HELSINGIN YLIOPISTO
HELSINGFORS UNIVERSITET
UNIVERSITY OF HELSINKI

Master's thesis
Geography
Urban geography

Strategic districting for the mitigation of educational segregation: A pilot model for school district optimization in Helsinki

Hertta Sydänlammi

2019

Supervisor: Venla Bernelius

UNIVERSITY OF HELSINKI
FACULTY OF SCIENCE
DEPARTMENT OF GEOSCIENCES AND GEOGRAPHY
GEOGRAPHY

P. O. Box 64 (Gustaf Hällströmin katu 2)
00014 University of Helsinki

HELSINGIN YLIOPISTO
HELSINGFORS UNIVERSITET
UNIVERSITY OF HELSINKI

Tiedekunta - Fakultet - Faculty Faculty of Science		Osasto - Institution - Department Department of Geosciences and Geography
Tekijä - Författare - Author Hertta Sydänlammi		
Työn nimi - Title Strategic districting for the mitigation of educational segregation: A pilot model for school district optimization in Helsinki		
Oppiaine - Läroämne - Subject Regional Studies		
Työn laji/ Ohjaaja - Arbetets art/Handledare - Level/Instructor Master's Thesis / Venla Bernelius	Aika - Datum - Month and year May 2019	Sivumäärä - Sidoantal - Number of pages 76 p + 18.

Tiivistelmä - Referat - Abstract

The social urban structure of Helsinki has experienced a significant rise in spatial differences during the last two decades. This development has reflected on schools as rising differences between schools' student compositions and learning outcomes. Additionally, signs of independent school effects have been observed in several studies. The differentiation of student compositions is feared to exacerbate residential segregation and differentiate schools' operating environments further. It is possible, however, to intervene this development by drawing the school attendance districts such that the social differences between schools' student compositions are effectively minimized. For this purpose, new machine learning based optimization tools are needed.

The main objective of this master's thesis study is to examine the possibility to optimize Helsinki's school districts toward more internally heterogeneous and externally homogeneous social compositions. For this purpose, I have developed an optimization model that minimizes the variance of social variables between school districts by iteratively redrawing the districts' borders. In a pilot application of the model I optimize the school districts of Helsinki by using the share of population with immigrant background as the optimization variable, while existing school infrastructure (the school locations and student capacities), spatial contiguity of the districts, and school-specific maximum travel distances are used as constraints restricting the shapes that the districts can take.

The core finding of this study is that in Helsinki, the social compositions of school districts can be significantly evened out by redrawing the school district borders. However, for the model to be suitable for district planning in practice it needs further development. At this stage, the main limitations of the model are related to the shapes of the optimized districts, the model's time complexity and the lack of a constraint or optimization parameter that accounts for the safety of children's school trips.

Avainsanat - Nyckelord

Keywords
Segregation, schools, school districting, optimization model, learning outcomes

Säilytyspaikka - Förvaringsställe - Where deposited
Helsingin University Library, Kumpula Campus Library

Muita tietoja - Övriga uppgifter - Additional information
Source code available in: https://github.com/herttale/School-district-optimization

HELSINGIN YLIOPISTO
HELSINGFORS UNIVERSITET
UNIVERSITY OF HELSINKI

Tiedekunta - Fakultet - Faculty	Osasto - Institution - Department	
Matemaattis-luonnontieteellinen	Geotieteiden ja maantieteen osasto	
Tekijä - Författare - Author Hertta Sydänlammi		
Työn nimi - Title Strategic districting for the mitigation of educational segregation: A pilot model for school district optimization in Helsinki		
Oppiaine - Läroämne - Subject Aluetiede		
Työn laji/ Ohjaaja - Arbetets art/Handledare - Level/Instructor Pro gradu -tutkielma / Venla Bernelius	Aika - Datum - Month and year Toukokuu 2019	Sivumäärä - Sidoantal - Number of pages 76 s + 18.

Tiivistelmä - Referat - Abstract

Helsingin kaupunkirakenne on eriytynyt viimeisten vuosikymmenien aikana merkittävästi sosiaalisilla mittareilla tarkasteltuna. Kehitys on heijastunut kouluihin oppilaspohjien ja oppimistuloksien erojen kasvuna, minkä lisäksi Helsingissä on löydetty viitteitä myös itsenäisistä kouluvaikutuksista. Koulujen eriytymiskehityksen pelätään mainevaikutuksen kautta kiihdyttävän alueellista segregaatiota ja siten oppilaspohjien eriytymistä entisestään. Oppilaspohjien eroihin on kuitenkin mahdollista vaikuttaa määrittämällä oppilasalueet uudelleen tavalla, joka minimoi oppilasalueiden välisiä sosiaalisia eroja mahdollisimman tehokkaasti. Tätä varten tarvitaan uudenlaisia, koneoppimiseen perustuvia optimointityökaluja.

Tämän opinnäytetyön päätavoitteena on tutkia mahdollisuutta optimoida Helsingin oppilasalueita väestöpohjiltaan sisäisesti heterogeenisemmiksi ja keskenään homogeenisemmiksi. Tavoitetta varten olen kehittänyt työssäni automatisoidun optimointimallin, joka minimoi sosiaalisten muuttujien varianssia oppilasalueiden välillä oppilasalueiden muotoa varioimalla. Mallin pilottisovelluksessa optimoin Helsingin oppilaaksiottoalueita tasaisemmiksi käyttäen optimoitavana muuttujana vieraskielisen väestön osuutta. Olemassa olevaa kouluverkostoa eli koulujen sijaintia, oppilasalueiden maantieteellistä yhtenäisyyttä, enimmäisoppilasmääriä koulukohtaisella marginaalilla sekä koulukohtaista koulumatkan enimmäispituutta on käytetty mallissa alueiden muodostamista rajoittavina tekijöinä.

Tutkimukseni keskeinen löydös on, että oppilasaluerajoja siirtelemällä oppilasalueiden sosiaalisen pohjan eroihin voidaan vaikuttaa Helsingissä merkittävästi. Malli vaatii kuitenkin vielä perusteellista jatkokehittämistä soveltuakseen aluejakojen käytännön suunnitteluun, ja tässä vaiheessa sen merkittävimmät kehityskohteet liittyvät optimoitujen alueiden muotoon, mallin laskennalliseen vaativuuteen ja koulumatkojen turvallisuutta mittaavan optimointiparametrin puuttumiseen.

Avainsanat - Nyckelord
Segregaatio, alueellinen eriytyminen, koulut, oppimistulokset, optimointimalli, oppilasalueet

Keywords

Säilytyspaikka - Förvaringsställe - Where deposited
Helsingin yliopiston kirjasto, Kumpulan kampus

Muita tietoja - Övriga uppgifter - Additional information
Mallin lähdekoodi: https://github.com/herttale/School-district-optimization

Table of contents

1. Introduction ... 1

1.1 Background and motivation ... 1

1.2 The objectives and design of this study ... 3

1.3 Outline ... 4

2. Theory ... 6

2.1 Introduction to automated districting ... 6

2.1.1 What is districting? ... 6

2.1.2 The rise of automation ... 8

2.2 School districting – a special districting case ... 11

2.2.1 School districting in practice: overview of OECD countries ... 11

2.2.2 General criteria and objectives in school districting ... 13

2.3 Districting algorithms: history and current state ... 16

2.3.1 Districting – a challenging combinatorial problem ... 16

2.3.2 Early districting algorithms ... 18

2.3.3 Recent and current approaches to automated districting ... 19

2.3.4 School districting models ... 20

2.4 School districting and educational segregation ... 22

2.4.1 The link between residential segregation and educational segregation ... 22

2.4.2 The role of school choice and selective migration in educational segregation ... 26

2.5 School districting and educational segregation in Helsinki ... 28

2.5.1 The legal framework for school districting in Helsinki ... 28

2.5.2 Educational segregation in Helsinki ... 31

3. Methods: a pilot model for school district optimization in Helsinki ... 37

3.1 Study setting & data ... 37

3.2 Model summary ... 38

3.3 Model logic ... 40

3.3.1 Logic overview ... 40

3.3.2 Rules concerning block adoption ... 41

3.3.3 The logic behind rule 6 ... 42

3.4 Model implementation & specifications ... 44

3.5 Specifications for the example optimizations ... 48

3.5.1 Data preparation ... 48

3.5.2 The example optimizations ... 49

4. Analysis / Results ... 51

4.1 Sensitivity analysis ... 51

4.2 The main example optimizations...57
5. Discussion 67
5.1 The quality of the results: compactness, fairness & legitimacy...........................67
5.2 Model limitations and future development targets...71
6. Conclusions...75
Bibliography..77
Appendix: model source code 89
classes.py 89
main.py 93

1. Introduction

1.1 Background and motivation

The spatial gap between people of different socio-economic backgrounds and ethnicities is widening in all capital cities across Europe, which is made visible by rising levels of residential segregation (Marcińczak et al. 2015; Musterd et al 2017). Recently, this development has raised a discussion about the less addressed but notably segregated contexts of life that are also partly affected by residential segregation. These contexts include places of education, workplaces, leisure and other places people visit in their daily routines – the segregation of which can result in overall minimal exposure to diversity in people's daily lives (Boterman & Musterd 2016; van Ham & Tammaru 2016). The effects of residential segregation are not only translated to these other domains of life by physical distances, but also by administrative divisions that separate neighborhoods into school districts, hospital districts, catchment areas of public transport stops, et cetera.

The influence and power of different district divisions is well acknowledged in electoral districting, since it's consequences are widely discussed in the public every few years – especially in cases, where the election result is somehow in conflict with the popular vote or seems to be somehow affected by the division of the state into voting districts. How other, more mundane local level districting divisions affect our lives, how they create and realize domains of segregation in our everyday lives, is a much less acknowledged and addressed issue.

Political parties have used strategic districting as a means of affecting election outcomes for their benefit for centuries – the first documented case, based on which the term gerrymandering was coined, dates back to 1810s (Hardy 1990). Because of this history, the term *strategic districting* is easily associated with corruption and manipulation that leads to unbalanced and unnatural outcomes, while neutrality is usually considered a desirable objective in districting. This association, though, is more related to the history of abuse (see e.g. Hardy 1990, Johnston 2002) than the method or practice of strategic districting itself. While in many cases neutrality and evenness are rational and appropriate takes on a districting problem, there are cases and situations where strategic districting could be used to hinder the segregation in important domains of life – and to benefit the society as a whole.

In Finland, one of the critical contexts of life that are becoming increasingly socially polarized are the places of education. Though the Finnish school system is internationally regarded as one of the most successful schooling systems on the scale of equality of educational opportunities (OECD 2004, OECD 2018b), lately the differences between schools' learning outcomes have been on rise, especially inside larger cities and most notably in the capital city Helsinki (Kuusela 2010; Bernelius 2013; Kupari et al. 2013). This development has been strongly linked to growing residential segregation, which is inevitably affecting the social compositions of both school districts and schools (Bernelius 2011).

Educational segregation is problematic, because it tends to lead to differentiation of overall student performance and learning outcomes on both school and individual level – a differentiation that cannot be explained only by the individual backgrounds of the students (e.g. OECD 2004; Rumberger & Palardy 2005; Mickelson & Bottia 2010, Perry & McConney 2010). This means that while the individual skills and backgrounds of students affect individual achievement, so does the characteristics of their schools and their student compositions. However, it is possible to influence the polarization of schools' student compositions and prevent educational segregation by redrawing the school district borders in a manner that optimizes the social evenness between school districts.

Unfortunately, the current GIS-tools are practically non-existent or inadequate for district of optimization – although there are several commercial and open-source GIS tools available for quick and easy evaluation of a particular partitioning of the city, they are generally not capable of solving the combinatorial challenge of finding an optimal or nearly-optimal partitioning of the city into school districts. With the current software, the amount of manual work that would be needed for comparing all the possible alternative plans makes the evaluation of the best or even nearly optimal district division practically impossible. For this purpose, more powerful tools are needed – tools that can model and produce multiple, differently weighted optimal districting plans effectively.

The motivation for this master's thesis study arises from the realization that a) differences between schools' student compositions produce asymmetric learning outcomes, b) the differences between schools are growing, and c) this development can be influenced by strategic districting. As mentioned, one significant barrier hindering the ability to influence school segregation is the lack of appropriate tools for redistricting and school district planning. While automation and artificial intelligence is penetrating to almost every aspect of people's lives and business, it is still largely underutilized in the decision supporting and information producing processes of the public sector. This aspect brings up the final realization d) – new tools are needed for effective strategic districting.

1.2 The objectives and design of this study

The main objective of this master's thesis study is to develop an algorithm for school district optimization and pilot its application in the context of Helsinki. The aim is to model a socially more even school district plan – a plan that would consist of internally more heterogeneous and externally more homogeneous school districts in terms of social compositions, but also to find out the scale in which the optimization would be theoretically possible in Helsinki. This scale is studied by producing alternative districting plans that maximize the social variance inside individual school districts and minimize it between the school districts, while accounting for travel times and schools' student capacities. The optimized districting plans are then compared to the district plan of the current school year (2018–2019).

The methods used in the model development have been inspired by the large body of studies made on automated electoral districting. The model developed utilizes local search techniques to iteratively improve the current school district division until the optimal division under specific restriction rules is found. In previous studies, this approach to district optimization has been found to be among the most feasible, and consequently also one of the most utilized. The approach chosen is experimental, and does not pursue to find a single, objectively optimal redistricting solution directly applicable to real world, but rather to pilot a possible approach to the existing school districting problem in Helsinki and demonstrate the need for automated districting approaches and initiatives.

This study is also an initiative towards a data-oriented and machine learning based districting procedure. A step towards a decision-making system, where technology is utilized for extracting information of different options, and for comparing plans and their consequences in an effective manner. In the Helsinki context, the modeling of school district optimization has never been done before, which makes this study the first one of its kind.

1.3 Outline

This thesis is divided into the following five parts: (1) introduction, (2) theory and literature, (3) methods, (4) results, (5) discussion and (6) conclusions.

Chapter 2 presents relevant theories, concepts and previous research. The topics covered include the history of districting and district optimization both on a general level and from the perspective of school districting. The previous district optimization initiatives and methods are presented, and the problem defined with some mathematical formalism. In this chapter, I also go through the aspects and research related to educational segregation, its causes and consequences, and introduce the context of the pilot study: the city of Helsinki.

In chapter 3, I give a detailed explanation of the optimization algorithm developed in this study, as well as its implementation and piloting in the Helsinki context. First, I summarize the model’s features, after which I explain the logic of the model through a simple metaphor. Last, I explain the details related to the piloting and implementation of the model, as well as the details of applying the model to Helsinki school district division.

In chapter 4, I present the key findings from the empirical part of this study: the application of the school district optimization model on Helsinki school district division. The resulting optimized district plan is also compared to the original district division to measure the scale in which the districts are optimizable with chosen parameters.

Chapter 5 provides a critical analysis of the developed model and the empirical results. The findings are evaluated, and the limitations are discussed. Also, recommendations for future research are given.

Chapter 6 summarizes the findings of this thesis.

2. Theory

2.1 Introduction to automated districting

2.1.1 What is districting?

Districting, also known as zone design is a process in which an administrative area – political unit of any size – is divided into smaller regions for administrative purposes (e.g. Williams 1995; Vanneschi et al. 2017). It is used at all levels of administration, ranging from international (f.ex. EU NUTS categorization) and state-level (e.g. electoral district planning) to for example land use, rescue service or school attendance zoning practiced by municipality officials. Though districting can simply be seen as a practical measure of organizing public services for citizens, it's also an allocation of real-world physical, social and human resources: space, people, communities and neighborhoods. The allocation outcomes may define – depending on the society and the districting objective–, which public and social resources are within the citizens' reach: which school their children may attend, which public health care unit they have access to, who they can vote for in parliamentary elections, what kind of public transport is offered and how much a trip costs, or how far is the closest fire station. Districting is a political process that shapes the social environment and people's activity spaces, and thus has real social and human influences. Some of these influences may lead to residential choice patterns and direct effects in housing prices (e.g. school quality's effect on the school's attendance zone's housing prices, see Cheshire & Sheppard 2004; Brunner et al. 2012; Danielsen et al. 2015), while others may have major state-level political effects (e.g. the effect of electoral districting on parliament compositions, see Johnston 2002).

Districting decisions, like other resource allocation decisions in the public sector, are complex problems both politically and practically. The political complexity arises from the number of conflicting interests and interest groups at stake: the criteria for different districting initiatives reflect contestable objectives, and not all even seemingly neutral objectives benefit all demographic groups equally. The practical complexity arises from the difficulty of "measuring, assessing and evaluating the quality and quantity of impacts associated with alternative allocation patterns" (Malczewski & Jackson 2000). And not only the impacts of this multi-dimensional problem are difficult to verify, but it's also practically impossible for anyone to imagine all possible alternative districting solutions and compare them. Even a small-scale districting problem, due to its combinatorial nature, has easily thousands of alternative solutions that are impossible to enumerate explicitly (Keane 1975; Garey & Johnson 1979; Altman 1997).

Electoral district planning is one of the best known and most studied examples of districting, with gerrymandering being among the most infamous – yet popular – examples. Gerrymandering as a term denotes the "*deliberate manipulation of district borders in order to give a particular party, candidate or group a distinct advantage or disadvantage in an election*" (Williams 1995). The term was originally coined in a *Boston Weekly Messenger* cartoon after the Democratic-Republican party and its governor Elbridge Gerry manipulated Massachusetts' electoral districts' borders to win the state election against popular vote in 1812 (Hardy 1990; Bozkaya et al. 2003; Hunter 2011; Liu et al. 2016), and it became widely used to refer to politically biased districting. The case caused a vast amount of criticism calling for political neutrality and fair representation in electoral districting. In one of the earlier studies on fair minority representation, Stanwood (1871) writes:

> *"The alleged failure of the district system, and the apparent unfairness to minorities which it involves, have led thinkers in this country, and to a still greater extent in Europe, to cast about for a new plan which shall both do greater justice to the several factions in a constituency and secure a higher order of talents in the representative."*

Since then, while maintaining its popularity in the field of political studies and operations & management science, discussion of fair, unbiased and optimal districting procedures has emerged in also other fields like geography, social sciences and economics. One of the most important empirical twists to the research and discussion of districting, however, has come along with the improvements in computer technology.

2.1.2 The rise of automation

The idea of generating algorithmically optimized district plans developed in 1960s, during a decade when the foundations for mathematical theory of artificial intelligence were established – or which Russell & Norvig (2003) call the era of “early enthusiasm, great expectations” in the history of AI. Since, automated districting through computer modeling has been applied to a plethora of zoning task. Some of the examples include:
- Defining electoral districts of equal population and compact form for a state or a country (e.g. Nagel 1964; Garfinkel & Nemhauser 1970; Horn 1995; Hojati 1996; Mehrotra et al. 1998; Bozkaya et al. 2003; Rincon-Garcia et al. 2013; Vanneschi et al. 2017).
- Determining an optimal partitioning of two-level telecommunication network into local networks (clusters) and deciding a hub location for each cluster (Park et al. 2000)
- Redistricting urban police departments’ command boundaries in order to both maximize an effective use of patrol cars and balancing the workload between officers in different districts (D’Amico et al. 2002; Camacho-Colladosa et al. 2015).
- Defining sales districts for traveling salesperson team to balance workloads between employees (Hess & Samuels 1971; Easingwood 1973; Shanker et al. 1975; Zoltners & Sinha 1983)
- Assigning students to schools (forming school districts) while both minimizing travel or busing costs and balancing student distribution by social attributes and school capacity (Heckman & Taylor 1969; Ferland & Guénette 1990; desJardins et al. 2007)
- Optimizing tickets pricing zones to correspond better to the use of the public transportation network (Tavares-Pereira et al. 2007)
- Maintaining Census output geographies in order to make them suitable for a new publication of Census data (Cockings et al. 2011)

Initially, the goal of automated districting was to produce unbiased, politically neutral electoral district plans, which computer modeling was seen as a means to achieve (Vickrey 1961; Williams 1995). Throughout history (and in many cases today), district plans had been designed and drawn manually, and the partitioning decisions had been based on local knowledge, intuition, and the designers’ subjective insight (Cockings et al. 2011). These processes usually lacked objectivity while being time consuming and resource intensive. The automation was argued to remove these problems, as it would work as a veil of ignorance (in the Rawlsian manner), since the objectives and criteria would be agreed upon beforehand while the eventual, actual districting solution would be unknown to all parties until the algorithm was run (Vickrey 1961; Browdy 1990; Altman 1997). Eliminating human intervention in the actual calculation process was seen as promoting fair outcomes.

This idea has since been criticized for ignoring that the original problem of conflicting interests does not disappear with automation – goal-setting, choosing parameters for objective functions and choices regarding utilized methodologies contain conflicting values and outcomes, which makes the process of districting inherently political (Altman 1997). District optimization approaches have further been blamed for being incomprehensible to interested parties (like administrators, taxpayers, parents, teachers, students), for implicitly assuming a social consensus, for frequently failing to implement criteria or goals that decision-makers keep important (Malczewski & Jackson 2000) and for often being computationally too heavy. The criticism aims to underline that there is rarely one, objectively most optimal solution for districting, and stresses the magnitude of the obviously enormous solution spaces of potential real-world use cases.

Some of the early developers of automated districting already saw the role of automation not as replacing human influence but rather as a tool for decision making, comparison and generating sophisticated information about existing alternatives. As Heckman & Taylor state in their 1969 article on school district optimization:

> *"In practice, one always wants to compute optimal solutions under a variety of racial balance assumptions and examine the trade-off between integration achieved and cost of doing so. This point is crucial. Our view is that the role of the management scientist is not to provide one optimal plan under a given set of assumptions, but to provide a set of efficient options or alternatives together with some measure of their feasibility or cost."*

This has since become the prevalent sentiment among researchers especially what comes to less obvious and more political districting problems, like school districting.

Since the beginning of the millennium, the focus in districting literature has further shifted from trying to create one, objectively optimal plan to generating several possible and better options:

> "*The goal is to create a decision support system that allows one to contextualize a particular redistricting map by creating an ability to understand a range of possible redistricting maps for a particular geographic area. — No single plan will satisfy every interested stakeholder – there is no perfect plan. If there is no perfect plan for every constituent, then it makes sense that we will instead be choosing from a bounded set of "reasonably imperfect plans". The discrete optimization framework is ideal as a vehicle for identifying large sets of "reasonably imperfect" redistricting plans.*" (Liu et al. 2016 about their method for electoral districting)

> "*The basic idea of this work is to provide the decision makers, in this case politicians, with good viable plans (solutions to the problem). These plans should not be affected by political criteria and should be seen as options which can be considered before a final decision is made.*" (Bação et al. 2005 about their method for electoral districting)

> "*Because of the complexity of the decision-making problem, tools are needed to help end users generate, evaluate, and compare alternative school assignment plans. A key goal of our research is to aid users in finding multiple qualitatively different redistricting plans that represent different trade-offs in the decision space.*" (desJardins et al. 2007 about their method for school attendance zoning)

Those nowadays reproducing the idea of theoretical neutrality achieved via automation also do admit that the algorithm's objectivity and performance in "fairness" is strongly dependent of the method's ability to model real-world phenomena and parametrize different criteria into the objective function:

> "*Automated procedures offer more efficient, systematic, and objective methodologies for designing optimised zoning systems than manual methods, although their success is still dependent on the extent to which it is possible to*

> *model real-world phenomena, whether it is feasible to parameterise the required design criteria, and the effectiveness of the zoning algorithm(s) employed.*"
> (Cockings et al. 2011)

Inevitably, with the current computing power and optimization techniques, computers are capable of producing enormous amounts of viable information in a form of different and differently weighted districting options, which can be used to support planning and political decisions. The manually produced districting solutions tend to be haphazard because of the very limited amount of different solutions they can evaluate – on the other hand, a computer model can compare several different objective functions in a relatively short time and with minimal resources. But – to be realistic –, as some districting criteria are not easily reduced to a mathematical formula, manual post-processing may be necessary to couple with the current optimization techniques to make the plans feasible. This reminds the enthusiast of the role of computational districting models, which inevitably is of a tool's, not of a decision-maker's.

2.2 School districting – a special districting case

2.2.1 School districting in practice: overview of OECD countries

School districting is the activity of dividing an administrative area to school attendance zones, also called school catchment areas or school districts. The division defines, which schools the children living inside the administrative area can attend. Student allocation systems based on school districting are widely used in OECD countries, while the level of the plans' authority and the body responsible for the allocation varies (OECD 2018a). In most countries, the decisions regarding school districting is done on local level, usually by a municipal authority or a regional school board. In some schooling systems the school admissions are completely determined by the districting procedures of local authority, and the parents either have no direct way to influence it (like in Israel) or the possibilities are very limited and uncertain (like in Norway) (OECD 2018a).

In most systems, though, while districting guarantees a school place to every child, the parents are free to choose another school if available places exist (e.g. in Finland, Sweden, Denmark, Hungary, Iceland and Japan). Usually districting is done based on the nearest school -principle, meaning that children have a right to attend a school that is nearest to their residence. In some countries, on the other hand (like in Latvia, Belgium, Czech Republic and Italy), the parents can freely choose any school without regional restrictions, and the final admission is usually dependent on the availability of places in the chosen school (OECD 2018a). While the decision-making procedures are versatile, so are the methods and criteria used in school admissions. In countries where law protects parental free choice, the authorities' possibilities to influence for example schools' student compositions and travel distances to schools are very limited. In many schooling systems, though, the admission plans are carefully crafted to meet certain objectives.

In the systems utilizing districting, school district plans have traditionally been created manually to determine the nearest school for each neighborhood. Historically, the most usual objective of school districting has been minimizing total travel distances to schools while considering some other criteria on the side, but probably the most large-scale school attendance zoning initiative was sparked by social factors and equality aspirations. In the US, drawing school attendance zones based on geographical distances was made mandatory for public schools by Supreme Court in 1968 in order to force racial desegregation in public schools (Robertson et al 2010). This legal-political process introduced the novel methods of management science to the field of educational systems planning (e.g. Heckman & Taylor 1969), and algorithmic school district planning has since been utilized to differing scales around the world, but most extensively in the US.

As in electoral districting, the initial goal of algorithmic approach in school districting was to remove certain biases from the process and in this case base the admissions purely on legal requirements for racial balance and overall general utility of the system. But, nowadays, despite of the already long history of school districting in the legal framework of racial balance, the school district plans in US have been accused of emphasizing racial segregation of schools or at least not effectively preventing it

(Chang 2018). Among researchers, though, manipulating the districts to promote equality of learning outcomes and societal coherence seems to be sparking a growing interest.

2.2.2 General criteria and objectives in school districting

While school districting problems include many of the elements that electoral districting problems consist of, they usually differ quite fundamentally from them in terms of optimization criteria. Most frequently, the optimization criteria used in electoral districting include equal population (by some marginal), compactness and contiguity (e.g. Bozkaya et al. 2003). These criteria are more concerned of large-scale geometries of the districts than the human geographies of the real world. On the contrary, school districting usually must consider many physical, infrastructure-related aspects like the existing school infrastructure (representing natural district centers), road-network based travel distances (school accessibility) and sometimes infrastructure-related safety aspects regarding school trips (rivers, railways, hi-traffic roads etc. crossing school districts).

desJardins et al. 2007 also list the following elements differentiating school districting from other districting problems: Firstly, while compactness is desirable both for the sense of community and for minimizing travel distances, it is not as important as in other fields of districting. Secondly, redistricting of schools occurs more frequently than in most districting cases (sometimes annually), while the district division also affects the daily lives of families and school-aged children in the area. Shifting school district borders may affect the students negatively if it leads to situations where the children are required to change school repeatedly during their school career. This makes minimizing the annual changes in a districting plan an important criterion. In most desirable case, annual redistricting should only concern the children starting their first grade. Thirdly, according to desJardins et al., the nature of the decision-making process in school districting makes generating multiple plans for comparison representing different trade-offs particularly desirable.

Caro et al. (2004) have identified 7 desirable properties that a good school district plan should satisfy. The criteria have been defined in the context of Philadelphia, but can

largely be generalized to school districting everywhere:
1) Each block or neighborhood needs to be assigned to exactly one school.
2) School assignments may not exceed the school capacities.
3) Each school district must be contiguous in a way that allows for travelling from any unit inside the district to any other unit belonging to the same district without leaving it on the way.
4) The school district boundaries cannot cut across railroads, rivers, or streets with heavy traffic.
5) The total distance traveled by students is minimized with a restriction that none of the school trips may exceed a specified maximum distance.
6) All students living in the same block must go to the same school. If this is not possible due to school grade capacities, at least all students on the same grade must go to the same school. The district division should not go against the residents' sense of neighborhood.
7) A new redistricting plan must maintain a certain similarity to the former/existing plan. The yearly changes in the districts should stay moderate in the long run.

This example requirement listing reflects the nature of school attendance zoning as a practical and political challenge. The first criterion is purely practical and reflects the main objective of school districting: every child must be pointed to a school unequivocally. The second criterion is related to the practical limitations set by the existing infrastructure. One must note that the infrastructure related limitations are usually only short-term hindrances, since school buildings can often be refurbished to meet the desired capacities. The third and the sixth criterion are related to preservation of neighborhood communities and to political legitimacy, i.e. the sense of justified attendance division in the eyes of public. In the districting literature, the third criterion *contiguity* is usually not justified with anything; its justification is seen as self-evident. Technically, though, the districts could be scattered as well if it would help in achieving other, more complex districting objectives (like racial balance). At the same time, travel times to schools could still be optimized.

The fourth criterion is related to the safeness of children's school trips – which again, like the fifth criterion for average school trip length, is related to school accessibility. In general, the districting systems and models weigh either *efficiency* or *equity* criteria when evaluating accessibility from students' homes to places of education (Malczewski & Jackson 2000). Efficiency criterion is usually formalized as minimization of total costs of travel, while equity criterion is more concerned about minimizing the range (or variance) of travel costs – regardless of whether they are measured in time, distance or money (Malczewski & Jackson 2000). The fifth criterion of this example listing regards both of these. The seventh criterion is related to the coherence and convenience of children's school careers and district planning's overall legitimacy in the eyes of public.

What is notable in the above requirement listing is that it lacks the requirement for geographical compactness, which is a common requirement in other fields of districting. Geographical compactness means that a district should have a smooth and somewhat circular form. While it could be easily seen as a desirable criterion with the same justification used for contiguity and neighborhood preservation, it's quite often sidelined or seen as internalized in accessibility criteria. The reason for this is that quite often compactness is conflicting with other objectives regarded as *more important*. What comes to segregation, for example, or other spatially polarized phenomena, a zoning plan that maximally minimizes polarization is hard to implement with geographically compact districts, which may actually make compactness a somewhat *undesirable* objective in that particular environment. For example, Bouzarth et al. (2018) noticed that in their bi-objective model for school redistricting, the objectives for minimizing total school busing distance and balancing the socioeconomic compositions of the schools were in direct conflict. As Bouzarth et al. note, some trade-offs must be made with distance- or compactness-related objectives if there is will to reduce the impact of spatially polarized phenomena on districts. On the other hand, Caro et al. (2004) argue that contiguity as a constraint may pose a similar problem: in their school redistricting case study in Philadelphia they found that "there is a clear trade-off between achieving certain racial balance and keeping contiguous (and compact) districts." For this reason, some studies focused on school district optimization have decided fully giving up the requirement for *both* contiguity and compactness (e.g. Clarke & Surkis 1968), while others have ended up generating multiple different districting alternatives weighing the conflicting objectives differently in each one (desJardins et al. 2007).

Many of the already early school districting studies emphasize that they regard full automation of school redistricting process infeasible, since there are so many less tangible social and acceptability aspects that are not easily reducible to mathematical functions (Ferland & Guénette 1990; Caro et al. 2004). This might be true, as school districting is a more complicated problem than many other districting problems - the requirements are more versatile and dimensional (like school trips' safeness). Some have even suggested that automated optimization methods would be only used as the very first step in districting, after which there would be a phase of manual post-processing (e.g. Ferland & Guénette 1990). Though this conflicts with the idea of unbiasedness, it might be necessary in the actual cases. In addition, machine learning could be utilized in first tracking and later predicting the post-processing patterns, and these predictions could be again used to train the optimization model itself to produce more sophisticated results.

2.3 Districting algorithms: history and current state

2.3.1 Districting – a challenging combinatorial problem

Districting problem is a close relative of more widely known clustering problem, and they share the same basic constraints. Clustering means the identification of natural groups in an environment, usually referring to groups of objects that are similar to each other and different from others in some specified terms (e.g. Naderi & Kilic 2016). On the contrary, districting can represent either a partition of space to as homogeneous groups as possible (like is usually the case in electoral districting and school districting) or to "natural groups" that are internally homogeneous and externally heterogeneous in the specified measure (like in land use planning and landscape classification).

Bação et al. give the following definition of the shared constraints in their 2005 article *Applying genetic algorithms to zone design*. Let $X = \{x_1, x_2, \dots, x_n\}$ be a set of small, basic geographic units. Let's the assume that there are k districts/clusters, and that the district/cluster Z_i is a subset of X. Now the following rules apply for both clustering and zoning problems:

1. $Z_i \neq \{\}$ for $i = 1, 2, \ldots, k$, which means that all of the clusters/districts are nonempty sets,

2. $Z_i \cap Z_j = \{\}$ for $i \neq j$, which means that intersections of the sets are always empty (none of the districts/clusters are overlapping, so each x_n can belong to only one Z_i),

3. $\bigcup_{i=1}^{k} Z_i = X$, meaning that together the districts/clusters contain all the geographic units in set X and nothing else.

In districting problems, at least the following two constraints (or objectives depending on the problem definition) are generally considered in addition: contiguity and some degree of geographical compactness (e.g. Williams 1995; Bação et al. 2005; Ricca et al. 2013; Rincón-García et al. 2017; Vanneschi et al. 2017). Contiguity means that all geographical units belonging to a zone must share a border with another unit in the same zone in a way that allows traveling from any point inside the zone to any other point in the zone without leaving it on the way (Williams 1995). This rule ensures that the zones don't become disconnected and scattered, and that they feel intuitive, justified and unbiased from the citizen's perspective (Williams 1995). Geographical compactness means that a zone's perimeter may not grow too much in relation to its area. This simplifies the forms the zones can take, as the more complex thin, meandering tentacle formations are not possible. As in the previous constraint, the aim of this is to make the zones more "acceptable" by making them seem less biased and politically manufactured.

Districting, just like clustering, is a discrete problem in a sense that the search space is not continuous but divided into discrete *base units*, which makes the problem finite and theoretically solvable (Williams 1995). It's also a NP-complete combinatorial problem with a solution space that grows very rapidly as the number of districts and their base units increase: the amount of possible solutions grows both as a function of the number of base units in the model - and to a more limited extent - as a function of the number of districts being formed (Altman 1997). The time needed for finding a solution to a districting problem grows exponentially as a factor of both districts and geographic base units, which makes the problem computationally intractable (Garey & Johnson 1979; Altman 1997). Usually the real-world problems that would benefit from automated districting are so large and multidimensional that explicitly enumerating all the possible solutions becomes extremely difficult if not impossible (Williams 1995; Altman 1997).

Some exact optimization algorithms have been proposed for solving the districting problem (for example Garfinkel & Nemhauser 1970, Caro et al. 2004). Unlike heuristic methods that only go through a fraction of all the possible solutions, exact methods are exhaustive in the sense that they search the whole search space and end up in a provably optimal solution. Due to the size of the real-world problems' solution space, they have either shown to be infeasible or applicable to only small or medium-sized problems (Browdy 1990; Williams 1995; Altman 1997; see also Bação et al. 2005 for a definition of a medium-sized problem's solution space). Although heuristic methods cannot guarantee finding the absolute optimal solution to a districting problem, they have been able to offer largely better partitions than traditional methods in many districting cases (see e.g. D'Amico et al. 2002; Bozkaya et al. 2003; desJardins et al. 2007; Ricca & Simeone 2008). As Altman (1997) notes regarding the real-world districting cases, "political solutions may not require the precision that theory demands."

2.3.2 Early districting algorithms

The idea of using an algorithmic approach in districting dates back to 1961, when William Vickrey suggested in his article *On prevention of gerrymandering* that "the human element must be removed as completely as possible from the redistricting process. In part, this means that the process should be completely mechanical, so that once set up, there is no room at all for human choice" (Vickrey 1961). Just a few years later in 1963, Weaver & Hess developed the first heuristic procedure for redistricting electoral districts in FORTRAN. Their procedure aimed to maximize compactness while forming districts close to equal in population size (Williams 1995, Weaver &

Hess 1963). Their heuristic is a local search algorithm that iteratively improves the created partition and stops when it's overall compactness stops improving. To avoid getting stuck to a bad local optimum, the algorithm is rerun with different "initial guesses" that work as a seed for the partition and represent the centroids of the districts. The initial guesses are formed based on the researchers' subjective insight (Weaver & Hess 1963).

After Weaver & Hess, numerous different heuristics have been developed and proposed for the zone design problem in different fields. Williams (1995) and Bação, Lobo and Painho (2005) divide these methodology proposals into three categories based on their approach: in the first the individual districts are built one by one and then aggregated to form a plan. In the second approach, the districts/zones are formed by simultaneously assigning the base geographic units to predefined district centers. In the third approach, an existing plan is modified by swapping base geographic units between the districts (local search methods). According to Williams (1995) these approaches tend to have the following drawbacks: the first may not succeed in forming districts that fit well together and may leave problematic or infeasible enclaves. The second approach, on the other hand, while avoiding the problems of the first, may be unable to ensure that any particular district will meet particular criteria, since the optimization function operates on the plan-level rather than on the district-level. The third approach, as Williams notes, may be politically most feasible since it only *improves* the existing plan by some measure, and the deviation from the original plan can be minimized. This, on the other hand, may generate path-dependent, locally optimal solutions that only marginally improve the original plan.

2.3.3 Recent and current approaches to automated districting

After 1995, many new approaches have been adopted and proposed for solving the problem. A good review of these more recent districting algorithms are given in Ricca, Scozzari & Simeone's meta-analysis from 2013. Some of the most interesting approaches include evolutionary algorithms (like genetic algorithm, see Bação et al. 2005; Liu et al. 2016), tabu search (a version of local search, e.g. Bozkaya et al. 2003; Ricca & Simeone 2008), Lagrangian relaxation combined with either local search, network modeling or fuzzy zoning (e.g. Hojati 1996; Naderi & Kilic 2016), graph partitioning models (Park et al. 2000), a discrete adaptation of artificial bee colony (DABC) and discrete method of musical composition (DMMC) (Rincon-Garcia et al. 2017), et cetera. The above-mentioned approaches have been modified to address most of the inevitable criticism and issues raised on the earlier implementations. Already at the beginning of the millennium, several metaheuristics such as simulated annealing, tabu search and genetic algorithms had shown to be successful in finding reasonably good solutions for many combinatorial optimization problems (e.g. D'Amico et al. 2002; Bação et al. 2005).

Some challenges, however, still remain. Some of the seemingly well-performing newer models are computationally too heavy to be applicable on most of the real-life problems (e.g. Minimum Cost Centered Partition Problem, see Ricca et al. 2013), some struggle with the objective function formulation, balancing criteria and result quality (for example many genetic algorithm implementations are not able to guarantee contiguity, see e.g. Ricca et al. 2013), some require extensive manual post-processing to guarantee the meeting of some requirements, and even of the very promising methods most are only tested on a single, relatively small and specific case (Ricca et al. 2013). All in all, despite of great advancements in the field, there has been no miracle invention that would work flawlessly on all districting problems. All methods tend to lead to solutions that largely depend on the algorithm implementation, objective function encoding and the solution seed (the initial starting point) and are not very easily generalized and applied to other districting problems. Still, interest on automation is keen, and new models are suggested frequently. As the computational technologies advance, more systematic development and research is needed.

2.3.4 School districting models

School district plan optimization for racially balancing the student compositions and for lowering the overall cost of transportation (e.g. school busing costs) has a long reaching history as well (see e.g. Clarke & Surkis 1968; Heckman & Taylor 1969), but in the history of districting automation largely dominated by electoral districting research it's barely an offshoot. This may be due to its local nature – while electoral districting is done on the state-level, school districting is largely done on the municipality level with less resources and less political and public interest. The existing school districting papers largely refer to electoral districting automation studies (e.g. desJardins et al. 2007) or to school districting models created before 1990s (e.g. Bouzarh et al. 2018). As Caro et al. (2004) note in their districting method review, the attempts to automatically create school districting plans have been in a decline after 1980's. At the same time, electoral and other districting models have been developed quite continuously.

Most of the earlier school district optimization models consider a linear, single attribute objective function (for example minimizing total busing distances to schools), while the other objectives (like not exceeding school capacities) are constrained to fall within some acceptable range (Caro et al. 2004, Bouzarth et al. 2018). Also, goal programming approaches (extension of linear programming that permits more than one objective to be integrated in the model) have been adopted (Knutson et al. 1980; Brown 1987). According to Caro et al. (2004), most of the earlier research achieve optimal or near-optimal solutions, but due to the methods' computational limitations, the case studies are either small or non-real example cases. Another limitation of these approaches is that they can only optimize linear parameters like compactness and travel distances, which are easily reduced to linear functions, while many of the interesting and relevant parameters (like social compositions of schools) are by nature non-linear: their values vary unpredictably in the search space. This is why most of these models have reduced the more complex parameters to simple constraints (e.g. the share of non-white students may not exceed a certain limit in any school).

Despite of having less degrees of freedom than districting problems without natural district centers, the solution spaces of school districting problems remain extremely large. While the simplest school district optimizations (like total busing cost/distance optimization) can be formulated as linear models, the more complex objective formulations are usually nonlinear, multi-objective, NP-complete and well beyond the capabilities of standard GIS tools (Caro et al. 2004). The objective of this work is to optimize a nonlinear social measure while constraining both linear and nonlinear attributes. In this manner the model developed here differs from the mainstream of automated school districting studies, while approaching the tradition of electoral districting. While school districting fundamentally differs from electoral districting, method-wise there is a lot of underutilized information in the political districting field to be applied on school districting.

2.4 School districting and educational segregation

2.4.1 The link between residential segregation and educational segregation

Educational segregation or *School segregation* denotes a phenomenon where schools are disproportionately composed of students of different ethnicities or/and socioeconomically different backgrounds relative to the community's composition. In other words, the students of different backgrounds go to different schools, meaning that the schools have relatively high concentrations of students with similar backgrounds (e.g. Mickelson & Bottia 2010). School segregation is strongly predictive of performance differences between schools: socioeconomically less advantaged students and schools tend to perform less well on measures of academic achievement than their more socially advantaged peers and counterparts (Perry & McConney 2010). In the majority of OECD countries, the average socioeconomic background of students in a school is also strongly linked to the learning outcomes on individual student level (OECD 2004).

Educational segregation is largely an outcome of school market function (e.g. school choice patterns), school admission policies and *residential segregation* (Karsten 2010). Residential segregation means “the degree to which two or more groups live separately from one another, in different parts of the urban environment” (Massey & Denton 1988). It may take multiple spatial forms, appear on any level of spatial organization (e.g. house-level, block-level, neighborhood-level, city-level, country-level), and appear on many levels and forms at the same time. Massey and Denton (1988) divide the dimensions of residential segregation to the following five categories: evenness, exposure, concentration, centralization, and clustering. Evenness means the uniformity of distribution of different social groups across the city. Exposure refers to the degree of social contact, i.e. the possibility for everyday interaction between members of different social groups within the city. Concentration means the relative amount of space occupied by a group in the urban environment – if a group occupies a small share of the total area in a city, it's residentially concentrated. Centralization refers to the degree to which a group occupies the central areas of an urban area, while clustering, the last dimension, refers to the extent to which the houses, blocks or neighborhoods occupied by a particular group adjoin or cluster in the space (Massey & Denton 1988). Because schools typically get a great share of their students from their neighboring areas regardless of the existing student allocation policies, segregation inevitably ends up affecting their student compositions.

Recently, there has been more discussion about the exposure dimension: the form of segregation that reaches beyond the neighborhood, i.e. the residential context. Until recently, segregation was mostly studied from the viewpoint of residential structure, because it's easily measurable and observable. However, there is an increasing academic interest towards the segregation that occurs in different domains of daily life – in the workplaces, schools, hobbies, transport, social media, and others – the residential domain being just one of them (Boterman & Musterd 2016; van Ham & Tammaru 2016). On the level of an individual, these domains together form an activity space, consisting of the locations that individual visits to perform her daily routines (Schnell & Yoav 2001; Wong and Shaw 2011; Boterman & Musterd 2016; van Ham & Tammaru 2016). Activity space is an important concept for understanding the scale of exposure to diversity that people face in their daily lives. Not surprisingly, individual level ethnicity and income are strongly related to the exposure to diversity: according to a Dutch study, highest and lowest income groups have least exposure to diversity in their neighborhoods, workplaces and daily transport, which together form a great deal of these people's activity spaces (Boterman & Musterd 2016). As Boterman & Musterd argue, non-exposure to diversity will quite certainly block different groups of people the opportunities to come closer to each other, while also creating a risk of *estrangement* between different groups of people.

While school districting based on distance or transportation costs typically provides convenience for families and seems like a utility-maximizing choice, it effectively zones by socioeconomic status (Bouzarth et al. 2018). This happens because people of similar economic backgrounds tend to occupy the same neighborhoods. In other words, school district boundaries are mechanisms for translating residential demographic patterns into the ethnic and socioeconomic compositions of schools – a districting plan can either exacerbate, duplicate or challenge the homogeneous ethnic and socioeconomic compositions of many residential neighborhoods in the partition (Mickelson & Bottia 2010). The extent to which it is possible to challenge the patterns of residential segregation by careful school districting is dependent on the scale and dimensions of segregation listed before. If for example clustering is strong but at the same time the clusters are small and quite evenly scattered around the city, school districts can be made socially fairly even quite easily. But if the distribution of social groups is very uneven, and e.g. the city is divided into wealthy south and poor north, accomplishing a non-segregated school district plan becomes far more difficult.

What makes educational segregation such a troubling issue is a group of phenomena generally referred to as *school effects*. A large body of studies have shown that differences in student compositions tend to lead to differentiation of overall student performance and learning outcomes on both school and individual level (e.g. OECD 2004; Rumberger & Palardy 2005; Mickelson & Bottia 2010; Perry & McConney 2010; Bernelius 2011; Schwartz 2011; Rothwell 2012). This means that while the individual backgrounds of students can affect individual student achievement, so can the compositional characteristics of their school's student body. The compositional effects are revealed when the aggregate of person-level variables (including the school they attend) are related to outcomes even after controlling for the effects of individual characteristics (Rumberger & Palardy 2005).

The issue with school segregation is not small. The scale of the phenomenon is exposed in PISA-report from 2004, which concluded that while differences between countries are large, on average in OECD countries the schools with higher than average socioeconomic status perform better than would be predicted by their actual socioeconomic intake and schools with lower than average socioeconomic status perform below their expected value in PISA tests. According to the 2004 report,

> *"in the majority of OECD countries the effect of the average economic, social and cultural status of students in a school – in terms of performance variation across students – far outweighs the effects of the individual student's socio-economic background"* (OECD 2004).

On average in OECD countries,

> "*differences in the performance of 15-year-olds between schools account for 34 per cent of the OECD average between-student variance*" (OECD 2004).

In the light of these findings, one can quickly conclude that the situation is problematic from the viewpoint of equality of opportunity. If a student has no possibility or ability to affect which school they attend, e.g. for financial, social, educational or administrative reasons, the effect that a school independently produces on their performance is not in their power. Still, whether the problem – variation in learning outcomes produced by schools – needs actions to be taken against school segregation per se is another question. If the social composition affects students' performance because of its relationship to such aspects about the schools that are at least seemingly alterable, like resources (e.g. teacher-pupil ratios, physical school spaces, education materials) or structures and practices, then the problem may not be the segregation itself (Rumberger & Palardy 2005). In that case, increasing school resources and reforming school structures and practices may be sufficient to address the gap in equality of opportunities between students in low and high performing schools (Rumberger & Palardy 2005).

However, if the effects of student base's social composition cannot be explained by such school characteristics that can be changed or fixed, then the effects of segregation per se become the problem. Or, as Rumberger and Palardy (2005) explain,

> *"— even if the effects of segregation can be explained by seemingly alterable school characteristics but such characteristics appear to be triggered by the social*

> *makeup of the students served for example, educators and school officials consistently respond to high concentrations of poor minority students with lower expectations and a less challenging curriculum then segregation is again problematic."*

Peer effects are widely believed to influence children's school performance. According to peer effects theories, socially and educationally advantaged students positively influence the "pro-education beliefs, values, and behaviors of other students with whom they attend school through a variety of sociological, cultural, and psychological dynamics" (Mickelson & Bottia 2010). In other words, students who attend diverse schools benefit from social interactions with their more advantaged peers through their social capital and networks, cultural habits, values, beliefs, motivations and other pro-education attitudes and behaviors. But the effect of student composition can also come in other ways: in their 2005 study Rumberger & Palardy found four significant explaining variables behind the school effect: (1) teachers' expectations about students' ability to learn, (2) the average hours of homework that students completed per week, (3) the average number of advanced (college prep) courses taken by students in the school and (4) the percentage of students who reported feeling unsafe at school. All of these four aspects are seemingly alterable in theory, but besides being unsure in effectiveness, the actions and resources needed for making the changes in practice are likely to be far beyond just mixing the student compositions between schools by careful redistricting.

2.4.2 The role of school choice and selective migration in educational segregation

School choice is the process of active school selection practiced by parents and students to either access or avoid particular schools. According to multiple studies, parents tend to regard the schools with a high average socioeconomic status of the student base and with ethnic composition resembling their own as *most suitable* schools for their children (Henig 1990; Karsten et al. 2003; Bernelius 2013). In other words, the motivation behind school choice is often the attempt to access "superior" peer groups and avoid less privileged peer groups (Saporito 2003; Allen 2007; Musset 2012; Bernelius 2013; Bernelius & Vaattovaara 2016). Parents tend to make conclusions about the quality of school environments based on the imagined quality of their surrounding neighborhoods in terms of ethnicity and social status: according to Bernelius & Vaattovaara (2016), in Helsinki school choices have a clear link to the socio-economic and ethnic characteristics of the school's catchment area, and especially the rejection of particular schools seems largely consistent with socio-spatial segregation (Bernelius & Vaattovaara 2016).

While school admission policies in most OECD countries allow some flexibility in school choice within municipality borders, it's usually very limited by the availability of free places, since the schools must by default guarantee a place for each school-aged child living in their attendance zone (Musset 2012; OECD 2018). Because of this, in many countries the only way for families to guarantee a place in a particular school for their children is via *selective migration*, i.e. making residential choices based on the perceived quality factors of schools. The perceived quality of a school and its student composition have been shown to affect migration decisions of middle-class and wealthy families in particular (Brunner et al. 2012; Kosunen 2014; Danielsen et al. 2015). Practically this means that educational segregation is not just the outcome of residential segregation and school choice, but the link works also the other way around: school segregation actually tends to further exacerbate residential segregation via families' migration decisions. In systems with a determinative districting policy and limited school choice, access to high-performing schools is unequal by social status because of housing availability and cost: it's very difficult – if not financially impossible – for families of lower economic status to get access to the highest-performing schools via migration, while this opportunity always stays open for wealthy families (Rothwell 2012).

The association between perceived school quality and residential prices have been proved before in several studies (see e.g. Cheshire & Sheppard 2004; Gibbons & Machin 2008; Brunner et al. 2012; Rothwell 2012; Harjunen et al. 2018). For example, across the 100 largest metropolitan areas in US, the housing costs in areas near high-performing public schools have been shown to be on average 2.4 times as much as in areas near low-performing public schools, which translates to average home prices being $205,000 higher near well-performing schools (Rothwell 2012). The same has been also been discovered inversely: free school choice programs have been found to reduce residential segregation and enhance the attractiveness and prices of residences located near poorly performing schools (Brunner et al. 2012; Danielsen et al. 2015). According to a common argument by school choice advocates, free school choice can weaken the effects of residential segregation on educational segregation by loosening the link between the schools' and the neighborhoods' social compositions, and by allowing more educational opportunities for residents of lower-status neighborhoods (see e.g. Musset 2012).

Unfortunately, the free school choice policy does not seem to be able to tackle educational segregation but have even shown to exacerbate it (Saporito 2003; Musset 2012; Bernelius & Vaattovaara 2016; Yang Hansen & Gustafsson 2016; Kosunen et al. 2016). The reason behind this is that active school selection is mostly exercised by advantaged and achievement-oriented families (Musset 2012). Many studies show that middle-class families use school choice as a strategy for avoiding disadvantaged peer groups, which means that free school choice actually ends up enhancing the opportunities of the already well-off, while potentially worsening school segregation and harming families of lower socio-economic status (Saporito 2003; Allen 2007; Musset 2012; Malmberg et al. 2014; Bernelius & Vaattovaara 2016; Kosunen et al. 2016). While expanding free school choice policies don't seem to provide a shortcut to educational desegregation, careful districting may be a more feasible strategy to reduce the perceived quality differences between schools.

2.5 School districting and educational segregation in Helsinki

2.5.1 The legal framework for school districting in Helsinki

According to the Finnish Basic Education Act (1998/628), a municipality is obliged to arrange basic education for children living inside its borders. The compulsory education (grades 1 to 9) is entirely publicly funded, typically part of the funding coming from municipal taxes and part from government subsidies (OECD 2018a). General and vocational upper secondary education are also completely publicly funded, and the financial responsibilities are shared between the state and the municipalities. Even though private schools exist, they are also funded publicly and are not allowed to charge tuition fees or other fees. In Finland generally, only a fraction of comprehensive school pupils attends a private or state school.

According to the Basic Education Act's 6th section, the municipality is obliged to point a school for each school-aged child living inside its borders in a way that the child's school trips become as short and safe as possible:

> *"Education shall be arranged in municipalities so as to make pupils' travel to and from school as safe and short as possible in view of the habitation, the location of schools and other places of education, and public transportation. In the arrangement of preprimary education, account shall additionally be taken of the participating children's access to day-care services." (Amendment 1288/1999)*

Each child has a right to attend the exact school they are pointed to. Alternatively, a child has a right to attend another appropriate school pointed by the municipality, if they want to study a language that is not offered in the closest school's curriculum. Additionally, a child may also apply to any other school within the municipality's borders, and the education provider may accept the application if the school's capacity is not exceeded. In practice, the applications are always accepted if the school's capacity allows it.

In the framework set by the national regulations, the municipalities have extensive autonomy what comes to organizing basic education. The local education committees define the principles regarding schools' student intake and attendance planning that are not stated in the law (see e.g. Helsinki's Educational Administration's Rules of Procedure 2010, 5 §). These principles include, for example, the rules for setting school-specific student capacities for students starting their first grade and creating a school districting plan according to which the first graders are allocated to specific schools. According to a decision document of Helsinki's Education Committee (2014), the capacity-setting is done in order to guarantee regionally and educationally balanced development.

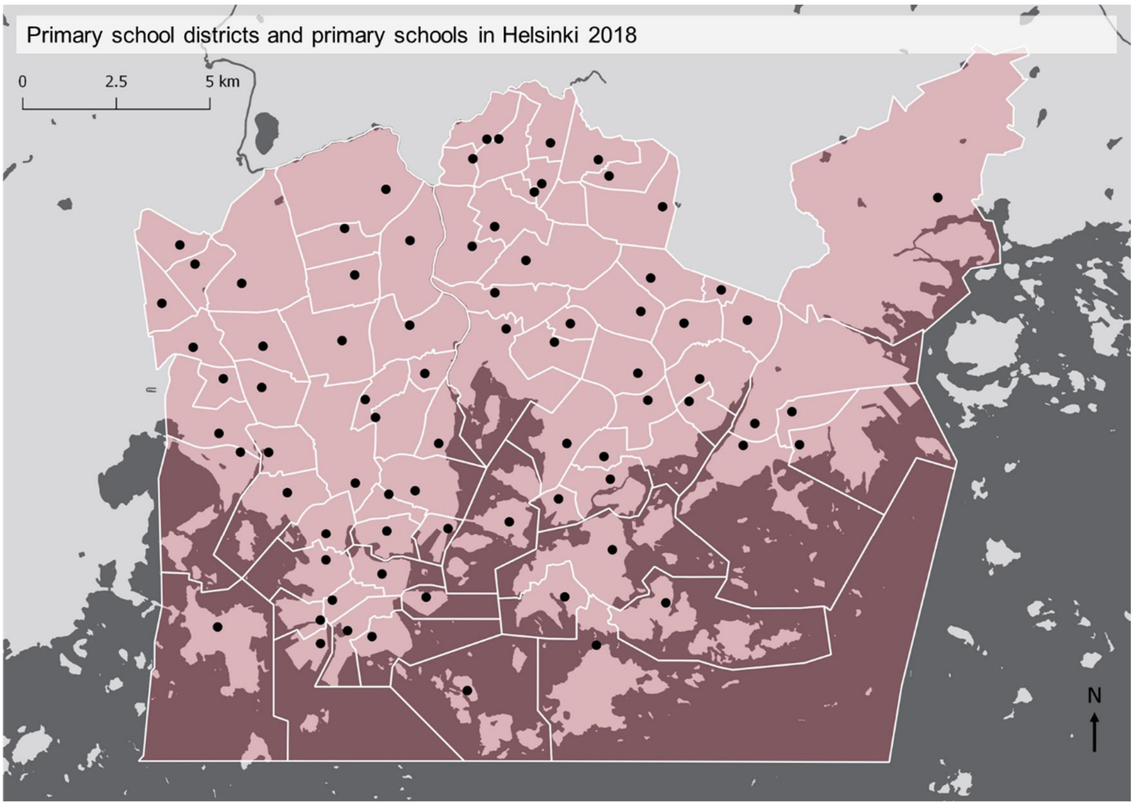


Figure 1. Finnish-language primary school districts (white borders) and corresponding primary schools (black dots) of school year 2018–2019.

The 2nd section of the Basic Education Act, which defines the objectives of education, states that "Education shall promote civilisation and equality in society and pupils' prerequisites for participating in education and otherwise developing themselves during their lives." While the law states that fostering equality in society is one of the main objectives of education, it does not define any objectives or guidelines related to e.g. the evenness of social compositions between schools (this kind of statement could be for example, that the school districts must be drawn in a way that fosters diversity within schools as much as possible). Instead, treating these issues belong to the scope of municipal autonomy and local authorities. From this point of view, the law weights efficiency (e.g. the school trips must be as short as possible) more than aspects related to the equality of educational outcomes. This may be partially because a) in Finland, school segregation is more local than a national problem, and b) the divergence of learning outcomes and schools' student bodies has been discovered quite recently in Helsinki (Bernelius 2011; Välijärvi et al. 2015).

2.5.2 Educational segregation in Helsinki

Finland has a reputation of excellent educational equality, and for a reason: the variance in student performance has been – and still is – low by international standards. Precisely, the variance has been more than 15 percent below the OECD average variance (OECD 2004). Also, the proportion of between-school performance variance has been about 10 per cent of the OECD average level, which means that on international standards, the effect of school on individual student's performance has seemed low in the light of PISA studies (OECD 2004). The problem with these numbers, though, was that they were based on the results of only a few schools and thus were inadequate to depict the variance of school and student performance inside the Helsinki metropolitan area (Bernelius 2011; Nissinen 2015). While on average the effect of school on individual students' performance in other parts of Finland has been around 3–10%, in Helsinki it has been as high as 18% (Kuusela 2006).

The more recent PISA studies have shown a drop in Finnish students' results, and what is most worrying, the bottom decile of schools saw a markedly larger drop in student performance than most of the schools, while some even fell below the OECD average (Kupari et al. 2013). At the same time, signs of independent school effects in Helsinki have been found in several studies (Kauppinen 2008; Bernelius 2011; Nissinen 2015). This has been largely attributed to a growing educational segregation, more specifically the differentiation of student compositions between schools (Bernelius 2011). These differences have been shown to form specifically as a consequence of residential segregation and active school choices by families (Bernelius & Vaattovaara 2016)

Since 1970s, the city of Helsinki has actively been mixing social socio-spatial patterns by forcing diversification of different housing types in its planning of neighborhoods. Part of this policy has been a practice of locating the city's own affordable housing projects in upscale residential areas. When combined with a fairly even distribution of wealth across the population, these policies have kept the differentiation of neighborhoods' social compositions on a relatively low level across Helsinki (Bernelius & Vaattovaara 2016). Despite of these efforts and low levels of segregation on international standards, ethnic and income differences between neighborhoods have been on a steady growth for the last 20 years (Kortteinen & Vaattovaara 2015; Bernelius & Vaattovaara 2016). More specifically, some neighborhoods have seen a substantial rise in wealth and income, while others have been quite stagnant in terms of income development (Vilkama et al. 2014). At the same time, the immigrant population has grown significantly and concentrated strongly on certain areas of the city (Vilkama et al. 2014).

These developments have led to significant spatial concentration of advantage and disadvantage by national standards, creating propitious conditions for growing educational segregation. Studies reveal that the ethnic differentiation among children in Helsinki is greater than ethnic differentiation among the whole population, while the differentiation between schools is even greater than the differentiation among children between residential or school districts (Riitaoja 2010; Bernelius 2013). According to Vilkama (2011) in certain neighborhoods “the share of the immigrant population is close to 40%, exceeding 50% for school-aged children”. In addition, these neighborhoods cluster close to each other, which also signals about a larger regional pattern. As an example can be mentioned the clustering of schools getting additional financial support (for a challenging operative environment) into the Eastern and Northeastern major districts of Helsinki (Bernelius 2013).

The picture below (figure 2) presents the share of population with immigrant background in statistical grid cells in Helsinki as measured in 2018. The black lines represent the school district division for school year 2018–2019 (as in figure 1).

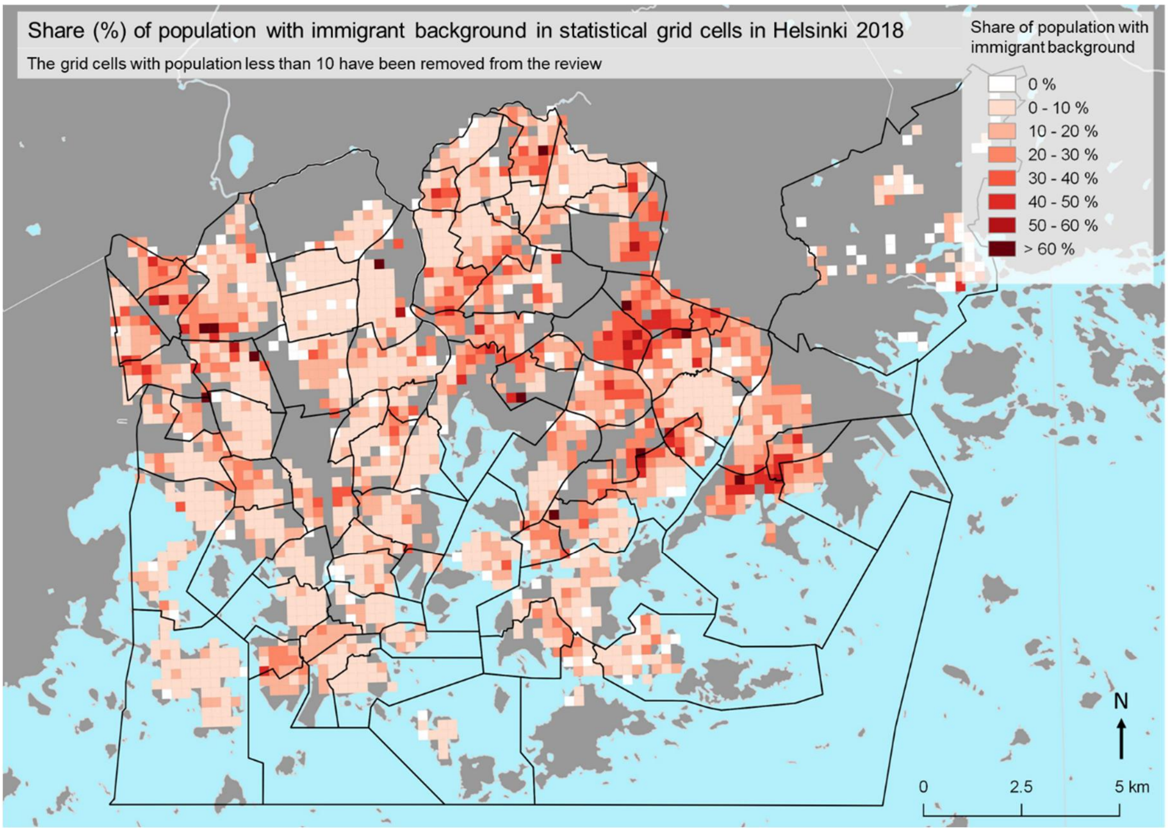


Figure 2. Share (%) of population with immigrant background in statistical grid cells (250m) in Helsinki, as measured in 2018. The black lines represent the school district borders of school year 2018 – 2019. The grid cells with population less than 10 have been removed from the review. Data source: Register data, SeutuCD'18.

One can quickly notice, that the spatial distribution of population with immigrant background varies locally but also contains a clear macro pattern. The next image (figure 3) presents the translation of this distribution into population patterns on school district level.

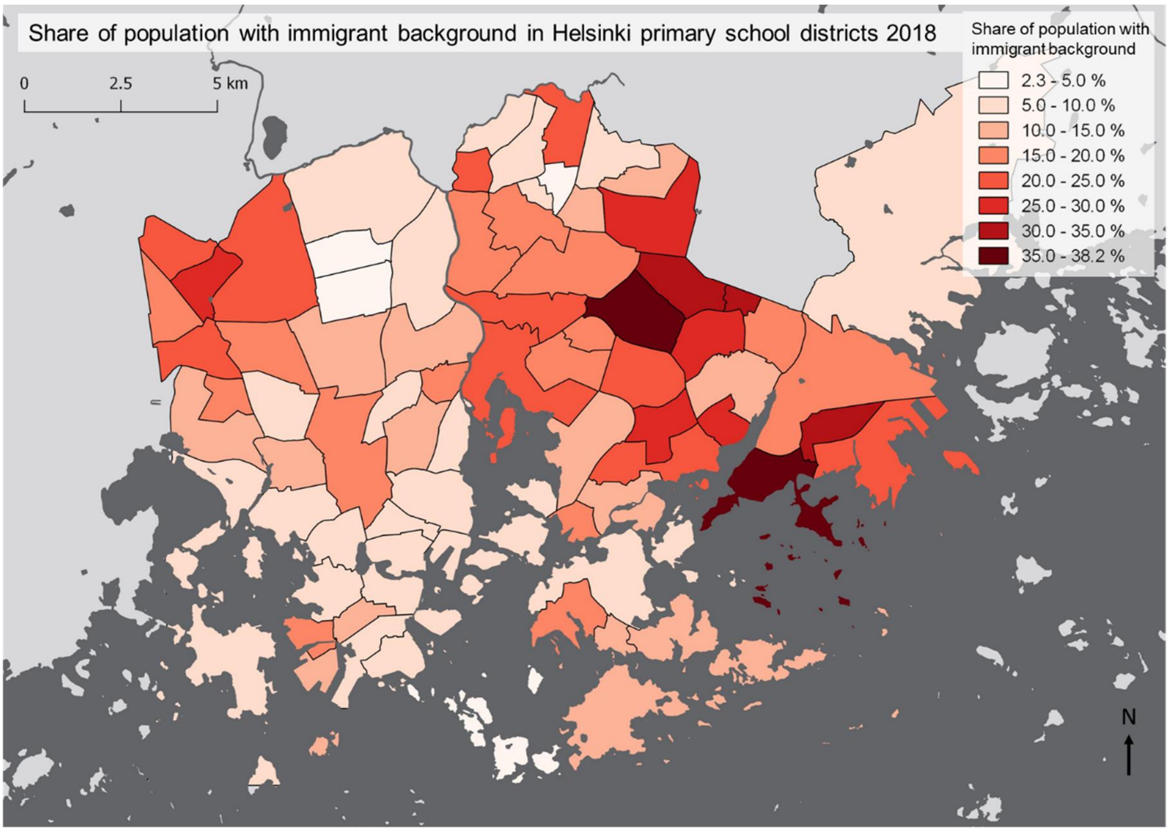


Figure 3. Share (%) of population with immigrant background in primary school districts of Helsinki as measured in 2018. Data source: Register data, SeutuCD'18.

In Helsinki, residential segregation is the single most important reason for differentiation of student compositions between schools (Bernelius 2013). Residential spatial patterns are reflected strongly on schools because the majority of Helsinki's school aged children attend the school which they are allocated to by the school districting plan. In 2016, a bit over two-thirds of primary school students and nearly half of the secondary school students in Helsinki attended their ”proximity school”, i.e. the school they were allocated to by the districting plan (Bernelius 2013; Bernelius & Vaattovaara 2016). Residential segregation is further exacerbated by families making migration choices based on perceived school qualities (especially the assumed pupil composition). The patterns of school rejection are also reflected in housing prices in Helsinki (Harjunen et al. 2018). For families with children, a large proportion of immigrants has been shown to be among the major reasons to reject a residential area in migration choices across Helsinki metropolitan area (Kuukasjärvi 2013).

Another reason behind educational segregation in Helsinki is the free school choice policy: as stated before, in Finland parents are free to apply for a place for their child in any school within the municipality borders, and in most cases a school approves an application if free places exist. In Helsinki, the school choices have been shown to follow the internationally familiar pattern of middle class parents avoiding less advantaged peer groups (Kosunen 2014; Bernelius & Vaattovaara 2016). This habit further aggravates school segregation: in their 2016 analysis Bernelius & Vaattovaara discovered that school choice produces an independent effect for student compositions between schools, differentiating them further from each other. They found that

> *"a high percentage of immigrants (R = 0.40) and poorly educated adults (R = 0.30) are particularly statistically significantly linked to the school being rejected in school choices, whereas high income level and a high education level of adults (R= -0.35) are linked to a smaller likelihood of the school being rejected in the choices."*

According to Kuusela (2006), drawing the school districts such that the school capacities would be filled, i.e. the schools would have minimal number of extra places for students applying from other districts, would be the most effective way to control and prevent school segregation alongside with residential planning policies.

Helsinki has long been working for improving the equality of education, both by careful school district drawing and directing targeted funding for schools located in less advantaged areas. The schools receiving additional targeted funding are defined based on an index of Positive Discrimination, which is calculated based on a school catchment area's social features and features of the school's student base. These features include share of population with immigrant background (measured as the percentage of non-native Finnish or Swedish speakers attending the school), average level of parental education (the percentage of adults without education past basic schooling in the catchment area), average income level (household's average income in the catchment area), and popularity of the school (number of students rejecting the school compared to the number of students attending the school but living in some other school's catchment area) (Silliman 2017). The targeted funding policy have turned out to have a large positive impact on low-performing native students and students of immigrant backgrounds (Silliman 2017). More specifically, according to Silliman, “native students are 3 percentage points less likely to drop out of education after middle school and that students from an immigrant background are both 6 percentage points less likely to drop out of education as well as 7 percentage points more likely to attend the academic track of upper-secondary school as a result of the [positive discrimination] policy”.

Despite of the promising effects of the positive discrimination policy, the differences between school districts and schools’ social compositions remain large in Helsinki. While targeted funding may weaken the effects of school segregation on students, its effectiveness in preventing or reversing school segregation has not been proved. Instead, redrawing school districts can have an instant effect on the student composition differences between schools. On the other hand, inequality of educational outputs may require inequality of inputs even, when the school districts are drawn as carefully as possible regarding student compositions. Careful district planning and targeted funding are not mutually exclusive, but rather complementary strategies in the prevention of school segregation and negative school effects, as the benefits from exposure to diversity remain significant.

3. Methods: a pilot model for school district optimization in Helsinki

3.1 Study setting & data

The aim of this study is to (1) pilot school district optimization in the context of Helsinki, and (2) find out, how much the social variation between the current primary school districts in Helsinki could be reduced if constraints of compactness, travel times and student capacity were relaxed, and the districts reshaped. For this purpose, I have developed a computational optimization approach and tested it on Helsinki school district plan. The resulting district division demonstrates the scale in which the school districts of school year 2018-2019 are optimizable in terms of social compositions. The optimization parameter I have chosen to use in this pilot model is *the share of population with immigrant background*, since while it has been found to correlate significantly with socioeconomic measures of disadvantagedness in Helsinki (Vilkama 2011, Kortteinen & Vaattovaara 2015), it is also particularly significantly linked to a school being rejected in school choices (e.g. Bernelius 2013; Bernelius & Vaattovaara 2016). Also, as an already binary variable it is the simplest one, at this phase, to base the optimization function on. The approach chosen is experimental, and does not pursue to find a single, objectively optimal redistricting solution directly applicable to real world.

The model developed in this study utilizes geographic data from the following sources:
1) SeutuCD 2018, which is a GIS dataset published annually by Helsinki Region Environmental Services Authority (SeutuCD 2018). The data consists of municipal register data of Helsinki Metropolitan Area's municipalities, presented as unit-specific geographic information. The information extracted from SeutuCD 2018 for this study are population, population with immigrant background (measured as population with some other mother tongue than Finnish or Swedish), and the amount of 7-year-olds, all presented as a building-level geographic information.
2) The school districting plan for school year 2018–2019, provided as a GIS dataset by the Education division of Helsinki and Urban Environment Division via Helsinki

Region Infoshare (The Education division of Helsinki 2018). The dataset consists of primary school district polygons (78 pcs).
3) MetropAccess distance matrices (Toivonen et al. 2014), a geographic dataset consisting of travel times and distances along the road network from each statistical grid cell (YKR cell, 13,230 pcs) to all other grid cells inside Helsinki metropolitan area. The modes of transport cover travel by foot, public transport and private car. The model utilizes walking distance data from this dataset for measuring accessibility between schools and places of residence.

The model developed in this study can be run on different weights and levels of relaxation of the original districts' parameters, and in chapter 4 (results chapter) I will demonstrate how altering of the weights will affect the optimization result, i.e. the optimized school districting plan. In order to separate these alternative runs with alternative weights from the two main runs of the model, I will call the main runs simply the *main example optimizations*. In the following sections, I will first give general-level overview of the model and it's functioning, and then explain it's working and data generation process in more technical detail.

3.2 Model summary

The model developed in this study, inspired e.g. by the work by Farmer et al. (2004), utilizes a version of heuristic optimization technique known as local search. Local search techniques (e.g. simulated annealing, see Browdy 1990; Altman 1997; D'Amico 2002; and hill climbing, see Altman 1997; desJardins et al. 2007) have been regarded as one of the most feasible approaches to districting and other combinatorial optimization problems, not only due to their simplicity and relatively low computational complexity but also because of their good performance on discrete, nonlinear, multi-attribute problems (Ingber & Rosen 1992; D'Amico 2002; Russell & Norvig 2003; Ricca & Simeone 2008; Rincon-Garcia et al. 2013). In practice, the algorithm applied in this study is a mix of hill climbing and simulated annealing: it advances by searching and choosing the best values in the neighborhood (as in hill climbing, see desJardins et al. 2007), but on each iteration also has a diminishing probability for choosing a random value instead (as in simulated annealing, see e.g. Ingber & Rosen 1992; Russell &

Norvig 2003). By sometimes choosing a random value, the procedure is trying to avoid getting stuck in a bad local optimum. This, though, is not enough for solving the optimization problem, as with declining randomness included the algorithm is stochastic and likely ends up finding different local optimums on each run. To approximate the global optimum, the algorithm is run multiple times with diminishing amount of random selection, and the best outcome is selected among the set of possible solutions. When the amount of times the procedure is run is sufficient, the final selection from the plethora of districting assignments should represent a nearly optimal configuration. In both main example optimizations of this study, the model was run 100 times, meaning that in both the “optimal” districting plan was chosen from 100 semi-optimal divisions.

As the optimization piloted is a real-world case, many mundane aspects must be taken into account in the calculation. While the aim is to produce a district division that is socially as even as possible, it cannot happen completely on the cost of travel distances to schools, ignore the current school infrastructure or break the school districts into scattered blocks. As a contrast to the term optimized by the objective function – the share of population with immigrant background –, these mundane aspects are treated in the model as hard constraints (the constraint method, see Eiselt & Laporte 1987; Malczewski & Jackson 2000), which is a common approach in multi-criteria optimization problems (Eiselt & Laporte 1987; for example applications see Bozkaya et al. 2003 and Bação et al. 2005). The constraints of the model on the general level are as follows:

- Travel distances are restricted to not surpass a district’s individual, original maximum travel distance multiplied by a chosen factor (e.g. 1.15 times the original maximum distance).
- The areal contiguity of the districts are preserved, meaning that the districts must be traversable from any point to any other point inside the district without crossing the district’s border on the way.
- The district’s student base may not exceed the district’s initial student base multiplied with a chosen factor (e.g. 1.15 times the original student base).

The optimization is not started from random seeds or even from the schools' locations, but instead the model uses the original school district configuration as a starting point, trying to gradually improve the districts by swapping small areal units, blocks, between them. The requirement for circumferential compactness has been relaxed in the model at this stage as it – regarding Helsinki's residential segregation – very likely conflicts with the model's main objective.

3.3 Model logic

3.3.1 Logic overview

The logic of the optimization model can be best understood through a metaphor. Let's imagine that the optimization model is a board game, played by the school districts. The game consists of rounds, and as in board games usually, during a round each district has one turn. The board game board consists of small residential blocks that are divided between the districts as the game starts. During its own turn, a district can choose one residential block in its neighborhood, and dissolve that block to itself – or in other words, adopt one block from another district. The adoption is always made according to a set of specific rules. The game continues until none of the districts want to adopt new blocks, and the situation stabilizes. When this happens, the game ends, and the result – the division of residential blocks between the school districts – is the optimized district plan.

The objective of the game from the districts' point of view is trying to bring their own, initial optimization value (the share of population with immigrant background) closer to the global mean, so on each turn they aim to choose a new block that has the greatest averaging impact on their own value. There are a few twists in the game, though. In the beginning of their turn, a district must first roll a dice. If the district has "bad luck" and the dice gives a large number, it must adopt a random block instead of "the best" block in its neighborhood. But on each round, the dice is replaced with a new one, and the new dice always has a smaller largest number than the previous one. This means that probability of having to adopt a random block is larger when the game starts, but it gradually diminishes and reaches 0 as the game advances, so towards the end of the game the districts can adopt only best blocks. The randomness factor makes the game stochastic, which means that each game ends differently, even though the players are always the same and they always start with the same initial division of residential blocks. If we step out of the game metaphor for a moment, this means that on each run, the model finds a different local optimum. But because the aim of the optimization is to approximate the global optimum, the model is run multiple times and the best result chosen among the multiple resulting divisions. So, the game is played multiple times in a row, and only the best result of all games is chosen among the multiple end results.

3.3.2 Rules concerning block adoption

Whether the district can adopt a new block on its turn, a random or the best, depends on a set of rules that define which blocks the districts can dissolve to themselves. A district can’t adopt a new block if none of the neighboring blocks satisfy the rules. As adopting a block results in removing the block from another district, the rules cover also aspects related to the consequences of removing a block from another district. To minimize the amount of calculation needed, the computationally most simple rules are verified first, and only the blocks satisfying the rule are selected on the next round of rule-verifying. During a district’s turn, the rules are verified in the following order:

1. According to the first rule, a district can only adopt a block that it touches. This refers to the blocks that the district shares a border – a line segment – with (“rook” contiguity). However, the block must not already belong to the district (the district can’t adopt the blocks that already belong to it).
2. According to the second rule, a district cannot adopt a block that contains a school building. This ensures that the school buildings stay in their original districts.
3. The third rule states that a district can’t adopt a block that would make it exceed its individual student limit by a chosen marginal. For example, in the two *main example optimizations*, this marginal is 12.5 % in the first run and 20 % in the second run, and it’s calculated based on the initial amount of 7-year-olds living inside the district.
4. The fourth rule ensures that a district can't adopt a block that is too far away. What is too far away, is measured as a maximum road-network based walking distance. For example, in the two *main example optimizations*, the maximum walking distance is

based on the district's original maximum walking distance from a block to the district’s school, multiplied by a factor of 1.125 in the first run and 1.20 in the second run.
5. The fifth rule ensures that a district can't adopt a block that would break some other district's contiguity. If adopting a block would result in cutting another district into pieces, adopting the block is forbidden.

If the district is to adopt a random block, it adopts it at this point by drawing it among the blocks that satisfied the above rules. The drawing is made by generating a random number and choosing the block according to it. If none of the blocks satisfied the rules listed above, the district doesn’t adopt a new block. On the other hand, if the district is to adopt *a best block*, it will further to testing the following two rules:

6. The sixth rule is that the district can only degrade another district's value (take another district’s value further away from the mean) when taking a block from it, if it will result in decreasing the numerical distance between the values of these two districts. As this rule is the most complex one to understand, it will be further explained in the next chapter (3.3.3).
7. The seventh rule is that from the blocks that satisfy all the rules listed above, the district always chooses the block that improves its own value the most, i.e. has the greatest averaging impact on its own value.

After selecting a block that satisfies the 6th and 7th rule, the district dissolves the block to itself, which results in the block being removed from its previous host. If none of the blocks satisfied the 6th and 7th rule, the district doesn’t adopt a new block.

3.3.3 The logic behind rule 6

The sixth rule is that the district can only degrade another district's value (take another district’s value further away from the mean) when taking a block from it, if it will result in *decreasing the numerical distance between the values* (share of population with immigrant background) of these two districts. The point of this rule is ensuring in a Rawlsian manner (e.g. Browdy 1990) that the worst-off districts enjoy a kind of a veto right – they may improve themselves on the cost of others and the whole plan, unless it would result in other districts big losses while bringing very little help for the worst-off districts trying to improve themselves. More concretely, when a district is trying to improve its own social value, it cannot do it on the cost of a district that is even further away from the mean (and on the same side of the mean). On the other hand, a district trying to improve its own value may do it on the cost of another district, if afterwards the two districts have a smaller distance between them in terms of the optimization value.

To demonstrate this rule, let's imagine there are two districts, A and B, that are on the same side of the mean value (10). Let's say that the district A has value 12 and B has value 14. Rule 6 makes the situation asymmetric in the sense that now district B can improve its value on the cost of district A to a defined extent, but district A cannot do the same. Let's first imagine that A (the better-off district) has a turn. Now, A cannot choose any block from the district B (the worse-off district) that would result in B's value shifting further away from the global mean, because it would increase the numerical distance between them. For example, if adopting a block from B would result in A’s value shifting from 12 to 11, but as a result B's value would change from 14 to 15, the numerical distance between them would grow from 2 to 4, which is against the rules. Now let’s imagine the same initial starting point of A having value 12 and B having value 14, but this time it's district B's turn to adopt a block. Now, B can actually adopt a block from A, even if this would result in A's value shifting from 12 to 13 and B's value shifting from 14 to 12, which would actually make B the "better off" district on the cost of district A. This is because numerical distance between 12 and 13 is smaller than between 12 and 14.

The 6th rule is tested with the following formula:

$$|(I_{oc} - \mu_g) - (I_{sc} - \mu_g)| > |(I_{on} - \mu_g) - (I_{sn} - \mu_g)|$$

where

μ_g is the global mean of districts’ social indexes
I_{oc} stands for the current social index of the block’s current district (”other” district)
I_{sc} stands for the current social index of the district in turn
I_{on} stands for the social index of the block’s current district (”other” district) after

adoption

I_{sn} stands for the social index of the district in turn after adoption

If the formula's left side is strictly greater than the right side, the rule is satisfied, and the district can move on to test the final (seventh) rule. In the above example, the case of district A adopting a block from B would not satisfy the rule, as $|(14 - 10) - (12 - 10)| \ngtr |(15 - 10) - (11 - 10)|$, but district B adopting a block from A would satisfy the rule, as $|(12 - 10) - (14 - 10)| > |(13 - 10) - (12 - 10)|$. Using the numerical distance in rule 6 confirms that extremities have a reasonable weight in the optimization. This means that the districts cannot greedily optimize themselves on the cost of a single district, but rather the model is also optimizing the range of the districts on the global level.

3.4 Model implementation & specifications

On the code level, the model consists of classes and functions. The classes define the data structures, attributes and class functions that construct the core geographic objects used in the optimization: the school districts and the blocks. In the model, the blocks are formulated as instances of class Block, which is a representation of residential block unit. The school districts are formulated as class SchoolDistr, which represents a primary school attendance district. In this particular implementation of this model, the instances of class SchoolDistr are polygons representing Helsinki's school districts of school year 2018 – 2019, and the residential blocks that make up the game board are 250 meters * 250 meters polygons based on Finnish census grid (YKR). The model should preferably be implemented with a more natural representation of residential blocks than a census grid but defining such a division and producing the data for it would require a study of its own.

Both classes have class attributes that contain the information required in optimization. The block instances contain information about their geography (coordinates), total population, amount of 7-year-olds, population with immigrant background, current school district and whether they contain a primary school building. The districts contain information about their geography (coordinates), a dictionary of residential blocks currently belonging to them, a road-network based distance matrix, information about the original student base size and original maximum travel distance to the district's own school, as well as their current social structure (total amount of 7-year-olds and the share of population with immigrant background living inside them).

The model's main function consists of three nested loops (figure 4). The outermost loop is responsible for running the individual games. The actual optimization happens in an inner while-loop that runs the game and its rounds until the situation stabilizes and no more blocks are swapped between the districts. The rounds make up the third loop, inside which the districts make their turns and adopt new blocks.

On each round, the district in turn will check its neighboring blocks and choose among them either a random or the best block and adopt it, depending on the value of the *dice*. The dice is a random number generated between *a ceiling and a floor value* (in a classic dice the ceiling value is 6 and the floor value is 1). The district will have to adopt a random block if the random number generated is *larger than the break value* (in a classic dice the break value could be for example 4). The break and the floor values are fixed values throughout all games and rounds, but the ceiling value changes in two dimensions.

The first dimension is inside a single game – the model's second loop. When the game starts, the ceiling value is always at its highest, and in the end of each round, the ceiling value drops. In a classic dice, this would mean that on the second round the dice would only have numbers from 1 to 5, on the third from 1 to 4, et cetera. As the break value, 4, keeps constant, the probability for having to choose a random block drops on each round. While on the first round the probability was 2/6 (= 1/3), on the second round it drops to 1/5. On the third round, as the ceiling value drops to 4 – to the level of the break value – the probability for having to choose a random value drops to zero, since it is no longer possible to get values over 4. After this, the ceiling keeps dropping further and the districts deterministically improve their values until a stable state is reached and no more adoptions are made.

The second dimension in which the ceiling value changes is inside the outermost loop. The outermost loop runs the game for a fixed amount of times (in the main example optimization 100 times) with a ceiling value that grows every time a new game is started. This means, for example, that in the first game the ceiling value

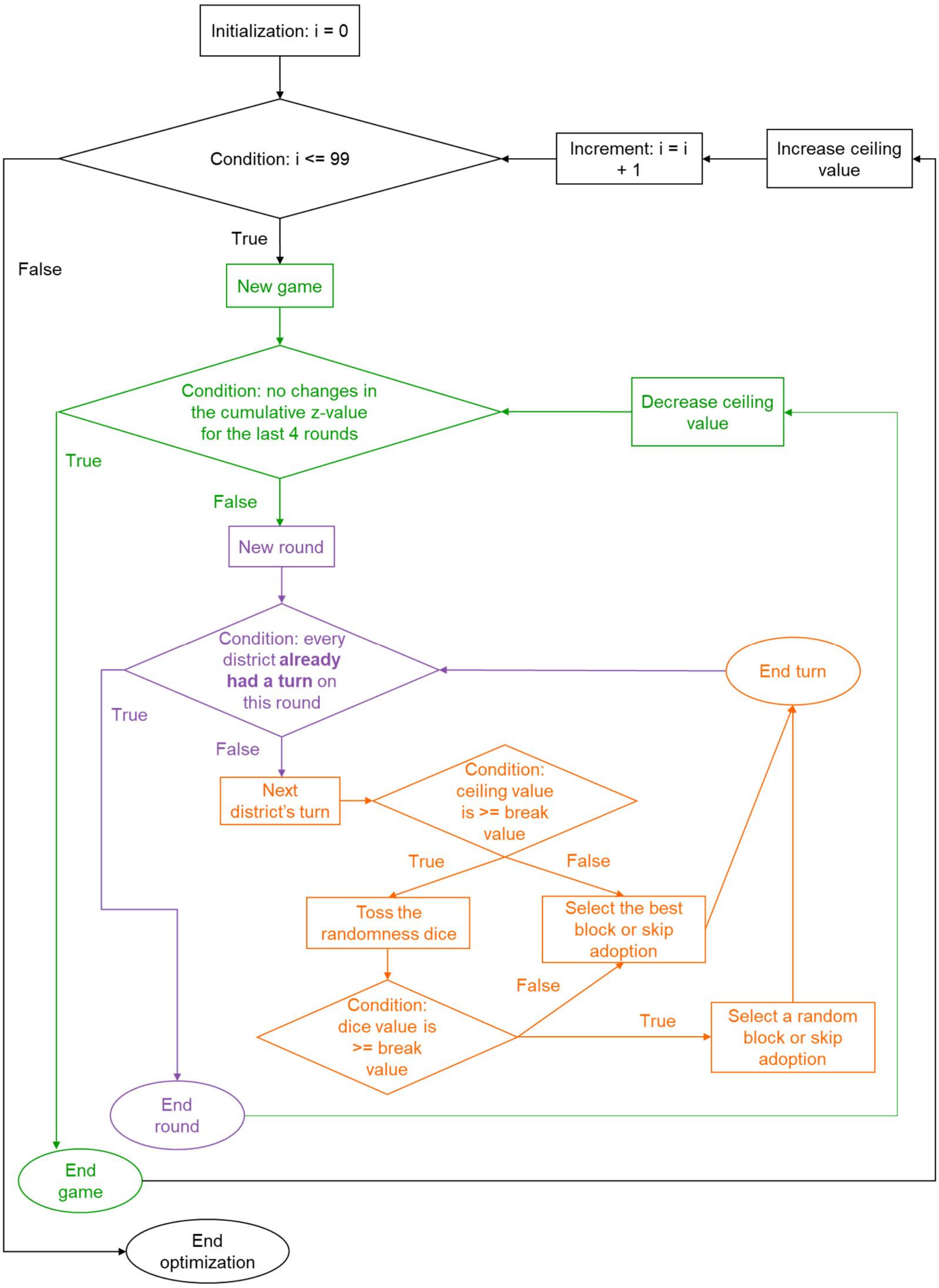


Figure 4. Flow chart of the optimization's main function. In the chart, black color represents the runs, green color represents a single optimization (game), violet color represents a round and orange color represents a turn. The graph structure presents how the processes are nested: a turn happens inside a round, a round happens inside a game, and the game is run 100 times by the optimization.

starts its ascend from value 6, but in the second game it starts its ascend from value 7, in the third game from value 8, et cetera. The first game is always deterministic: the ceiling value starts descending from the level of the break value. But as more games are played, the diminishing randomness parameter always starts from a higher value, and the following games have more rounds with a possibility of having to adopt a random block. For example, in the main example optimizations, in which the game is played 100 times, the ceiling value is 50 on the first game's first round. Also, the break value is 50, and as there is no possibility of getting values over 50 by tossing a dice with values from 0 to 50, no random adoptions are made. As the rounds inside the first game further, the ceiling value keeps dropping lower. Each time a new game is started, however, the ceiling value starts lowering from a higher value. When the second game starts, the ceiling value is 51 on the first round. And when the 50th game is starts, the ceiling value is 87 on the first round. In the last game, the 100th game, the ceiling value starts lowering from value 124. This means that as the ceiling value drops by 5 on each round, the last game has total 15 rounds where random adoptions can happen, but the probability for random adoption is diminishing as the rounds in this game further. When the ceiling value reaches the break value after 15 rounds, the rest of the game is played without random adoptions.

A game is ended when no more districts adopt new blocks on their turns. The value tracking this and the advancement of the optimization, hereafter referred to as the *cumulative z-value*, is calculated as the sum of the differences between the of share of population with an immigrant background of each zone and the average share of population with immigrant background of all the zones, standardized with the standard deviation. This function can be expressed in mathematical terms as follows:

$$\sum_{j} \left| \frac{P_j - \mu}{\sigma} \right|$$

where P_j represents the share of population with immigrant background in the j^{th} district, μ represents the average share of population with immigrant background per district, and σ represents the districts' *population standard deviation* in share of population with immigrant background. When this value stops improving, i.e. the optimization reaches a plateau and stays on it for four rounds, the game is ended.

The model has a constant space complexity, as it only keeps in memory the current state of the districts and the currently best game result – the paths to these are not stored. As a new game is played, it's result is only saved if it's better than the previous best result – and the previous best result is erased when a better one is generated. The time complexity, on the other hand, is a function of games, rounds, districts and blocks. The number of rounds needed in a single game depends on the model restriction factors (i.e. the maximum travel times), the amount of randomness included (how many rounds there is with random adoptions), and the number of blocks in the game. The execution time grows steeply as the number of blocks and randomness are increased and as the restrictions are relaxed, as does the problem's solution space.

3.5 Specifications for the example optimizations

3.5.1 Data preparation

To create suitable dataset for instances of class block, the school district IDs were first joined spatially to seutuCD points from school district polygons. The seutuCD points were then joined to fishnet to form a spatially contiguous polygon pattern. As the available distance matrices are based on Finnish census grid (YKR-grid) polygons, the fishnet was created with matching dimensions to the census grid. After that, the IDs from the census grid were joined to the fishnet. Finally, only the polygons covering Helsinki and land (not sea) were selected in the dataset.

The instances of class SchoolDistr (the school districts) were generated by grouping the blocks by their original school district ID's. Because some of the original school districts have complex forms, a few districts did not generate as contiguous in the district generation process. The non-contiguity was partly caused by a few small islands, and partly by small complex forms in the original districts causing some blocks to become imperfectly attached to their respective districts. The problem was solved by removing 17 blocks from the block data and generating the districts without these blocks. In the final dataset, there are 3,703 blocks and 78 school districts. The specs of the removed 17 blocks are presented below.

Table 1. This table presents properties for the 17 blocks that were removed from the district configuration. Data source: Register data, SeutuCD'18

	Population	Population with immigrant background	Number of 7-year-olds
	505	154	1
	540	44	10
	442	40	4
	11	4	0
	125	3	2
	40	2	0
	24	1	1
	134	1	0
	0	0	0
	0	0	0
	0	0	0
	0	0	0
	0	0	0
	0	0	0
	0	0	0
	0	0	0
	0	0	0
Total:	1821	249	18

3.5.2 The example optimizations

First, a sensitivity analysis was run on the model to compare alternative outputs generated with alternative restriction weights and to test, which levels of randomness produce the best outputs. The districts were optimized 16 times with mutually growing restriction weights for maximum travel time and student capacity, ranging from 1.025 times the original to 1.40 times the original, adding 0.025 between the optimizations. Inside each optimization, the model was run 20 times. The ceiling randomness parameter ranged from 50 (the first run value and the break value) to 126 (the last run), growing by 4 on each run. Inside the individual runs, the ceiling value decreased by 5 on each round.

Based on the results from the sensitivity analysis, the parameters for the main example optimizations of the model were decided. In the two main example optimizations, the model was run 100 times. The ceiling randomness parameter in the runs ranged from 50 (the first run) to 124 (the last run). The break value used was 50, and the ceiling value increased by 0.75 after each game. Inside the individual runs, the ceiling value decreased by 5 on each round. The maximum distance was restricted to not surpass 1.125 times the district’s original maximum distance in run A and 1.20 times in run B, and the amount of 7-year-olds was restricted to not surpass 1.125 times the original amount of 7-year-olds in run A and 1.20 times in run B.

The runs of the model were executed on a laptop with an Intel Pentium 2020M processor at 2.4 GHz and 6.4 GB DDR3 RAM. Running the optimization 100 times took approximately 20 hours, which means that on average, a single run on the model with this computer took approximately 12 minutes.

4. Analysis / Results

4.1 Sensitivity analysis

In the sensitivity analysis the optimization was run multiple times with gradually relaxing restriction weights for maximum travel time and student capacity (ranging from 1.025 to 1.4 times the districts' original values). The purpose of this analysis was to test, how sensitive the optimization result is for the level of constraint relaxation. The results from sensitivity analysis are presented in the enclosed table (table 2).

Table 2. This table presents the results for the sensitivity analysis. The presented values are absolute values, i.e. the values represent the different district divisions' actual statistics before and after optimization, with gradually growing levels of constraint relaxation.

Level of constraint relaxation	Cumulative z-value	Min %-value (without Suomenlinna)	Max %-value	Range (without Suomenlinna)	Standard deviation
(Original districts)	63.46	3.76%	38.96%	35.20%	8.11%
1.025	59.03	3.65%	36.95%	33.29%	7.63%
1.050	56.25	3.71%	35.37%	31.66%	7.29%
1.075	54.08	4.20%	36.16%	31.96%	7.08%
1.100	53.87	4.84%	35.81%	30.97%	7.12%
1.125	51.79	5.40%	35.13%	29.73%	6.75%
1.150	51.98	5.38%	35.61%	30.23%	6.91%
1.175	51.41	5.35%	35.28%	29.93%	6.78%
1.200	51.31	5.78%	33.36%	27.57%	6.71%
1.225	51.18	4.85%	35.62%	30.77%	6.78%
1.250	50.25	5.26%	34.97%	29.71%	6.74%
1.275	49.25	5.02%	33.17%	28.15%	6.49%
1.300	48.41	6.18%	33.61%	27.42%	6.41%
1.325	48.09	5.29%	33.91%	28.62%	6.40%
1.350	48.60	5.29%	33.17%	27.87%	6.43%
1.375	47.67	5.15%	33.31%	28.15%	6.38%
1.400	47.55	5.01%	33.23%	28.22%	6.28%

The table's values describe the variation in the share of population with immigrant background within a particular district division. In the table's first row are presented the values for original, unaltered district division and in the other rows are the values for optimized district divisions accordingly. The first column, *level of constrain relaxation*, states the used constraint factor which is used to multiply the districts' original maximum walking distances to their schools and original amount of 7-year-olds to form the reference values that cannot be surpassed in the optimization. *Cumulative z-value* shows the districts' values' cumulative standardized distance from the global mean. The minimum and maximum values refer to the district division's minimum and maximum values, i.e. the districts having smallest and greatest shares of population with immigrant background. The range refers to the absolute difference between the values of the districts having smallest and greatest shares of population with immigrant background, and the standard deviation means simply the standard deviation of the districts' shares of population with immigrant background. Suomenlinna island, which has the lowest share of population with immigrant background of all the districts has been excluded from the calculation of minimum and range values, as it is not connected to the mainland via land route and thus could not be optimized.

The following table (table 3) shows the proportional change of the optimized district divisions' values compared to the original district division as the level of constraint relaxation is increased.

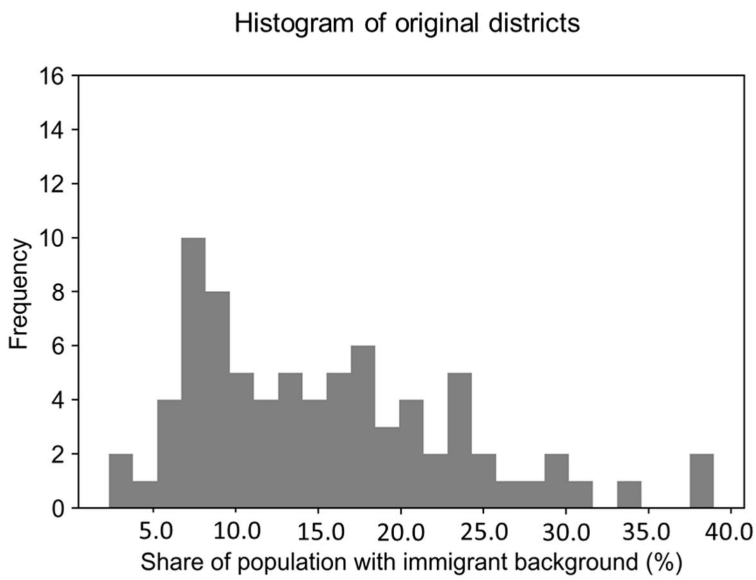
Table 3. This table presents the results for sensitivity analysis as a proportional change from the original district division as the level of constraint relaxation is increased.

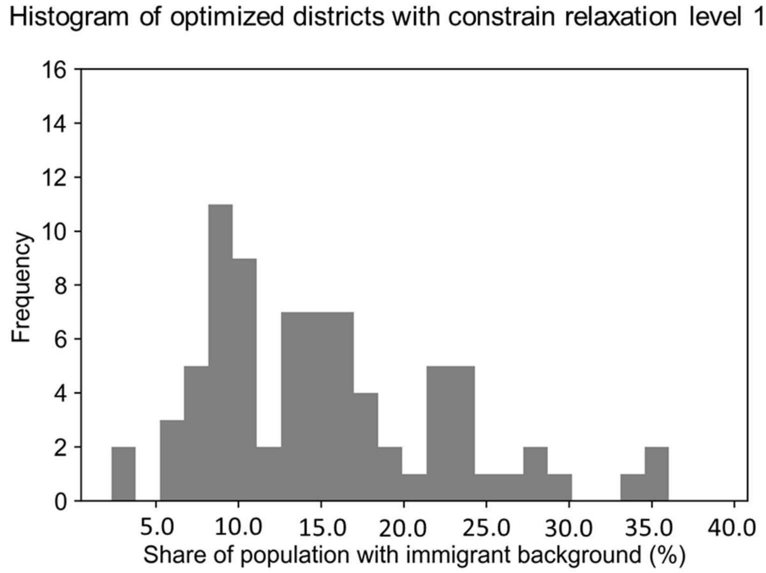
Level of constraint relaxation	Change of cumulative z-value	Change of Min %-value (without Suomenlinna)	Change of Max %-value	Change of Range (without Suomenlinna)	Change of Standard deviation
(Original districts)	0%	0%	0%	0%	0%
1.025	-7%	-3%	-5%	-5%	-6%
1.050	-11%	-1%	-9%	-10%	-10%
1.075	-15%	+12%	-7%	-9%	-13%
1.100	-15%	+29%	-8%	-12%	-12%
1.125	-18%	+44%	-10%	-16%	-17%
1.150	-18%	+43%	-9%	-14%	-15%
1.175	-19%	+42%	-9%	-15%	-16%
1.200	-19%	+54%	-14%	-22%	-17%
1.225	-19%	+29%	-9%	-13%	-16%
1.250	-21%	+40%	-10%	-16%	-17%
1.275	-22%	+33%	-15%	-20%	-20%
1.300	-24%	+64%	-14%	-22%	-21%
1.325	-24%	+41%	-13%	-19%	-21%
1.350	-23%	+41%	-15%	-21%	-21%
1.375	-25%	+37%	-15%	-20%	-21%
1.400	-25%	+33%	-15%	-20%	-22%

As the table shows, a relatively large change in the cumulative z-value, maximum value, range and standard deviation happens already at the smallest constraint relaxation level. This suggests that significant improvements could be already made with relatively small changes to the districts. After the first rounds, the values keep optimizing as the level of constraint relaxation is increased, but the change slows down.

Also, the optimization results' values don't improve deterministically as relaxation is increased, but despite of the clear improving trend they show significant random variation, also to worse. This is a consequence of both the stochastic nature of the model and the increasing number of possible solutions as the constraints are relaxed, but it also confirms that running the model 20 times isn't a sufficient number for the main example optimizations. Theoretically, if the model is run sufficient number of times the results should only improve (though the magnitude of improvement can vary significantly) as the constraints are relaxed further.

Particularly notable is, that compared to the original districts and the districts optimized with a constraint relaxation level of 1.40, nearly half of the optimization has happened in terms of cumulative z-value already with the relaxation level of 1.05. The following histograms show how the distribution of the districts change with relaxation levels 1.05 and 1.40 compared to the original districts.





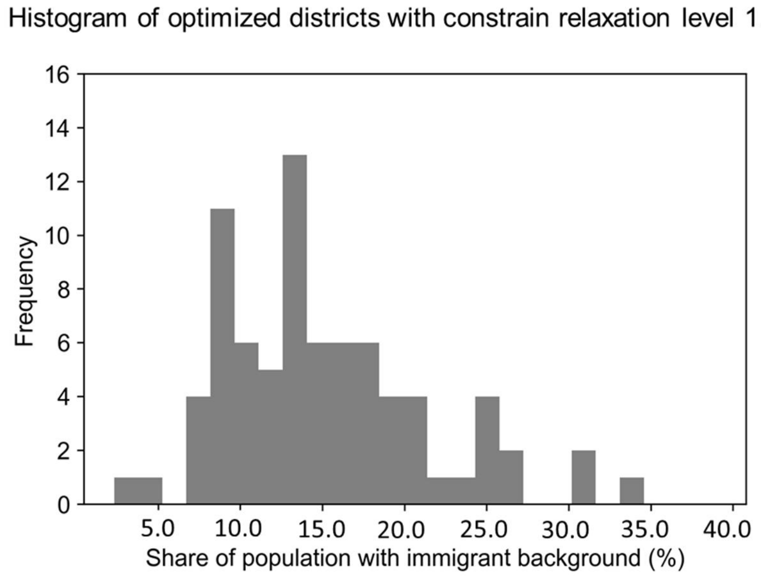


Figure 5. The histograms of original district division and optimized district divisions with constraint relaxation levels of 1.05 and 1.40.

The histograms demonstrate how the distribution of districts narrows as a result of optimization, and the districts' shares of population with immigrant background start to center between 10% and 20%. While three districts in the right tail clearly separate from other districts in both the original and optimized divisions, the worst-off districts also see a clear improvement in both optimization results. The same convergence can be seen in the maps below, presenting the forms that the optimized districts take on both levels of constraint relaxation.

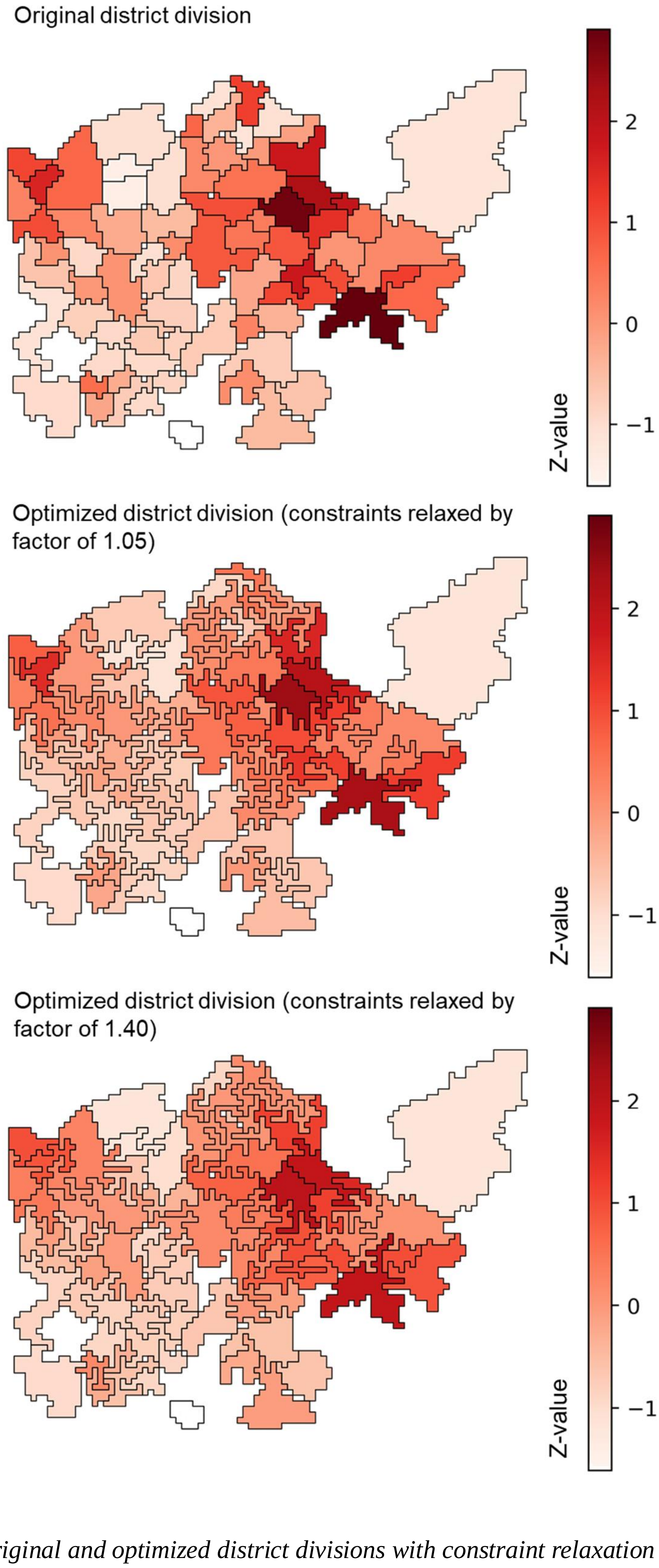
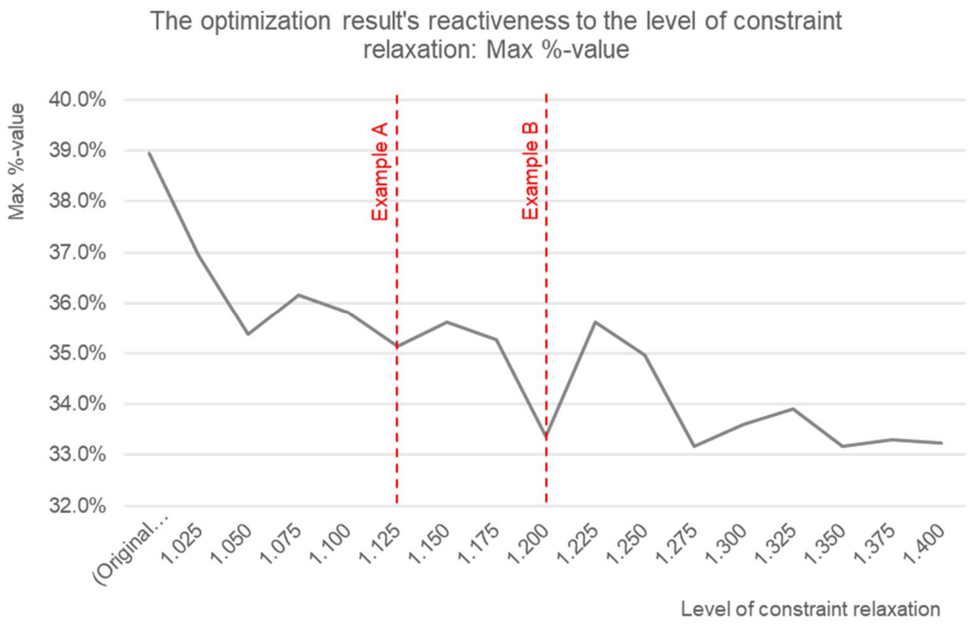


Figure 6. Maps of original and optimized district divisions with constraint relaxation levels 1.05 and 1.40. The coloring of the maps represents the z-value of the districts, calculated as the district's values difference from the global mean and standardized by the standard deviation. Data source: Register data, SeutuCD'18.

The following graphs (figure 7) are representations of the previous tables' second and fourth columns, and they describe the optimization result's reactiveness to the level of constraint relaxation from the cumulative z-value's perspective and maximum value's perspective. In both graphs, the form of the line shows a slowing level of optimization as the constraint relaxation level increases.



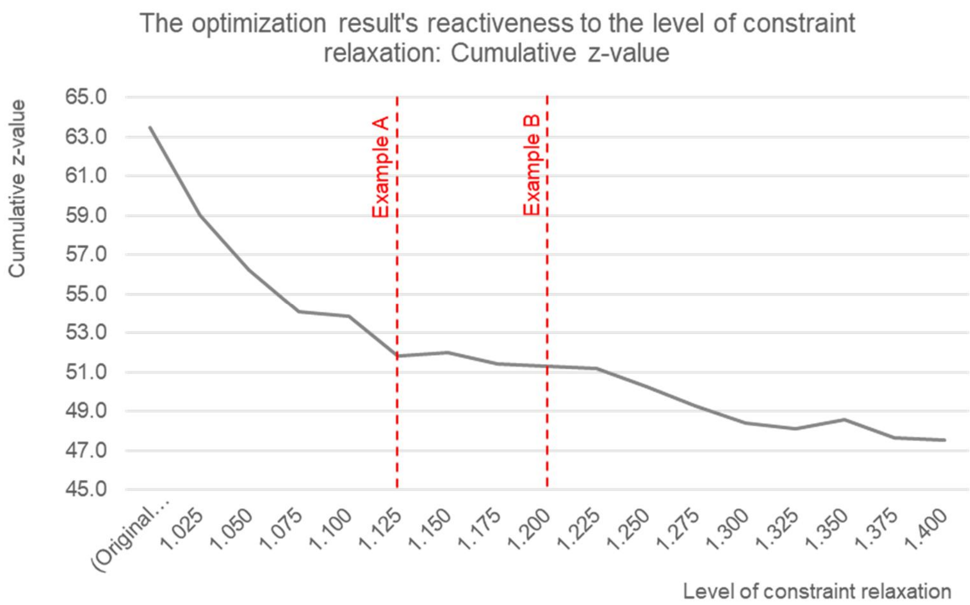


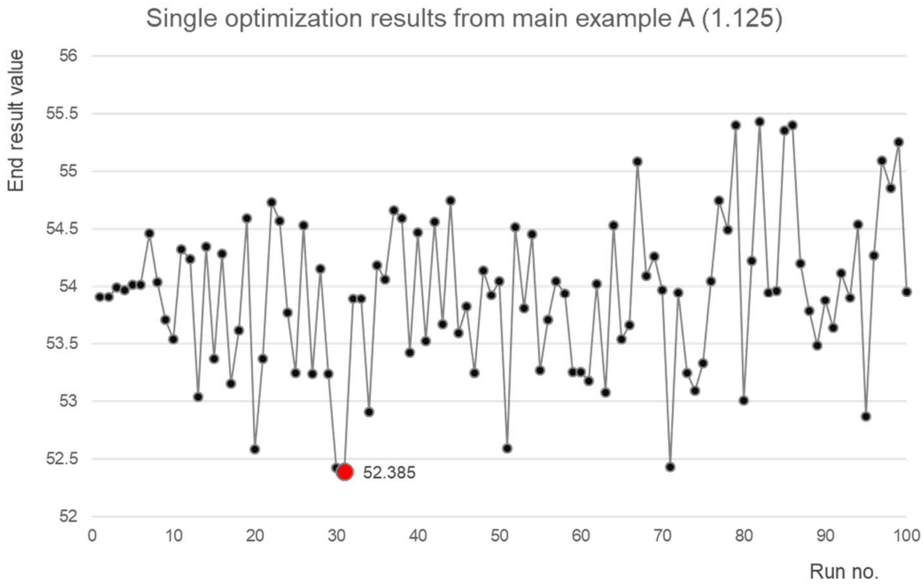
Figure 7. The optimization result's reactiveness to the level of constraint relaxation from maximum value's perspective (the topmost graph) and the cumulative z-value's perspective.

As the graphs and the tables show, the optimizations with constraint relaxation levels 1.125 (=12.5%) and 1.20 (=20 %) provide relatively effective results. These two relaxation levels produce small but significant optimization peaks compared to the next and previous relaxation levels. While this is partially the result of randomness and small amount of runs inside single optimization, these optimization peaks also suggest that these levels of relaxation can provide good results in relation to the compromised walking distances and school capacities. Both relaxation levels are also somewhat realistic regarding the real-world context, and thus these relaxation levels, highlighted in the above graphs with red vertical lines, were chosen for the two main example optimizations. The example optimization with relaxation level 1.125 is hereafter referred to as example optimization *A* and the other optimization with relaxation level 1.20 is hereafter referred to as example optimization *B*.

4.2 The main example optimizations

As specified in the chapter 3.5, both main example optimizations A and B include 100 individual runs of the model with altering levels of randomness. The following graphs (figure 8) present the progress of the optimization as the final cumulative z-values of individual runs. In example A, the best result was generated on 31th run, and in example B, the best result was generated on 28th run. In the graphs, the runs that generated the best result in terms of cumulative z-values are highlighted with a red point.

As can be seen in the graphs below, the variability in the 100 run results inside both optimization examples is significant but not extremely large. Additionally, none of the run result values in the graphs are significant outliers, which shows that the number of runs included in the optimizations have most likely been sufficient. Compared to the first, deterministic result (there is no randomness included in the first round), the best runs’ end results are significantly smaller, showing that random adoptions have pushed the districts beyond bad local optimums.



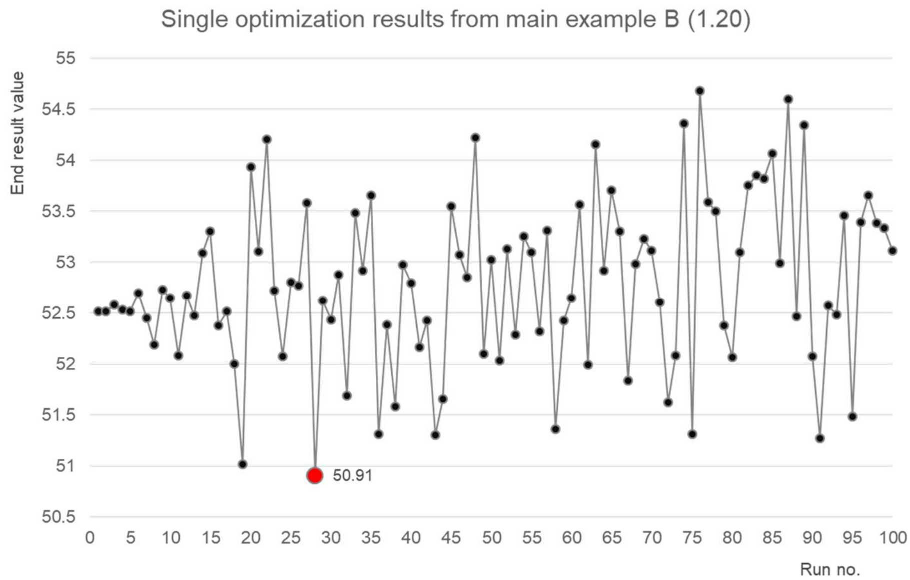
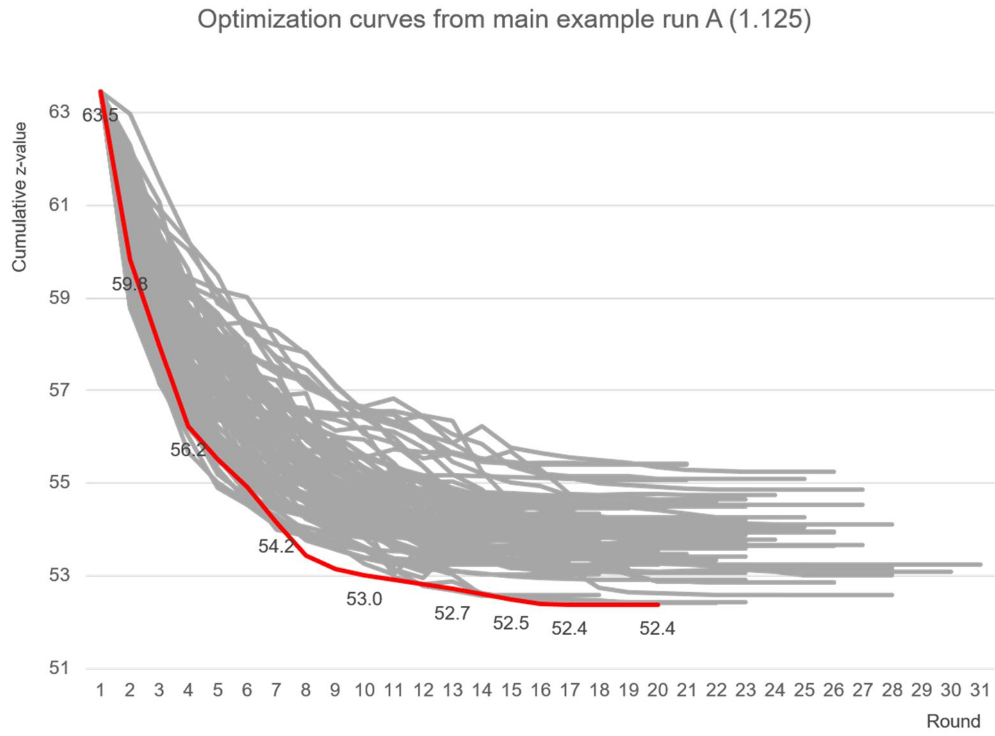


Figure 8. In example A, the 31st run generated the best result, while in example B, the 28th run generated the best result.

In the graphs below are also presented all the optimization curves for both example optimizations, and the highlighted curve is the optimization curve of the best run. In example A, the best run diverges from the other runs fairly early (around round 8), whereas in example B the divergence happens fairly late (around round 15).



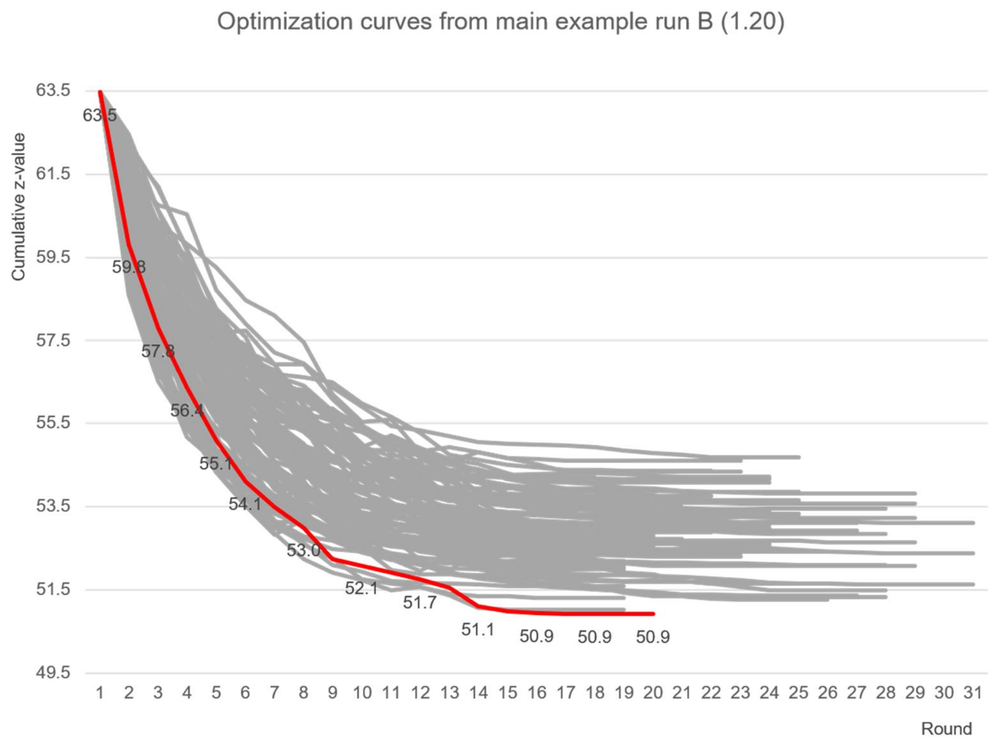


Figure 9. Optimization curves of all 100 runs in example optimizations A and B. The red line shows the optimization curve of the run generating the best result.

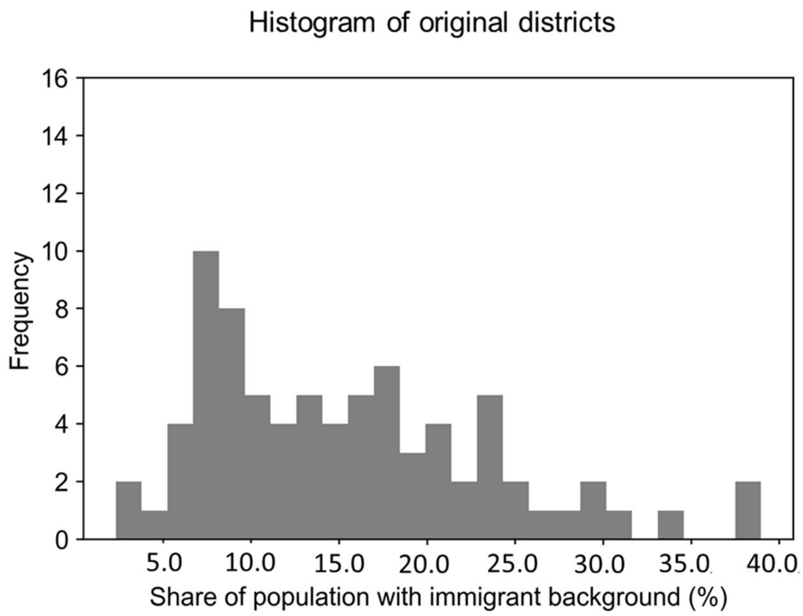
The optimization curves are formed based on the cumulative z-value of the districts on each round. Cumulative z-value is used as a tracking value of the model, and its function is further explained in chapter 3.4.

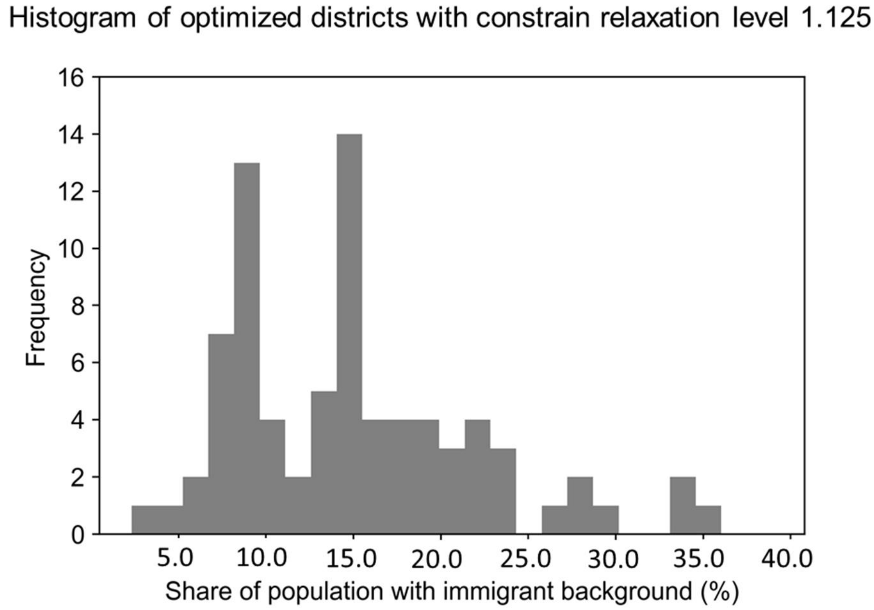
Table 4. Key values from the main example optimization results and the original district division. The topmost section presents the absolute values, the middle section presents the absolute differences (A's and B's values minus the original values) and the bottom section presents these differences as proportional to the original district division.

	Original	Example run A	Example run B
Absolute values			
Min value	3.8%	4.2%	5.3%
Max value	39.0%	35.5%	34.8%
Range	35.2%	31.3%	29.5%
Standard deviation	8.0%	6.9%	6.7%
Cumulative z-value	61.85	50.77	49.30
Absolute difference from the original			
Min value	-	0.5%	1.6%
Max value	-	-3.4%	-4.1%
Range	-	-3.9%	-5.7%
Standard deviation	-	-1.1%	-1.3%
Cumulative z-value	-	-11.07	-12.55
% difference from the original			
Min value	-	12.7%	42.2%
Max value	-	-8.8%	-10.6%
Range	-	-11.1%	-16.2%
Standard deviation	-	-13.6%	-16.6%
Cumulative z-value	-	-17.9%	-20.3%

The key values of the original district division and the optimized district divisions A and B are presented above in table 4. As in previous tables, Suomenlinna island has been removed from the review. The values of the original district division are presented in the second column. As can be observed in the table, the original district division's disparities in the share of population with immigrant background are large, ranging from 3.8% to 39.0%. In the example optimization A, the range of the share of population with immigrant background drops from 35.2 percentage points to 31.3 percentage points, equivalent to 11.1% drop. The cumulative z-value, which is the aggregate measure of district's variation from the global mean, drops from 61.85 to 50.77, which equals 17.9 % proportional decrease. The standard deviation drops by 1.1 percentage points, equivalent to 13.6% proportional decrease, and the maximum value decreases by 3.4 percentage points, which equals a proportional drop of 8.8%. These improvements were achieved with 12.5% relaxation to the original district division's maximum travel times and school student capacities.

With the 20% relaxation level in example B, the values optimize further, but the drop is not proportionally as large in relation to the relaxation level, as it is in example A. In example B, the range of the districts's values drops from 35.2 percentage points to 29.5 percentage points, meaning a 16.2% proportional drop. The standard deviation of the districts drop from 8.0 percentage points to 6.7 percentage points, equivalent to 16.6% proportional decrease, and the maximum value drops from 39.0% to 34.8%, equivalent to 10.6% proportional decrease. The cumulative z-value drops from 61.85 to 49.30, which equals 20.3% proportional decrease.





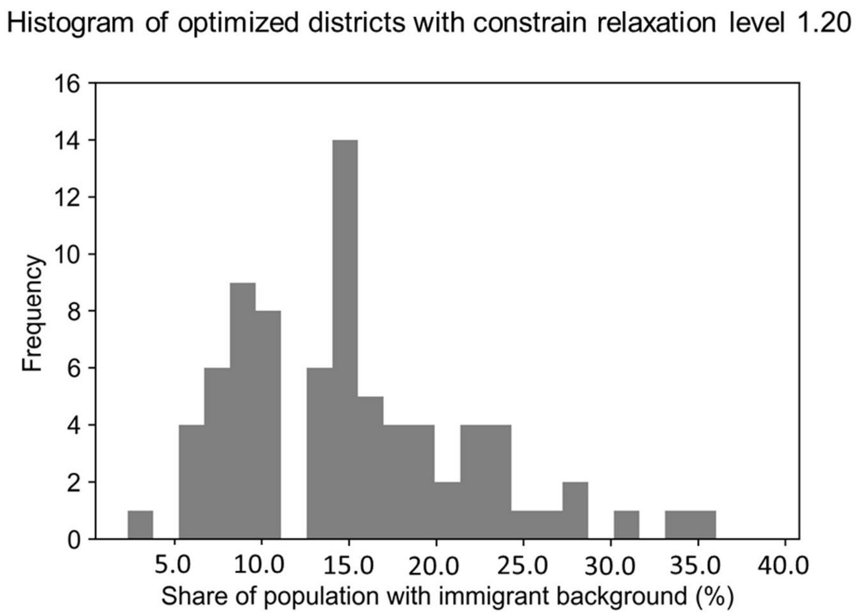


Figure 10. The histograms of original district division and optimized district divisions. The first histogram shows the frequency distribution of the original district division, and the second and the third show the frequency distributions for example A and example B respectively.

The above histograms show the frequency distribution of the districts in terms of the share of population with immigrant background. The variance of the districts is visually smaller in the optimized district divisions as compared to the original, as the values of the optimized districts accumulate towards 10 and 15 percent. Still, three districts holding the right tail persist clearly separated from the rest of the districts in both optimization results, despite of their values' clear improvement.

In the following figures (11 and 12) are presented maps of the optimized district divisions together with a map of the original district division. The value plotted with a red color in the maps are the districts' z-values, calculated as a difference between a district's share of population with immigrant background and the average share of population with immigrant background based on all districts, standardized with the standard deviation. As can be observed in the map of the original division, the districts with large share of population with immigrant background are clustered in the North-Western and Eastern parts of Helsinki. The regional dimension of segregation in Helsinki, i.e. the regional disparity of the placement of population with immigrant background, makes the optimization difficult especially from the most disadvantaged districts' point of view, as they are surrounded by other disadvantaged districts. Despite of this, significant improvements are achieved in the values of these districts in example optimizations A and B, and also the differences across all districts are visibly flattened.

In both example optimizations, the forms of the optimized districts turn up complex and tentacular. This happens because the compactness of the districs is not restricted in the optimization at this stage, but instead the optimization is allowed to happen on its cost. Some of the districts turning up most complex are also regionally the most evened-out: for example in the northernmost area, a group of districts with relatively high differences in their original values have blended and optimized substantially. For some districts, however, the non-inhabited areas seem to form a barrier that prevents further optimization: the maximum travel time constraint is reached before inhabited blocks with desired impact are found. This seems to have happened at least in the Norteastern and the Northwestern regions, where differences between some districts have remained surprisingly large.

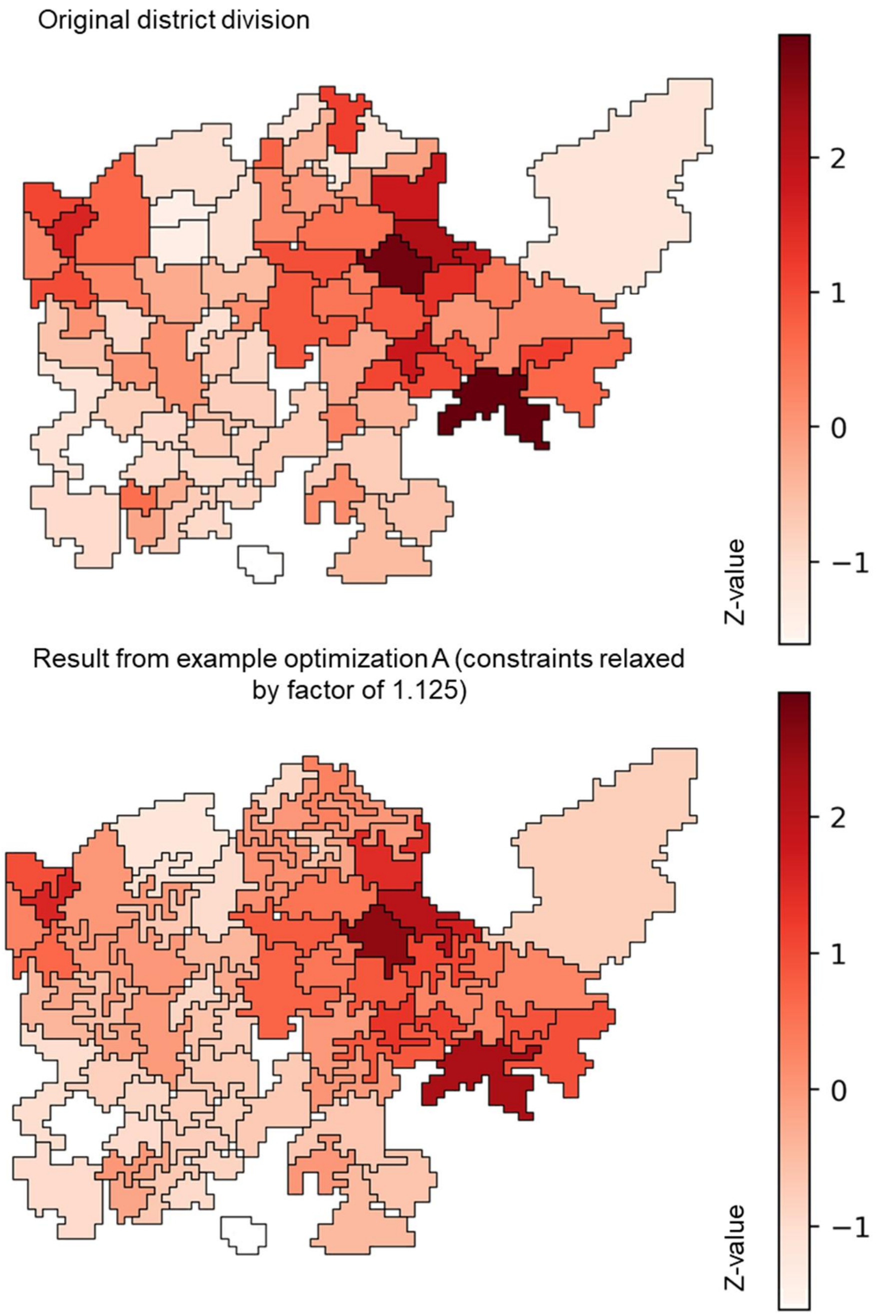


Figure 11. Maps of original district division and optimized district division A. The coloring of the maps represents the z-value of the districts, calculated as the district's values difference from the global mean and standardized by the standard deviation. Data source: Register data, SeutuCD'18.

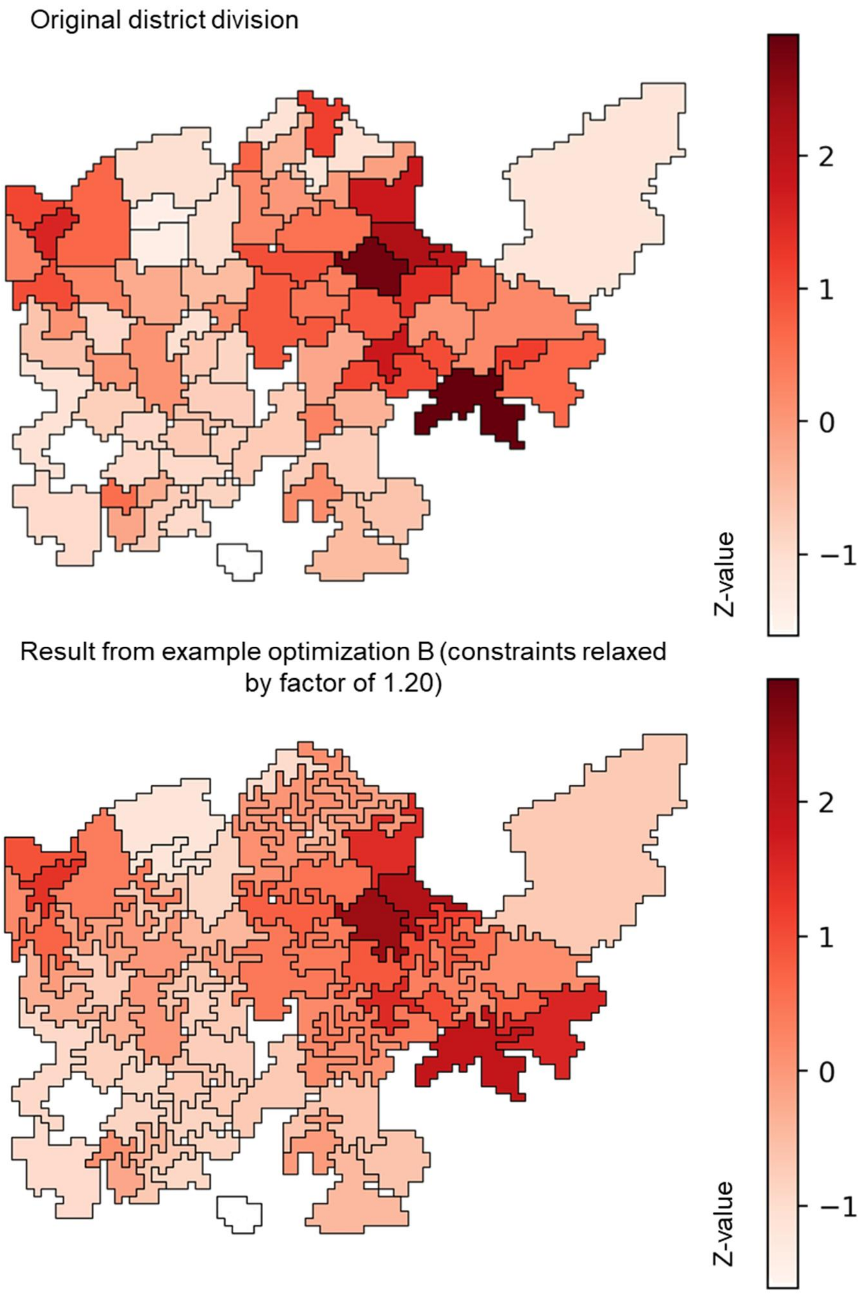


Figure 12. Maps of original district division and optimized district division B. The coloring of the maps represents the z-value of the districts, calculated as the district's values difference from the global mean and standardized by the standard deviation. Data source: Register data, SeutuCD'18.

The effect of the uninhabited areas combined with travel time restriction can be observed by comparing the map of Helsinki presenting inhabited statistical grid cells and the share of population with immigrant background (figure 2) to the maps presenting the absolute change (percentage points) in district's respective values after optimization (figure 13). As can be observed, the greatest improvements in terms of reverting towards the global mean happen inside regions that don't have large spatial discontinuities in inhabitation (e.g. the northernmost region).

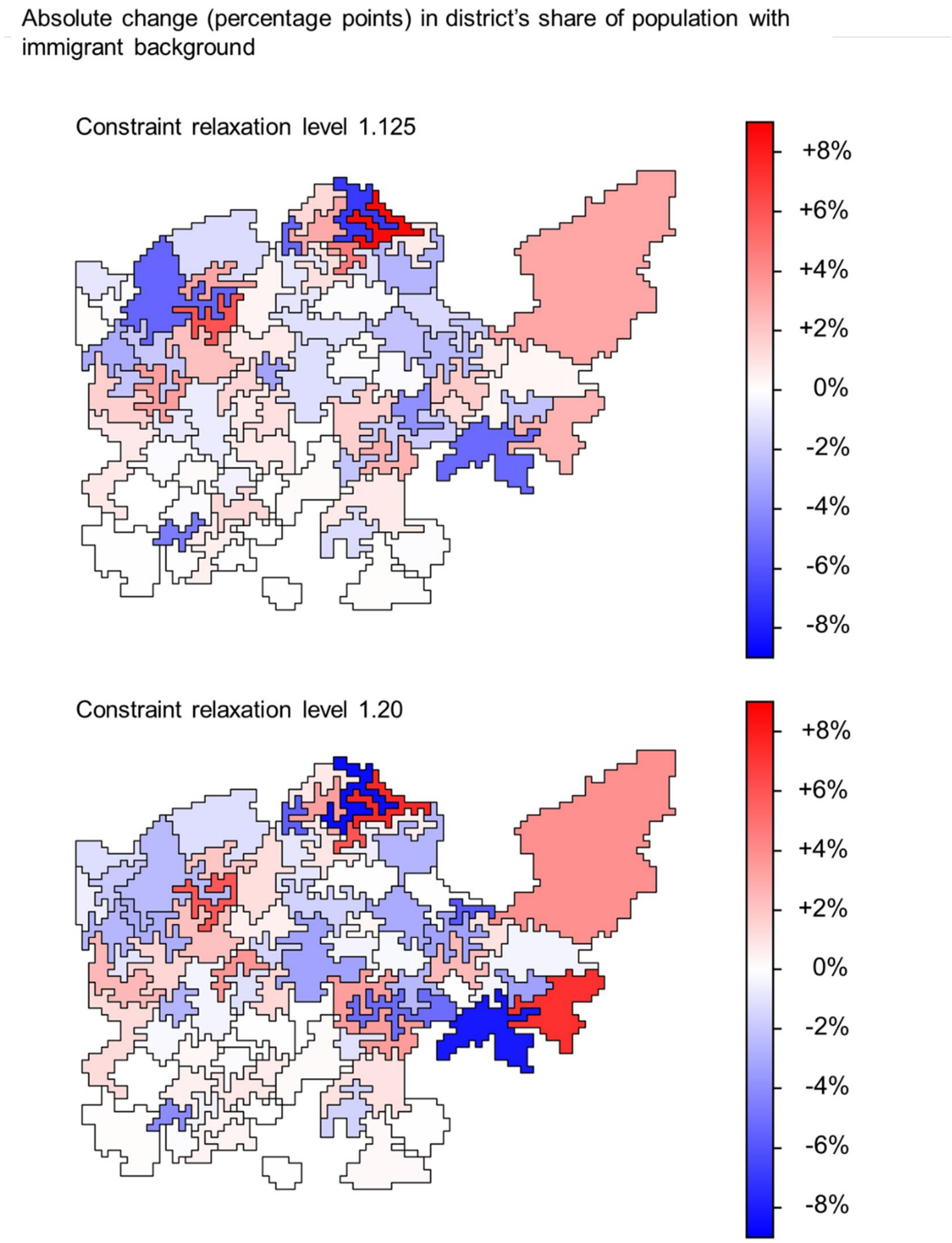


Figure 13. The absolute change in districts' share of population with immigrant background, measured in percentage points. The red color represents increase, and the blue color decrease, while the intensity of the color represents the magnitude of the change. Data source: Register data, SeutuCD'18.

5. Discussion

5.1 The quality of the results: compactness, fairness & legitimacy

The results of this study show that optimization of school districts in Helsinki is possible, and substantial improvements can be reached with relatively small changes to the districts from travel distances' and existing school network's point of view. When looking at the results, though, the inevitable question that rises is about the districts' complex forms – or in other words, their lack of compactness. The question about whether compactness is a desirable feature in a school district or not and whether compactness should be included in the optimization as a parameter or not isn't simple. There is no absolute need for compactness, neither from the legal, the technical, nor from the systemic point of view. The complex forms don't break anything on the technical or administrative level. Actually, the complex and tentacular district forms are the most classic example about gerrymandering: the strategic manipulation of a district division for achieving a particular objective - which is exactly the aim of this study.

Nevertheless, a geometry-oriented vision of a school district as a compact, circular catchment area makes the tentacular districts feel unnatural. The desire for compactness is in great part psychological – we prefer easily understandable solutions over those that can't be figured out at a glance. Another aspect related to this is that complex forms are, compared to compact forms, more likely to split the natural community borders and functional areas in multiple directions. Preserving the natural communities in school district division is usually considered desirable (see e.g. Caro et al. 2004). But again, breaking socio-spatial patterns is exactly the aim of the optimization, and in many cases splitting communities is the only way to achieve a socially mixed district division. In a socio-spatially diverged urban environment, complex and tentacular school districts have a greater potential than compact districts to radically alter the socio-spatial interactions that schools generate.

Regardless of whether the desire for compact districts is mainly irrational, it can make a complex and non-compact district plan infeasible: if the district plan is commonly not seen as legitimate and justified, it can be politically impossible to implement. As stated earlier, none of the districting solutions benefit everyone equally, and in school districting there is always a complex set of interests at stake. While in principle everyone is an advocate of educational equality, in practice families exercise school choice diligently and are willing to pay more for a home near a particular school – as a proof of which can be mentioned discontinuity in home prices at the school district borders in Helsinki (Harjunen et al. 2018). As the parents are usually not very willing to compromise their children's perceived benefit for common good, some of them are very likely to strongly oppose changes to the school districts, especially if those changes would result in longer distances between homes and the assigned school, and a larger and more diverse (on average more disadvantaged) student composition.

When looking at the results generated by the pilot optimizations in this study, another question that rises instantly is about the significance of optimization achieved. While the optimized district divisions are notably more even than the original district division, is the difference significant enough considering the effort of redistricting and the losses of efficiency what comes to travel times? Are the benefits from this scale of optimization sufficient to cover the inconvenience caused and the political turbulence stirred? As noted, evidence regarding the benefits of socially diverse schools on learning outcomes is extensive both on the level of individual and the system (Mickelson & Bottia 2010), but studies that measure benefits for learning outcomes gained from mixing the student compositions of advantaged and disadvantaged schools (by e.g. granting children from disadvantaged neighbourhoods places in schools with higher social status) are case-specific results that are not directly generalizable to other cities and educational systems. Additionally, the effects of school district optimization are largely dynamic, most likely affecting patterns of school choice and selective migration. This means, that exact predictions regarding the effects of optimization on student achievement are impossible.

However, Bernelius & Vaattovaara (2016) have found that in Helsinki, only 30% of parents are happy with their child attending a school where the proportion of students with immigrant background is between 20–50%. This means, that when the proportion rises over 20% in a school, the school will most likely be increasingly rejected in school choices, which again aggravates school segregation. The optimization piloted in this study was able to decrease the amount of schools crossing this 20% "tolerance limit" on the district level and bring those still crossing it closer to the lower end of the 20–50% range. In this way, the optimization of school districts can be predicted to at least decrease rejection of schools by making the schools' *perceived* quality differences notably smaller.

Predicting the effects of school district optimization on learning outcomes is also difficult because the divergence of schools' student compositions is only one aspect affecting the equality of educational opportunities in schools. Even if all social variation between schools' student compositions could be erased with school district optimization, educational segregation would still exist as within-school segregation related to weighted-curriculum education and students' social networks in the schools (Rumberger & Palardy 2005). It's impossible to completely erase these forms of segregation with any justifiable intervention. Still, a schooling system with strong within-school segregation in otherwise socially diverse schools create better possibilities for interaction between students of different backgrounds than a schooling system with strong segregation between schools. So, while it's impossible to eliminate the inequalities of educational opportunities just by optimizing school districts, the bases for pursuing such goals are better in socially diverse schools.

One may still wonder, whether the optimized plans that compromise travel times, school capacities and compactness are fair and utility-maximizing on the system level. This depends largely on the chosen viewpoint. If only judged from the travel time and infrastructure capacity aspects, the optimized district plans are not efficient. One may also point out that they may be unlawful – the Finnish law requires that children's school trips are as short as possible. Despite of that, even now the school districts in Helsinki are not perfect catchment areas based on distances and school capacities, so the definition of "as short as possible" is in reality somewhat flexible in relation to other objectives set by the local authorities. School district planning is all about value choices and priorities – about how aspects of equal opportunity and efficiency are weighted in the policy choices in relation to each other. From the equality of opportunities point of view, a socially more even district division is provenly fairer than a polarized division, but from the efficiency and travel-time point of view, it might not be.

One may also ask, why not only focus on targeted funding of lowest-performing schools instead of massive redistricting efforts, if it has shown to be effective in improving students' learning outcomes. The question is good, but it forgets the larger context of school segregation: the bidirectional link between residential and school segregation (e.g. Bernelius 2013), importance of social networks and other peer effects, the benefits from exposure to diversity, et cetera. Additionally, while targeted funding has been able to mitigate the effects of school segregation in Helsinki (Silliman 2017), it has not proven to be able to decrease school segregation itself. So, what comes to the prevention of divergence in the equality of educational opportunities, the targeted funding alone is most likely an alleviation of symptoms – but not a cure to the disease of segregation itself, and as such largely an answer to a wrong question.

The significant factors affecting the optimization result of a heuristic model, in addition to optimization parameters and logic, include the spatial structure of the optimization environment (e.g. the regional distribution of population), and the state from which the optimization is started from (i.e. the "seed"). This is also one of the most important things to note about the results of this study: as the optimization in this study was based on improving the existing school district plan with somewhat realistic parameters, the results only show how the current division could be improved with minimal changes to the system. If instead the optimization was started from a random amount of seeds with random locations (in that case seeds being the school buildings), the results would more accurately represent the theoretically optimal school district division and likely show much higher scales of evenness. In that case, however, the result would be only suitable for purely theoretical review, and likely have little practical significance.

In addition to the results obtained by piloting the application of the model, the other main result of this study is the model itself. Next, I will go through the main limitations of the model from technical point of view and present some future avenues for study.

5.2 Model limitations and future development targets

The main limitations of the model developed in this study include time complexity and scalability aspects. While local search algorithms are generally known for their relative efficiency (e.g. Ingber & Rosen 1992; Russell & Norvig 2003), the combinatorial nature of the problem requires special emphasis on powerful and efficient data structures. Though the model developed in this study already utilizes many powerful data structures, it's development has been mostly logic-oriented. This said, the implementation of the developed algorithm could very likely be optimized further to reduce its overall run time. Reducing the run time would enable improving the result accuracy further: the randomness included in the model could be increased, as more model runs and result evaluations could be executed in a reasonable time.

One concrete example of potential inefficiency in the implementation is that the algorithm developed in this study can be argued to be "wasteful" of function evaluations: a district does not remember which blocks it has already evaluated and therefore evaluates partially the same blocks on each of its turns (see e.g. Wolpert et al. 1997 for prevention of wastefulness). From the model logic's point of view, however, re-evaluating the same blocks on each turn is at least partially necessary, as the point is to always search for the *current* best move in the neighborhood. One possible improvement, though, could be made by storing some information about the blocks when first evaluating them. Based on this stored information the district could quickly sort out the most likely "bad" blocks in the beginning of its turn, and thus avoid doing expensive calculations on those no-go blocks on every turn. However, in order to avoid getting stuck on plateaus and bad local optimums, some of these no-go blocks should still be regularly re-evaluated and also sometimes randomly accepted.

Another technical benefit could be acquired by dividing the execution of the model for multiple processors. This would enable running multiple versions of the model parallelly, with differing amounts of randomness. This would also help to mitigate the poor upward scalability of the model – as noted, the running time required for the execution of the model grows as a factor of the number of blocks and districts in the model. This means that the running time grows steeply as the number of blocks and districts in the model increase. While distributing the model execution to different processors does not change the demanding combinatorial nature of the problem or remove its computational intractability, it would make the runs with larger problems feasible on a somewhat average computer.

In addition to time complexity issues, another clear target for further development of the model is the incorporation of an important optimization parameter – the safety of children's school trips – to the model. The safety aspect could be included in the model for example as a cost distance matrix. Another possible approach would be breaking the blocks into discontinuous regions by major roads and railways and run the optimizations separately for each region. The question about school trips' safety is, however, more complex than this. Not all railways or major roads are dangerous for children, and not all small roads are safe. A single crossing on the same road can be hazardous, while the next one may be not. The question is about the quality of urban planning and traffic planning, which can vary greatly even within a small region.

Also, while compactness is a debatable objective, testing its inclusion in the model either as a constraint or optimization parameter would definitely be an important step towards increasing its practical value and applicability in real-life contexts. However, as noted, the inclusion of compactness parameter may in practice be a significant obstacle for optimization in a segregated environment, since it would significantly reduce the number of possible solutions by preventing the formation of complex forms. This conflict has already been discovered in the previous studies (Caro et al. 2004; desJardins et al 2007; Bouzarth et al. 2018).

Another interesting development target for this work would be changing the model from "hard" constrained single-objective optimization to "soft" multi-objective optimization. In multi-objective optimization, the constraints could be included in the actual optimization function and weighted with different powers and multipliers. In this way, the optimization could be made more flexible, as no moves would be prohibited. With this logic, the optimization would move from happening only between the district to also happening between the optimization parameters: a district could choose a block that is far away if choosing it would significantly improve its social composition, but leave a block that would only minimally diversify its social composition while deteriorating its other optimization measures.

In addition to the development targets listed above, the optimization parameter measuring social composition could be extended to include also other measures, like household income and adult population’s level of education. While these measures have been found to strongly correlate in space with the share of population with an immigrant background (Vilkama 2011, Kortteinen & Vaattovaara 2015), they are also important, independent measures of social diversity. In the next phases of development, the model should preferably also be implemented with a block data that represents the natural neighborhoods better than the census grid units – or even with building-level polygons.

After proving that significant homogenization of the school districts can be achieved with local search approach, it would be both interesting and beneficial to test also other sophisticated methods for solving the problem. For example, approaches including genetic algorithms with nested local search have shown excellent results in other fields of research and could significantly increase efficiency compared to only local search-based approaches (Bação et al. 2005; Vanneschi et al. 2017). Also, more novel approaches like artificial bee colony (ABC) and method of musical composition (as in Rincon-Garcia et al. 2017) have given very promising results in electoral districting when compared to other automated districting approaches. However, to my knowledge, these novel methods have not yet been applied to school districting problems.

As a complex problem with multiple, sometimes conflicting objectives, school districting will probably always stay a somewhat ill-defined problem. As Caro et al. (2004) note, it’s very likely that new criteria emerge and existing criteria require modification during a districting process. Because of this, interactive districting tools utilizing integrated optimization algorithms are needed in the future. These kinds of tools could also cater for manual post-processing, easy exclusion of certain areal units from the optimization and calculation of different starting points (seeds) for optimization. Most importantly, adoption of machine-learning based approaches for school districting would be a step towards ensuring the quality of the district plans – a step that would make comparison of alternative plans easy, and a step towards ensuring that the generated output actually matches the defined goals in the best possible manner.

6. Conclusions

Angel Gurria, the OECD Secretary-General wrote in the latest *Education at a glance* the following:

> "*The conditions and social environments we are allotted at birth may seem as random as a lottery draw, yet they will define our starting position on the path of life by affecting not only the opportunities available to us, but also the social and emotional capital needed to ease our way*." (OECD2018b)

How the positions people are born to affect their educational careers and social interactions early in life is not, however, completely beyond our reach. In the light of numerous studies, one important tool for building equality of educational opportunities is ensuring that school districts are drawn to maximize social diversity within and evenness between the schools' student compositions.

The aim of this study was to develop a model for automated school districting and pilot its application in the context of Helsinki. While Helsinki is still only moderately segregated compared to other European capital cities, on the level of schools' student compositions social divergence is clear. The results generated with the pilot model show that significant homogenization of school districts could be achieved in Helsinki even with relatively small changes to student intakes and travel distances. While the model developed in this study can provide valuable insight into school district optimization, the approach still needs further development before being able to provide optimized school district divisions readily applicable to the real world.

As Mickelson and Bottia (2010) write, an integrated and diversity-enhancing educational system is not a weather-like phenomenon, that is largely beyond the reach of human efforts to affect or create. On the contrary, our educational systems are built based on highly conscious policy choices. This is best demonstrated by the diversity of different educational systems across the OECD countries, showing that wealthy societies are capable of creating a great variation of more and less effective, more and less inclusive, more and less socially integrated and more and less equal learning outcomes producing educational systems (OECD 2018a). Schools are contexts of life, where the foundations for future social life, educational career and social class are laid. If this important life domain becomes increasingly segregated, it both threatens the equality of educational opportunity and has potential of increasing segregation in other important contexts of life. Hindering this development, however, is neither out of our reach nor an overly complicated effort. In the end, the way school districts are drawn is simply a matter of values and knowledge put into practice.

Bibliography

Allen, R. (2007). Allocating pupils to their nearest secondary school: The consequences for social and ability stratification. *Urban Studies*, 44(4), 751-770.

Altman, M. (1997). The computational complexity of automated redistricting: is automation the answer? *Rutgers Computer & Technology Law Journal* vol. 23.

Bação, F; Lobo, V; Painho, M (2005). Applying genetic algorithms to zone design. *Soft Computing*, 9(5), 341-348.

Basic Education act 628/1998. English translation available at: <https://www.finlex.fi/en/laki/kaannokset/asiasanat.php?start=a&id=580163>

Bernelius, V. (2011). Osoitteenmukaisia oppimistuloksia? Kaupunkikoulujen eriytymisen vaikutus peruskoululaisten oppimistuloksiin Helsingissä. *Yhteiskuntapolitiikka* 76 (2011):5.

Bernelius, V. (2013). Eriytyvät kaupunkikoulut: Helsingin peruskoulujen oppilaspohjan erot, perheiden kouluvalinnat ja oppimistuloksiin liittyvät aluevaikutukset osana kaupungin eriytymiskehitystä. *Research series* 2013:1, City of Helsinki Urban Facts 2013.

Bernelius, V.; Vaattovaara, M. (2016). Choice and segregation in the ‘most egalitarian’ schools: Cumulative decline in urban schools and neighbourhoods of Helsinki, Finland. *Urban studies*, 53(15), 3155-3171.

Boterman, W. R.; Musterd, S. (2016). Cocooning urban life: Exposure to diversity in neighbourhoods, workplaces and transport. *Cities*, 59, 139-147.

Bouzarth, E. L.; Forrester, R.; Hutson, K. R.; Reddoch, L. (2018). Assigning students to schools to minimize both transportation costs and socioeconomic variation between schools. *Socio-Economic Planning Sciences* 64 (2018) 1-8.

Bozkaya, B.; Erkut, E.; Laporte, G. (2003). A tabu search heuristic and adaptive memory procedure for political districting. *European Journal of Operational Research*, 144, 12–26

Browdy, M. H. (1990). Simulated annealing: an improved computer model for political redistricting. *Yale Law & Policy Review*, 8(1), 163-179.

Brown, P. J. (1987). Trade-off and compromise in secondary school reorganisation schemes: a goal programming approach. *Papers in Regional Science*, 61(1), 115-130.

Brunner, E. J.; Cho, S. W.; Reback, R. (2012). Mobility, housing markets, and schools: Estimating the effects of inter-district choice programs. *Journal of Public Economics*, 96(7-8), 604-614.

Camacho-Colladosa, M.; Liberatore, F.; Angulo, J.M. (2015). A multi-criteria police districting problem for the efficient and effective design of patrol sector. *European Journal of Operational Research*, Vol. 246 No. 2, pp. 674-684.

Caro, F.; Shirabe, T.; Guignard, M.; Weintraub, A. (2004). School redistricting: embedding GIS tools with integer programming. *Journal of the Operational Research Society* (2004) 55, 836–849.

Chang, A. (2018). We can draw school zones to make classrooms less segregated. This is how well your district does. *VOX*. Vox media 2018. <https://www.vox.com/2018/1/8/16822374/school-segregation-gerrymander-map>

Cheshire, P.; Sheppard, S. (2004). Capitalising the value of free schools: the impact of supply characteristics and uncertainty. *The Economic Journal*, 114(499), F397-F424.

Clarke, S.; Surkis, J. (1968). An operations research approach to racial desegregation of school systems. *Socio-Economic Planning Sciences*, 1(3), 259-272.

Cockings, S.; Harfoot, A.; Martin, D.; Hornby, D. (2011). Maintaining existing zoning systems using automated zone-design techniques: methods for creating the 2011 Census output geographies for England and Wales. *Environment and Planning A*, 43(10), 2399-2418.

D'Amico, S.; Wang, S.; Batta, R.; Rump, C. (2002). A simulated annealing approach to police district design. *Computers & Operations Research*, 29(6), 667-684.

Danielsen, B.; Fairbanks, J.; Zhao, J. (2015). School choice programs: The impacts on housing values. *Journal of Real Estate Literature*, 23(2), 207-232.

desJardins, M; Bulka, B.; Carr, R.; Jordan, E.; Rheingans, P. (2007). Heuristic search and information visualization methods for school redistricting. *AI Magazine*, 28 (3), p. 59.

Easingwood, C. (1973). A heuristic approach to selecting sales regions and territories. *Operational Research Quarterly*, 24(4), 527–534.

Eiselt, H. A.; Laporte, G. (1987). Combinatorial optimization problems with soft and hard requirements. *Journal of the Operational Research Society*, 38(9), 785-795.

Farmer, M. E.; Bapna, S.; Jain, A. K. (2004). Large scale feature selection using modified random mutation hill climbing. *International Conference on Pattern Recognition (ICPR)*, Cambridge UK, p. 287-290.

Ferland, J. A.; Guénette, G. (1990). Decision support system for the school districting problem. *Operations Research*, 38(1), 15-21.

Garey, M. R.; Johnson, D. S. (1979). *Computers and intractability: a guide to the theory of NP-completeness*. W.H. Freeman and Co, NY, 1979.

Garfinkel, R. S.; Nemhauser, G. L. (1970). Optimal political districting by implicit enumeration techniques. *Management Science*, 16(8), B495–B508

Gibbons, S.; Machin, S. (2008). Valuing school quality, better transport, and lower crime: evidence from house prices. *Oxford review of Economic Policy*, 24(1), 99-119.

Hardy, L. C. (1990). The Gerrymander: Origin, Conception and Re-emergence. Rose Institute of State and Local Government.

Harjunen, O.; Kortelainen, M.; Saarimaa, T. (2018). Best education money can buy? Capitalization of school quality in Finland. *CESifo Economic Studies*, 2018, 150–175

Heckman, L.; Taylor, H. (1969). School rezoning to achieve racial balance: a linear programming approach. *Socio-Economic Planning Sciences*, 3(2), 127-133.

Helsinki Education Committee (2014). Lukuvuoden 2015 - 2016 yleisopetuksen enimmäisoppilasmäärät 1. luokille. Finnish sub-committee, decisions, 13/2014(16.12.2014). <https://dev.hel.fi/paatokset/asia/hel-2014-013483/skj-2014-13/>

Helsinki’s Educational Administration’s Rules of Procedure (2010). <http://docplayer.fi/71050591-Helsingin-kaupungin-opetustoimen-johtosaanto-kaupunginvaltuuston-hyvaksyma-huhtikuun-21-p-na-2010.html>

Henig, J. R. (1990). Choice in public schools: An analysis of transfer requests among magnet schools. *Social Science Quarterly*, 71(1), 69.

Hess, S. W.; Samuels, S. A. (1971). Experiences with a sales districting model: criteria and implementation. *Management Science*, 18(4), 41–54.

Hojati, M. (1996). Optimal political districting. *Computers & Operations Research*, 23(12), 1147–1161.

Horn, M (1995). Solution Techniques for large Regional partitioning. *Geographical Analysis*, Vol. 27, No. 3, Ohio State University Press.

Hunter, T. R. (2011). The First Gerrymander?: Patrick Henry, James Madison, James Monroe, and Virginia's 1788 Congressional Districting. *Early American Studies: An Interdisciplinary Journal* 9(3), 781-820. University of Pennsylvania Press.

Ingber, L.; Rosen, B. (1992). Genetic algorithms and very fast simulated reannealing: A comparison. *Mathematical and computer modelling*, 16(11), 87-100.

Johnston, R. (2002). Manipulating maps and winning elections: measuring the impact of malapportionment and gerrymandering. *Political geography*, 21(1), 1-31.

Karsten, S. (2010). School segregation. *Equal opportunities*, 193-209, OECD.

Karsten, S.; Ledoux, G.; Roeleveld, J.; Felix, C.; Elshof, D. (2003). School choice and ethnic segregation. *Educational policy*, 17(4), 452-477.

Kauppinen, T. M. (2008). Schools as mediators of neighbourhood effects on choice between vocational and academic tracks of secondary education in Helsinki. *European Sociological Review*, 24(3), 379-391.

Keane, M. (1975). The size of the region-building problem. *Environment and Planning A*, 7(5), 575-577.

Knutson, D. L.; Marquis, L. M.; Ricchiute, D. N.; Saunders, G. J. (1980). A goal programming model for achieving racial balance in public schools. *Socio-Economic Planning Sciences*, 14(3), 109-116.

Kortteinen, M.; Vaattovaara, M. (2015). Segregaation aika. *Yhteiskuntapolitiikka*, 80 (2015):6. p. 562-574.

Kosunen, S. (2014). Reputation and parental logics of action in local school choice space in Finland. *Journal of Education Policy*, 29(4), 443-466.

Kosunen, S.; Bernelius, V.; Seppänen, P.; Porkka, M. (2016). School choice to lower secondary schools and mechanisms of segregation in Urban Finland. *Urban Education* (2016): 0042085916666933.

Kupari, P.; Välijärvi, J.; Andersson, L.; Arffman, I.; Nissinen, K.; Puhakka, E.; Vettenranta, J. (2013). PISA 12 ensituloksia [PISA 12 preliminary results]. Helsinki, Finland. Ministry of Education and Culture.

Kuukasjärvi, K. (2013). Asuinalueiden maineiden merkitys pääkaupunkiseudun alueelliseen erilaistumiskehitykseen – tarkastelussa lapsiperheet. Master's thesis, University of Helsinki. <http://hdl.handle.net/10138/40074>

Kuusela, J. (2006). Temaattisia näkökulmia perusopetuksen tasa-arvoon. *Oppimistulosten arviointi* 6/2006. Opetushallitus. Yliopistopaino, Helsinki 2006.

Kuusela, J. (2010). *MetrOP–alueen kouluista toisen asteen yhteishakuaineiston perusteella*. In Rimpelä, M.; Bernelius, V. (2010). Peruskoulujen oppimistulokset ja oppilaiden hyvinvointi eriytyvällä Helsingin seudulla: MetrOP–tutkimus 2010–2013. *Geotieteiden ja maantieteen laitoksen tutkimuksia* B1, 38–43.

Liu, Y. Y.; Cho, W. K. T.; Wang, S. (2016). PEAR: a massively parallel evolutionary computation approach for political redistricting optimization and analysis. *Swarm and Evolutionary Computation*, 30, 78-92.

Malczewski, J.; Jackson, M. (2000). Multicriteria spatial allocation of educational resources: an overview. *Socio-Economic Planning Sciences* 34 (2000) 219-235.

Malmberg, B.; Andersson, E. K.; Bergsten, Z. (2014). Composite geographical context and school choice attitudes in Sweden: A study based on individually defined, scalable neighborhoods. *Annals of the Association of American Geographers*, 104(4), 869-888.

Marcińczak, S.; Musterd, S.; van Ham, M.; Tammaru, T. (2015). Inequality and rising levels of socio-economic segregation: Lessons from a pan-European comparative study. *Regions and cities vol* 89, Routledge.

Massey, D. S.; Denton, N. A. (1988). The dimensions of residential segregation. *Social forces*, 67(2), 281-315.

Mehrotra, A.; Johnson, E. L.; Nemhauser, G. L. (1998). An optimization based heuristic for political districting. *Management Science*, 44(8), 1100–1114.

Mickelson, R. A.; Bottia, M. (2009). Integrated education and mathematics outcomes: A synthesis of social science research. *The North Carolina Law Review*, 88, 993.

Musset, P. (2012). School Choice and Equity: Current Policies in OECD Countries and a Literature Review. *OECD Education Working Papers*, No. 66, OECD Publishing.

Musterd, S.; Marcińczak, S.; Van Ham, M.; Tammaru, T. (2017). Socioeconomic segregation in European capital cities. Increasing separation between poor and rich. *Urban Geography* 38(7), 1062-1083.

Naderi, S.; Kilic, K. (2016). Fuzzy Zoning: A Lagrangean Relaxation Approach. *IEEE International Conference on Systems, Man, and Cybernetics* (SMC) (pp. 001714-001718).

Nagel, S. S. (1964). Simplified bipartisan computer redistricting. *Stanford Law Review*, 17, p. 863-899.

Nissinen, K. (2015). Ovatko koulut eriytymässä? In Välijärvi, J.; Kupari, P.; Ahonen, A. K.; Arffman, I.; Harju-Luukkainen, H.; Leino, K.; Niemivirta, M.; Nissinen, K.; Salmela-Aro, K.; Tarnanen, M.; Tuominen-Soini, H.; Vettenranta, J.; Vuorinen, R. (2015). Millä eväillä uuteen nousuun? PISA 2012 tutkimustuloksia. *Opetus- ja kulttuuriministeriön julkaisuja* 2015:6.

OECD (2004). Learning for Tomorrow's World: First Results from PISA 2003 (Vol. 659). Programme for International Student Assessment, *OECD Publications* 2004.

OECD (2018a). Education at a glance 2018. OECD indicators. Annex 3: sources, methods and technical notes. OECD Publications 2018.

OECD (2018b). Education at a glance 2018. OECD Publications 2018.

Park, K.; Lee, K.; Park, S.; Lee, H. (2000). Telecommunication node clustering with node compatibility and network survivability requirements. *Management Science*, 46(3), 363-374.

Perry, L.; McConney, A. (2010). Does the SES of the school matter? An examination of socioeconomic status and student achievement using PISA 2003. *Teachers College Record* 112(4), 1137-1162.

Ricca, F.; Scozzari, A.; Simeone, B. (2013). Political districting: from classical models to recent approaches. *Annals of Operations Research*, 204(1), 271-299.

Ricca, F.; Simeone, B. (2008). Local search algorithms for political districting. *European Journal of Operational Research*, 189(3), 1409-1426.

Riitaoja, A. L. (2010). Asuinalueiden ja koulujen eriytyminen Helsingissä: Yhteiskunnalliset ja kasvatukselliset haasteet. *Terra* 122 (2010): 3.

Rincón-García, E. A.; Gutiérrez-Andrade, M. Á.; de-los-Cobos-Silva, S. G.; Mora-Gutiérrez, R. A.; Ponsich, A.; Lara-Velázquez, P. (2017). A comparative study of population-based algorithms for a political districting problem. *Kybernetes*, 46(1), 172-190.

Rincón-García, E.; Gutiérrez-Andrade, M.; De-los-Cobos-Silva, S.; Lara-Velázquez, P.; Ponsich, A. S.; Mora-Gutiérrez, R.; (2013). A multiobjective algorithm for redistricting. *Journal of applied research and technology*, 11(3), 324-330.

Robertson, A., Morrison, M. A., Shade, W. G., Johnston, R., Zieger, R., Langston, T., & Valelly, R. (2010). *Encyclopedia of US political history*. CQ Press.

Rothwell, J. (2012). Housing costs, zoning, and access to high-scoring schools. Brookings Institution, Ministerio de Educación del Perú 2012. <http://hdl.handle.net/123456789/3510>

Rumberger, R.; Palardy, G. (2005). Does segregation still matter? The impact of student composition on academic achievement in high school. *Teachers college record*, 107(9), 1999.

Russell, S; Norvig, P (2003). *Artificial intelligence: a modern approach*. Pearson Education Limited.

Saporito, S. (2003). Private choices, public consequences: Magnet school choice and segregation by race and poverty. *Social problems* 50(2), 181-203.

Schnell, I.; Yoav, B. (2001). The sociospatial isolation of agents in everyday life spaces as an aspect of segregation. *Annals of the Association of American Geographers*, 91(4), 622-636.

Schwartz, H. (2011). Housing policy is school policy: economically integrative housing promotes academic success in Montgomery county, MD. *The Education Digest*, 76(6), 42.

SeutuCD 2018. Rekisteriaineistot: Väestö 2018. Esri Shapefile. HSY 2018.

Shanker, R.; Turner, R.; Zoltners, A. (1975). Sales territory design: an integrated approach. *Management Science*, 22(3), 309–320.

Silliman, M. (2017). Targeted Funding, Immigrant Background, and Educational Outcomes: Evidence from Helsinki's “Positive Discrimination” Policy. *VATT Working Papers 91.* VATT Institute for Economic Research, Helsinki 2017.

Tavares-Pereira, F.; Figueira, J.; Mousseau, V.; Roy, B. (2007). Multiple criteria districting problems. The public transportation network pricing system of the Paris region. *Annals of Operations Research,* 154, 69-9.

The Education division of Helsinki 2018. School service areas of the city of Helsinki. Esri Shapefile.

Toivonen, T., M. Salonen, H. Tenkanen, P. Saarsalmi, T. Jaakkola & J. Järvi (2014). MetropAccess -Travel Time Matrix. DOI: 10.13140/2.1.2449.8884

van Ham, M.; Tammaru, T. (2016). New perspectives on ethnic segregation over time and space. A domains approach. *Urban Geography*, 37(7), 953-962.

Vanneschi, L.; Henriques, R.; Castelli, M. (2017). Multi-objective genetic algorithm with variable neighbourhood search for the electoral redistricting problem*. Swarm and Evolutionary Computation*, 36, 37-51.

Vickrey, W. (1961). On the prevention of gerrymandering. *Political Science Quarterly*, 76(1), 105-110.

Vilkama, K. (2011). Yhteinen kaupunki, eriytyvät kaupunginosat? Kantaväestön ja maahanmuuttajataustaisten asukkaiden alueellinen eriytyminen ja muuttoliike pääkaupunkiseudulla. *Tutkimuksia,* 2011:2. Helsingin kaupungin tietokeskus 2011.

Vilkama, K.; Lönnqvist, H.; Väliniemi-Laurson, J.; Tuominen, M. (2014). Erilaistuva pääkaupunkiseutu: sosioekonomiset erot alueittain 2002-2012. *Tutkimuksia* 2014:1. Helsingin kaupungin tietokeskus 2014.

Weaver, J; Hess, S (1963). A Procedure for Nonpartisan Districting: Development of Computer Techiques. *The Yale Law Journal*, vol. 73, p. 288-309.

Williams, J. (1995). Political redistricting: a review. *Papers in regional science: The journal of the RSAI 74*, I: 13-40.

Wolpert, D. H.; Macready, W. G. (1997). No free lunch theorems for optimization. *IEEE transactions on evolutionary computation*, 1(1), 67-82.

Wong, D. W.; Shaw, S. L. (2011). Measuring segregation: An activity space approach. *Journal of geographical systems*, 13(2), 127-145.

Yang Hansen, K.; Gustafsson, J. E. (2016). Causes of educational segregation in Sweden–School choice or residential segregation. *Educational Research and Evaluation*, 22(1-2), 23-44.

Zoltners, A.; Sinha, P. (1983). Sales territory alignment: a review and model. *Management Science*,
29(3), 1237–1256.

Appendix: model source code

Model source code also available in https://github.com/herttale/School-district-optimization

classes.py

```
from shapely.ops import cascaded_union
from copy import deepcopy
import random
from shapely.geometry import LineString

class SchoolDistr:
    """ The class representing the school districts """

    def __init__(self, school_id, blocks, td_matrix):

        # class attribute 1: the school id number
        self.school_id = school_id
        # class attribute 2: the blocks belonging to the district (as a dict,
            # with keys corresponding to the td_matrix keys).
        self.blocks = blocks
        # class attribute 3: distance matrix (as a dict, with keys
            # corresponding to the blocks keys).
        self.td_matrix = td_matrix
        # class attribute 3: the geometry of the district (shapely polygon)
        self.geometry = None
        # class attribute 4: the maximum allowed distance from block to the
            # district's school
        self.max_distance = None
        # class attribute 5: the amount of 7-year-olds living inside the
            # district
        self.students = None
        # class attribute 6: the maximum amount of 7-year-olds that the
            # district can host
        self.student_limit = None
        # class attribute 7: the current value of the optimization parameter
        self.optimization_value = None
        # function call: initiate district attributes
        self.initiate_distr_attrs()

    # Method for initializing attributes
    def initiate_distr_attrs(self):
        self.geometry = self.calculate_geometry()
        self.max_distance = self.calculate_max_distance()
        self.students = self.calculate_student_base()
        self.student_limit = self.students*1.20
        self.optimization_value = self.calculate_optimization_value()

    # Method for updating attributes
    def update_distr(self):
        self.geometry = self.calculate_geometry()
        self.students = self.calculate_student_base()
        self.optimization_value = self.calculate_optimization_value()

    # Method for calculating the district's geometry as cascaded union of the
        # block geometries
```

```
    def calculate_geometry(self):
        geom_list = []
        for key, block in self.blocks.items():
            geom_list.append(block.geometry)
        return cascaded_union(geom_list)

    # Method for calculating the district's maximum distance constraint. The
        # travel time data must not include infinite distance values.
    def calculate_max_distance(self):
        maxt = 0
        for key, block in self.blocks.items():
            ttime = self.td_matrix[key]['walk_d']
            if ttime > maxt:
                maxt = ttime
        return maxt * 1.20

    # Method for calculating the current value of the optimization parameter
    def calculate_optimization_value(self):
        majority_pop = 0
        minority_pop = 0
        for key, block in self.blocks.items():
            majority_pop += block.lang_majority
            minority_pop += block.lang_other
        return minority_pop/(minority_pop + majority_pop)

    # Method for calculating the current amount of 7-year-olds living
        # inside the district
    def calculate_student_base(self):
        student_sum = 0
        for key, block in self.blocks.items():
            student_sum += block.student_base
        return student_sum

    # Method for calculating the district's neighbourhood: which blocks
        # the district shares a line segment with
    def touches_which(self, blocks_dict):
        neighbors = []
        for key, block in blocks_dict.items():
            if type(self.geometry.intersection(block.geometry)) == LineString:
                if key not in self.blocks:
                    neighbors.append(block)
        return neighbors

    # Method for calculating whether a block is too far for adoption
        # Returns True if the block is too far
    def is_too_far(self, block):
        dist = self.td_matrix[block.block_id]['walk_d']
        return dist > self.max_distance

    # Method for adopting a selected block
    def add_block(self, block):
        if block == None:
            return
        else:
            block.school_id = self.school_id
            self.blocks[block.block_id] = block

    # Method for removing an adopted block
    def remove_block(self, block):
        if block == None:
            return
        else:
            del self.blocks[block.block_id]

```

```
    # A method for testing if adopting a block would break another district's
        # contiguity. Returns True if contiguity would break.
    def break_contiguity(self, block):
        blocks_copy = deepcopy(self.blocks)

        geom_list = []
        for key, item in blocks_copy.items():
            geom_list.append(item.geometry)
        geom1 = cascaded_union(geom_list)
        del blocks_copy[block.block_id]

        geom_list = []
        for key, item in blocks_copy.items():
            geom_list.append(item.geometry)
        geom2 = cascaded_union(geom_list)

        return type(geom1) != type(geom2)

    # A method for selecting the best block in neighbourhood
    def select_best_block(self, blockset, districts, global_mean,
                          global_st_dev):
        majority_pop = 0
        minority_pop= 0

        for key, value in self.blocks.items():
            majority_pop += value.lang_majority
            minority_pop += value.lang_other

        best_block = None

        for block in blockset:
            # test for rule 2
            if block.contains_school == False:
               # test for rule 3
                if (block.student_base + self.students) <= self.student_limit:
                    # test for rule 4
                    if self.is_too_far(block) == False:
                        current_district = districts[block.school_id]
                        # test for rule 5
                        if current_district.break_contiguity(block) == False:
                            # calculate specs for the block's current district
                            current_district_majority_pop = 0
                            current_district_minority_pop= 0

                            for key, value in current_district.blocks.items():
                                current_district_majority_pop += \
                                        value.lang_majority
                                current_district_minority_pop += \
                                        value.lang_other

                            current_d_new_value = ((current_district_minority_pop
                                                    - block.lang_other)/
                                                    (current_district_minority_pop
                                                     - block.lang_other +
                                                     current_district_majority_pop
                                                     - block.lang_majority))
                            current_d_current_value = ((current_district_minority_pop)/
                                                         (current_district_minority_pop
                                                         + current_district_majority_pop))

                            # test the adoption outcome in relation to current state
                            if best_block == None:
                                own_new_value1 = ((minority_pop + block.lang_other)/
                                                  (minority_pop + block.lang_other +
```

```
                                                   majority_pop + block.lang_majority))

                                # test for the rule 6
                                if (abs(current_d_new_value - global_mean) <=
                                    abs(current_d_current_value - global_mean) or
                                    abs((current_d_current_value - global_mean) -
                                        (self.optimization_value - global_mean)) >
                                        abs((current_d_new_value - global_mean) -
                                            (own_new_value1 - global_mean))):

                                     if (abs(own_new_value1 - global_mean) <
                                         abs(self.optimization_value - global_mean)):
                                         best_block = block

                            # test the adoption outcome in relation to the current best_block
                            else:

                                own_new_value2 = ((minority_pop + block.lang_other)/
                                                  (minority_pop + block.lang_other +
                                                   majority_pop + block.lang_majority))
                                current_best = ((minority_pop + best_block.lang_other)/
                                                (minority_pop + best_block.lang_other +
                                                 majority_pop + best_block.lang_majority))

                                # test for the rule 6
                                if (abs(current_d_new_value - global_mean) <=
                                       abs(current_d_current_value - global_mean) or
                                       abs((current_d_current_value - global_mean) -
                                           (self.optimization_value - global_mean)) >
                                           abs((current_d_new_value - global_mean) -
                                               (own_new_value1 - global_mean))):

                                    if (abs(own_new_value2 - global_mean) <
                                        abs(current_best - global_mean)):
                                        best_block = block
        # return the best block
        return best_block

    # A method for selecting a random block in neighbourhood
    def select_random_block(self, blockset, districts):
        blocklist = []
        for block in blockset:
            # test for rule 2
            if block.contains_school == False:
                # test for rule 3
                if (block.student_base + self.students) <= self.student_limit:
                    # test for rule 4
                    if self.is_too_far(block) == False:
                        current_district = districts[block.school_id]
                        # test for rule 5
                        if current_district.break_contiguity(block) == False:
                            blocklist.append(block)

        if len(blocklist) > 0:
            # generate a random number for selecting a block
            randomindx = random.randint(0,len(blocklist)-1)
            # return a random block according to the random number generated
            return blocklist[randomindx]

class Block:
    """ The class representing the residential blocks """
    def __init__(self, geometry, block_id, lang_majority, lang_other, student_base,
                 school_id, contains_school):

```

```
        # class attribute 1: the geometry of the block (shapely polygon)
        self.geometry = geometry
        # class attribute 2: block id
        self.block_id = block_id
        # class attribute 3: the amount of population with Finnish or Swedish as
            # their mother tongue
        self.lang_majority = lang_majority
        # class attribute 4: the amount of population with other languages than Finnish
            # or Swedish as their mother tongue
        self.lang_other = lang_other
        # class attribute 5: the amount of 7-year-olds living in the block
        self.student_base = student_base
        # class attribute 6: the id of the school district the block currently
            # belongs to
        self.school_id = school_id
        # class attribute 7: True if the block contains a school, otherwise False
        self.contains_school = contains_school
```

main.py

```
# the main optimization function
def main(districts_orig, blocks_dict_orig):

    import numpy as np
    import statistics as st
    import random
    from classes import Block, SchoolDistr
    from copy import deepcopy

    current_best_cumul_zvalue = None
    current_best_distr_division = None
    current_best_curve = None
    all_optimization_curves = []

    for iteration in range(0,100):

        print(iteration)
        districts = deepcopy(districts_orig)
        blocks_dict = deepcopy(blocks_dict_orig)

        # create a list for tracking the change in cumulative z-value
        cumulative_zvalues_list = []
        # create a variable for tracking the iterations inside while-loop
        main_iteration = 0
        # set the ceiling value for probability calculation (now it ranges from
            # 50 to 124 adding 0.75 on every iteration
        ceil = np.floor(0.075 * iteration * 10 + 50)

        # calculate the global mean and standard deviation for original
            # districts' optimization values
        districts_values_list = []
        for key, item in districts.items():
            districts_values_list.append(item.optimization_value)
        global_mean = sum(districts_values_list)/len(districts)
        global_st_dev = np.std(districts_values_list, ddof = 0)

        while True:

            # calculate the current cumulative z-value
            cumulative_zvalue = 0
            for key, distr in districts.items():
```

```
                cumulative_zvalue += abs((distr.optimization_value -
                                          global_mean)/global_st_dev)
            cumulative_zvalues_list.append(cumulative_zvalue)

            # test whether the optimization can be terminated - if yes, return
                # optimized district division and corresponding optimization curve
            if main_iteration >= 12:
                checkvalue = st.mean([cumulative_zvalues_list[main_iteration],
                                      cumulative_zvalues_list[main_iteration-1],
                                      cumulative_zvalues_list[main_iteration-2],
                                      cumulative_zvalues_list[main_iteration-3]]) \
                                        - cumulative_zvalues_list[main_iteration]

                if round(checkvalue, 5) == 0 or main_iteration > 40:
                    break

            # increase iteration
            main_iteration += 1
            print("main_iteration round:", main_iteration,
                  ', current cumulative z-value:', cumulative_zvalue)

            # iterate the districts
            for key in list(districts.keys()):
                # generate a random number for defining whether a best or a random
                    # block will be chosen on this turn
                if ceil >= 50:
                    random_int = random.randint(0, ceil)
                else:
                    random_int = 0

                # check what blocks the district in turn touches
                neighbors = districts[key].touches_which(blocks_dict)
                # select best or random block based on random_int
                if random_int > 50:
                    block_to_add = districts[key].select_random_block(neighbors,
                                                        districts)
                else:
                    block_to_add = districts[key].select_best_block(neighbors,
                                                        districts, global_mean, global_st_dev)

                if block_to_add != None:
                    # remove block from its previous owner and update values
                    districts[block_to_add.school_id].remove_block(block_to_add)
                    districts[block_to_add.school_id].update_distr()

                    # add block to the new district
                    block_to_add.school_id = key
                    districts[key].add_block(block_to_add)
                    districts[key].update_distr()

            # decrease ceiling value
            ceil -= 5

        all_optimization_curves.append(cumulative_zvalues_list)

        if (current_best_cumul_zvalue == None or
            cumulative_zvalue < current_best_cumul_zvalue):
            current_best_cumul_zvalue = cumulative_zvalue
            current_best_distr_division = districts
            current_best_curve = cumulative_zvalues_list

    return({"current_best_distr_division":current_best_distr_division,
                            "current_best_curve":current_best_curve})
```