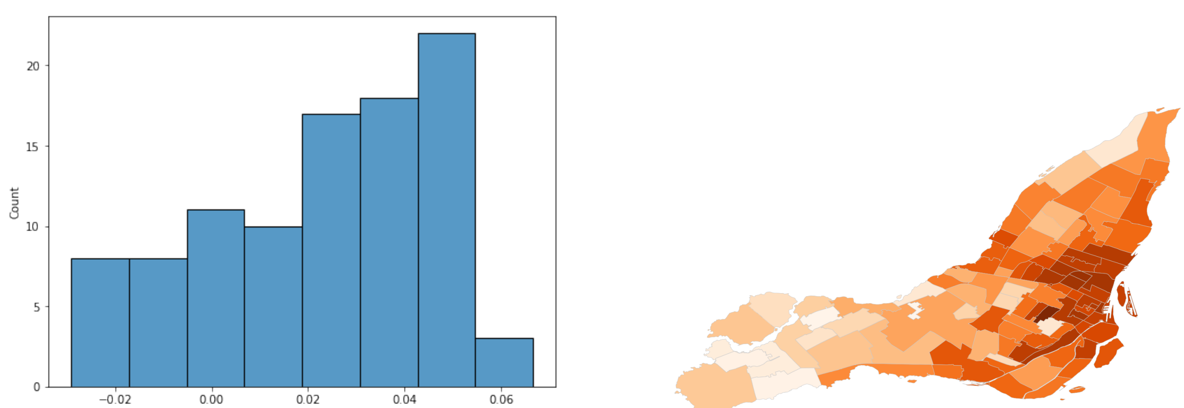
(b) Heatmap

Figure 4.2. Distribution of FSA's Smoke Index

rates of poverty result in higher material deprivation, notably in the sub-neighborhoods of Parc-Extension and Saint-Michel. Finally, Figure 4.5 indicates how the NDVI measure varies through FSAs and neighborhoods. The heatmap shows that the west side of Montreal has very large green spaces coverage, while the east side suffers from it. The urban central areas are more subject to varying level of coverage due to uneven distribution of large parks.

Neighborhood	Percentage of park area	Ratio of park area per capita
Ahuntsic-Cartierville	8.5	15.3
Anjou	3.7	11.9
Côte-des-Neiges-Notre-Dame-de-Grâce	3.7	4.8
L'Île-Bizard-Sainte-Geneviève	1.7	22.5
Lachine	4.6	18.4
LaSalle	6.5	14.0
Le Plateau-Mont-Royal	8.6	6.7
Le Sud-Ouest	12.4	24.7
Mercier-Hochelaga-Maisonneuve	7.3	13.7
Montréal-Nord	3.3	4.4
Outremont	4.5	7.2
Pierrefonds-Roxboro	5.0	19.5
Rivière-des-Prairies-Pointe-aux-Trembles	4.8	19.1
Rosemont-La Petite-Patrie	6.1	6.9
Saint-Laurent	3.1	13.4
Saint-Léonard	6.2	10.6
Verdun	23.0	32.3
Ville-Marie	3.7	6.7
Villeray-Saint-Michel-Parc-Extension	5.4	6.2

Table 4.2. Neighborhood parks statistics: Percentage of park area in neighborhood and ratio of park area (m^2) per capita.

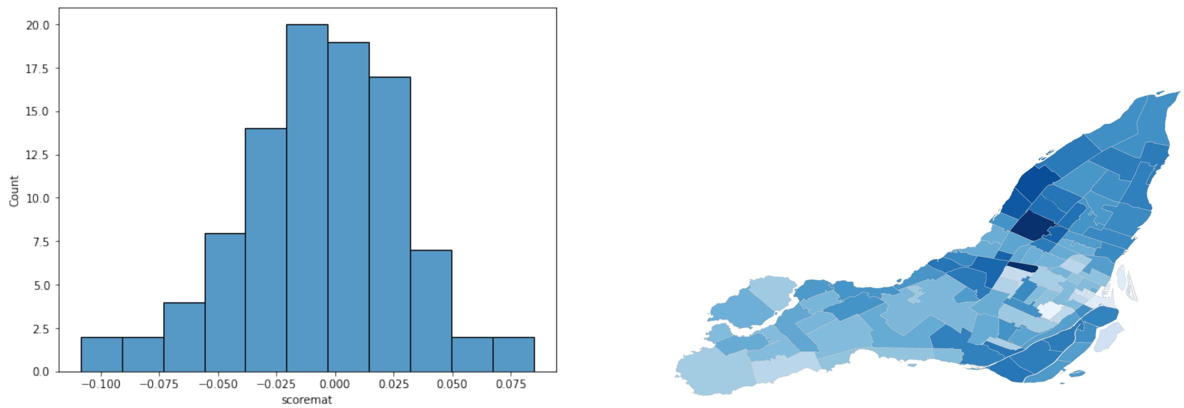


(a) Histogram (b) Heatmap

Figure 4.3. Distribution of FSA's Social Deprivation Index

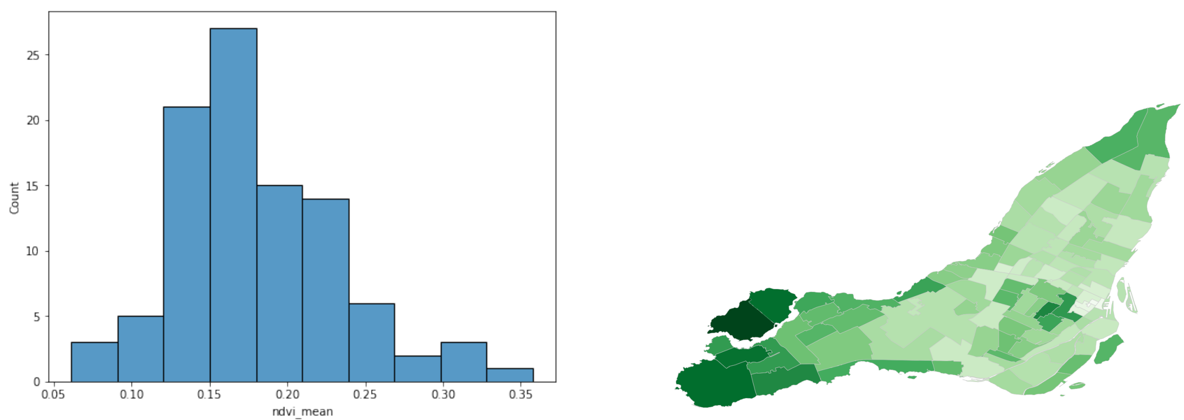
4.2. Instance Generation

In this section, we detail the instance generation process by introducing the procedure for setting the model's parameters of Chapter 3. As reflected in what follows, not all parameters of our model (2SFFLD) can be rigorously determined for our case study due to the



(a) Histogram (b) Heatmap

Figure 4.4. Distribution of FSA's Material Deprivation Index



(a) Histogram (b) Heatmap

Figure 4.5. Distribution of FSA's NDVI

lack of data. Namely, the utilities discussed in Section 3.2 should be determined through robust statistical techniques. Nevertheless, by taking advantage of the interpretation of the SIM model for the utilities, we discuss the aspects supporting the values considered in our instances. This allows us to generate instances mimicking the real-world problem topologies. In this way, the primary objective of this research, which is to suggest a model formulation to optimize the UGSs location and design, while guaranteeing solvability, can be demonstrated. On the other hand, the results from the model using the following instances, although realistic, cannot be used as a direct recommendation for urban planning in the city of Montreal. To ensure applicability of the 2SFFLD model, we suggest robust statistical methods and surveys that justify the choice of parameters.

This section is segmented such that we list how each component of our model formulation is defined. This includes in Section 4.2.1 the weighting fairness parameters of the first-stage fair budget allocation, as well as the baseline budget. It follows in Section 4.2.2 with the demand points aggregation methods, their corresponding weight, the park facilities location, the baseline budget, the maintenance costs, the design options, the utility function, and the distance of the second-stage location and design model.

4.2.1. First-Stage Parameters

Weighting fairness parameters. The weighting factors ρ_n of neighborhoods $n \in N$ are a result of multiplicative factors for different fairness considerations. The first fairness attribute ρ_{n1} is the population density. Indeed, in very dense areas, the city should plan on installing more green spaces, given the lack of private gardens. We compare each neighborhood's density to Montreal's average population density, and use this ratio as the multiplicative adjustment, with a maximum adjustment of $\pm 10\%$ to prevent unbalanced budget allocation. The second attribute is the material deprivation factor ρ_{n2} and the social deprivation factor ρ_{n3}. In the same manner as for the population density, we compare each neighborhood's level of material and deprivation index to Montreal's average, and use this ratio as a multiplicative factor with a maximum adjustment of $\pm 5\%$. We process in the same manner for the third fairness component, namely the smoke pollution factor ρ_{n4}. The resulting multiplicative factor ρ_n is set as

$$\rho_n = \rho_{n1} \times \rho_{n2} \times \rho_{n4} \ .$$

A table, with the full details of the fairness weighting multiplicative factors, is available in Appendix C.

Baseline budget. The baseline budget of the first-stage fair budget allocation model is derived from a predefined budget per capita. Therefore, the total yearly baseline budget $\bar{b}_n$ of a neighborhood n varies with respect to the population, and ensures at the very least fairness of budget between individuals. The most recent published budgets of six central neighborhoods are used to estimate the average amount invested in parks per capita for the city of Montreal: Villeray-Saint-Michel-Parc-Extension [Ville de Montréal, 2021d], Rosemont-La Petite-Patrie [Ville de Montréal, 2021b], Montréal-Nord [Ville de Montréal, 2019] and Mercier-Hochelaga-Maisonneuve [Ville de Montréal, 2018]. We justify this choice given their large size and central location, and also the availability of recent information about the investment made specifically in parks. We use the total park budget, divided by the estimated population of the corresponding neighborhood, as an estimation of the ratio of budget per capita. In this manner, we achieve a final amount of \$42 per capita. Furthermore, since parks' management has a planning horizon over many years, we presume a budget period of five years. Although this model does not yield a solution for each year, it suggests an UGS planning that can be achieved over a period of five years, which is more realistic than a one-year horizon. Hence, for each $n \in N$, $\bar{b}_n$ is determined by multiplying 42$ $\times$ 5 by the population size available in Table 4.1. Using this definition of the fixed baseline budget $\bar{b}_n$, the total available budget B_T is defined as the sum of the $\bar{b}_n$ in Equation 3.1.1 of Section 3.1 from Chapter 3. [3] Finally, the threshold δ_n, defined as the maximal budget percentage deviation from the fixed baseline budget, is set equal to $\pm 30\%$. This threshold value is arbitrary, and is set with the intention to avoid an unbalanced budget allocation between neighborhoods.

4.2.2. Second-Stage Parameters

Demand points aggregation. We start by introducing three different demand points I aggregation levels, and the final recommended choice is made according to the results of the model and the computation time presented in Chapter 5.

After careful analysis of the sociodemographic and geographic attributes of the neighborhoods and their corresponding parks' distribution and accessibility, the baseline neighborhood is set to Rosemont-La Petite-Patrie), and is used to test the different aggregation levels of the set of demand points. This borough is one of the largest of the central boroughs, having the second-largest population density in the city of Montreal, and displaying a reasonably low ratio of park area per capita. Using a baseline neighborhood of such large size, with a potential for large number of demand points, allows us to establish if the second-stage model can perform well with larger instances. In particular, it serves our purpose of identifying the right balance between demand granularity and model solvability.

The three aggregation levels considered are the postal codes, clusters of postal codes, and FSAs. Our baseline neighborhood has a total of 2,331 postal codes, but only a total of five FSAs. Given the significant increase of the number of variables with the linearization method presented in Section 3.2 of Chapter 3, the computation time resulting from the postal codes' method is expected to be unreasonably large. Therefore, we suggest the clusters' method to address the over-simplified FSA basis and the over-complex method using the postal codes. Using the cluster methodology described in Section 4.1.1, we obtain 200 clusters in the neighborhood of Rosemont, each grouping the postal codes with the geographic coordinates that are the nearest to the average coordinates of the cluster. Each cluster is presumed to be located at its centroid, namely the mean latitude and longitude of the postal codes belonging to the cluster. The number of clusters is chosen such that a significant reduction in clustering runtime is reached for a specific number of clusters. Detailed experimental results validating our cluster methodology, namely in terms of the optimal objective value sensitivity to the demand aggregation level and the computation times, are presented in Chapter 5.

[3] In the next section, a potential adjustment of the $\bar{b}_n$ value is discussed so that maintenance costs are guaranteed to be covered.

Weight of the demand points. Once the aggregation level of the demand points I is set, each demand point i is segmented in the following age groups S: 0-14 years (children), 15-64 years (teenagers and adults) and 65 years or more (elderly). We create these three age classes given an increased sensitivity to distance to parks for younger children and for the elderly. The weight w_{is} associated to each demand point i and its segment s is based on the population of each FSA from the Statistics Canada 2016 census [Statistics Canada, 2016], and we use the age distribution to estimate the population of each age group. When using the postal code aggregation level, we make the simplistic assumption that the FSA population is distributed uniformly among its corresponding postal codes.

Park (facility) location. The set of existing parks $\bar{J}$ is simply equal to the set of neighborhood parks, excluding the large parks managed by the city or other entities than the borough. The set of new parks $\tilde{J}$ is simplified by assuming a potential new location at each FSA centroid. In real-life application, the decision-maker has knowledge about locations that can be converted in UGSs, and their exact location. Our approximation allows us to cover a neighborhood almost uniformly, with a reasonable number of locations.

Maintenance costs. Regarding the approximation of the costs of park management, including planning and maintenance, we use the report on Montreal's 2020 performance indicators. In this publication, the total amount related to planning, managing and maintaining parks is established at \$80 per capita, or \$31,000 per hectare. For full access to the report, the reader is referred to Ville de Montréal [2021a]. For simplicity, we use the estimated cost per hectare and convert it to the equivalent cost of \$3.10 per m^2 of park land. For new parks, we approximate the acquisition and installation cost to a total of \$15 per m^2. This gross estimation deserves refinement, given that acquisition prices are very sensitive to location and size. For this reason, our model presumes a hypothetical price, but it can be replaced with an appropriate amount depending on the locations of potential new UGSs. As a side note, the neighborhood baseline total budget is adjusted in exceptional cases where the budget per capita is insufficient to cover the minimal maintenance costs. This scenario happens when the neighborhood has a large park coverage. To address this, we set the baseline neighborhood budget as the maximum of the budget per capita formula, and the minimal budget required for maintenance adjusted with a 1.05 factor. The decision of adding a 5% budget increase to the minimal maintenance budget is to ensure that areas in the neighborhood with lower green spaces accessibility have the opportunity for new installations.

Design options. The number of design options for each location $j \in J$ is fixed to a number of three alternatives, whether it is an existing park or a new park location. In real-life applications, urban planning involves a varying number of design options according to each location, with an estimated budget according to each scenario. Here, we suggest a more simplistic approach. The associated cost c_{jr} for each design option is defined as follows:

$$c_{jr} = c_{\text{maintenance}} \times area_j \times (1 + 0.8 \cdot (r - 1)),$$

where $c_{\text{maintenance}}$ is the maintenance cost of \$3.10 per m^2, $area_j$ is the size of the park in m^2, and $(1 + 0.8 \cdot (r - 1))$ is a multiplicative factor that increases the total cost of each design option with an additive percentage of 80%. For new parks located at the FSA centroid, we presume a land area of 50,000 m^2, which is equivalent to a large neighborhood park. As a reminder, the first design option for an existing park is the basic maintenance scenario where no improvement is made, while for new parks, the first scenario corresponds to the smallest investment option. If no new park is to be located at the location $j \in \tilde{J}$, then the variable x_{jr} would be equal to 0. For existing parks, constraints of the model formulation ensure that at least one design option is selected.

Utility function. Now, we define the parameters of the model specific to the SIM. We remind the reader of the following utility function for each $i \in I, s \in S, j \in J, r \in R(j)$:

$$u_{isjr} = \frac{A_{sjr}}{(1 + d_{ij})^{\beta_s}} \ ,$$

where

$$A_{sjr} = \alpha_j \cdot (1 + \theta_{sjr}) \ .$$

The basic attraction level α_j is set equal to the ALE score (see Section 4.1.1 for the definition) of location $j \in \bar{J}$ for already installed parks, and to the ALE score of FSA j for a new park $j \in \tilde{J}$. The attraction increase θ_{sjr} for existing parks $j \in \bar{J}$ is set equal to a percentage increase in line with the suggested design options:

$$\theta_{sjr} = 0 \text{ for } r = 1, \quad \theta_{sjr} = 0.5 \text{ for } r = 2, \quad \theta_{sjr} = 1 \text{ for } r = 3.$$

As noted above, the attraction increase for design options $r \in R(j)$ is fixed for all age segments $s \in S$ and facility points $j \in J$, but the model allows for more realistic parameters. Indeed, different age segments should have different perception of different design options and type of parks. Given our lack of real-life data, we suggest this simple attraction increase assumption. Furthermore, we reiterate that θ_{sj1} is set equal to 0 because it is associated to the maintenance option $r = 1$. The attraction increase for a new park j is set equal to:

$$\theta_{sjr} = 0.75 \text{ for } r = 1, \quad \theta_{sjr} = 1.5 \text{ for } r = 2, \quad \theta_{sjr} = 3 \text{ for } r = 3.$$

As noted, we presume a large attraction increase for newer parks, given individuals' preference for new installations. This assumption can be refined to account for long term decrease of attraction.

The distance sensitivity parameter β_s is set according to different age groups. This parameter should be estimated using robust statistical methods introduced in Huff [2021], but are roughly approximated in our research, given the unavailability of data on park usage by the population. For the children and elderly age groups, we set β_s to 1.5, and to 1.0 for teenagers and adults. Indeed, as suggested earlier, children have a higher sensitivity to distance given that park visits are usually supervised by parents, while elderly require closer facilities given reduced mobility.

Finally, since our optimization model presumes non-negative utilities, to prevent errors in the solving step, we shift all utilities u_{isjr} to positives values. Indeed, negative utility values can exist due to negative Can-ALE scores. Thus, to maintain the scale of utilities, we simply shift them instead of normalizing them. Furthermore, we define the no-choice utility u_{is}^0 as defined in Section 3.2 of Chapter 3, and use α_j and β_s's definition above.

Distance. The distance d_{ij} from demand point $i \in I$ to existing parks $j \in \bar{J}$ is calculated using the Euclidean distance formula, adjusted with a multiplicative factor of 1.3 to approximate the actual travel distance. This adjustment factor is derived using an average of the Google Distance Matrix distances divided by the corresponding Euclidean distances for a set of points. For new parks, the distance from the demand point belonging to the same FSA as the park is assumed equal to 500 meters, and 1000 meters for demand points outside the FSA where the park is located. In real-life application context, one needs to compute the actual travel distance for the selected demand point to the exact new park location. Given that the location of new park is set at the centroid of the FSAs for simplicity, we opt for this assumption. Next, given the context of the UGSs planning, we define a maximum distance threshold. Indeed, according to the literature, neighborhood parks should be located close enough to residential areas so that individuals have a minimum incentive to visit them. A discussion about this subject can be found in City Parks [2017]. For this reason, we set the maximum distance to parks for children at 500 meters. For all other age groups, maximum distance for parks less than 50,000 m^2 is set to 500 meters, and to 800 meters for larger parks. Indeed, people have more incentive to walk longer for larger UGSs. In the model, we set the utility u_{isjr} to 0 for distances above these predefined thresholds.

Chapter 5

Computational Experiments

We now present the computational results from the implementation of the 2SFFLD model described in Chapter 3 using the case study of the city of Montreal detailed in Chapter 4. Note that the instances used for the implementation of the model were set using the limited data available as described in Chapter 4; however, these should be derived using statistical methods and adapted according to the context for which the model is being used. Here, we do not prescribe an UGSs planning for the city of Montreal, but rather present a flexible two-stage optimization model and prove its applicability and performance using a large real-life based instance. For this reason, we use the resulting objective function value and the computing time to compare different demand point aggregation methods and the budget allocation methods, but make no claims about the reliability of the actual results to the city of Montreal. However, we expect our framework to be useful for practitioners and to promote the discussion on the use of optimization tools for city decision-making. We further discuss the potential future and practical impact of our work in the last chapter of the thesis.

We start this chapter with a layout of the experimental setup, describing the implementation details of our framework in Section 5.1. Section 5.2 shows how the different aggregation methods compare in the selected baseline neighborhood, and we recommend a clustering methodology to improve performance and accuracy of the results. We finalize with Section 5.3 and discuss how the city-level UGS distribution compares under the baseline and the fair budget allocation methods.

5.1. Experimental Setup

In this section, we review the experimental setup to compute the results presented in the following sections. To start, we note that our methodologies are implemented in Python 3.9.7. Optimization models are solved with Gurobi 9.5.0 using 6 cores. The experiments run on a Dual-Socket Intel(R) Xeon(R) Gold 6226 clocked at 2.70GHz (12 Cores per Socket, 24 Cores total) and is equipped with 376 GB RAM. Below, we describe the predefined model's parameters, and how to interpret the different model's attributes. The details below are specific to the second-stage optimization, given that the first-stage consists of a fairly simple linear program, and it does not require additional model parametrization.

First, we introduce the different model's parameters as defined in Gurobi [1]: the pre-solve level, the degenerate simplex moves limit, and the time limit. For the clusters and the postal codes' methodologies, namely the larger instances, the pre-solve level is set to the option 2, the maximum value available, which implies a much longer and complex pre-solving step but whose preliminary results indicated to compensate due to tighter final optimality gaps. The degenerate simplex moves limitation parameter is set to 0, as recommended in Gurobi for problems resulting in an important number of *Total elapsed time* messages in the log. These choices were made to speed up the computing time and follow Gurobi's guidelines in this context. We also set the time limit for solving the UGS planning in each neighborhood to 3,600 seconds (one hour) for both the FSA and the clusters' methodology, and to 18,000 seconds (five hours) for the postal codes' method. The reason for these time limits is to ensure that the application of our solution can easily be used in a context of city-decision making and that multiple tests can be done in reasonable time. The comparison of these methods is addressed in the next section. These different time limits according to the demand aggregation level is simply to adapt the computing time depending on the problem size, although it is known that the time to solve an NP-hard problem is expected to increase exponentially with its instance size. In Section 5.3, we use the clusters' methodology for the aggregation of the demand; the number of clusters used in each neighborhood (i.e., instance) is available in Appendix B.

In the next sections, Gurobi's key model's attributes[2] are discussed to compare methodologies: the runtime, the model status, the GAP, and the objective value. While the first-stage model is a simple linear program, the second-stage program solved for each neighborhood is a mixed-integer linear program, which can be slow to solve. Regarding runtime and model status, for smaller neighborhoods, Gurobi's solver is able to reach an optimal solution with a small runtime, but larger instances usually reach the predefined time limit. In the latter case, we simply retrieve the best solution found by the solver within the time limit. The resulting model status encountered in our experiments are presented in Table 5.1 and are referred to in the results' discussion. The GAP is defined as the absolute value of the gap percentage between the objective value reached by the best feasible objective value found so far and the tightest bound computed, and serves as a good indicator of the improvement of the solution at each iteration over the branch-and-bound process. Therefore, when the solver reaches optimality, the final GAP is 0, while in the opposite case, it is strictly greater than 0; the latter will occur for runs where the time limit is reached. Finally, the discussed

[1] https://www.gurobi.com/documentation/9.5/refman/parameter_descriptions.html
[2] https://www.gurobi.com/documentation/9.5/refman/attributes.html

Status code	Value	Description
OPTIMAL	2	Model was solved to optimality (subject to tolerances), and an optimal solution is available.
TIME_LIMIT	9	Optimization terminated because the time expended exceeded the value specified in the TimeLimit parameter.

Table 5.1. Gurobi model status code definition

objective values relating to Objective (3.2.5a) and discussed in the following sections should be interpreted as the percentage of the population that is using parks. Therefore, the aim is to be close to 100%.

In what follows, we present tables of results. The meaning of the columns in these tables are as follows: "Status" provides Gurobi status, "GAP (%)" provides the optimality gap as a percentage, "RunTime (s)" provides the time in seconds to solve an optimization problem, "ObjVal (%)" provides the value of the best objective computed by the solver.

5.2. Enhancing the accuracy and scaling of the second-stage optimization

Here, we detail the different aggregation level methods of the demand points to show how it impacts the model's implementation performance in the neighborhood of Rosemont, and recommend an effective clustering method to improve the performance and the total runtime. The different aggregation methods are tested using the baseline budget described in Section 4.2.1 at the neighborhood level, while the fair budget allocation method is discussed in Section 5.3 at the city-level. Given that we impose a maximum deviation of the fair budget per neighborhood in relation to the baseline, the conclusions on the computational performance presented here are not expected to differ if we use the fair budget. As underlined in Section 4.2.2, the selected neighborhood of Rosemont is chosen for the comparative analysis given its central location, its very large residential area and a significantly low park ratio per capita.

The first demand aggregation we used is the FSA level, given its use by Statistics Canada for the population's census. As mentioned in previous sections, the FSA consists of the first 3 characters of the postal code, which consists of six characters. The first character refers to the province, territory or region of residence, and corresponds to "H" for the Metropolitan region of Montreal. The second character indicates whether the region is urban or rural, and any number different from 0 is associated to an urban region. The third character identifies, with the first 2 characters, the unique subarea of the metropolitan region of Montreal. In the neighborhood of Rosemont, there are only five FSA and a total of 55 neighborhood parks. This suggests that the computed traveling distance from the centroid of the FSA to the

Method	Status	GAP (%)	RunTime (s)	ObjVal (%)
FSA	2	0.0	4.00	84.0
Postal Code	9	14.8	18,000.00	75.7
Postal Code with FSA solution				75.2
Postal Code with cluster solution				82.9
Clusters	9	2.4	3,600.00	83.4
Clusters with FSA solution				75.1

Table 5.2. Model results for Rosemont-La Petite-Patrie with baseline budget

parks and the assumed preference of the demand points are extremely approximate, should this aggregation level be used. Indeed, this method mistakenly attributes more preference to the parks near the centroid of each FSA and neglects parks closer to the FSA's zonal delimitations. Therefore, postal codes further from the center of the FSA are disadvantaged. Nevertheless, this method is a decent starting scenario and is helpful to compare more granular methods discussed below. Table 5.2 displays the performance of the solver at the FSA level. The solver reaches an objective value of 84% with a 0% GAP.

With the aim to obtain more accurate results, we originally tested the postal code methodology, which consists of the smallest and most granular aggregation level available to us. Figure 5.1 illustrates the comparison of the number of demand points for the FSA and the postal code method in Rosemont. Our first attempt at running the model using the postal code methodology generated a very long runtime. The attempt was to obtain preliminary results according to this methodology to eventually select an appropriate aggregation method for the final model suggestion. For this reason, we have set the time limit to five hours, which resulted in an objective value of 75.7%, with a GAP value of 14.8%. To compare this result to the FSA methodology described above, we test how the FSA solution performs under the postal code aggregation level objective. The result is shown in Table 5.2, with an objective value of 75.2%. This suggests a potential overestimation done by the FSA aggregation level objective, and that the postal code method increases the percentage of population's park visits by 0.5% compared to the FSA solution, which should be improved with a larger runtime limit or with an improvement of our methodology as follows.

To address the unreasonable runtime resulting from the postal code method, and to ensure that the suggested model in this work can be applied to realistic problems while obtaining a good-quality estimate of the expected probability of individuals visiting parks, we recommend a clustering method of the demand points. Indeed, aggregating the postal codes to clusters using the *k-means* method on the geographic coordinates can help to reduce the problem size significantly, such that conclusive results are reached under a reasonable runtime. This method iteratively computes the centroid for each of the k clusters, and assigns each point, in this case each postal code, to the nearest centroid based on the geographic coordinates. The centroid of each cluster is defined as the mean of the coordinates of all points assigned to the cluster, which is why it is called the *k-means* method. The algorithm ends when optimality is reached, meaning when all points are in the cluster with the nearest average. As a reminder, we assign a number of 200 clusters to the neighborhood of Rosemont, given its total of 2,331 postal codes. Figure 5.2 illustrates how the cluster methodology compares to the postal code method, which suggests that this demand aggregation still displays a fair precision in a context of urban planning. After testing this method, we achieve a high expected probability of individuals visiting parks within one hour.

Although several papers mentioned in the literature review of the FLPs in Section 1.2 address the NP-hard component using heuristic algorithms, we instead suggest a slight reduction of the model's accuracy of demand points' location with the postal code model by using clusters, which remains a very precise aggregation method such that it covers uniformly the neighborhood, as shown in Figure 5.2. To justify this conclusion, Table 5.2 shows the clustering method results. This method yields an objective value of 83.4% in one hour, while the postal code methodology reaches an objective value of 82.9% using the clusters' solution.

Therefore, we manage to increase the postal code's result from 75.7% to 82.9% using a methodology that requires one hour instead of five hours. As a validation step, the cluster aggregation level objective yields a value of 75.1% using the FSA solution, which emphasizes the lack of accuracy of the FSA method. Therefore, our final model recommendation not only finds its strength in its performance, which is a direct consequence of the demand points aggregation methodology, but also in its application to a realistic instance size. The next section compares how the objective value varies for each neighborhood for the baseline budget and the fair budget allocation method under this clustering framework.

5.3. Implementing the first-stage optimization to improve fairness at the city-level

We now discuss how the fair budget allocation, determined in Section 3.1, compares with the baseline budget allocation. Namely, *(i)* we analyze how the available budget per neighborhood affects the optimal value of the second-stage objective, *(ii)* we discuss the performance of our approach, *(iii)* we demonstrate that the adoption of a SIM goes beyond the use of distance as an element to guide UGS planning through L_1-norm metric, and *(iv)* we discuss the importance of properly estimating the no-choice utility. As a reminder, the baseline budget method is based on a fixed budget for green spaces per capita ratio, with a minimum neighborhood budget set to cover maintenance expenses with an addition of 5% budget for flexibility. For more details, the reader is referred to Section 4.2. The fair budget allocation aims to improve the UGSs distribution at the city-level by accounting for the varying population's density, the social and material deprivation index and the pollution

Figure 5.1. Rosemont-La Petite-Patrie FSA versus Postal Code aggregation

level. Indeed, fairness does not necessarily yield equal results, such that neighborhood requiring more investments are heavily advantaged by such a method while neighborhoods with sufficient green coverage should at least be able to cover for the maintenance expenses of their parks. Tables 5.3 and 5.4 show how the baseline budget method performance compares with the fair budget allocation methodology.

Objective value. First, the neighborhoods that benefit the most from a budget increase with the fair budget allocation method include Côte-des-Neiges-Notre-Dame-de-Grâce and

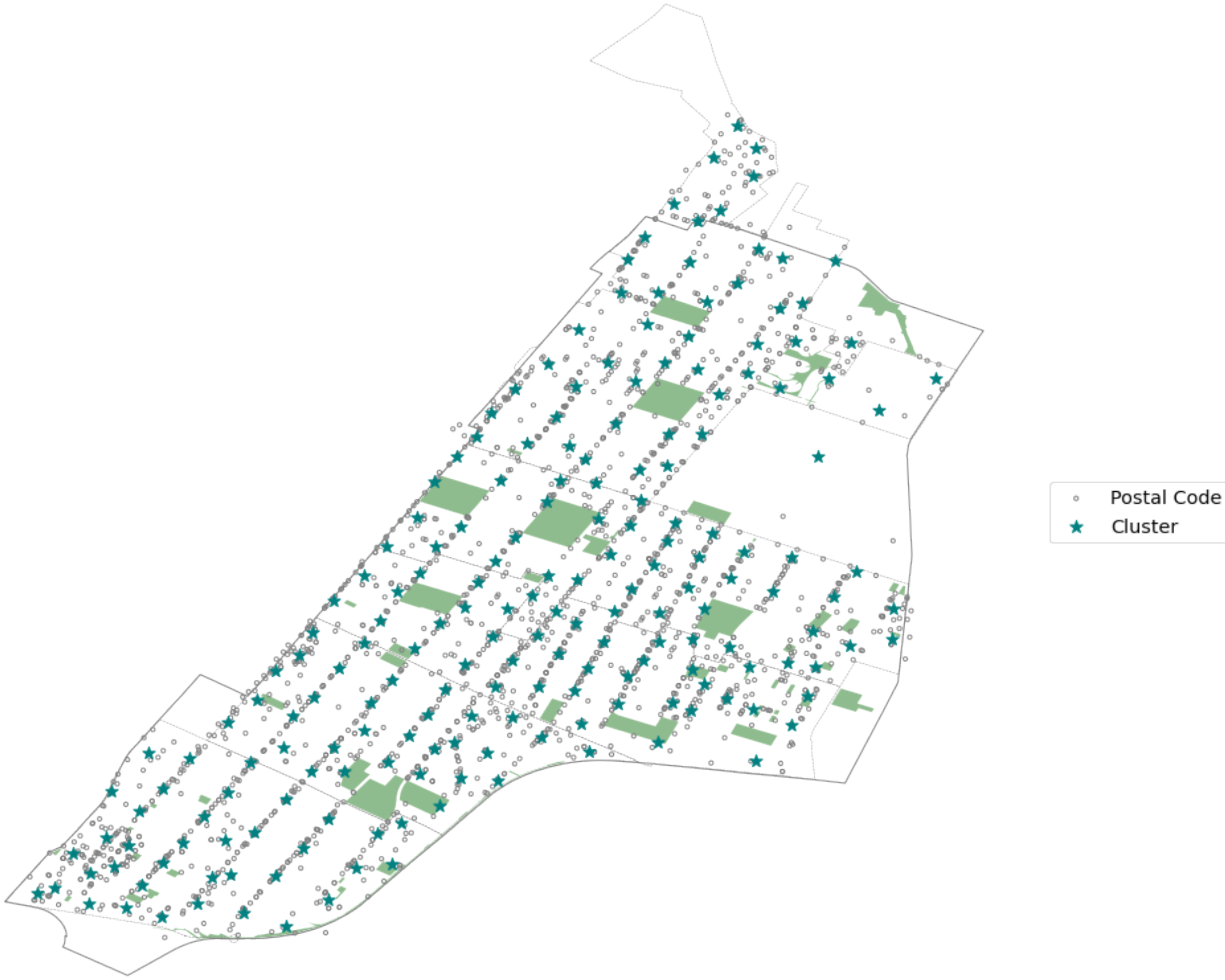


Figure 5.2. Rosemont-La Petite-Patrie Cluster versus Postal Code aggregation

Villeray-Saint-Michel-Parc-Extension, with an objective increase of 2.1% and 1.2% respectively. These boroughs are deemed very large residential areas, and differ from other neighborhoods mostly due to higher population density. We underline one of the model's shortcomings here, in that we ignore the large parks of the city of Montreal in our problem. Indeed, although these neighborhoods do display a lack of borough's parks when compared to wealthier neighborhoods such as Outremont or more remote and greener neighborhoods such as Pierrefonds-Roxboro or L'Île-Bizard-Sainte-Geneviève, they do have access to large parks including Frédéric-Back Park, Jarry Park and Mont-Royal Park to name a few. In Figure 5.3, we display the large parks that are managed at the city-level and that are excluded of the optimization problem.

Neighborhood	Budget (M$)	Status	GAP (%)	RunTime (s)	ObjVal (%)
Ahuntsic-Cartierville	33.3	9	0.7	3,600.00	99.0
Anjou	9.5	2	0.0	29.00	93.8
Côte-des-Neiges-Notre-Dame-de-Grâce	27.3	9	2.5	3,600.00	82.1
L'Île-Bizard-Sainte-Geneviève	6.4	2	0.0	12.00	95.9
Lachine	10.4	2	0.0	1,880.00	97.5
LaSalle	17.2	9	1.2	3,600.00	97.3
Le Plateau-Mont-Royal	18.6	9	1.2	3,600.00	94.4
Le Sud-Ouest	17.7	9	0.1	3,600.00	99.0
Mercier-Hochelaga-Maisonneuve	28.9	9	0.7	3,600.00	99.0
Montréal-Nord	17.8	2	0.0	1,313.00	66.7
Outremont	6.6	2	0.0	28.00	84.3
Pierrefonds-Roxboro	19.1	9	0.5	3,600.00	98.5
Rivière-des-Prairies-Pointe-aux-Trembles	33.1	9	0.0	3,600.00	99.4
Rosemont-La Petite-Patrie	29.0	9	2.4	3,600.00	83.4
Saint-Laurent	22.4	9	0.3	3,600.00	98.4
Saint-Léonard	15.7	2	0.0	3,068.00	63.0
Verdun	36.2	2	0.0	2,457.00	70.9
Ville-Marie	18.8	9	3.7	3,600.00	92.7
Villeray-Saint-Michel-Parc-Extension	30.4	9	3.0	3,600.00	83.1
Average					89.2

Table 5.3. Model results using the clusters method and baseline budget

Neighborhood	Budget (M$)	Status	GAP (%)	RunTime (s)	ObjVal (%)
Ahuntsic-Cartierville	33.3	9	0.7	3,600.00	99.0
Anjou	8.3	2	0.0	90.00	93.5
Côte-des-Neiges-Notre-Dame-de-Grâcece	35.5	9	1.7	3,600.00	84.2
L'Île-Bizard-Sainte-Geneviève	6.4	2	0.0	12.00	95.9
Lachine	10.4	2	0.0	1,877.00	97.5
LaSalle	12.1	9	0.8	3,600.00	97.2
Le Plateau-Mont-Royal	24.2	9	0.8	3,600.00	94.9
Le Sud-Ouest	16.2	9	0.6	3,600.00	99.0
Mercier-Hochelaga-Maisonneuve	26.3	9	0.7	3,600.00	99.0
Montréal-Nord	12.4	2	0.0	1,285.00	63.8
Outremont	4.6	2	0.0	32.00	80.7
Pierrefonds-Roxboro	19.1	9	0.5	3,600.00	98.5
Rivière-des-Prairies-Pointe-aux-Trembles	33.1	9	0.0	3,600.00	99.4
Rosemont-La Petite-Patrie	33.3	9	4.1	3,600.00	83.2
Saint-Laurent	21.4	9	0.7	3,600.00	98.4
Saint-Léonard	13.0	2	0.0	625.00	56.9
Verdun	36.2	2	0.0	2,164.00	70.9
Ville-Marie	13.1	9	1.9	3,600.00	91.6
Villeray-Saint-Michel-Parc-Extension	39.5	9	1.6	3,600.00	84.3
Average					89.0

Table 5.4. Model results using the clusters method and fair budget

As a result, the suggested model is deemed sufficiently flexible to include these considerations should it be done. Again, the purpose of this work is not to make an UGS planning recommendation, but to prove the applicability and the performance of our methodology. Alternatively, the neighborhoods displaying the largest sacrifice under the fair budget allocation are Montréal-Nord, Outremont and Saint-Léonard, with a respective objective value decrease of 2.8%, 3.7% and 6.1%. These neighborhoods have the largest budget decrease, and suffer from a budget cut at the cost of multiple neighborhoods which are already at the minimum tolerable budget to cover the maintenance cost, even though they

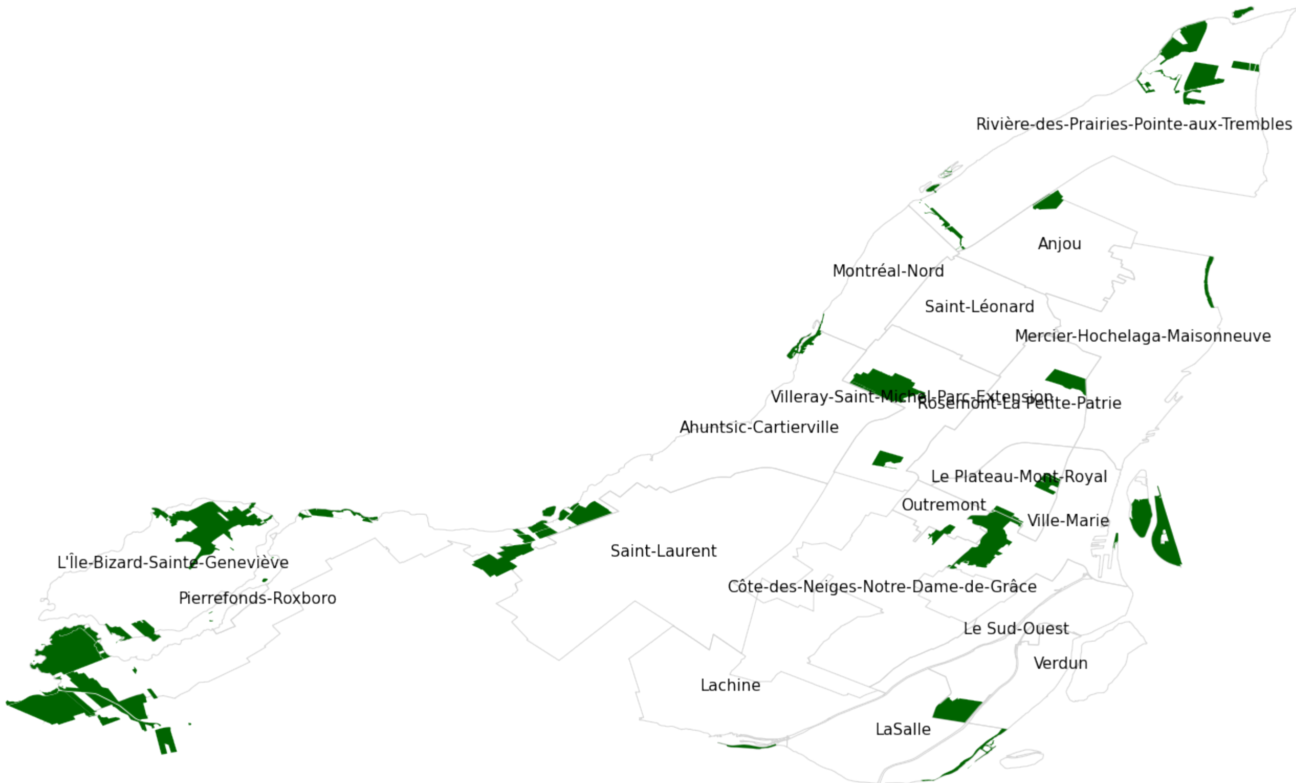


Figure 5.3. Large parks managed by the city of Montreal

should be penalized with a fair budget allocation.

We further question the model as to how does the fair allocation method impacts the total probability of individuals visiting parks. Tables 5.3 and 5.4 display the average objective value weighted with the population of each neighborhood. The average totals to 89.2% for the baseline budget, and to 89.0% for the fair budget, which suggests a slight decrease of the overall objective under the fair allocation method. This should be expected given that the budget is re-allocated to neighborhoods to advantage the areas with higher deprivation, which can be addressed with an improved green spaces coverage. Therefore, this can have the impact of reducing the objective value of greener neighborhoods at the cost of the most deprived neighborhoods.

Performance. Regarding performance, around 40% of the second-stage optimizations are solved to optimality within the time limit, and the highest optimality gap is 4.1%. The most demanding neighborhoods with respect to total runtime are Ahuntsic-Cartierville, Mercier-Hochelaga-Maisonneuve and Ville-Marie, which have the largest number of postal codes, and thus, of clusters. We also observe no monotonic relation between the budget and the optimal GAP (a proxy of the problem difficulty). For instance, the budget increase of Rosemont-La Petite-Patrie results in a larger GAP but the budget increase of Le Plateau-Mont-Royal results in a lower GAP. This is logic since the size of our mathematical programming model is not affected by the budget and thus, an increase in the budget does not necessarily increase the time to solve the problem. Indeed, very low budgets result in small sets of feasible solutions and high budgets result in very large feasible sets; these two extremes are likely to result in easier to find solutions while "medium" budgets complexify the search for the optimal solution.

L_1-Norm. We now analyze how the fair allocation affects the average distance from the demand points to the park locations from the final solution. To this end, we use the L_1-*Norm* on the expected distance from individuals' to the surrounding parks using the probabilities p_{isjr} defined in Section 3.2. The values of this metric per neighborhood under the baseline and the fair budget allocation method are available in Appendix F. Lower values of the L_1-*Norm* would suggest an increase of fairness, with lower deviation from the average distance to parks. These results actually show that for some neighborhoods with an increase of budget when using the fair allocation, such as Côte-des-Neiges-Notre-Dame-de-Grâce, Le Plateau-Mont-Royal and Villeray-Saint-Michel-Parc-Extension, we observe an increase of the L_1-*Norm*. This counterintuitive observation shows that the use of utilities instead of distances to model the preferences of the population significantly changes the nature of the problem. As an example, Outremont is found with a significant decrease of budget with the fair allocation, but also results in a lower L_1-*Norm* measure. This stretches the importance of the definition of the neighborhoods' objective in the decision process and the (unsurprising) sensitivity of the solution according to the modeling of park visits.

No-choice option. We complete this discussion by reminding the impact of the quantification of the no-choice utility parameter u^0_{is} on the scale of the objective value. Indeed, the model's results show multiple neighborhoods with significantly high probability values near 100%, including in Ahunstic-Cartierville, Le Sud-Ouest, Mercier-Hochelaga-Maisonneuve and Rivière-des-Prairies-Pointe-aux-Trembles. Although these residential areas can evidently display a higher propensity for park visits, this could also be a consequence of the underestimation of the no-choice utility and this, reflects the importance of parametrizing a model with proper data estimation. To support this idea, Table 5.5 exemplifies how the objective value varies with respect to different no-choice utility in the neighborhood of Outremont (with the baseline budget). We chose this neighborhood (instance) since we can compute optimal second-stage solutions to it within the time limit and, thus, properly evaluate the sensitivity to the no-choice utility. The sensitivity analysis is made by applying a multiplicative factor to the no-choice utility parameter, u^0_{is}, $\forall\ i,s\ \in\ I,S$ as indicated in Table 5.5.

Multiplicative factor	ObjVal (%)
110%	83.4
100%	84.3
90%	85.3
80%	86.4
70%	87.5

Table 5.5. No-choice utility sensitivity in the neighborhood of Outremont

Final remarks. For a sample solution, Appendix D and Appendix E, respectively, display the design options selected for the existing and new parks, under the baseline and the budget allocation method in the neighborhood of Rosemont. As noted in these results, new parks hypothetically located at the centroid of the FSAs are also subject to change of design options when subject to a change of neighborhood budget. Based on the previous results, we can safely suggest that the model yields reasonably realistic results, even though the parameters' are based on assumptions from existent statistical indexes. Refining such assumptions, and thus, parameters estimation, should improve the credibility of urban planning recommendations based on the model's results.

Chapter 6

Conclusions and Future Work

In this work, we proposed a two-stage fair facility location and design model to assist public decision-makers concerned with urban green spaces planning. The two-stage formulation allows us to encompass fairness considerations in the first-stage through the budget allocation among neighborhoods, based on fairness measures and a maximal baseline budget deviation. The second-stage of the problem is a special case of the facility location problem, and illustrates realistic considerations of urban planning by including a choice model for users' behaviors and using justified assumptions to define the underlying parameters. In particular, through the use of choice models, the second-stage objective mimics the likelihood of visits to parks, representing a significant improvement over the traditional distance minimization objective. We have further proved that a competitive facility location problem can be extended to non-competitive frameworks by re-defining the no-choice option as the element to compete against. We used linearization techniques to obtain a second-stage mixed-integer linear program that can be solved with standard solvers, while increasing the complexity of traditional competitive location problems with the inclusion of design options.

Next, we generated instances of our two-stage fair facility location problem based on data available about the city of Montreal, allowing us to apply our model. Due to the large second-stage problems for the boroughs of Montreal, we devised a clustering method with the intent to reduce the size of the second-stage mixed-integer linear programs, while ensuring the performance and applicability of the resulting solution. Indeed, this methodology proves sufficient accuracy in a context of urban green spaces design and location problem.

Through the case-study of the city of Montreal, we show the applicability of our model, discussing the effects of budget allocation and parks coverage. This work is a first step towards the development of decision-aid tools for the planning of UGSs by cities.

One of the main shortcomings in our case-study comes from the lack of use of statistical techniques for estimating the parameters of the choice model. For this reason, we recommend a model calibration using GIS methodologies as suggested in Huff [2021] to improve the accuracy of the model's solution. Nevertheless, we were able to prove the performance of our model under reasonable time constraints, and obtained results that are intuitive given then underlying parameters' assumptions. Refining such parameters should therefore only improve the solution's accuracy. Another shortcoming of the model comes from the consideration of long-term planning. Indeed, we considered a static model using a total budget estimated for a period of five years, but we recommend an extension of this model to yield a budget planning per year, and potentially account for a dynamic response at each period, i.e., a change on demographics and also on the utilities of the SIM.

Regarding future works, we recommend a thorough theoretical characterization of the clustering methodology. Improving the process for defining the exact number of clusters required to achieve optimal results under an acceptable time limit and gain knowledge on the trade-offs between the exact method and the clusters' method should significantly improve the reliability of the results. We also suggest a comparison of the different decision-making structures. Indeed, in the City of Montreal, decisions are taken in a two-stage fashion. It would be interesting to compare this methodology to the alternative where the city also takes the second-stage decisions; in this way, we could see the advantages and limitations of a coordinated second-stage. This method should increase substantially the number of variables of the model, which could be addressed with heuristics or a more aggregate clustering approach than the aggregation level used in this thesis.

An interesting direction for this work could also be to integrate green paths in the model. These paths are intended to connect parks to each other and display a sufficient amount of trees and vegetation to be considered as a "green" path. Another addition is to account for blue spaces, defined as the space allocated to water bodies or watercourses. In the 2021 WHO report "Green and blue spaces and mental health: new evidence and perspectives for action", green spaces are considered simultaneously with blue spaces as a mean to address more recent concerns about climate change and mental health [Braubach et al., 2021].

We are hopeful that this work can prove its application in the city of Montreal, and that it can contribute to the stream of literature that puts forward methodologies for applying theoretical concepts to the benefit of public health and environmental initiatives. We further underline the great importance of encouraging multidisciplinary work to achieve solutions in response to public challenges, while maximizing the use of the substantial amount of available data, and consequently aim to improve the collective quality of life.

References

Robert Aboolian, Oded Berman, and Dmitry Krass. Competitive facility location model with concave demand. *European Journal of Operational Research*, 181:598–619, 09 2007a.

Robert Aboolian, Oded Berman, and Dmitry Krass. Competitive facility location and design problem. *European Journal of Operational Research*, 182(1):40–62, 2007b.

Robert Aboolian, Oded Berman, and Dmitry Krass. Optimizing facility location and design. *European Journal of Operational Research*, 289(1):31–43, 2021.

Fauzy Ammari, Keiichi Ogawa, and Toshihiko Miyagi. Spatial interaction model in health-care facility location-allocation. *Infrastructure Planning Review*, 17:219–228, 2000.

Miguel F. Anjos, Bernard Gendron, and Martim Joyce-Moniz. Increasing electric vehicle adoption through the optimal deployment of fast-charging stations for local and long-distance travel. *European Journal of Operational Research*, 285(1):263–278, 2020.

Felipe Aros-Vera, Vladimir Marianov, and John E. Mitchell. p-Hub approach for the optimal park-and-ride facility location problem. *European Journal of Operational Research*, 226 (2):277–285, 2013.

Georg Bechler, Claudius Steinhardt, and Jochen Mackert. On the linear integration of attraction choice models in business optimization problems. *SN Operations Research Forum*, 2:12, 02 2021.

Stefano Benati. The maximum capture problem with heterogeneous customers. *Computers & Operations Research*, 26(14):1351–1367, 1999.

Oded Berman and Dmitry Krass. Flow intercepting spatial interaction model: a new approach to optimal location of competitive facilities. *Location Science*, 6(1):41–65, 1998.

Oded Berman and Dmitry Krass. Locating multiple competitive facilities: spatial interaction models with variable expenditures. *Annals of Operations Research*, 111(1):197–225, 2002.

Oded Berman, Tammy Drezner, Zvi Drezner, and Dmitry Krass. Modeling competitive facility location problems: New approaches and results. In *Decision Technologies and Applications*, pages 156–181. INFORMS, 2009.

Dimitris Bertsimas, Vivek F. Farias, and Nikolaos Trichakis. The price of fairness. *Operations Research*, 59(1):17–31, 2011.

Christopher G. Boone, Geoffrey L. Buckley, J. Morgan Grove, and Chona Sister. Parks and people: An environmental justice inquiry in baltimore, maryland. *Annals of the Association of American Geographers*, 99(4):767–787, 2009.

Sylvain Bouveret, Yann Chevaleyre, and Nicolas Maudet. Fair allocation of indivisible goods., 2016.

Matthias Braubach, Vladimir Kendrovski, Dorota Jarosinska, Pierpaolo Mudu, Maria Beatrice Andreucci, Femke Beute, Zoe Davies, Sjerp de Vries, Julie Glanville, Hans Keune, et al. Green and blue spaces and mental health: new evidence and perspectives for action. Technical report, World Health Organization, 2021.

Hsueh-Sheng Chang and Chin-Hsien Liao. Exploring an integrated method for measuring the relative spatial equity in public facilities in the context of urban parks. *Cities*, 28(5): 361–371, 2011.

City Parks. Pedestrians and Park Planning: How Far Will People Walk? https://www.smartcitiesdive.com/ex/sustainablecitiescollective/pedestrians-and-park-planning-how-far-will-people-walk/24937/, 2017. Accessed: 2022-04-14.

Charlotte Collins, Dagmar Haase, Stefan Heiland, and DNadja Kabisch. Urban green space interaction and wellbeing – investigating the experience of international students in berlin during the first COVID-19 lockdown. *Urban Forestry & Urban Greening*, page 127543, 2022.

Alexis Comber, Chris Brunsdon, and Edmund Green. Using a GIS-based network analysis to determine urban greenspace accessibility for different ethnic and religious groups. *Landscape and Urban Planning*, 86(1):103–114, 2008.

Emma Coombes, Andrew P. Jones, and Melvyn Hillsdon. The relationship of physical activity and overweight to objectively measured green space accessibility and use. *Social Science & Medicine*, 70(6):816–822, 2010.

Nicola da Schio, Amy Phillips, Koos Fransen, Manuel Wolff, Dagmar Haase, Silvija Krajter Ostoić, Ivana Živojinović, Dijana Vuletić, Jakob Derks, Clive Davies, Raffaele Lafortezza, Dennis Roitsch, Georg Winkel, and Rik De Vreese. The impact of the COVID-19 pandemic on the use of and attitudes towards urban forests and green spaces: Exploring the instigators of change in belgium. *Urban Forestry & Urban Greening*, 65:127305, 2021.

Dajun Dai. Racial/ethnic and socioeconomic disparities in urban green space accessibility: Where to intervene? *Landscape and Urban Planning*, 102(4):234–244, 2011.

Tammy Drezner. Locating a single new facility among existing, unequally attractive facilities. *Journal of Regional Science*, 34(2):237–252, 1994a.

Tammy Drezner. Optimal continuous location of a retail facility, facility attractiveness, and market share: An interactive model. *Journal of Retailing*, 70(1):49–64, 1994b.

Tammy Drezner. Location of multiple retail facilities with limited budget constraints — in continuous space. *Journal of Retailing and Consumer Services*, 5(3):173–184, 1998.

Tammy Drezner. Gravity models in competitive facility location. In *Contributions to Location Analysis*, pages 253–275. Springer, 2019.

Tammy Drezner, Zvi Drezner, and P Kalczynski. A cover-based competitive location model. *Journal of the Operational Research Society*, 62(1):100–113, 2011.

Tammy Drezner, Zvi Drezner, and Dawit Zerom. Competitive facility location with random attractiveness. *Operations Research Letters*, 46(3):312–317, 2018.

H. A. Eiselt and G. Laporte. Competitive spatial models. *European Journal of Operational Research*, 39(3):231–242, 1989.

José Fernández, Boglárka G.-Tóth, Juana L. Redondo, Pilar M. Ortigosa, and Aránzazu Gila Arrondo. A planar single-facility competitive location and design problem under the multi-deterministic choice rule. *Computers & Operations Research*, 78:305–315, 2017a.

José Fernández, J.L. Redondo, Pilar Ortigosa, and Boglárka G.-Tóth. *Huff-Like Stackelberg Location Problems on the Plane*, pages 129–169. Springer, 04 2017b.

C. Filippi, G. Guastaroba, and M.G. Speranza. On single-source capacitated facility location with cost and fairness objectives. *European Journal of Operational Research*, 289(3):959–974, 2021.

Aránzazu Gila Arrondo, José Fernández, J.L. Redondo, and Pilar Ortigosa. An approach for solving competitive location problems with variable demand using multicore systems. *Optimization Letters*, 8, 02 2014.

Milad Gorji. Competitive location: a state-of-art review. *International Journal of Industrial Engineering Computations*, 6:1–18, 09 2015.

Christine Haaland and Cecil Konijnendijk van den Bosch. Challenges and strategies for urban green-space planning in cities undergoing densification: A review. *Urban Forestry & Urban Greening*, 14(4):760–771, 2015.

Knut Haase and Sven Müller. Management of school locations allowing for free school choice. *Omega*, 41(5):847–855, 2013.

Knut Haase and Sven Müller. A comparison of linear reformulations for multinomial logit choice probabilities in facility location models. *European Journal of Operational Research*, 232(3):689–691, 2014.

S.Louis Hakimi. On locating new facilities in a competitive environment. *European Journal of Operational Research*, 12(1):29–35, 1983.

Donald A Hay. Sequential entry and entry-deterring strategies in spatial competition. *Oxford Economic Papers*, 28(2):240–257, 1976.

Elaine Hoffimann, Henrique Barros, and Ana Isabel Ribeiro. Socioeconomic inequalities in green space quality and accessibility—evidence from a southern european city. *International Journal of Environmental Research and Public Health*, 14(8), 2017.

Harold Hotelling. Stability in competition. *The Economic Journal*, 39(153):41–57, 1929.

David Huff and Bradley M McCallum. Calibrating the Huff model using arcGIS business analyst. *ESRI White Paper*, pages 1–33, 2008.

David L. Huff. Defining and estimating a trading area. *Journal of Marketing*, 28(3):34–38, 1964.

David L. Huff. Parameters estimation in the Huff model. *Esri*, page 3, 12 2021.

Dorothy C. Ibes. A multi-dimensional classification and equity analysis of an urban park system: A novel methodology and case study application. *Landscape and Urban Planning*, 137:122–137, 2015.

INSPQ. Material and Social Deprivation Index. https://www.inspq.qc.ca/en/deprivation/material-and-social-deprivation-index, 2019. Accessed: 2022-04-14.

Christopher Jung, Sampath Kannan, and Neil Lutz. A center in your neighborhood: Fairness in facility location. *ArXiv*, abs/1908.09041, 2019.

Nadja Kabisch and Dagmar Haase. Green justice or just green? provision of urban green spaces in berlin, germany. *Landscape and Urban Planning*, 122:129–139, 2014.

H Kucukaydin, N Aras, and I K Altinel. A discrete competitive facility location model with variable attractiveness. *Journal of the Operational Research Society*, 62(9):1726–1741, 2011.

Landsat Missions. Landsat Normalized Difference Vegetation Index. https://www.usgs.gov/landsat-missions/landsat-normalized-difference-vegetation-index, 2022. Accessed: 2022-04-14.

G. Laporte, S. Nickel, and F.S. da Gama. *Location Science*. Springer International Publishing, 2015.

Li Li and Ma. Spatial optimization for urban green space (UGS) planning support using a heuristic approach. *Applied Geography*, 138:102622, 2022.

Arika Ligmann-Zielinska. Spatial optimization. *International Encyclopedia of Geography: People, the Earth, Environment and Technology: People, the Earth, Environment and Technology*, pages 1–6, 2016.

Yun Hui Lin and Qingyun Tian. Branch-and-cut approach based on generalized Benders decomposition for facility location with limited choice rule. *European Journal of Operational Research*, 293(1):109–119, 2021.

Ivana Ljubić and Eduardo Moreno. Outer approximation and submodular cuts for maximum capture facility location problems with random utilities. *European Journal of Operational Research*, 266(1):46–56, 2018.

Bianca Lopez, Christopher Kennedy, Christopher Field, and Timon McPhearson. Who benefits from urban green spaces during times of crisis? perception and use of urban green spaces in new york city during the COVID-19 pandemic. *Urban Forestry & Urban Greening*, 65:127354, 2021.

R. Duncan Luce. *Individual Choice Behavior: A Theoretical analysis*. Wiley, New York, NY, USA, 1959.

Tien Mai and Andrea Lodi. A multicut outer-approximation approach for competitive facility location under random utilities. *European Journal of Operational Research*, 284(3):874–881, 2020.

Michael T. Marsh and David A. Schilling. Equity measurement in facility location analysis: A review and framework. *European Journal of Operational Research*, 74(1):1–17, 1994.

Marcela Maury-Mora, María Teresa Gómez-Villarino, and Carmen Varela-Martínez. Urban green spaces and stress during COVID-19 lockdown: A case study for the city of madrid. *Urban Forestry & Urban Greening*, 69:127492, 2022.

Carolina Mayen Huerta and Ariane Utomo. Evaluating the association between urban green spaces and subjective well-being in mexico city during the COVID-19 pandemic. *Health & Place*, 70:102606, 2021.

Daniel McFadden et al. Conditional logit analysis of qualitative choice behavior. *Frontiers in econometrics*, 1973.

Sven Müller and Knut Haase. Customer segmentation in retail facility location planning. *International Journal of Business Research*, 7:235–261, 10 2014.

Masao Nakanishi and Lee G Cooper. Parameter estimation for a multiplicative competitive interaction model—least squares approach. *Journal of marketing research*, 11(3):303–311, 1974.

M.N. Neema and A. Ohgai. Multi-objective location modeling of urban parks and open spaces: Continuous optimization. *Computers, Environment and Urban Systems*, 34(5): 359–376, 2010.

Roland Ngom, Pierre Gosselin, and Claudia Blais. Reduction of disparities in access to green spaces: Their geographic insertion and recreational functions matter. *Applied Geography*, 66:35–51, 2016.

Tomasz Noszczyk, Julia Gorzelany, Anita Kukulska-Kozieł, and Józef Hernik. The impact of the COVID-19 pandemic on the importance of urban green spaces to the public. *Land Use Policy*, 113:105925, 2022.

Kyushik Oh and Seunghyun Jeong. Assessing the spatial distribution of urban parks using GIS. *Landscape and Urban Planning*, 82(1):25–32, 2007.

Edward C Prescott and Michael Visscher. Sequential location among firms with foresight. *The Bell Journal of Economics*, pages 378–393, 1977.

Juana L. Redondo, José Fernández, José Domingo Álvarez Hervás, Aránzazu Gila Arrondo, and Pilar M. Ortigosa. Approximating the pareto-front of a planar bi-objective competitive facility location and design problem. *Computers & Operations Research*, 62:337–349, 2015.

William John Reilly. *The law of retail gravitation*. WJ Reilly, 1931.

Charles Revelle. The maximum capture or "sphere of influence" location problem: Hotelling revisited on a network. *Journal of Regional Science*, 26:343 – 358, 07 1986.

ServiceObjects. ZIP and Postal Code Database with GeoCoordinates for US and Canada. https://www.serviceobjects.com/blog/free-zip-code-and-postal-code-database-with-geocoordinates/, 2020. Accessed: 2022-04-14.

Adri Stark, Jake Tobin Garrett, and Nahomi Amberber. The 2021 Canadian city parks report: Centring equity & resilience. *Park People*, 2021.

Statistics Canada. Census Profile, 2016 Census. https://montreal.ca/en/boroughs, 2016. Accessed: 2022-04-14.

Kenneth E. Train. *Discrete Choice Methods with Simulation*. Cambridge University Press, 2 edition, 2009.

M. Vallejo, D. Corne, and P. Vargas. Online/offline evolutionary algorithms for dynamic urban green space allocation problems. *Journal of Experimental & Theoretical Artificial Intelligence*, 29(4):843–867, 2017.

Ville de Montréal. Budget 2018 et PTI 2018-2020 de l'arrondissement de Mercier–Hochelaga-Maisonneuve. https://montreal.ca/articles/documents-financiers-de-mhm-20896, 2018. Accessed: 2022-04-14.

Ville de Montréal. Présentation budgétaire 2019. https://montreal.ca/articles/documents-financiers-de-montreal-nord-8044, 2019. Accessed: 2022-04-14.

Ville de Montréal. Limite administrative de l'agglomération de montréal. https://donnees.montreal.ca/ville-de-montreal/polygones-arrondissements, 2020. Accessed: 2022-04-14.

Ville de Montréal. A look at Montréal's performance indicators. https://montreal.ca/en/articles/look-montreals-performance-indicators-21776, 2021a. Accessed: 2022-04-14.

Ville de Montréal. Budget 2021 et PDI 2021-2030 de rosemont-la petite-patrie. https://montreal.ca/articles/documents-financiers-de-rosemont-la-petite-patrie-14986, 2021b. Accessed: 2022-04-14.

Ville de Montréal. Construction in parks:creating spaces that are more inviting and more accessible, 2021c. URL https://montreal.ca/en/articles/construction-parkscreating-spaces-are-more-inviting-and-more-accessible-14432.

Ville de Montréal. VSP budget de fonctionnement 2022 et PDI 2022-2031. https://montreal.ca/articles/documents-financiers-de-vsp-14354, 2021d. Accessed: 2022-04-14.

Ville de Montréal. Municipal administration boroughs. https://montreal.ca/en/boroughs, 2022a. Accessed: 2022-04-14.

Ville de Montréal. Grands parcs et parcs-nature. https://montreal.ca/lieux?mtl_content.lieux.tags.code=PR015&orderBy=dc_title, 2022b. Accessed: 2022-04-14.

Ville de Montréal. Installations récréatives, sportives et culturelles extérieures. https://donnees.montreal.ca/ville-de-montreal/installations-recreatives-sportives-et-culturelles, 2022c. Accessed: 2022-04-14.

Ville de Montréal. Grands parcs, parcs d'arrondissements et espaces publics. https://donnees.montreal.ca/ville-de-montreal/grands-parcs-parcs-d-arrondissements-et-espaces-publics, 2022d. Accessed: 2022-04-14.

WHO. Urban green space interventions and health: A review of impacts and effectiveness, 2017.

Jennifer Wolch, John Wilson, and Jed Fehrenbach. Parks and park funding in los angeles: An equity-mapping analysis. *Urban Geography*, 26:4–35, 05 2013.

Changdong Ye, Lingqian Hu, and Min Li. Urban green space accessibility changes in a high-density city: A case study of macau from 2010 to 2015. *Journal of Transport Geography*, 66:106–115, 2018.

Eun Joo Yoon, Bomi Kim, and Dong Kun Lee. Multi-objective planning model for urban greening based on optimization algorithms. *Urban Forestry & Urban Greening*, 40:183–194, 2019. Urban green infrastructure – connecting people and nature for sustainable cities.

Yuhan Yu, Wenting Zhang, Peihong Fu, Wei Huang, Keke Li, and Kai Cao. The spatial optimization and evaluation of the economic, ecological, and social value of urban green space in shenzhen. *Sustainability*, 12(5), 2020.

Wenting Zhang, Shan Li, Yunxiang Gao, Wenping Liu, Yuankun Jiao, Chen Zeng, Lin Gao, and Tianwei Wang. Travel changes and equitable access to urban parks in the post COVID-19 pandemic period: Evidence from wuhan, china. *Journal of Environmental Management*, 304:114217, 2022.

Åsa Ode Sang, Igor Knez, Bengt Gunnarsson, and Marcus Hedblom. The effects of naturalness, gender, and age on how urban green space is perceived and used. *Urban Forestry & Urban Greening*, 18:268–276, 2016.

Appendix A

Dataset Columns

Neighborhood dataset

- neighborhood_id
- name
- type (neighborhood or independent suburb)
- coordinates: well-known text (WKT) description of neighborhood geometry (polygon and multi-polygon)
- centroid: WKT representation of neighborhood centroid
- area_m2: area of neighborhood in m^2
- n_ct: number of census tracts intersecting this neighborhood
- pop2016: total 2016 population within neighborhood
- pop_density_km2: Neighborhood population density
- age_0_14: % of population aged 0-14 years
- age_15_64: % of population aged 15-64 years
- age_65more: % of population aged 65 years or more
- household: total 2016 number of households within neighborhood
- avg_income: average household income ($)
- n_parks: number of parks within neighborhood
- list_parks_id: list of parks within neighborhood

FSA dataset

- fsa: FSA id
- coordinates: well-known text (WKT) description of FSA geometry (polygon and multi-polygon)
- centroid: WKT representation of FSA centroid
- area_m2: area of FSA in m^2
- n_ct: number of census tracts intersecting this FSA
- pop2016: total 2016 population within FSA
- pop_density_km2: FSA population density
- age_0_14: % of population aged 0-14 years
- age_15_64: % of population aged 15-64 years
- age_65more: % of population aged 65 years or more
- household: total 2016 number of households within FSA
- avg_income: average household income ($)
- scoresoc: weighted social deprivation score
- scoremat: weigthed material deprivation score
- ndvi_ncells: number of NDVI raster cells within FSA
- ndvi_mean: mean NDVI
- ndvi_min: min NDVI
- ndvi_max: max NDVI
- ale_ncells: number of ALE raster cells within FSA
- ale_mean: mean ALE
- ale_min: min ALE
- ale_max: max ALE
- uheat_ncells: number of Unusual heat score raster cells within FSA
- uheat_mean: mean Unusual heat score
- uheat_min: min Unusual heat score
- uheat_max: max Unusual heat score
- smoke_ncells: number of smoke pollution score raster cells within FSA
- smoke_mean: mean smoke pollution score
- smoke_min: min smoke pollution score
- smoke_max: max smoke pollution score
- n_parks: number of parks within FSA (valid only if neighborhood completely falls within Ville de Montréal boundaries)
- list_parks_id: list of parks within neighborhood

Park dataset

- park_id: Park id
- coordinates: well-known text (WKT) description of park geometry (polygon and multi-polygon)
- centroid: WKT representation of park centroid
- area_m2: area of park in m^2
- park_type_1: Park main category
- park_type_2: Park subcategory
- fsa_ids: FSA intersection park
- n_install: Number of installations within park
- install_types: Types of installation withing park
- ndvi_ncells: number of NDVI raster cells within park
- ndvi_mean: mean NDVI
- ndvi_min: min NDVI
- ndvi_max: max NDVI
- ale_ncells: number of ALE raster cells within park
- ale_mean: mean ALE
- ale_min: min ALE
- ale_max: max ALE
- uheat_ncells: number of Unusual heat score raster cells within park
- uheat_mean: mean Unusual heat score
- uheat_min: min Unusual heat score
- uheat_max: max Unusual heat score
- smoke_ncells: number of smoke pollution score raster cells within park
- smoke_mean: mean smoke pollution score
- smoke_min: min smoke pollution score
- smoke_max: max smoke pollution score

Appendix B

Number of Clusters

Neighborhood	Number of clusters
Ahuntsic-Cartierville	250
Anjou	100
Côte-des-Neiges-Notre-Dame-de-Grâce	150
L'Île-Bizard-Sainte-Geneviève	50
Lachine	100
LaSalle	150
Le Plateau-Mont-Royal	150
Le Sud-Ouest	200
Mercier-Hochelaga-Maisonneuve	250
Montréal-Nord	150
Outremont	50
Pierrefonds-Roxboro	150
Rivière-des-Prairies-Pointe-aux-Trembles	200
Rosemont-La Petite-Patrie	200
Saint-Laurent	200
Saint-Léonard	100
Verdun	100
Ville-Marie	250
Villeray-Saint-Michel-Parc-Extension	200

Appendix C

Weighting Fairness Parameters

The factors below are used to derive the final weighting fairness parameter ρ_n for the neighborhood $n \in N$:

$$\rho_n = \rho_{n1} \times \rho_{n2} \times \rho_{n3} \times \rho_{n4} \ .$$

Neighborhood n	Density ρ_{n1}	Social ρ_{n2}	Material ρ_{n3}	Smoke ρ_{n4}	Total ρ_n
Ahuntsic-Cartierville	0.90	0.95	1.05	1.00	0.90
Anjou	1.03	0.95	1.05	1.00	1.02
Côte-des-Neiges-Notre-Dame-de-Grâce	1.10	1.05	0.95	1.01	1.11
L'Île-Bizard-Sainte-Geneviève	0.90	0.95	0.95	0.96	0.78
Lachine	0.90	1.05	1.05	0.99	0.98
LaSalle	0.90	0.95	1.05	1.01	0.90
Le Plateau-Mont-Royal	1.10	1.05	0.95	1.00	1.10
Le Sud-Ouest	0.90	1.05	0.95	1.01	0.90
Mercier-Hochelaga-Maisonneuve	0.90	1.05	1.05	0.99	0.99
Montréal-Nord	1.10	0.95	1.05	1.00	1.09
Outremont	1.10	0.95	0.95	1.01	1.00
Pierrefonds-Roxboro	0.90	0.95	1.05	0.97	0.87
Rivière-des-Prairies-Pointe-aux-Trembles	0.90	0.95	1.05	0.99	0.89
Rosemont-La Petite-Patrie	1.10	1.05	0.95	1.00	1.10
Saint-Léonard	1.10	0.95	1.05	1.00	1.10
Saint-Laurent	0.92	0.95	1.05	1.01	0.92
Verdun	0.90	1.05	0.95	1.02	0.91
Ville-Marie	1.10	1.05	0.95	1.00	1.10
Villeray-Saint-Michel-Parc-Extension	1.10	1.03	1.05	1.01	1.19

Appendix D

Design Option Solution of Existing Parks

Park ID	Baseline budget	Fair budget
0082-000	1	1
0082-001	2	2
0126-000	1	1
0127-000	3	1
0187-000	1	1
0190-000	3	3
0196-000	2	1
0197-000	1	1
0200-000	3	2
0202-000	1	1
0204-000	3	3
0205-000	3	3
0206-000	1	1
0207-000	1	1
0209-000	1	1
0210-000	1	1
0211-000	1	1
0212-000	1	1
0216-000	3	3
0217-000	3	3
0218-000	3	2
0219-000	1	1
0220-000	1	1
0221-000	2	1
0241-000	1	1
0243-000	3	3
0244-000	3	1
0291-000	3	2

Park ID	Baseline budget	Fair budget
0292-000	3	1
0300-000	1	3
0303-000	1	1
0589-000	3	3
0590-000	1	1
0785-000	1	1
0878-000	1	1
0879-000	2	1
1017-000	3	3
1037-000	1	1
1038-000	1	1
1064-000	1	1
1133-000	1	1
1166-000	1	1
1167-000	1	1
1168-000	1	1
1169-000	1	1
1170-000	1	1
1171-000	2	1
1172-000	2	2
1173-000	1	1
1174-000	3	1
1186-000	3	1
1209-000	3	3
1272-000	3	3
1273-000	3	1
7000-000	1	1

Appendix E

Design Option Solution of New Parks

Park ID	Baseline budget	Fair budget
H1T	3	3
H1X	3	0
H1Y	1	3
H2G	1	3
H2S	1	1

Appendix F

L_1-Norm of the distance per neighborhood under the baseline and the fair budget

The L_1-Norm for a set of demand points I and segments S is defined as follows:

$$L_1 = \sum_{i \in I} \sum_{s \in S} w_{is} \cdot |\bar{d}_{is} - \bar{d}|,$$

where

$$\bar{d}_{is} = \sum_{j \in J} \sum_{r \in R(j)} p_{isjr} d_{ij}$$

and

$$\bar{d} = \frac{\sum_{i \in I} \sum_{s \in S} w_{is} \bar{d}_{is}}{\sum_{i \in I} \sum_{s \in S} w_{is}}.$$

Neighborhood	L_1 norm with baseline budget	L_1 norm with fair budget
Ahuntsic-Cartierville	1,164,820	1,164,820
Anjou	189,600	196,100
Côte-des-Neiges-Notre-Dame-de-Grâce	31,602	33,325
L'Île-Bizard-Sainte-Genevièveve	249,174	249,174
Lachine	243,233	243,177
LaSalle	269,140	274,292
Le Plateau-Mont-Royal	34,906	35,506
Le Sud-Ouest	368,463	369,811
Mercier-Hochelaga-Maisonneuve	447,265	447,360
Montréal-Nord	28,229	27,052
Outremont	32,268	28,502
Pierrefonds-Roxboro	504,732	504,732
Rivière-des-Prairies-Pointe-aux-Trembles	634,183	634,183
Rosemont-La Petite-Patrie	34,258	34,084
Saint-Laurent	328,378	330,393
Saint-Léonard	23,032	19,023
Verdun	21,577	21,577
Ville-Marie	30,888	30,124
Villeray-Saint-Michel-Parc-Extension	31,135	31,656