Expanding the Secondary School Network in Sub-Saharan Africa using Meta-Heuristic Facility-Location Techniques

Darragh Minogue
ID: 21253383

School of Computer Science
University of Galway

Supervisor: James McDermott

In partial fulfilment of the requirements for the degree of MSc in Artificial Intelligence

September 7, 2022

Acknowledgements

Firstly, thank you to my thesis supervisor James McDermott for his guidance and support throughout this Capstone Project. I am extremely grateful for his patience in understanding. I also really enjoyed the time spent talking through potential solutions and problem solving.

Thank you also the program directors for an excellent course. I have learned a lot during this year.

Finally, thank you to my wife and my family for their continuous support during this master's and every other endeavour.

Declaration

I, Darragh Minogue, hereby declare that this thesis, titled “Expanding Secondary School Networks in Sub-Saharan Africa using a Meta-Heuristic Facility-Location Techniques”, and the work presented in it are entirely my own except where explicitly stated otherwise in the text, and that this work has not been previously submitted, in part or whole, to any university or institution for any degree, diploma, or other qualification.

Signature:

Abstract

This project adopts facility-location approaches to improve planning around the expansion of secondary school networks in Sub-Saharan Africa. Access to secondary education is limited; children often face extremely long distances to attend secondary education, and many drop out as a result. As resources are limited, phased expansions require strategic decision making that can benefit from optimisation models. This research responds to this need by applying meta-heuristic optimisation methods to find optimal locations to construct new secondary schools that best serve the demand from middle schools in Sub-Saharan Africa where walking is the primary mode to accessing school. The aim is to minimise walking distance for children when proposing new locations whilst also maximising expected enrolment by prioritising areas that currently have high middle school enrolment. To solve this problem, a single objective location model is developed and applied to secondary school expansion in Ethiopia. Most of the existing literature on school location problems use the geographic distribution of the population as the foundation, but this can often be unavailable, contested and/or unreliable. Instead, this project makes use of readily available data that is not politicised: the GPS locations of middle schools. The models developed are therefore pragmatic and can be easily adapted to other countries that also face similar data constraints.

Keywords: *School Location Problem, Facility-Location, Meta-Heuristics, Optimisation*

Word Count: 12,332 (not including title page, Declaration of Authorship, Acknowledgements, and Bibliography).

GitHub Repository: https://github.com/minogud2/Capstone

Table of Contents

Chapter 1

1. Introduction

This project adopts Facility-Location approaches to improve planning around the expansion of secondary school networks. The classic Facility-Location Problem (FLP) is an optimisation problem that identifies suitable locations for services or industries based on geographical demands, capacity, distances, or cost. The general aim is to maximise profits for businesses or minimise response times or distances to services for communities. Examples of FLP range from deciding where to locate factories and warehouses to fire stations or hospitals. In this research, FLP optimisation methods are applied to find optimal locations to construct new secondary schools that best serve the demand from middle schools in Sub-Saharan Africa where walking is the primary mode to accessing school. Key to solving the problem is understanding that when a school is located within a village, there is no impact of distance on enrolment nor dropout. However, beyond 1-2km, distance affects initial access to school but also creates barriers to retention, completion, and transition to higher level (Theunynck, 2009). As such, there is a need to minimise walking distance for children when proposing new locations whilst also maximising expected enrolment by prioritising areas that currently have high middle school enrolment, due to limited budgets for school construction. To solve this problem, a single objective location model is developed and applied to secondary school expansion in Ethiopia.

1.1. Motivation

Expansion to secondary education is a major priority, yet funding is limited

While Gross Enrolment Rates[1] (GER) for primary education in Sub-Saharan Africa (SSA) reached 100% by 2020, GER for secondary stands at 44% (The World Bank, 2022). Most children don't access secondary education. In 2015, the UN General Assembly adopted the Sustainable Development Goals, of which most SSA countries are participating States (UN, 2022). As part of the educational goal, a key performance indicator is Target 4.1: "By 2030, ensure that all girls and boys complete free, equitable and quality primary and secondary education leading to relevant and effective learning outcomes" (UNESCO, 2019). This involves the provision of 12 years of publicly funded education for all (Grades 1-12), of which 9 years are compulsory. To meet this target, widespread expansion of secondary education is required. National Governments then, with support from the donor community, need to decide how this expansion should be rolled out.

As Garfield et al. (1994) point out, expanding access to secondary education can occur in multiple ways. Firstly, demand can be enhanced by increasing enrolment in primary schools or middle schools and improving retention. A larger cohort of students are therefore ready to

[1] "The gross enrolment rate (GER) is the total enrolment in a specific level of education, regardless of age, expressed as a percentage of the population in the official age group corresponding to this level of education. The GER can exceed 100% because of early or late entry and/or grade repetition" (INEE, 2022).

join secondary. Secondly, over-capacity existing secondary schools can be expanded through the construction of new classrooms and hiring of more teachers, thereby increasing the number of children that can attend secondary school. Thirdly, new schools can be constructed which serve the unmet demand from primary schools or middle schools located too far away from existing secondary schools. Since the first two solutions offer limited scope for increasing access to schooling, this paper looks to draw more attention to the third solution, which is essentially a FLP.

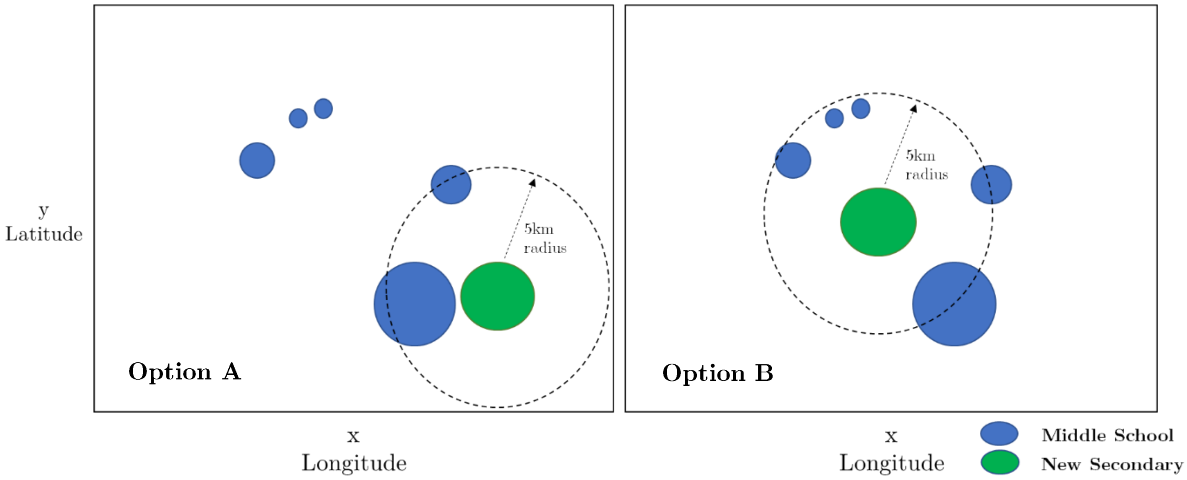
Despite the significant investment needed to create more access to secondary education, donors are more reluctant to get involved in school construction (Theunynck, 2014). They are concerned largely with the quality of education currently being delivered. As Theunynck (2018) points out: “the needed completion of school networks is actually challenged by the increased focus on investing on the existing network rather than completing it” (pg 7). Attention and funding have shifted towards ensuring retention and quality learning. Yet, if countries in SSA are to meet their SDG targets, this won’t suffice to address the sheer volume of children not accessing secondary education (66%). New schools need to be built and as funding for such projects will be limited, finding optimal locations for new secondary schools is therefore extremely relevant.

Current approaches towards planning where to construct new schools can be improved

To identify new school locations, planners typically “divide up the country into school catchment areas by applying norms regarding the minimum population required to establish an economically viable school and a maximum distance that children should be expected to travel from home to school” (Theunynck, 2009: p:15-16). While debates can be had over the recommended distance to school, the main challenge in SSA is finding accurate census data to help plan. In 2013, only 12 of 49 countries in SSA held a census within the last 10 years (The Guardian, 2014). With demographics changing rapidly due to urbanisation and climate change, planning school construction using population estimations won’t yield reliable results.

Instead, what is often used as a proxy during planning is the location of primary or middle schools as demand nodes. It is assumed that these schools are constructed within a village and new secondary schools can then be constructed nearby to best serve the community. Planners then, whether centralised or decentralised, identify locations with the highest demand in middle schools (i.e., largest enrolment in middle schools) and seek to construct secondary schools nearby to best serve that school population. This can produce good results as the proposed locations tend to be in urban areas close to multiple feeder schools. However, the approach has not been sufficiently examined. In some cases, a secondary school could be constructed between multiple feeder schools and offer a higher expected enrolment, despite having a smaller individual middle school enrolment. (See Figure 1.1. for a visual description of this problem. Option A constructs close to the main middle school and while it is close to one other middle school it is too far from three other potential middle schools. Option B secondary school serves all 5 schools within a 5km radius. The size of the school is denoted by the size of the bubble).

Figure 1.1 Visual Description of Key Challenge of School Location Problem



The goal of this project is therefore to improve planning in a way that identifies the potential locations of secondary schools across a large search space that serve multiple middle schools that may otherwise be overlooked. If their collective enrolment is sufficient and distance is minimal, the proposed location can be included within a solution. Indeed, the feasibility of construction requires due consideration to land availability and topography, inter-ethnic or intra-ethnic tensions, all of which require more local knowledge which is out of scope for the project. This approach does not deny these factors, but instead, looks to present an improved and more objective way of initially identifying the potential locations for secondary school construction.

1.2. Research Question

The main research question addressed in this paper is: *Given a budget for the construction of n secondary schools in Ethiopia, where should they be located?* Variants such as n schools in a particular region are also considered. To answer this question, the project explores whether meta-heuristic optimisation techniques offer improvements on the current planning methods mentioned above.

1.3. Structure of Thesis

This thesis is structured as follows. Chapter 2 provides specific background to the case study of this research: Ethiopia. Chapter 3 provides a comprehensive literature review of related work on FLPs and school location problems. Chapter 4 provides an overview of the data used in the project and its limitations. Chapter 5 conducts some analysis on the datasets to gain a better understanding on how to formulate the problem. Chapter 6 describes the methodology, the benchmarks, and experimental settings. Chapter 7 outlines the results, while Chapter 8 provides the conclusion.

Chapter 2

2. Background

Ethiopia was chosen as the case study for this project due to the availability of enrolment and school location data and the scale of infrastructural investment needed to expand the secondary school network over the next 8 years to meet SDG targets. This section provides a brief overview of the country and its educational indicators, describes its recent plans for increasing access to secondary education, and discusses why this research can help improve how Ethiopia meets national and international targets.

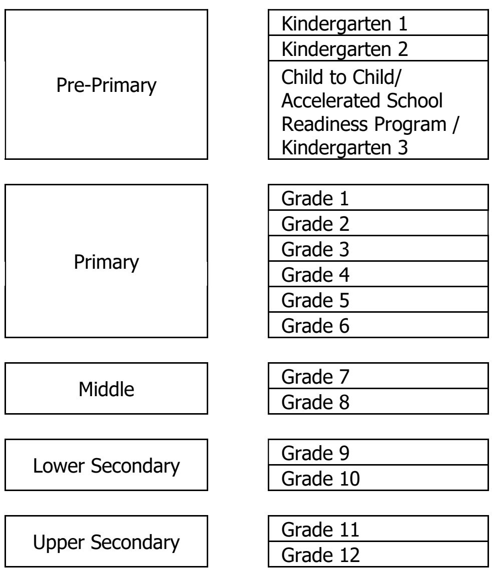
2.1. Country Overview

Table 2.1 Profile

1. Full country name	The Federal Democratic Republic of Ethiopia
2. Area	1,104,300 km2
3. Population	113,656,596 (40% < 15 years old)
4. Regions	9 regions (Afar, Amhara, Benishangul-Gumuz, Gambella, Harari, Oromia, Southern Nations, Nationalities and Peoples Region (SNNPR), Somali, Tigray) and 2 city administrations (Addis Ababa and Dire Dawa)
5. Zones	63
6. Woredas (Districts)	832
7. Urbanisation rate	Less than 20% urban
8. Languages	93
9. Religion	"Ethiopian Orthodox 43.8%, Muslim 31.3%, Protestant 22.8%, Catholic 0.7%, traditional 0.6%, other 0.8%"
10. Ethnic Groups	More than 90
11. Population below poverty line	24% (15% in urban areas, 26% in rural)
12. Major Employer	70% in the agricultural sector

Sources: (CIA World Factbook, 2022), (Ethnologue, 2022), (Reuters, 2021).

Figure 2.1 Ethiopia Educational Ladder



Ethiopia is the second most populous country in Africa and one of the fastest growing economies in the world, sustaining an average GDP growth rate of 9.8% between 2004-2021, compared to the SSA average of 3.8% for the same period (The World Bank, 2021). Alongside this economic progress, significant gains have been made in the educational sector. Since 2004/2005, the GER has increased from 2.1% to 46.4% in pre-primary, 67.8% to 109.9% in primary & middle school (grades 1-8), and from 12.4% to 38.9% in secondary (grades 9-12) in 2019/2020 (The World Bank 2022). See Figure 1 for a breakdown of the educational system in Ethiopia and the corresponding grades[2]. These improvements are a direct result of sustained Government and donor investment. However, with 37,750 primary and middle schools and only 3,688 secondary schools (FGoE, 2020), supply remains the major challenge.

2.2. Prioritising Access to Secondary Education in Sector Plans

Ambitious targets were set in the country's five-year Education Sector Development Plan (ESDP-V) in 2015 to help expand access to secondary education. The Government sought to increase GER in lower secondary from 38.9% to 74% (FGoE, 2015). However, the plan did not specify how many schools that would require. By 2020, the actual GER was 51.05%, a more modest increase of 12.15% than 35.1%. 1,355 new secondary schools were constructed during this period, with many existing schools also expanded to accommodate more students. This type of target setting is problematic. It means that gross enrolment targets are set without a comprehensive understanding as to how many schools are required to meet a target, where the schools should be approximately located, or how construction should be phased in based on the existing demand as funding becomes available. Recognising these issues, the Ministry of Education commissioned the World Bank to GPS map all the schools in the country in 2019. School mapping was last conducted in 2011. The results aimed to guide infrastructure development, policy analysis and decision making within the Ministry of Education's new five-year ESDP-VI (2020-2025). However, in the new sector plan, while a target increase of 24% was set (FGoE, 2020), in the key performance indicator "Number of newly established secondary schools", the target is "TBD"- for "To Be Determined". GIS is not being used to facilitate a more analytical method for determining school locations and setting targets.

2.3. Improving target setting

Irrespective of what the agreed upon target is, there is a need to identify potential locations of schools that bring large returns on future enrolment. It allows for the Government to understand how many schools are needed to meet a target, how much funding is required for construction and where to prioritise when funding becomes available, or proposals for new funding are submitted. However, finding optimal locations is extremely difficult to do manually given the size of the continuous search space. It can result in biased selections, lengthy consultation processes, or a focus only on individual schools with high enrolment rather than clusters of primary schools that are close together. As UNESCO (2022), point out, while "GIS [alternatives] exist which are open-source and free, it requires a certain level of technical expertise. Using GIS often means that actors at decentralized levels are dependent upon these resources and systems at centralized levels, which can cause strains in planning processes". This research proposes a response to this challenge.

[2] Previously, Primary Education in Ethiopia consisted of 8 grades, implemented in two cycles, Primary 1st cycle (Grades 1-4), and Primary 2nd cycle (Grades 5-8). In 2019/2020, the system has been changed to Primary (Grades 1-6) and Middle school (Grades 7-8).

Chapter 3

3. Literature Review

Many different approaches exist concerning the planning of locations of new public services such as schools. This section summarises the related work and although not exhaustive, provides an overview of the current and historical approaches to school planning including the norms relating to distance to school. It then provides an understanding of the broad features of FLPs and finally, reviews some of the literature on FLP approaches used for school location planning.

3.1. Review of School Location Planning

School location planning was originally developed as a process of diagnosing the existing distribution of educational opportunities and allocating resources against a set of criteria or norms (Hallack 1977, Theunynck 2009). It became popular from the 1960s in Africa, building largely on a process of secondary school expansion from France in 1959 (Hallack, 1977: p14). The process, also known as school mapping, has been rolled out across multiple countries with the assistance of international donors and NGOs. Serge Theunynck, one of the leading sources for school location planning in developing countries notes that school mapping has typically been a desk exercise that draws on collective data, estimated enrolment ratios, and limited feedback of local authorities (Theunynck, 2014: p20). The process typically involves making top-down decisions, that are often politically motivated, and can result in the improper placement of new schools or classrooms. “For instance, 16% of the 15,600 classrooms in Guinea were recorded unused in 1999-2000” (*ibid*). Consequently, some norms have emerged across developing countries which help guide the allocation of resources and achieve more equitable access to education. The norms relate to 1) “accessibility and efficiency” 2) “school quality”, and 3) “construction technology” (Theunynck 2009: p15).

Accessibility and Efficiency

These norms refer to the guidance and criteria applied when determining whether to build a new school or expand an existing school and where. Planners typically establish catchment areas which are defined by “the maximum acceptable distance a child can travel between home and school, the size of the school, and the density of school-age-population” (Hallack, 1977: p149). New schools are then constructed in areas that have sufficient population for an economically viable school and are within a certain distance threshold. But what is an acceptable population size to cover or an appropriate distance between home and school?

Size

Countries typically plan the construction of larger schools to yield the benefits of economies of scale. Feeder schools with high enrolment are therefore prioritised or clusters of primary or middle schools that serve as feeder schools to secondary schools. However, given that 58% of

SSA live in rural settings (The World Bank, 2021), there aren't always clusters of feeder schools within an appropriate distance from each other. This raises questions relating to the phased prioritisation of school construction because the "average school model for rural areas needs to be the smallest that is economically feasible, rather than largest" (Theunynck, 2009: p17). Theunynck conducted an analysis of economies of scale across primary schools in 4 countries (Chad, Lesotho, Ghana, and Burundi) and found the optimal minimum size of a primary school to be 200. This means the "larger schools have no significant marginal cost advantage over schools with 200 students" (*ibid*). For secondary education, few sources exist. Ndiritu (2012) recommends the optimal size for secondary schools in Nyandarua District, Kenya and found the optimal size to be 400 students. This ultimately relates to the accumulative minimum capacity of feeder schools in this school location problem. If the accumulative expected enrolment from three feeder middle schools is 100 students, should that be considered a feasible solution? Ndiritu recommends, that new schools shouldn't be constructed in these locations until the feeder schools reach the optimal size (*ibid*). This means there should be more focus on expanding access to primary and middle school education before focusing on secondary. However, this may never happen if there are not enough inhabitants in an area. This is therefore clearly a constraint for this problem that requires a policy choice that considers cost and decisions around equitable access to education.

Distance

Distance-to-school is a key assumption when planning the construction of new schools. In the 1970s, it was commonly recommended to locate schools within a 3km radius or a 45-minute walk (Hallack 1977). However, by the 2000s, in response to a growing body of evidence suggesting the negative impact of distance on access to school, most Ministries of Education in Africa adopted 2km, or 30 minutes, as a planning norm (see Theunynck 2009 for a country level summary on the effects of distance on school enrolment and retention). However, this mostly concerned primary. For lower secondary, little information has been systematically collected across developing countries and "contrary to primary schools, the radius of the school catchment area, a good proxy of home to school, is almost never considered as the basis of a norm", except for India who in 1986 adopted 3km as the planning standard (Theunynck, 2018: p18). Yet, irrespective of the norm, Theunynck points out that "Ministries of Education do not know what the distance is between the schools and the population served by them. They have no household tool to collect this data. The information is found in national household surveys…or poverty surveys" (2018: p2) which are often based on samples and not comprehensive enough to define catchment areas. Simpler methods are therefore needed which rely on more readily available data.

School Quality

These norms concern the factors that create a conducive learning environment. It relates to issues like classroom size, water and sanitation, proximity to roads, boarding facilities, houses for teachers, libraries, health services etc. (Theunynck, 2009). All these determinants influence access and retention of children in school. However, building all these different factors into an

FLP requires a comprehensive map of auxiliary facilities, ground-water maps, a road network etc. which is unrealistic at a large scale. It much more likely to be more readily available from local authorities once initial approximate locations are selected.

Construction Technology

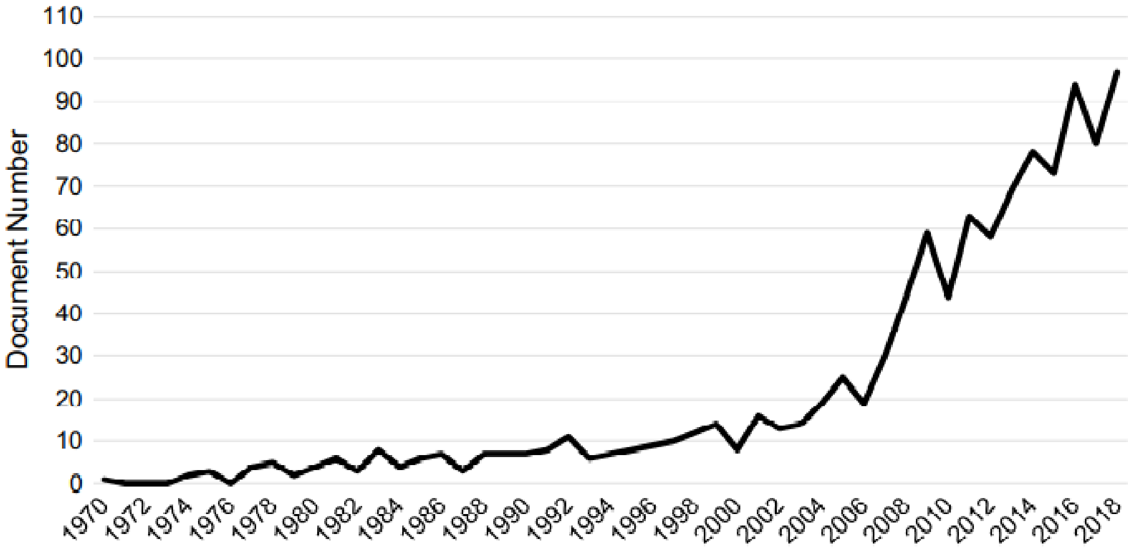
This relates to overall construction costs and procurement arrangements including what materials to use and what architectural standards to follow. Most SSA countries have their own standards, and some standards can vary across regions within countries as planning efforts get devolved (Theunynck, 2014). Standards can also vary across donors and NGOs. This variance affects planning efforts as the budget typically determines how many schools can be constructed. It is therefore a difficult constraint to build into this school location problem.

3.2. Review of Facility Location Problems

The Facility Location Problem (FLP), also known as location analysis, is a popular class of optimization problems in the field of logistics, operations and management science that has multiple applications across manufacturing and service industries. As Plastria (2002: p2) describes: "a location problem arises whenever a question is raised like: where are we going to put the thing(s)", which subsequently raises two follow-up questions: "Which places are available?" and on "What basis do we choose?". Answering these questions typically involves an objective function that seeks to "minimise the cost of satisfying some set of demands (of the customers) with respect to some set of constraints" (Farahani and Hekmatfar, 2009: p1). Given the high costs associated with facility construction, location decisions have long term strategic importance (Owen and Deskin, 1998) and require careful consideration, especially in circumstances where budgets are extremely limited.

While the early history of location science is debated among researchers (Farahani and Hekmatfar, 2009: p1), the study of location theory emerged in 1909 when Alfred Weber developed a model to identify the location of a single industrial plant by minimising transportation costs from two raw materials to a single market and adjusting the location to consider labour costs and the benefits of agglomeration economies (Church et. al. 2022). Distances in the Weber problem are typically the Euclidean distance, but almost all distance functions can be used (Farahani and Hekmatfar, 2009: p5). The objective is often formulated in two ways: 1) minisum, or 2) minimax. Minisum "minimizes the sum of the weighted distances between the new facility and the already existing facilities", while minimax "minimiz[es] the maximum distance between the newly placed facility and all existing facilities" (Litoff, 2015). The direct numerical solution for the Weber problem wasn't found until 1962 by Kuhn and Keunne. In the 1960s, Stollsteimer (1963) introduced the plant location problem which expanded the Weber problem to locate multiple facilities "recognising the importance of economies of scale and diseconomies of scale" to solve for optimal locations (Osleep and Cromley, 1977: p. 143). Since then, there has been a rise in the number of publications on the subject, especially since the late 2000s. See Figure 1, which shows the growing number of publications on service facility location problems between 1970-2018.

Figure 3.1 Number of documents on service facility location problems per year.



(Turkoglu and Genevois, 2019: p402).

3.3. Key Properties of FLP

Key properties do emerge across different facility location problems which can help determine what type of model is necessary. Turkoglu and Genevois (2020: p404-468) examined 90 different papers on service facility location problems and identify the following key properties across each the FLPs. The properties outlined below are relevant to this school location problem.

Table 3.3. Key Properties of FLP

Key Feature	Description	Relevance to School Location Problems
Space	Can the proposed locations be located anywhere on a plane (continuous) or can the facilities only be located from a given set of potential sites (discrete).	Schools can be located as longitude or latitude coordinates anywhere within a country boundary. If discretised, pre-determined catchment areas are ranked for selection.
Distance	For network models, the distance is the shortest path. In discrete or continuous location models distance metrics are typically Euclidean, Manhattan or Chebyshev.	Any method can be used, but limitations exist with respect to data availability and reliability.
Time	Is the problem static in that ‘there is just one period and parameters don’t change over time.’ Or is the problem dynamic “where there are discrete time planning horizons and the problem parameters change over these planning horizons” (ibid: p406).	Problem is static in that the problem is solved for a particular time. Each new year new schools are constructed, and solutions will need to be revised. But this also requires regular GPS mapping of new schools and marking of closed schools.
Parameters	Is the model deterministic or probabilistic? In deterministic models, the model inputs are “assumed to be known with certainty and fixed whilst in probabilistic models inputs are not	If the problem is continuous the model will need to be probabilistic, due to its np-hard nature. If discretised, deterministic models can be used. Meta-heuristics

	known with certainty and subject to uncertainty" (ibid: p406).	involve stochastic search methods which are probabilistic.
Capacity	Many location models such as maximum covering, p-median and p-centre models assume that facilities have unlimited capacity i.e., an un-capacitated problem. Others have limited capacity in that demand allocation is restricted at facilities i.e. a capacitated problem.	Minimum and maximum capacity can be considered, but this depends largely on budget.
Facilities	Are the number of facilities pre-determined? Or does the solution involve providing this data as output? Problems like p-median and p-center require the number of facilities as input for the problem.	Current targets for secondary are set using GER and population estimations. Building a model using GER would mean the number of facilities is not predetermined and the model would provide this as output. If budgets are known in advance, the number of facilities can also be determined in advance.
Number of Facilities	Is the problem single facility or multiple facility? If single facility all demand is served from one facility, whereas with multi-facility different demand points must be met which increases the complexity of the model.	This is a multi-facility problem.
Facility Type	Are all the facility types identical or do they differ? Facilities can often differ in terms of price and size.	Facility type and size per new school depending on the demand from feeder schools.
Number of Objectives	Does the problem seek to solve single or multiple objectives? "Single objective models aim to minimize or maximize a function such as maximizing total profit or minimizing total cost. However, multiple objective models intend to find compromise solutions by dealing with conflicting objectives"	Single objective functions would aim to maximise enrolment, minimise cost, or minimise the weighted distance, or maximise accessibility. Multiple objectives would involve two or more objective functions for optimisation.
Competition	Will there be other services within the market that act as competition for market share? For example, a fire station will be uncompetitive in that it will cover a general area, while a supermarket may be competitive in that other supermarkets within the area may compete over the same customers.	Competition could be factored in depending on issues such as commute time, over-crowding, or even more socio-economic issues like ethnicity. However, in areas of poor supply, it is not realistic. Instead, distance is the more appropriate determining factor.

3.4. Location Models

As is evident from the different properties for an FLP, variations of the problem have emerged with respect to key decision variables, assumptions, and constraints, which impact model design. Klose and Drexl (2005) summarise these into three broad categories: 1) continuous location models, 2) network location models, and 3) mixed-integer programming models. Continuous location models have two defining characteristics: a) the solution space is continuous, meaning that it's possible to locate facilities at every point on the plane and b) distance is measured using a suitable metric like the Manhattan distance or Euclidean distance. The aim is then to find the coordinates for the given number of facilities by minimising "the sum of distances between the facilities and [the] given demand points" (*ibid*: p6). When seeking a single facility, it reduces to the Weber problem. For multiple facilities, exact algorithms can be found, but it is NP-hard which makes it difficult to solve "large real instances in short times" (Armas et al. 2017: p1163). Approximation algorithms have been explored but they tend to be 'outperformed by straightforward heuristics with no performance guarantees' (*ibid*).

For discrete models, network location and mixed-integer programming models can be used. For the former, distance is taken to be the shortest path on the graph. "Nodes represent demand points and potential facility sites correspond to a subset of the nodes and to points on arcs" (Klose and Drexl, 2005: p6). Instead of locating facilities anywhere on a plane, the problem becomes discretised with "demands and travel between demand sites and facilities…. assumed to occur only……on the nodes or links of a network" (Daskin, 2013: p12). Of the network location models, the p-median model is "probably the most applied location model" (Segura et al. 2016: p 1140). The model looks to 'minimise the demand-weighted average distance between a demand node and the facility to which it is assigned' (*ibid*: p1139). Such models are often used in distribution planning when minimising the overall transport cost is required. The p-center model is also common and used in situations where the aim is to locate a given number of facilities such that "the maximum distance is minimised" (Klose and Drexl, 2005: p7). In both cases, "given demand and cost minimisation" is considered as the objective (*ibid*: p7).

For mixed integer programming, the goal is to determine the optimal placement of a discrete set of facilities over a finite set of possible locations. It differs from network location models in that these models "explicitly take the structure of the set of potential facilities and the distance metric into account while mixed-integer programming models just use input parameters without asking where they come from" (Klose and Drexl, 2005: p8). An example of mixed integer programming models includes the maximal covering problem. This seeks to determine the minimal subset of facilities such that demand is reached within a maximal distance i.e., the covering coefficient e.g., one hospital covering multiple demand points within a 50km radius. Facilities are then allocated to cover pre-determined demand points.

For a more detailed summary of the variations of the facility location problem see Klose and Drexl (2005) and Farahani and Hekmatfar (2009).

3.5. Application of FLP techniques to School Location Planning

In the 1970s, the merits of operational research techniques were explored for school location planning. As Hallack (1977 p222-223) describes, it would involve a linear procedure which seeks to optimise a cost function (e.g. cost of transport), or a provision of education function (e.g. a guarantee of inter-ethnic mixing), with weighting then applied to limitations e.g. travel time from home to school. When considering existing pre-determined catchment areas, the problem becomes a discretised allocation problem in which a set of population areas are allocated to fixed capacity schools. However, it was deemed infeasible back then due to computational limitations in the 1970s. Another important point was also noted by Hallack (1977) who stated that in finding the "economic optimum", its application often carries psychological, social, and political consequences. An objective method for school location planning is therefore not sufficient by itself. Location decisions are inherently normative and require careful consideration of the socio-economic consequences.

In the late 1980s and 1990s computational limitations gradually became less of an issue and a growing number of publications used FLP techniques to solve school location problems. Tewari and Jena in 1987, was one of the earliest publications. Across two cities in India, more than 1.5 million people, and a maximum distance of 8km to the high school, they developed a maximal coverage model which aimed to maximise the population covered for a fixed number of schools. Later, Pizzolo and da Silva (1997), developed an un-capacitated p-median model for large urban settings which sought to identify the locations of p schools which minimizes the sum of the overall residence-to-school distances. Today, as de Armas et al. (2022: p4) note, "most state-of-the-art approaches about school location" use a variation on these models: 1) maximum covering or 2) the p median. The variations include introducing capacity constraints, budget constraints, accounting for the existing locations of schools, expanding existing schools, including different transportation modes (e.g. school buses), introducing urban and rural parameterisation, and even including normative goals like ensuring racial balance per school (For a more detailed review, see de Armas et. al. 2022). What is clear from this review is that choosing an approach then depends largely on the properties of the problem, data availability, data reliability, key assumptions, and constraints.

3.6. Contribution of Project

Although very complex systems have been developed that can more accurately predict travel time to school from population centres, this project contributes to the wider literature by developing a model that uses data that is readily available within Ministries of Education in SSA. Population data in Ethiopia is not easily available and its reliability is questionable (see section 4.3). Middle schools' enrolment and location data is often used as a proxy for demand for secondary. This project therefore aims to gauge the merit of current practices and build meta-heuristic models which outperform these approaches. The models will be more easily adaptable to the context as the data used will garner more trust. Since school census data is collected annually, it can also be a useful tool for routine adaptation.

Chapter 4

4. Data

Three different sources of data from Ethiopia are used in this project. This section will describe the datasets, the steps taken during pre-processing and the limitations that surrounding their use.

4.1. Datasets

- **School Enrolment Data (2018/2019).** This contains school level data on 41,485 schools in Ethiopia from the 2018/2019 school year. It was obtained from the Ministry of Education. The dataset contains the unique code of each school, the level of education offered at the school (pre-primary, primary, secondary), gender disaggregated enrolment by grade level and the information on the administrative levels of the school i.e., the region, zone and *woreda* (district), of the school. (FGoE, 2019)

- **GIS Locational Data (2020).** This includes school locational data from 26,227 schools from the Ministry of Education, Ethiopia. The data was obtained in 2020. School mapping was postponed in 2020 due to the outbreak of COVID. The dataset contains the longitude and latitude coordinates of approximately 70% of primary and secondary schools in Ethiopia. (FGoE, 2019).

- **Subnational Administrative Boundaries (2020).** Regional, zonal and *woreda* (district) shape files were obtained from UNOCHA (2020). With these administrative boundaries, the models can ensure proposed locations are within agreed upon administrative boundaries. Should boundaries change, the system can use updated shape files and account for new administrative boundaries. Data is also more portable when generated with indexes that match the standardized region, zone or *woreda* codes. (HDX, 2020)

4.2. Pre-Processing

- **Keep only essential data from school enrolment dataset.** Of the 41,485 schools in the dataset, only middle and secondary schools are required for this project. The dataset is therefore filtered and reduced in size.

- **Cross check GIS Locational data for the correct administrative data.** New administrative boundaries are drawn in Ethiopia regularly. To ensure the locational and enrolment data correspond to the correct administrative boundaries, there is a need to check the longitude and latitude coordinates against the correct Ethiopia Shape files. Using GeoPandas, each of the coordinates are checked against the country, region, zone and *woreda* (district) polygons, and the correct boundary names are extracted for use.

- **Merge school enrolment dataset and GIS location dataset into a cleaned dataset.** The final sample of schools included is 16,157 middle schools and 1,658 secondary schools.

4.3. Limitations

- **School Mapping Network Incomplete.** Given that only 70% of the school network is used in this project, the results cannot be used as the comprehensive guide for planning secondary school construction across Ethiopia. It does however still provide useful results that can help planners identify potential locations for construction. Once the mapping is complete, the models can also be re-run to include the missing schools.

- **Population Estimations Omitted.** In Ethiopia, the last census was completed in 2007 and it has been widely considered inaccurate (for more information see: Bekele, 2017), largely due to the political nature of a census in Ethiopia. "Counting is so important to Ethiopians various ethnic groups because it influences how much power and money they get" (The Economist, 2019). Some ethnic groups estimate an undercounting of approximately 3 million in the census (*ibid*). Although population estimations do exist, they are based on projections made from the 2007 census. This influences educational planning. For example, the Net Enrolment Rate in primary[3] was 100.05% in 2017/2018, which is statistically impossible. The decision was therefore taken to omit population figures in this project and focus instead on using middle schools and their size as a proxy for where communities live. By 2013, the World Bank estimated that less than 13 percent of children in Ethiopia had to walk more than 2km to primary school (The World Bank, 2017). With an increase in the number of schools from 30,485 in 2013 to 34,342 in 2019, this proportion will have reduced substantially. As such, building a new secondary school close to an existing middle school gives a good, albeit inexact, indication of where communities can reach. It's also more readily available and can therefore provide quick that don't require politicised discussions around population figures.
- **Budget Estimations Omitted**. How many schools should be built is a question of budget. However, as pointed out in Section 3.1 budget depends on size, construction requirements, and school quality standards. This can differ from region to region, donor to donor, or community to community given the foreign exchange shortages that Ethiopia faces (Reuters, 2018), costs for construction materials are likely to change substantially from month to month and year to year. Building in variable costs is therefore extremely difficult and likely to affect the usability of the model. Instead, these costs can be pre-determined by local authorities. Feasibility can then be determined whether their budget covers the proposed *n* schools.

[3] "Total number of students of the official age group for a given level of education who are enrolled in any level of education, expressed as a percentage of the corresponding population." (UNESCO, 2022)

Chapter 5

5. Data Exploration

This section explores the data available to help describe the properties of the school location problem and gain a better understanding of the current distances children face in getting to school.

5.1. Establish a Distance Metric

To formulate the problem and explore existing distances, it's necessary to first establish a distance metric. The most common approach used is the Euclidean distance, but as UNESCO (2022: p4) point out, it can give a "false picture of geographic realities". Certain characteristics such as the slope of a terrain, and bodies of water can make a school appear close using a straight-line distance, but it may be unreachable. UNESCO (2022) propose an Isochrone alternative for less developed countries which charts an area that has the same travel time to a reference point considering cars, public transportation, walking and hiking and natural barriers. However, it assumes that walkers and hikers follow the same path as the road network and given that road and paths are not extensively mapped in Ethiopia, this is not an appropriate solution. Instead, this project uses the haversine distance which measures the angular distance between two points on a sphere in radians, which can then be converted to kilometres. The haversine distance faces much of the same criticisms as the Euclidean distance, but it is simple, widely used, and easy to understand. Only 12 per cent of the planning officers in the Ministry of Education in Ethiopia have a diploma in the domain (UNESCO, 2022: p4), and given the lack of technical expertise, the simpler the technique the better. The output can be easily understood without the need for outsourcing expertise in lengthy consultancy processes. Regardless of the output, all final locations will need local verification anyway.

5.2. Visualise and summarise the data

Table 5.1. Education Statistics

Key Statistics	Middle Schools	Lower Secondary Schools
School Buildings	16,157	1,658
Enrolment	2,316,324	1,459,749
Avg. Enrolment Per School	143	880
Min Enrolment Per School	2	27
Max Enrolment Per School	2,198	3,970

Overall, there are substantially more middle schools than lower secondary schools in Ethiopia, almost 10-fold. However, the enrolment difference isn't as pronounced which suggests two things: 1) students drop out before accessing lower secondary, and 2) lower secondary schools are clearly larger size buildings than middle school. The latter is supported by the much higher mean enrolment for lower secondary school. It's clear there is a preference for constructing larger schools and establishing economies of scale. This is likely due to budgetary constraints. Still, more than 137 lower secondary schools (8%) have less than 200 students enrolled, the minimum recommended size for an economically viable school according to

Theunyick (2009) and Ndiritu (2012). The largest lower secondary is almost 4000 students which could dictate the upper capacity constraint for a new school.

Figures 5.1 and 5.2 provide a visual representation of the distribution of schools across the country. One key limitation is quickly observable. Of the sample of schools mapped, less than 5% of the schools were mapped in two regions: Afar and Gambella. They are therefore excluded from model development in the subsequent chapters.

Figure 5.1. QGIS Satellite View of Middle School Sample

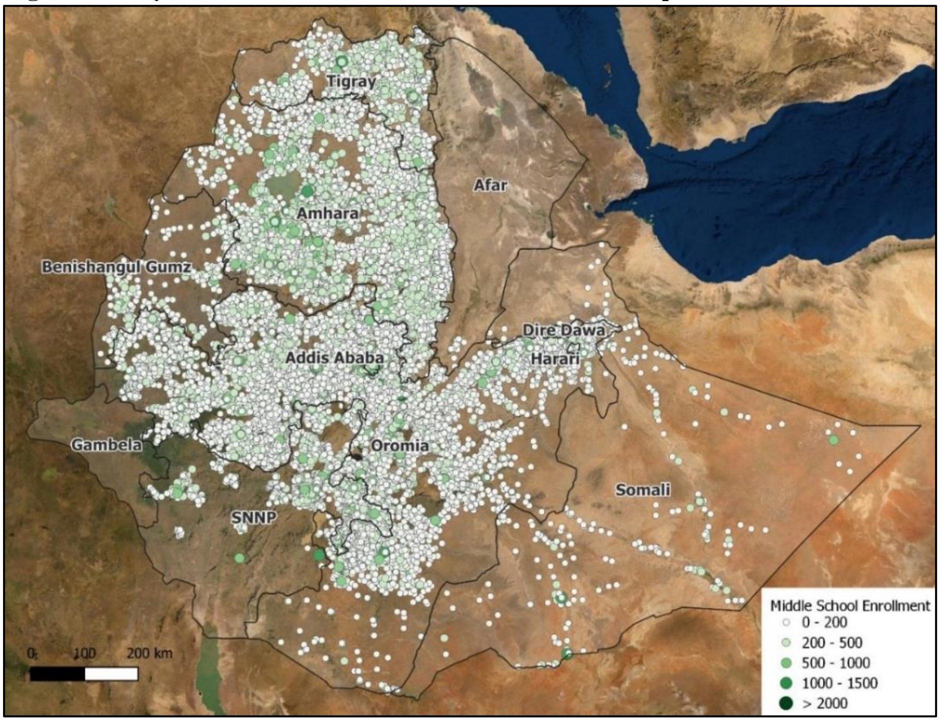
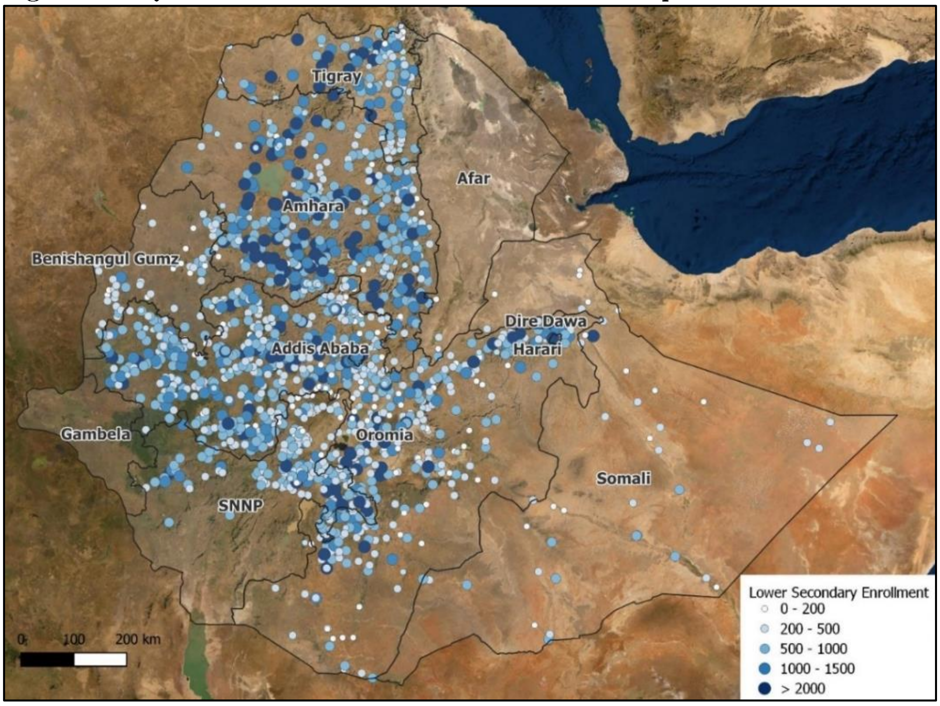


Figure 5.2. QGIS Satellite View of Middle School Sample



The average distance from middle school to nearest lower secondary school is 6.88km, and the range spans from 0 – 133.38km. For rural schools, the average distance increases to 7.46km. For urban schools, it's 4.1km. The pastoralist Somali region has the longest distances to travel for secondary education, but many outliers exist across other regions even within urban locations (See Figure 5.4). Most middle schools, 55%, are more than 5km away from the nearest lower secondary.

Figure 5.3. Histogram of the Distances of Middle Schools to Lower Secondary Schools

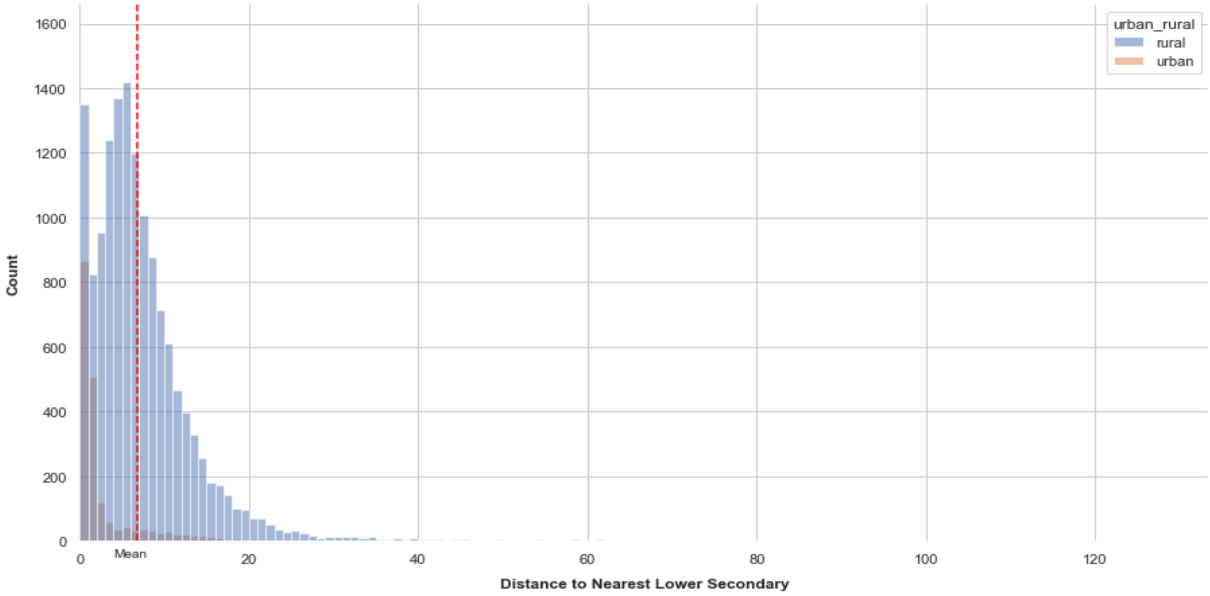
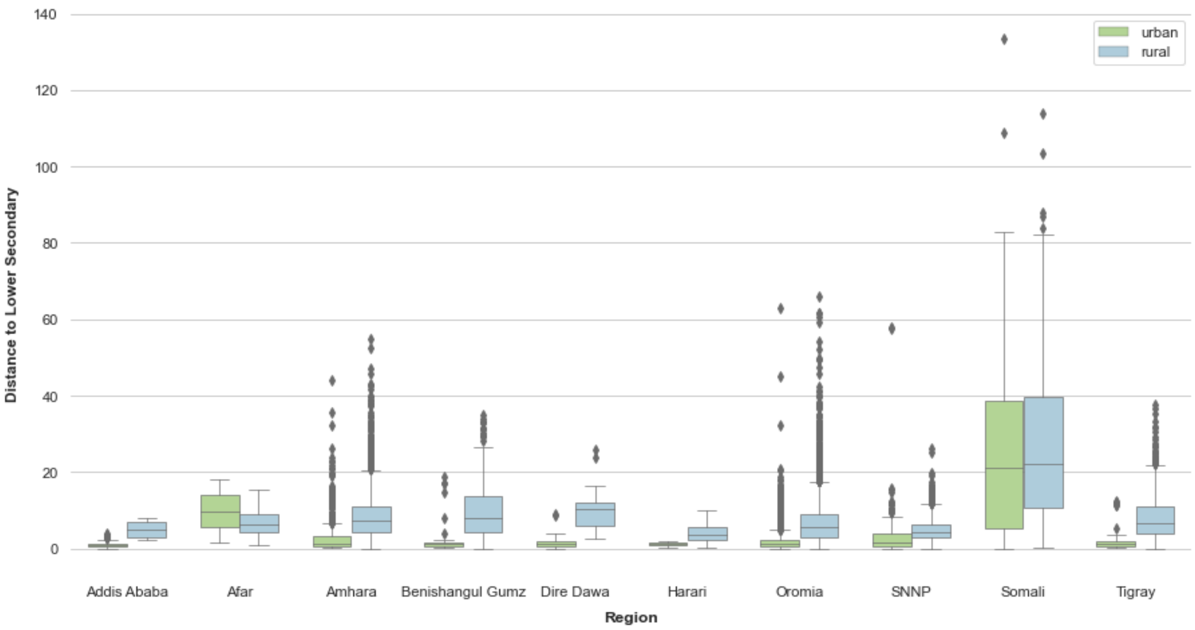


Figure 5.4. Box Plot of Current Distances of Middle Schools to Lower Secondary by Region



Conversely, 97% of lower secondary schools are located within 5km of at least 1 middle school. A quick inspection of some of the schools outside 5km using Google Earth would suggest that these schools are located in large settlements and that they may not be close to primary schools due to the incomplete mapping exercise. Since they are outside of what might be considered an acceptable distance range from middle schools, it won't impact the methodology of this project.

Each lower secondary school serves approximately 9.25 middle schools. The feeder schools are on average 7.6km away from each middle school. Taking the furthest middle school away as the maximum radius of the catchment, the average catchment size is 13.01. Most primary schools are therefore too far away from their closest secondary school.

Figure 5.5. Average no. of feeder Schools by region

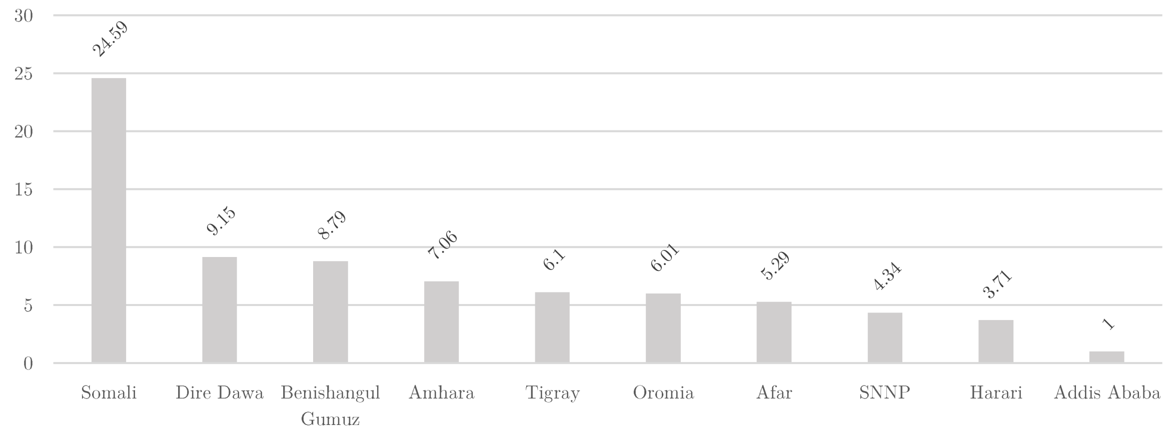
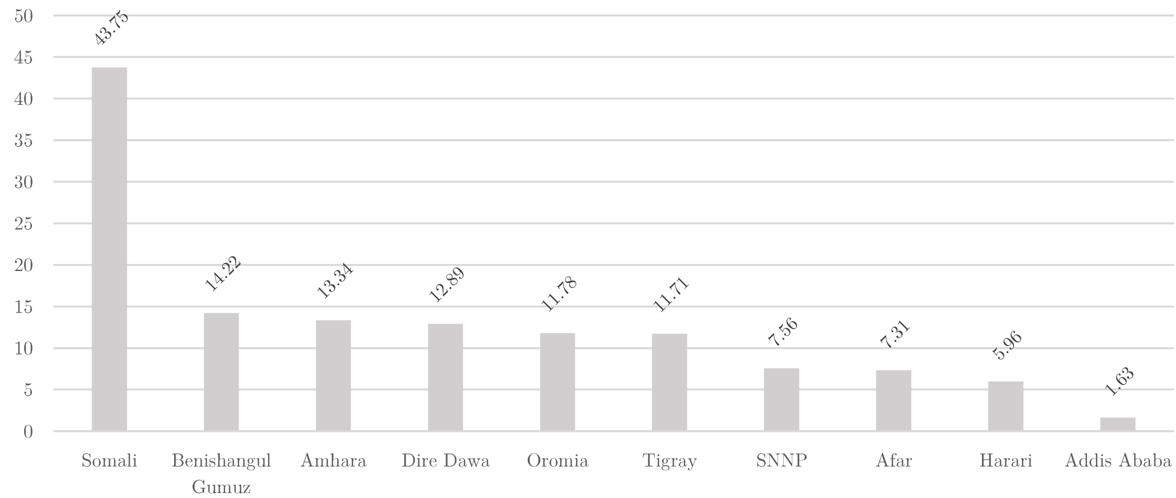


Figure 5.6. Average catchment size (radius) in km by region



Clusters of primary schools do exist. 65% of middle schools are within 3km of at least two other middle schools, 90.4% within 5km. Filtering this to only include middle schools that are more than 5km away from the nearest lower secondary schools (8,832 in total), the figures drop to 47% and 80% respectively. See Table 5.2 for a more detailed breakdown. It displays the size of the middle school clusters from 0 to greater than 10 and total distance in km the schools are from each other from 1km to 5km.

Table 5.2. Clusters of middle schools with more than 5km distance to lower secondary

No. of	1km		2km		3km		4km		5km	
Schools	count	%	count	%	count	%	count	%	count	%
0	7548	85.5%	6557	74.2%	4692	53.1%	2845	32.2%	1770	20.0%
1-2	1028	11.6%	1794	20.3%	3276	37.1%	4041	45.8%	3673	41.6%
3-4	161	1.8%	292	3.3%	566	6.4%	1356	15.4%	2117	24.0%
5-6	63	0.7%	108	1.2%	164	1.9%	350	4.0%	809	9.2%
7-8	19	0.2%	55	0.6%	83	0.9%	126	1.4%	246	2.8%
9-10	5	0.1%	9	0.1%	24	0.3%	74	0.8%	136	1.5%
> 10	8	0.1%	17	0.2%	27	0.3%	40	0.5%	81	0.9%

5.3. Key Observations

- Given the large proportion of middle schools clustered together, there is sufficient justification for exploring solutions that offer an improvement on selecting the top n middle schools and only constructing schools nearby. The large size of the clusters also shows that finding optimal locations is computationally expensive and therefore the project will add value.

- As the haversine distance is used, it represents the best-case expected travel time scenario. All proposed locations therefore need to be cross examined for feasibility. This can be done remotely using GIS software like QGIS or ArcGIS. If one of the proposed schools is in a lake or a ravine, the solution is void, and the next best solution can be reviewed. Similarly, the straight-line distances from middle school to the proposed location need to be observed for potential barriers e.g. inclines, rivers or lakes. Local knowledge can support this effort and help determine feasibility.

- The preference for building larger lower secondary schools supports the idea to focus on maximising expected enrolment from middle schools.

- Deciding the appropriate distance for travel is a normative and strategic decision. While children might already be commuting more than 6km to travel to lower secondary, it doesn't mean the network should expand based on factors like this, or it risks perpetuating inequity. As discussed in the literature review, distance has a negative effect on enrolment and there is a need to shape the expected enrolment to ensure that distance is minimised, but enrolment is maximised.

Chapter 6

6. Methods

This chapter outlines the overall methodology employed. It provides an overall description of the problem and key assumptions, describes the optimisation algorithms used and outlines how the model is formulated. Finally, the projects benchmarks and experimental settings are explained.

6.1. Problem Description

As described in previous sections, the problem seeks to expand the secondary school network by constructing new lower secondary schools that best serve the demand from the existing middle school network. The expectation is to construct new schools where expected enrolment is maximised thereby creating economies of scale. Distance between feeder middle schools and the new locations must also be minimised to ensure children are within an acceptable distance to attend the school.

In formulating this problem, several assumptions are considered:

1. Each middle school represents the demand node and therefore the population centre.
2. Regional Education Bureaus have a fixed budget for construction of new schools. This project assumes the budget is sufficient to cover 5 or 10 schools per region. Cost is therefore not a constraint of the model.
3. A new school can be located at any longitude or latitude if it's within a regional boundary i.e., the problem is continuous.
4. Distance affects children willingness to travel to school. 3km is considered an acceptable distance while 5km is considered the maximum acceptable distance. The accepted distance norm of 2km for primary is extended to 3km for secondary as the children are older and more likely to walk longer distances. This is the recommended distance for junior secondary proposed by Theunynck (2009). The 5km distance around a secondary school is therefore considered the radius of the catchment area. This assumption is made knowing that an acceptable distance is a normative decision. While some children will commute more than 5km to a secondary school, it doesn't mean that schools should be built with this in mind. Instead, the expansion of a school network should aim to close the distance due to the impact on expected enrolment and retention. This creates more equitable access to education.
5. Schools are considered uncompetitive given the poor supply of secondary education. Children attend the nearest lower secondary school.
6. New schools won't be constructed unless the expected enrolment is above a minimum capacity of 200 students. A maximum capacity is not set as it's determined by budget, which is not known in advance.
7. The number of feeder middle schools to a new secondary school is not limited by any policy.

8. Grade 8 and 9 enrolments for each middle school is the predictor for grades 10 and 11 of each new lower secondary. Although year to year dropouts could be considered, the construction of a new secondary school could also lead to higher enrolment and better retention thereby effecting dropouts. Given these unknown factors, the expected enrolment of new lower secondary schools is based only on the total enrolment of each nearby middle school.

6.2. Optimisation Algorithms

Two different algorithms are used in this project to find an optimal solution: Random Search (RS), and Covariance Matrix Adaptation Evolution Strategy (CMA-ES). RS is a simple optimisation algorithm which consists of randomly generating many solutions and choosing the one with the best results i.e., the minimum or maximum result. It provides a low benchmark for the project.

CMA-ES is an evolutionary black-box optimisation technique often used for continuous problems. The algorithm creates a sample population using a multivariate Gaussian distribution. A fitness function evaluates individuals, and the mean of the distribution is then calculated from the best performing individuals of the population. "The covariance matrix of the distribution is also updated to bias the generation of the next set of candidates toward directions of previously successful search steps" (Cai & Jiang, 2013: p131). Through subsequent generations, the sample distribution expands or contracts to get closer to the global optimum. The algorithm was chosen for several reasons. Firstly, CMA-ES is a meta-heuristic optimisation technique which performs well in finding approximate solutions and since the problem is NP-hard, finding exact solutions was not feasible. Secondly, the algorithm is considered reliable and performs well when compared with other black box optimisation algorithms (for more information on this, see Hansen et. Al. 2009). Thirdly, the algorithm is relatively parameter-free in that there are few hyper-parameters to tune. "Finding good (default) strategy parameters is considered as part of the algorithm design, and not part of its application" (CMAP, 2016). The algorithm is therefore more user-friendly when compared with other black-box optimisation techniques.

6.3. Model Formulation

Glossary

- MS: Middle Schools
- SS: Lower Secondary Schools
- EE: Overall Expected Enrolment
- EEI: Expected Enrolment Increase
- Feeder: a middle school that supplies a cohort of children to a secondary school.
- x: proposed longitude and latitude of new secondary schools (the genotype).
- d: distance in km between middle schools and the proposed new schools.
- n: total number of proposed new secondary schools.
- f(): objective function

The objective function, *f()*, aims to maximise *EE* given a set of feasible locations: x, of size n. The total number of new MS to construct, n, must be determined in advance. Feasibility of x is constrained using a box boundary which is based on the minimum and maximum MS longitudes and latitudes. Initial solution generations for x are randomly sampled from within the box boundaries using a *generate_random_sp()* helper function. Only randomly sampled solutions that fall within the regional polygon will be selected. The function *check_region()* is used to confirm this.

The fitness is evaluated using an *expected_enrolment()* function. This takes in four parameters: 1) a vector of GPS locations for all MS, 2) a vector of enrolment data for all MS, 3) x, and 4) a vector of the current distance between MS and their closest existing SS. The function then calculates the haversine distance in kilometers (km) from each new SS in x to all MS within the given region. The algorithm then addresses three key challenges:

1) **The need to identify the MS with the minimal distance to the new SS**. In this case, if an MS is close to two or more new SS, only the closest is selected. But if the MS is closer to an existing SS, then it's ignored as it's already a feeder school to the existing SS.

2) **The need to estimate the expected enrolment of the new SS, based on the distance to its feeder MS**. If a school is beyond a certain threshold distance in km, the MS is not considered a feeder school for an SS as it's considered too far to walk. To handle this, situation, a helper function, *shape()*, is used to estimate the expected enrolment from MS to the nearby SS.
   - *Shape Function.* The function takes in as parameters a) distance to nearest SS in km and b) MS enrolment. It then makes some assumptions about distance to return the expected enrolment per school. If a middle school is located within 3km of the newly proposed secondary school, it is expected that all children will attend the school. If the school is located more than 5km, zero children are expected to attend. If the school is located between 3-5km, a linear drop-off will affect the expected enrolment, e.g., for a middle school of 100 children at 4km, 50 students are expected to attend the new lower secondary school.

   The shape function is applied to all MS based on the distance to x.

3) **The need to divert expected enrolment from an MS if it is currently a feeder school to an existing SS but the new SS proposed is closer**. This is dealt with by subtracting the expected enrolment from the MS to the existing SS using the shape function, and then allocating it to the new SS which is closer.

The overall expected enrolment is then returned. EE includes the expected enrolment from x + the existing SS enrolment. Individuals within a population are evaluated on this figure. For CMA-ES, the objective function f() seeks the global minimum. As this problem is a maximisation problem, f() returns the inverse of the fitness function i.e.

expected_enrollment()*-1. See below a description of the fitness and objective functions in pseudo code.

```
# Fitness Function
def expected_enrollment(MS_location, x, MS_enrollment, current_distances):
    current_enrolment = sum(SS_Enrollment)
    d = calculate distance from MS_location to x
    d = Filter to only the closest SS in X
    d = Filter to only distances <5km and < current_distances
    d2 = [x, d, distance from MS to x, shape(enrollment of d)]
    d2 = Filter to only include the nearby SS if MS is close to 1 or more SS
    EEI = sum(d2[shape(enrollment of d)])
    # if new school is closer than old SS, subtract enrolment from that school
    dc = shape(current_distances, MS_enrollment)
    d_old -= sum(dc)
    EE = EEI + d_old
    return EE

# Objective Function
def f(x):
    return expected_enrollment(…)*-1 # inverse to maximise
```

The language of instruction for secondary is English across Ethiopia and therefore, in theory, SS could be constructed to serve MS across boundaries in two or more regions. However, this project has chosen to develop regional models rather than a national model. The main reason for this is that the Government of Ethiopia is heavily decentralised, and budgets are allocated directly to Regional Education Bureaus (REBs). Each REBs then decide how many schools can be constructed within their budget. Moreover, given the general poor supply of SS, the REBs will likely prioritise MS that fall within their own boundary lines. No cross-border exploration is therefore needed.

The results from each model are stored in a csv. As the haversine distance assumes no geographical constraints, the results are then further examined in QGIS for feasibility. The best results per region are surveyed over satellite imagery to look for barriers like water bodies, ravines, etc. If a barrier exists, the next best solution is selected.

6.4. Performance Benchmarks

Benchmarks provide a measure of success or failure. The overall performance is measured by finding a solution that provides the maximum expected enrolment from x. RS establishes the low benchmark for the optimisation algorithms. A second benchmark is the more intuitive manual approach typically followed by the Ministry of Education. This involves reviewing the MS enrolment by region, sorting it by the highest enrolment, and selecting the top n schools. The aggregate enrolment from only these n schools then represent the basic benchmark. Although potentially unbeknownst to the planner, the top n schools may also be close to other feeder MS. Schools with high MS enrolment are more likely to be in urban areas and therefore enrolment from these MS will also contribute to the new proposed school. To assess this, we assume the locations of the top n MS as the locations for the new school and apply *f()* to estimate the expected enrolment. This represents the more advanced benchmark as it provides a truer estimate than only the enrolment from the top n MS. The goal of this project is then to test whether meta-heuristic techniques can offer improvements on these benchmarks.

6.5. Experimental Settings

The project was implemented in python on a Windows 10 laptop with an Intel Core i7 10th Gen, and 16GB RAM. A custom random search algorithm was used while the CMA-ES algorithm was taken from the cma library[4].

Six main experiments are conducted in this project. The experiments provide regional comparison and gauge the effect of adjustments to 1) the algorithm, 2) the fitness evaluation budget, 3) the initial solution points (starting points), and 4) the main hyper parameter of CMA-ES: the sigma value. The regional comparisons provide useful information for how CMA-ES performs against the benchmarks and random search in large and smaller search spaces. Given the computational constraints, a fitness budget of 20,000 is tested and a fitness budget of 30,000 is then tested on the best sigma values. Thirty different starting points are considered for CMA-ES to demonstrate the stability of the results. For the sigma values in CMA-ES, 10 different values are explored. Sigmaσ in CMA-ES represents the initial standard deviation or step size. It requires scaling per region as the size of the search space differs. As such, 10 different evenly spaced-out values are tested between 0 and the standard deviation of the MS longitude and latitude coordinates, calculated as the square root of the variance: *square_root(std(longitudes)**2 + std(latitudes)**2).* All experiments are timed to measure the speed of the process. Average runtime (aRT) measures the time it takes to complete one run per algorithm. For CMA-ES this includes a run for each of the 30 different starting points. Experiment runtime or eRT is the overall time it takes to run a full experiment per region.

As shown during the data exploration stage, the minimum current distance to SS varies from region to region. As such, a 5km maximum distance can't be set for all regions because the regions of Addis Ababa, Dire Dawa and Harari, don't have sufficient sample of MS that were more than 5km distance away from the nearest SS. The minimum distance was therefore altered in these regions. See Table 6.5.1 for the breakdown of minimum distances per region that was used during experiments.

Table 6.5.1. Minimum Recommended Distance Per Region

Region	Minimum Distance	Comment
Addis Ababa	2	Only 3 MS > 5km, 9 MS > 3km, but 21 > 2km.
Amhara	5	3376 MS > 5km
Benishangul Gumz	5	204 MS > 5km
Dire Dawa	5	29 MS > 5km
Harari	3	Only 10 MS >5km, 21 > 3km
Oromia	5	5037 MS > 5km
SNNP	5	675 MS > 5km
Somali	5	272 MS > 5km
Tigray	5	686 MS > 5km

[4] Available at: https://cma-es.github.io/

Furthermore, there are many clusters of MS within 5km from each other whose aggregate expected enrolment does not exceed 200, the minimum capacity set in the project. As such for these regions, the value of n was reduced. See Table 6.5.2 for a breakdown of values of n used in the project.

Table 6.5.2. Maximum value of *n* used for experiments.

Region	N<=5	Comment
Addis Ababa	5	
Amhara	5	
Benishangul Gumz	5	
Dire Dawa	4	Insufficient MS meeting minimum capacity. Reduced to 4
Harari	2	Insufficient MS meeting minimum capacity. Reduced to 2
Oromia	5	
SNNP	5	
Somali	5	
Tigray	5	

Chapter 7

7. Findings

This section presents the results in figures and tables and provides summary analysis. Two different approaches were taken when adopting CMA-ES. The first, CMA-ES-1, runs the algorithm on the entire regional dataset. The second, CMA-ES-2, runs the algorithm on a subset of the data. The subset filters the regional datasets to include only MS that are within 5km of each other. It then aggregates the enrolment of these schools and returns only the top *n* schools with the highest enrolment. This reduces the size of the dataset which improves the efficiency of the algorithm. The results presented below are summarised by the best results by algorithm according to the size of *n*.

7.1. Results

Table 7.1 N<=5 Experiment results by region and algorithm (Max Iterations: 20,000)

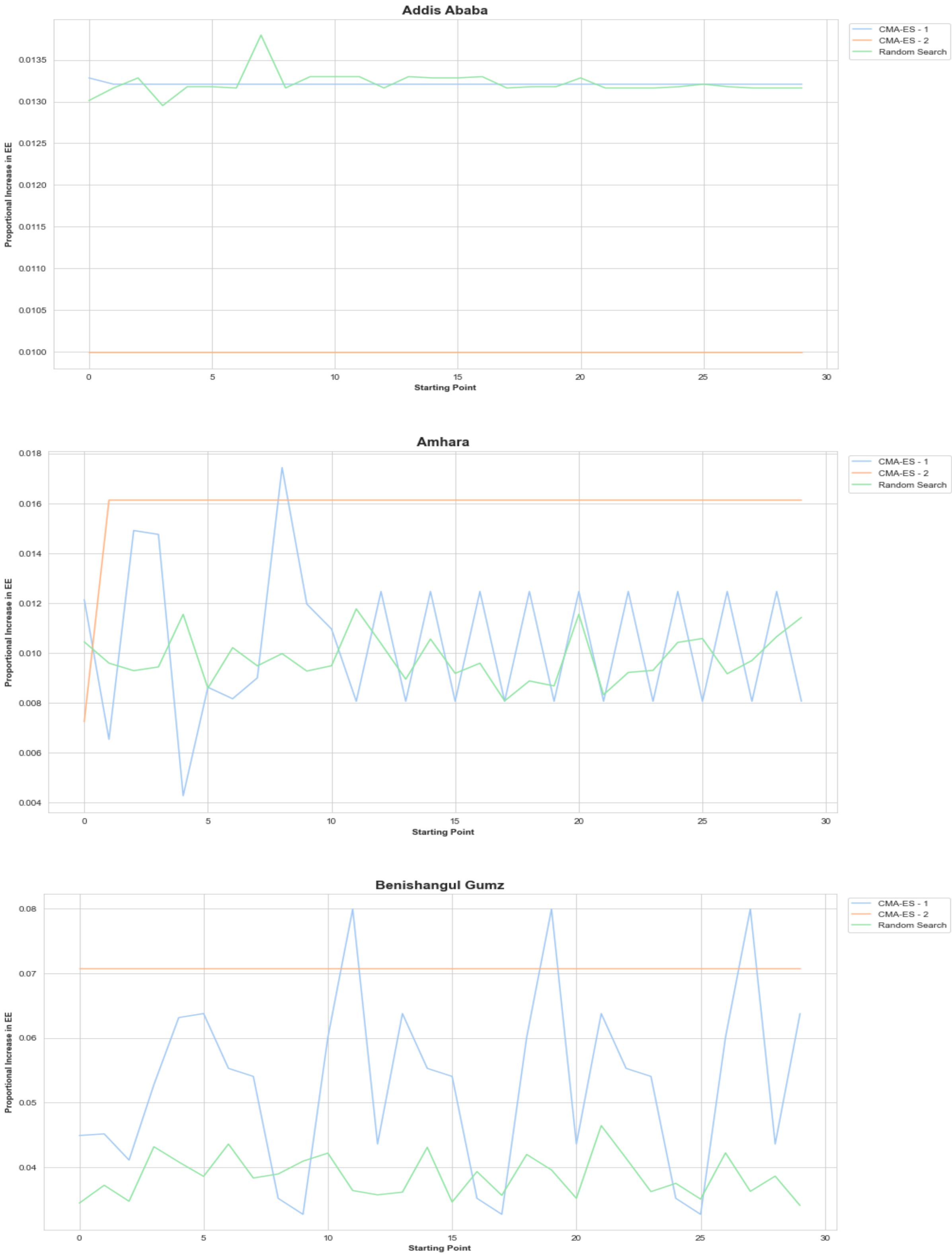
Region	Algorithm	Max EE	Max EEI	EE diff	Best sigma values & Count Per SP*	aRT**	eRT runtime **
Addis Ababa N=5	Basic Benchmark	**70,976**	4,810	7.3%	-	-	-
	Adv. Benchmark	66,827	661	1.0%	-	-	-
	Random Search	67,079	913	1.4%		49.6	1,489
	CMA-ES-1	67,045	879	1.3%	{0.078358: 1, 0.088152: 1}	6.5	341.31
	CMA-ES-2	66,827	661	1.0%	{0.078358: 30, 0.088152: 30}	0.9	10,138
Amhara N=5	Basic Benchmark	438,817	4,286	1.0%	-	-	-
	Adv. Benchmark	441,112	6,581	1.5%	-	-	-
	Random Search	439,646	5,115	1.2%	-	-	2,156
	CMA-ES-1	**442,107**	7,576	1.7%	{0.143655: 1}	113.0	16,005
	CMA-ES-2	441,540	7,009	1.6%	{0.477390: 29}	145.0	6,210
Benishangul Gumuz N=5	Basic Benchmark	25,854	1,654	6.8%	-	-	-
	Adv. Benchmark	25,911	1,711	7.1%	-	-	-
	Random Search	25,324	1,124	4.6%		46.6	
	CMA-ES-1	**26,134**	1,934	8.0%	{0.789055: 3}	109.8	4,595
	CMA-ES-2	25,912	1,712	7.1%	{0.337859: 30}	7.4	724
Dire Dawa N=4	Basic Benchmark	2,581	952	58.4%	-	-	-
	Adv. Benchmark	3,268	1,639	100.6%	-	-	-
	Random Search	3,661	2,032	124.7%		53.8	1,728
	CMA-ES-1	**3,813**	2,184	134.1%	{0.030970: 23, 0.072263: 1, 0.082586: 1, 0.092909: 1}	97.0	4,887
	CMA-ES-2	**3,813**	2,184	134.1%	{0.028045: 7, 0.037393: 7, 0.046741: 7, 0.056089: 15, 0.065437: 7, 0.074786: 14, 0.093482: 8}	76.8	5,617
Harari N=2	Basic Benchmark	4,656	390	9.1%	-	-	-
	Adv. Benchmark	4,510	244	5.7%	-	-	-
	Random Search	**4,979**	713	16.7%		47.5	1,474
	CMA-ES-1	**4,979**	713	16.7%	{0.037975: 30, 0.037975: 30, 0.056963: 30, 0.063292: 30, 0.012658: 29, 0.050634:	1.2	172

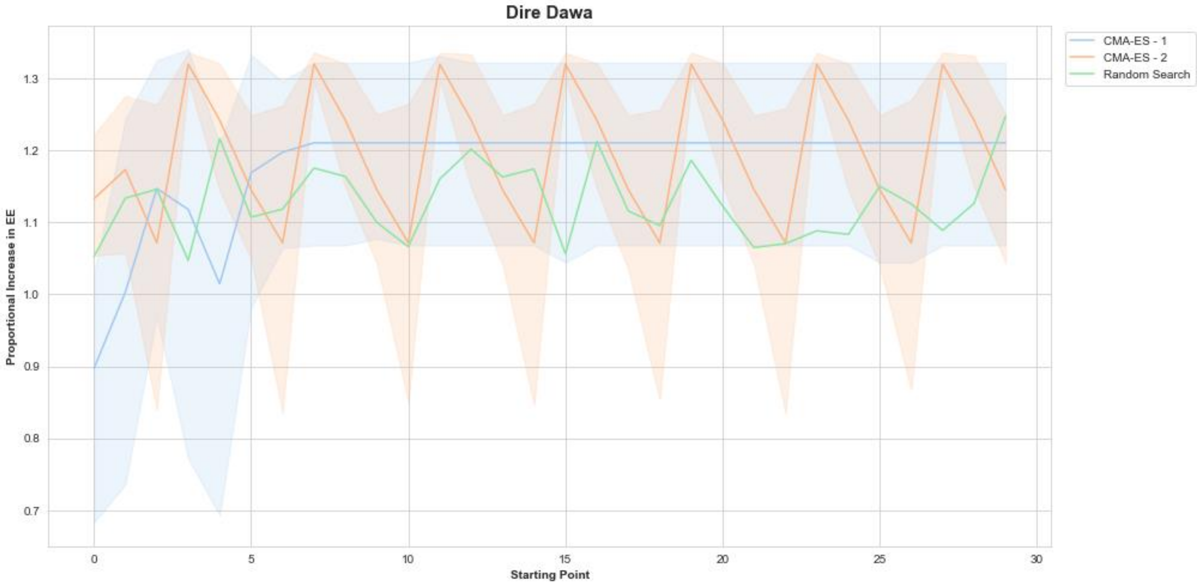
Region	Algorithm	Max EE	Max EEI	EE diff	Best sigma values & Count Per SP*	aRT* *	eRT runtime **
					29, 0.006329: 28, 0.018988: 1, 0.031646:1, 0.044305: 1}		
	CMA-ES-2	4,803	537	12.6%	{0.003690: 15, 0.007380: 15, 0.011070: 15, 0.014760: 30, 0.018450: 15, 0.022140: 15, 0.025830: 30, 0.029520: 15, 0.033209: 15, 0.036899: 30}	0.3	76
Oromia N=5	Basic Benchmark	628,831	3,583	0.6%	-	-	-
	Adv. Benchmark	**632,414**	7,166	1.1%	-	-	-
	Random Search	629,470	4,222	0.7%		100.8	3,088
	CMA-ES-1	631,688	6,440	1.0%	{0.448541: 3}	66.5	12,696
	CMA-ES-2	628,756	3,508	0.6%	{1.444516: 30}	49.1	2,214
SNNP N=5	Basic Benchmark	167,098	3,161	1.9%	-	-	-
	Adv. Benchmark	**171,016**	7,079	4.3%	-	-	-
	Random Search	168,384	4,447	2.7%		59.0	1,748
	CMA-ES-1	170,337	6,400	1.0%	{0.975340: 3}	305.7	8,740.65
	CMA-ES-2	165,934	1,997	1.2%	{0.093993: 1}	1.4	17
Somali N=5	Basic Benchmark	**19,666**	2,868	17.1%	-	-	-
	Adv. Benchmark	19,806	3,008	17.9%	-	-	-
	Random Search	18,368	1,570	9.3%		34.9	1,484
	CMA-ES-1	18,234	1,436	8.5%	{0.228912: 1}	0.2	55
	CMA-ES-2	17,352	554	3.3%	{0.180498: 1}	0.1	11
Tigray N=5	Basic Benchmark	123,941	2,520	2.1%	-	-	-
	Adv. Benchmark	126,047	4,626	3.8%	-	-	-
	Random Search	127,016	5,595	4.6%		54.7	1,766
	CMA-ES-1	129,495	8,074	6.6%	{0.740444: 5}	224.7	10,138
	CMA-ES-2	**130,588**	9,167	7.5%	{0.575634: 30}	141.5	5,495
Total N=41	Basic Benchmark	1,482,420	24,224	1.7%	-	-	-
	Adv. Benchmark	1,490,911	32,715	2.2%	-	-	-
	Random Search	1,483,927	25,731	1.8%	-	447	14,933
	CMA-ES-1	**1,493,832**	35,636	2.4%	-	925	57,629
	CMA-ES-2	1,485,525	27,329	1.9%	-	423	30,502
	Combination	**1,501,833**	**43,637**	**3.0%**			

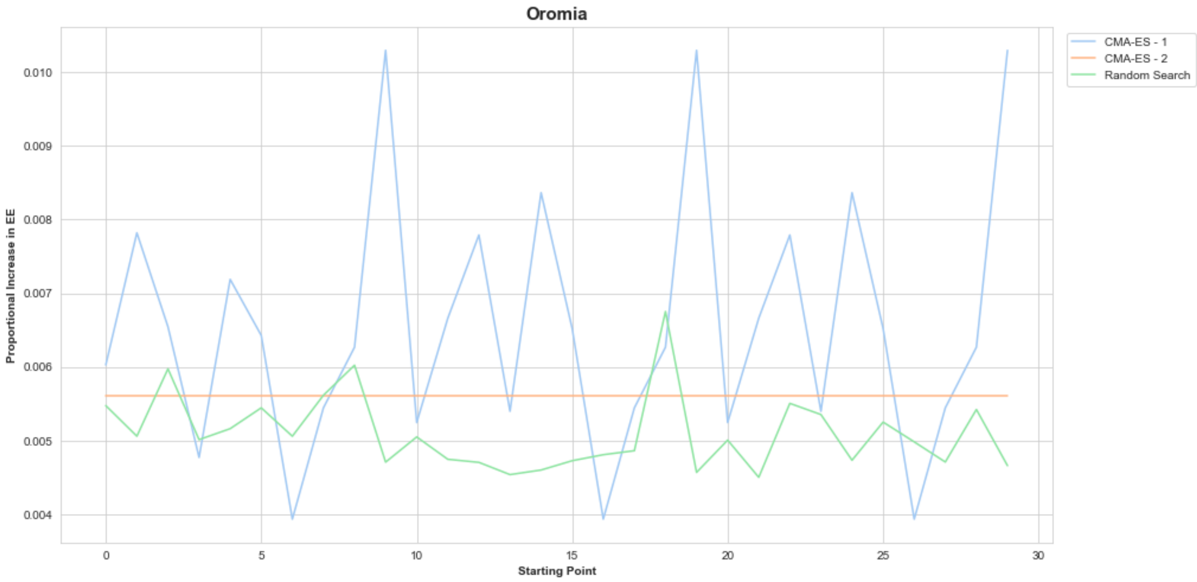
* Results are displayed in a dictionary per CMA experiment. The first number, the key, represents the sigma value, the second value represents the count of starting points that achieved the EE. Max possible starting points is 30.
** Both metrics are shown in seconds. aRT (Average RunTime and eRT (Experiment Run Time) for Basic Benchmark and Adv. Benchmark are negligible and therefore not included.

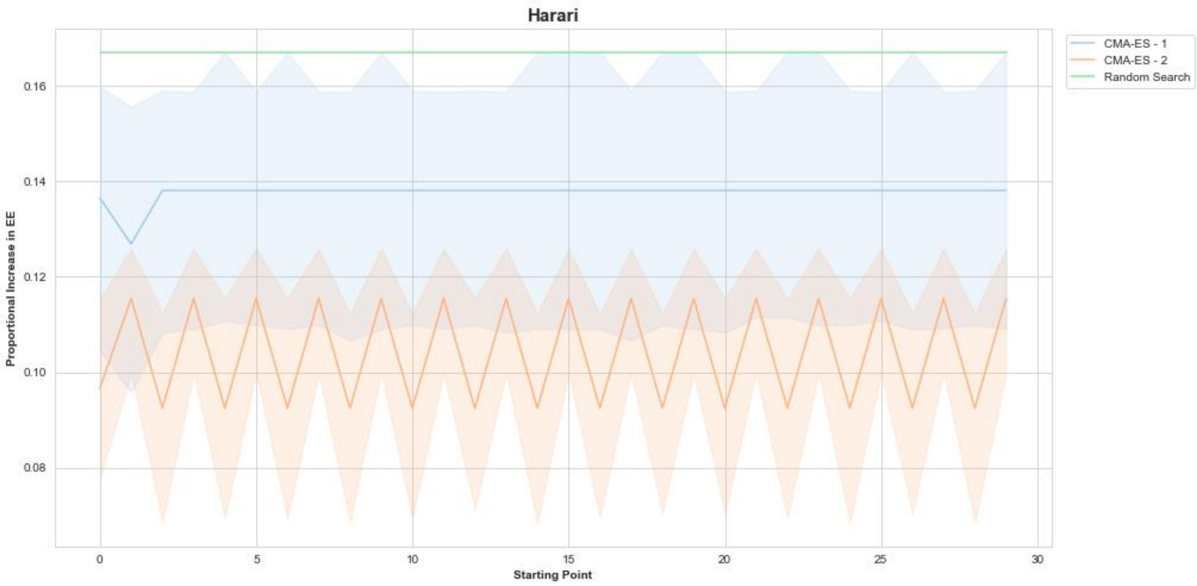
Overall, CMA-ES-1 provides the best aggregated result, leading to an expected increase in enrolment of 35,636 from 41 new schools. It finds an additional 2,921 students when compared with the advanced benchmark. In combining the highest EE results from the different methods, a further 8,001 students are found, with the majority increase stemming from the major difference in results for Addis Ababa and Tigray. The combined expected enrolment increases national secondary enrolment by 3%. CMA-ES (1 or 2) finds an improved solution to the benchmarks in 5 out of the 9 regions. It outperforms random search in 6 regions and finds an equal result in one region: Harari. The CMA-ES approaches vary in results and differ substantially in terms of runtime. CMA-ES-1 finds an improved solution in 7 out of 9 regions and 1 equal solution in Dire Dawa. However, to run the entire experiment, CMA-ES-1 took 16 hours 1 minute, while CMA-ES-2 took 8 hours 47 minutes. The size of the search space is substantially larger in CMA-ES-1 which increases the computational complexity and almost doubles the runtime. In terms of stability, CMA-ES-2 provides more consistent results across the 30 different randomised starting points. This provides a good indication for replicability. See Figure 7.1. for regional comparisons in line plots per region. Of note is CMA-ES-2 performing significantly worse than a random search in SNNP and Somali.

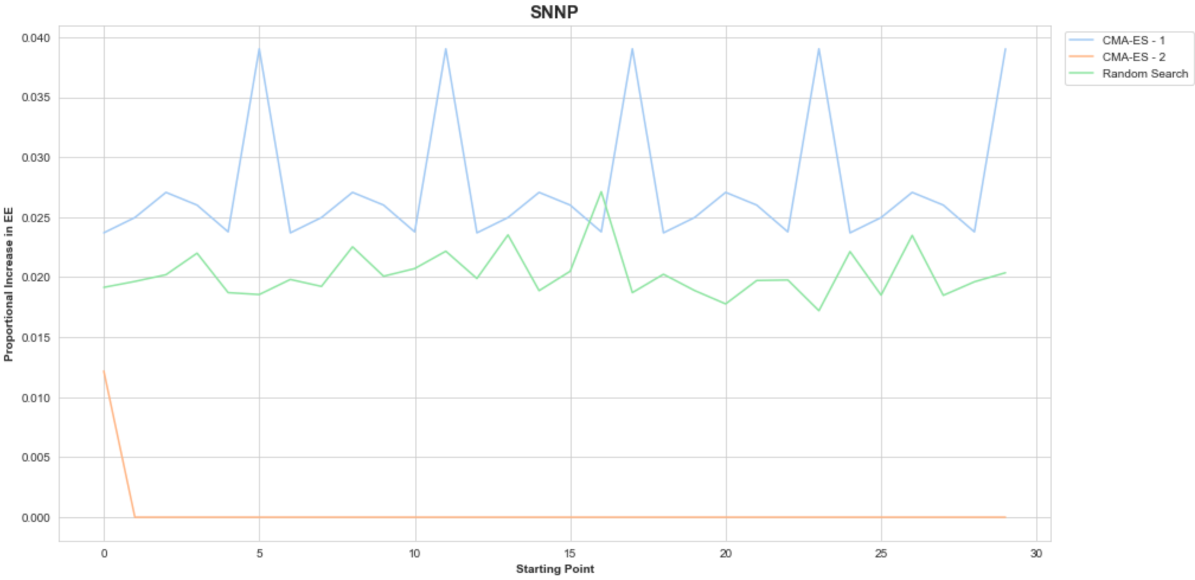
Figure 7.1. Line plots of CMA-ES algorithms by region compared with random search

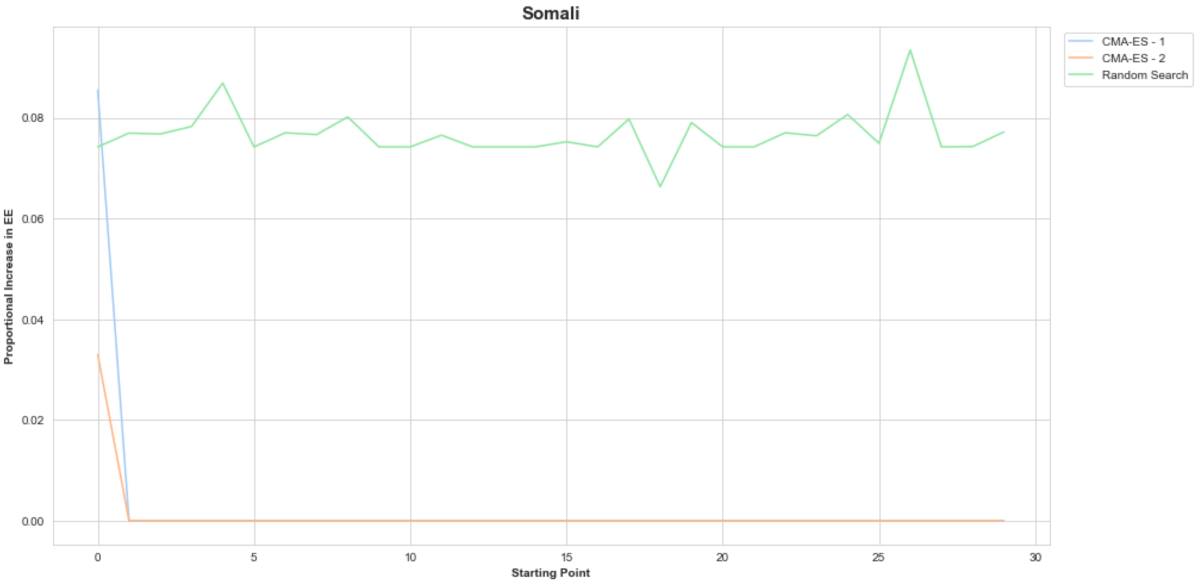


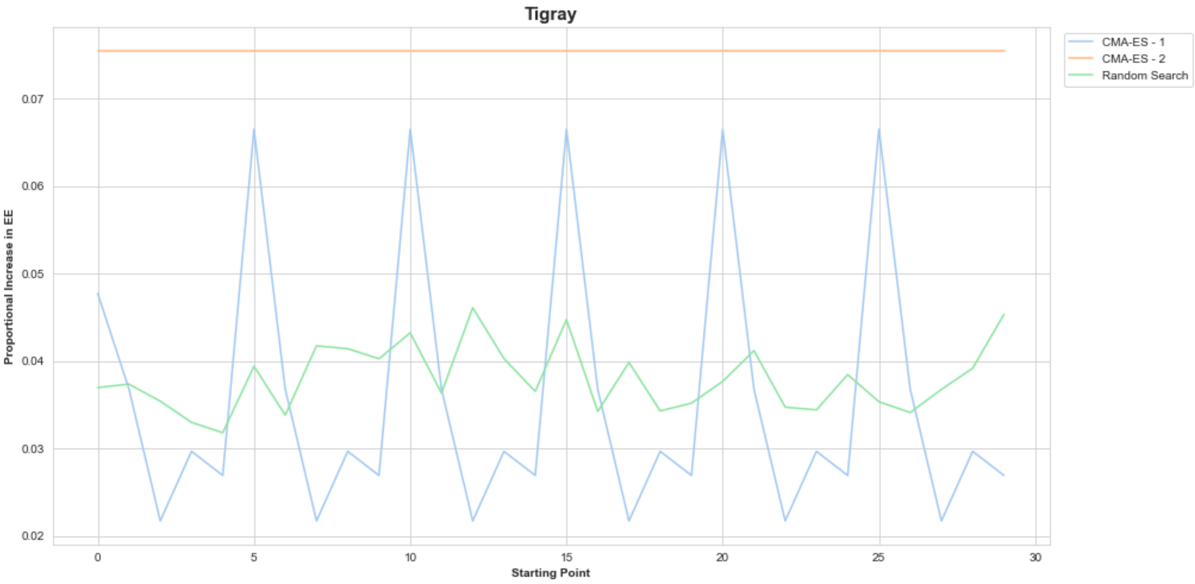

Dire Dawa
CMA-ES - 1
CMA-ES - 2
Random Search
Proportional Increase in EE
Starting Point


Oromia
CMA-ES - 1
CMA-ES - 2
Random Search
Proportional Increase in EE
Starting Point


Harari
CMA-ES - 1
CMA-ES - 2
Random Search
Proportional Increase in EE
Starting Point


SNNP
Proportional Increase in EE
Starting Point
CMA-ES - 1
CMA-ES - 2
Random Search


Somali
Proportional Increase in EE
Starting Point
CMA-ES - 1
CMA-ES - 2
Random Search


Tigray
Proportional Increase in EE
Starting Point
CMA-ES - 1
CMA-ES - 2
Random Search

To find further improvements with CMA-ES, two other experiments were conducted: the best sigma values were identified, and CMA-ES was re-run using these but with a higher maximum iteration. It was set to 30,000. This experiment was tested with n=5. See below the results.

Table 7.2. Experiment results with best sigma by region and algorithm (Max Iterations: 30,000)

Region	Algorithm	n=5					Sigma
		EE	EEI	EE Dif	aRT	eRT	
Addis Ababa	CMA-ES-1	66,988	822	1.2%	4.3	77.79	0.045941
	CMA-ES-2	66,827	661	1.0%	1.1	5.68	
Amhara	CMA-ES-1	440,680	6149	1.4%	33.6	1,215.51	0.143655
	CMA-ES-2	443,371	8840	2.0%	37.1	820.09	
Benishangul Gumuz	CMA-ES-1	25,886	1686	7.0%	59.5	1,668.39	0.789055
	CMA-ES-2	25,912	1712	7.1%	6.8	59.85	
Dire Dawa	CMA-ES-1	3,813	2184	134.1%	18.4	448.99	0.030970
	CMA-ES-2	3,813	2184	134.1%	17.2	483.46	
Harari	CMA-ES-1	4,979	713	16.7%	2.3	57.50	0.063292
	CMA-ES-2	4,803	537	12.6%	0.2	8.44	
Oromia	CMA-ES-1	629,778	4530	0.7%	176.3	2,989.82	0.448541
	CMA-ES-2	631,072	5824	0.9%	30.9	534.49	
SNNP	CMA-ES-1	170,316	6379	3.9%	141.5	2,909.48	0.975340
	CMA-ES-2	163,937	0	0.0%	0.0	1.11	
Somali	CMA-ES-1	17,267	469	2.8%	7.5	21.10	0.228912
	CMA-ES-2	16,798	0	0.0%	0.0	1.15	
Tigray	CMA-ES-1	130,410	8989	7.4%	49.6	1,044.13	0.740444
	CMA-ES-2	128,248	6,827	5.6%	77.2	614.26	
Total	CMA-ES-1	1,490,117	31,921	2.2%	493.0	10,432.71	-
	CMA-ES-2	1,484,781	26,585	1.8%	170.5	2,528.53	-

Max iterations didn't improve the overall expected enrolment for CMA-ES-1 or CMA-ES-2. Improvements can be seen in Amhara, but in Somali and SNNP a viable solution was not found. This is likely due to the rugged search space and the stochastic nature of CMA-ES results in these regions. More hyper-parameter tuning is required, especially for SNNP and Somali, as more exploration is likely needed, potentially outside of the first standard deviation. However, given the time limitations of this project, this was out of scope for this project.

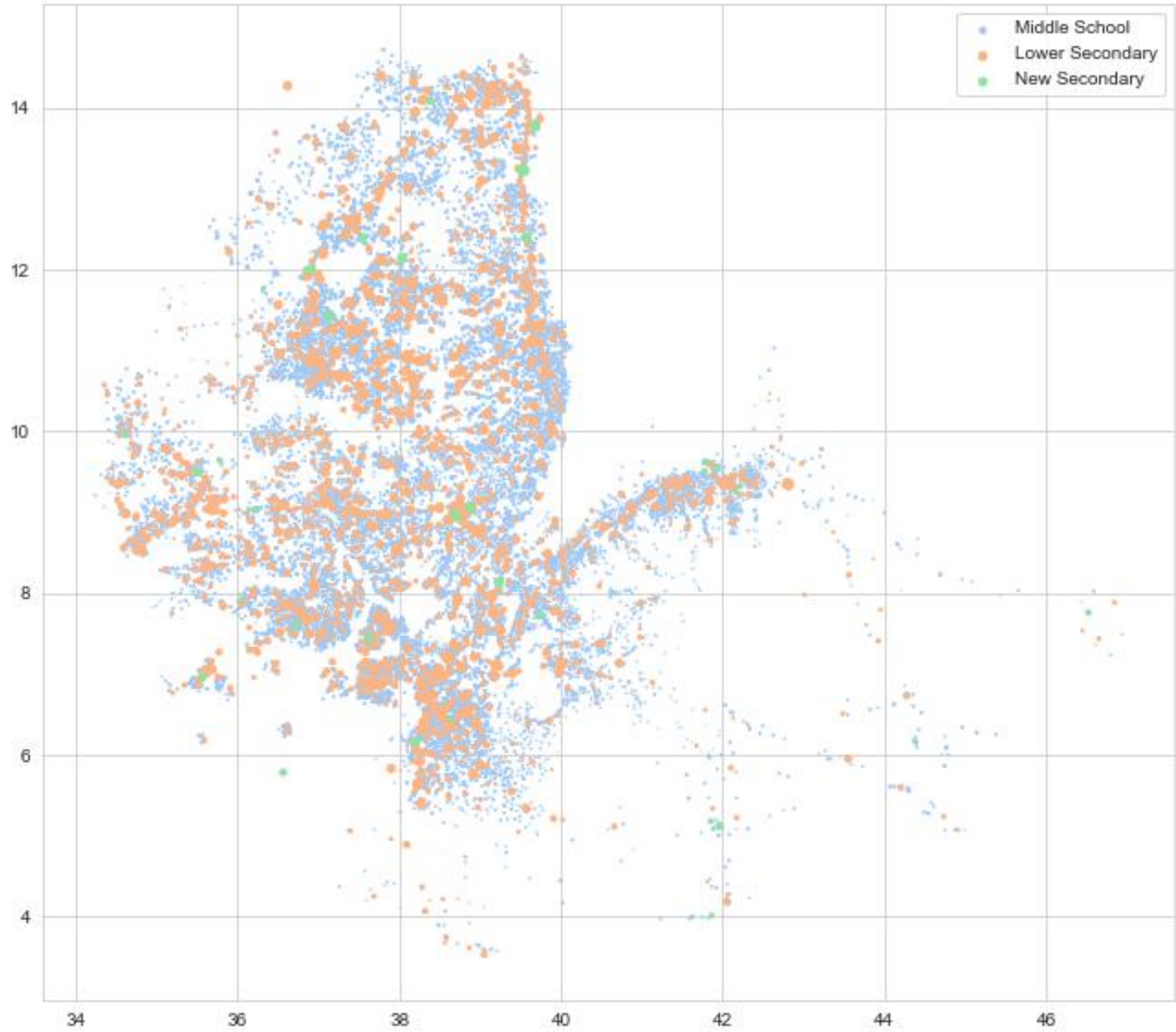
7.2. Best solution

Table 7.2 displays the best results by algorithm across all the experiments. Figure 7.2 provides a scatter plot image of these results. All final solutions were cross checked in QGIS to ensure there were no immediate barriers for construction like a ravine, river, or lake. None of these types of obstacles were noticed. However, in many cases, buildings already exist in the area. As such, the recommended location should be taken as a rough guide, and with more local knowledge, a suitable location could be found nearby.

Table 7.2 Best results by algorithm and region

Region	Algorithm	EE	EEI	EE Dif	Sigma	Max Iterations
Addis Ababa	Basic Benchmark	70,976	4,810	7.3%	-	-
Amhara	CMA-ES - 2	443,371	8,840	2.0%	0.47739	30,000
Benishangul Gumuz	CMA-ES - 1	26,134	1,934	8.0%	0.789055	20,000
Dire Dawa	CMA-ES - 1 & 2	3,813	2,184	134.1%	0.072263 & 0.056089	20,000
Harari	RS & CMA-ES - 1	4,979	713	16.7%	0.018988	0.167135
Oromia	Adv. Benchmark	632,414	7,166	1.1%	-	-
SNNP	Adv. Benchmark	171,016	7,079	4.3%	-	-
Somali	Adv. Benchmark	19,806	3,008	17.9%	-	-
Tigray	CMA-ES - 2	130,588	9,167	7.5%	0.575634	20,000
Total	-	1,503,097	44,901	3.1%	-	-

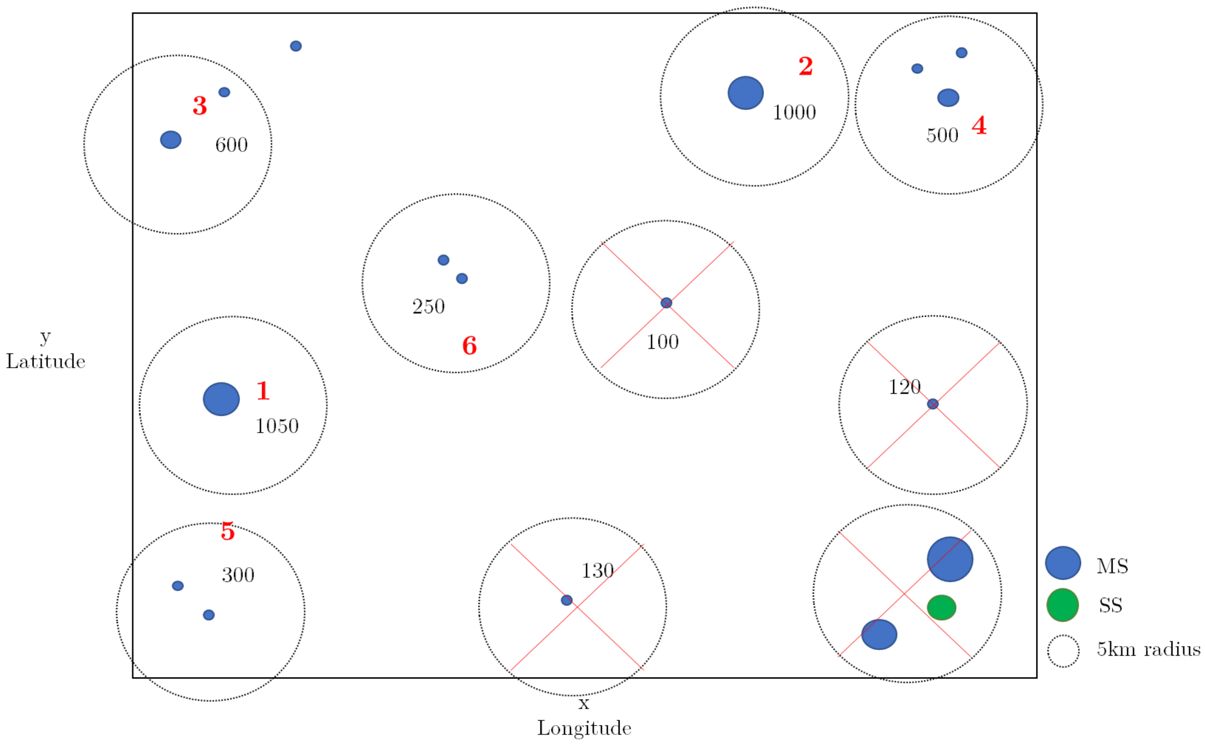
Figure 7.2. Scatter Plot of locations of new schools in addition to existing MS and SS.



7.3. Key observations and potential improvements

- The improved solutions found by the meta-heuristic models proves the value in using metaheuristic methods to improve school location planning.

- Often building a new SS where individual middle schools have the highest enrolment is the best strategy when looking to maximise enrolment. In these cases, CMA-ES didn't find an improved solution when compared with the benchmarks.
- Using the entire MS dataset is computationally expensive and using CMA-ES has relatively high time complexity. This is mainly due to calculating and updating the covariance matrix which is of order $O(n^3)$ (See Omidvar & Li, 2010 for more information on the use of CMA-ES for large scale global optimisation). While other evolutionary algorithms could be explored for improvements, further extensions of the project could make time improvements. For example, CMA-ES could be run only for each cluster of MS within 5km of each other. See Figure 7.3 for a visual representation of this problem. Total collective enrolment is shown within the 5km radius around middle schools, the rank of the cluster based on expected enrolment is shown in red, while red x marks indicate an invalid cluster as it doesn't meet the minimum capacity. In this model, CMA-ES could be run initially on the MS ranked 1 to find the optimal location. Then run again on rank=2, until rank=n. The search space would be more limited, and the results would likely be more stable. However, one limitation of such an approach is that it assumes existing MS as the centroids, and other potential locations for centroids could provide an improvement e.g., see rank 3.



- The assumptions made that concern the fitness function don't hold true for city administrations like Addis Ababa. Most MS are within 5km to the nearest SS. As such, the Basic Benchmark produces the highest result, but then due to the school's proximity to other secondary schools, (the key challenge 3 highlighted in Section 6.3,) the fitness function then reduces. CMA and the advanced benchmark therefore don't perform well. In these locations, the basic benchmark provides the best location, but the methodology isn't appropriate for Addis Ababa. In such locations, distance is clearly not the major impediment to attending school. Other factors might influence enrolment and other strategies for expanding access to secondary education would need to be explored (e.g. see Section 1.1. for a brief mention of some example strategies).

Chapter 8

8. Conclusion

This project has focused on one of the most basic public services facilities: educational facilities. Given the resource limitations that many countries face in SSA, planning where to construct new schools requires careful strategic decision making. Yet, many SSA countries face challenges relating to the availability and reliability of population data which inhibits planning and the ability to benefit from common approaches used to solve facility location problems. This project instead built a school location model that makes use of readily available data to Ministries of Education: the longitudes and latitudes of schools. A single objective location was built and applied to the expansion of secondary education in Ethiopia. It sought to find locations for SS that maximise expected enrolment from MS, considering distance as a major constraint and a minimum capacity.

Furthermore, the project examined whether the best location for a new SS is next to the individual MS with the highest enrolment. While this was proven to be true at times, the project has shown that in some cases it is not. Clusters of MS exist which can often provide higher expected enrolment. More strategic planning is therefore necessary when constructing new SS and meta-heuristic methods like the one applied in this project can support these processes, given the complexity of the search space.

Although these methods are useful, they can be problematic, as they can create an over-reliance on external support. An extension of this project would therefore be necessary to develop a product that could ensure these tools are easily available to Ministry of Education staff in SSA. For example, a web application could be developed whereby Ministry staff can upload key documents into an online form, mainly 1) school enrolment data, 2) school location data, and 3) administrative shape files. Once submitted, the analysis can take place on a server, with results emailed the next day to staff members. Trust in this process would be a problem. Explainable artificial intelligence is key when trying to deploy products with confidence from users. Technical support would likely be needed initially during routine planning processes and then once the tool is seen to improve planning, its utility will be understood. For example, in Ethiopia support could be offered during the planning for the five-year education sector development plans. It would allow for more realistic target setting, support the Government to make more efficient use of resources, and more importantly meet the existing demand. Utilizing such a tool during the comprehensive five-year sector development plan would provide an opportunity to widely engage the most important stakeholders across the entire planning process.

Although useful, this is a supplementary tool that can only support the initial selection of potential locations. Local knowledge is required to determine feasibility. In the future, this model could be extended to include other levels of schools e.g. (lower to higher secondary, or primary to middle school). A simplified product could also be developed which is not only limited to educational facilities, e.g., facility location for new health centres also based on demand from health posts. In this way, planning processes are improved using data that is reliable and readily accessible in SSA.

Bibliography

1. Bekele, S. (2017). The Accuracy of Demographic Data in the Ethiopian Censuses. Eastern Africa Social Science Research Review. 33. 15-38. 10.1353/eas.2017.0001, Accessed 22/7/2022, Available at: https://www.researchgate.net/publication/315883874_The_Accuracy_of_Demographic_Data_in_the_Ethiopian_Censuses
2. CIA World FactBook (2022), Ethiopia, Accessed, 24/7/2022, Available at: https://www.cia.gov/the-world-factbook/countries/ethiopia/
3. Church. R, Drezner. Z, Tamir. A, (2022) Extensions to the Weber problem, Computers & Operations Research, Volume 143, 105786, ISSN 0305-0548, https://doi.org/10.1016/j.cor.2022.105786
4. Daskin, M. (2013: p12), Network and Discrete Location: Models, Algorithms, and Applications, SN - 9781118536964, Wiley, Accessed 03/08/2022, Available at: https://books.google.ie/books?id=L9I5n0Sym4cC
5. de Armas, J., Juan, A.A., Marquès, J.M. et al. (2017) Solving the deterministic and stochastic uncapacitated facility location problem: from a heuristic to a simheuristic. J Oper Res Soc 68, 1161–1176 (2017). https://doi.org/10.1057/s41274-016-0155-6
6. de Armas J, Ramalhinho H, Reynal-Querol M (2022) Improving the accessibility to public schools in urban areas of developing countries through a location model and an analytical framework. PLoS ONE 17(1): e0262520. https://doi.org/10.1371/journal.pone.0262520
7. The Economist (2019), Why Ethiopia has postponed its census, Accessed 18/08/2022 at: https://www.economist.com/middle-east-and-africa/2019/03/29/why-ethiopia-has-postponed-its-census
8. Ethnologue (2022), Ethiopia Languages, Accessed, 24/7/2022, Available at: https://www.ethnologue.com/country/ET/languages
9. Farahani, Z. Reza & Abedian, Maryam & Sharahi, Sara. (2009, p1). Dynamic Facility Location Problem. 10.1007/978-3-7908-2151-2_15, Accessed 03/08/2022, Available at: https://www.researchgate.net/publication/227149037_Dynamic_Facility_Location_Problem
10. FGoE, (2020), Education Statistics Annual Abstract September 2019-March 2020 Federal Government of Ethiopia, Ministry of Education. Accessed 20/07/2022, Available at: https://moe.gov.et/storage/Books/Education%20Statistics%20Annual%20Abstract%20March%202019%20-%20September%202020.pdf
11. FGoE (2015) Education Sector Development Plan (ESDP V), Federal Government of Ethiopia, Ministry of Education. Accessed 20/07/2022, Available at: https://planipolis.iiep.unesco.org/sites/default/files/ressources/ethiopia_esdp_v_summary.pdf
12. FGoE (2019), School Enrolment Data, Federal Government of Ethiopia, Ministry of Education

13. FGoE (2020), GIS Locational Data, Federal Government of Ethiopia, Ministry of Education
14. FGoE (2020) Education Sector Development Plan (ESDP VI), Federal Government of Ethiopia, Ministry of Education. Accessed 20/07/2022, Available at: https://planipolis.iiep.unesco.org/en/2021/education-sector-development-programme-vi-esdp-vi-2013-%E2%80%93-2017-ec-202021-%E2%80%93-202425-gc-7239
15. Garfield R, Holsinger D., Ziderman, A. (1994), The Costs of Secondary Education Expansion World Bank, ESP Discussion Paper, Series #29. Washington, D.C. Accessed 20/07/2022, Available at: https://documents1.worldbank.org/curated/en/581301468765600027/pdf/multi-page.pdf
16. Hallack, J. (1977), Planning the location of schools : an instrument of educational policy, UNESCO Institute for Educational Planning, Accessed 3/08/2022, Available at: http://www.iiep.unesco.org/en/planning-location-schools-instrument-educational-policy-11481
17. HDX (2020), Ethiopia - Subnational Administrative Boundaries, Humanitarian Data Exchange, UNOCHA, Accessed: 01/06/2022, Available at: https://data.humdata.org/dataset/cod-ab-eth
18. INEE (2022), Inter-Agency Network for Education in Emergencies, Gross Enrolment Rate, Accessed 15/08/2022, Available at: https://inee.org/eie-glossary/gross-enrollment-rate-ger#:~:text=The%20gross%20enrollment%20rate%20(GER,entry%20and%2For%20grade%20repetition.
19. Klose, A., Drexl, A. (2015, p6-8) Facility location models for distribution system design, European Journal of Operational Research, Volume 162, Issue 1, 2005, ISSN 0377-2217, Accessed 06/08/2022, Available at: https://doi.org/10.1016/j.ejor.2003.10.031
20. Kuhn, H. W. and Kuenne, R. E. (1962). An efficient algorithm for the numerical solution of the generalized weber problem in spatial economics. Journal of Regional Science, 4. Accessed 4/08/2022, Available at: https://onlinelibrary.wiley.com/doi/10.1111/j.1467-9787.1962.tb00902.x
21. Litoff, A. (2015), Facility Location Problems, Northwestern University Process Optimization Open Textbook. Accessed 4/08/2022, Available at: https://optimization.mccormick.northwestern.edu/index.php/Facility_location_problems
22. Ndiritu, J. (2012), Optimal Size and existence of economies of scale in Public Secondary Schools: Nyandarua District, Kenya, MST-Department of Educational Management Policy & Curriculum Studies, Accessed 2/08/2022, Available at: https://ir-library.ku.ac.ke/handle/123456789/2564
23. Omidvar M.N. & Li X, 2010, A Comparative Study of CMA-ES on Large Scale Global Optimisation. In: Li, J. (eds) AI 2010: Advances in Artificial Intelligence. AI 2010. Lecture Notes in Computer Science(), vol 6464. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-17432-2_31

24. Osleep, J.P. and Cromley, R.G. (1977, p 143), The Location-Production-Allocation Problem With Nonlinear Production Costs. Geographical Analysis, 9: 142-159. https://doi.org/10.1111/j.1538-4632.1977.tb00568.x
25. Owen, S.H. Daskin, M.S.(1998) Strategic facility location: A review European Journal of Operational Research, 111 (3) (1998), Accessed 03/08/2022, Available at: https://reader.elsevier.com/reader/sd/pii/S0377221798001866?token=A21DEEACB7C2BEDBCF18C35394E19E145337881D2E806144189544A6A9C175C589BDDF547AE8B23A0D2D708A37156CC1&originRegion=eu-west-1&originCreation=20220829121046
26. Plastria F. (2002: p2), Continuous Covering Location Problems. 10.1007/978-3-642-56082-8_2, Accessed 3/08/2022, Available at: https://www.researchgate.net/publication/268245424_Continuous_Covering_Location_Problems
27. Pizzolato, N. and da Silva, H.F. (1997), The Location of Public Schools: Evaluation of Practical Experiences. International Transactions in Operational Research, 4: 13-22. https://doi.org/10.1111/j.1475-3995.1997.tb00058
28. Reuters (2018), Ethiopian foreign exchange shortage will last years: new premier, Accessed 20/08/2022, Available at: https://www.reuters.com/article/ozatp-uk-ethiopia-economy-idAFKBN1HO0OT-OZATP
29. Reuters, (2021), Ethopia's Ethnic Diversity, Accessed, 24/7/2022, Available at: https://graphics.reuters.com/ETHIOPIA-CONFLICT/TIGRAYANS/yzdpxbbzqpx/
30. Segura, E. Carmona-Benitez, R.B. Lozano,A. (2017, p1139-1140), Implications of the assumptions on which the p-median problem are based when distribution network design, Transportation Research Procedia, Volume 25, ISSN 2352-1465, Accessed 10/08/2022, Available at: https://doi.org/10.1016/j.trpro.2017.05.126
31. Stollsteimer J.F., (1963), A Working Model for Plant Numbers and Locations. American Journal of Agricultural Economics, 45: 631-645. https://doi.org/10.2307/1235442
32. Theunynck, S. (2009) School Construction for Universal Primary Education in Africa: Should Communities Be Empowered to Build Their Schools? Washington D.C.: The World Bank. Accessed 27/07/2022, Available at: https://documents1.worldbank.org/curated/en/109291468007863249/pdf/488980PUB0prim101Official0Use0Only1.pdf
33. Theunynck, S. (2018), Distance to School and the educational consequences, Affordable School Construction for Quality Education (ASCQE), The World Bank Group.
34. Theunynck, S. (2014), School Construction in Developing Countries What do we know? The World Bank, Accessed, 21/07/2022, Available at: https://www.humanitarianlibrary.org/resource/school-construction-developing-countries-what-do-we-know-0
35. The Guardian (2014), Development data: how accurate are the figures?, Accessed 23/07/2022, Available at: https://www.theguardian.com/global-development/poverty-matters/2014/jan/31/data-development-reliable-figures-numbers

36. Turkoglu D.C. and Genevois, M.E. (2019, p402-468) A comparative survey of service facility location problems, Annals of Operations Research, Springer, vol. 292(1) September. Accessed: 01/08/2022, Available at: https://ideas.repec.org/a/spr/annopr/v292y2020i1d10.1007_s10479-019-03385-x.html
37. UN (2022) Sustainable Development Knowledge Platform, States Members of the United Nations and States members of specialized agencies, Accessed: 17/07/2022, Available at: https://sustainabledevelopment.un.org/memberstates.html
38. UNESCO (2022) Policy Toolbox, School Location, International Institute for Educational Planning, Accessed: 2/08/2022, Available at: https://policytoolbox.iiep.unesco.org/policy-option/school-location/
39. UNESCO, (2022) Total net enrolment rate, Accessed: 18/08/2022, Available at: http://uis.unesco.org/en/glossary-term/total-net-enrolment-rate
40. UNESCO (2022), Metadata for the global and thematic indicators for the follow-up and review of SDG 4 and Education 2030, Accessed 18/07/2022, Available at: https://tcg.uis.unesco.org/wp-content/uploads/sites/4/2019/08/sdg4-global-indicators-4.1.pdf
41. UNESCO (2022), Isochrone-based catchment areas for educational planning, Gispo Ltd, International Institute for Educational Planning, IIEP Technical Note, Accessed 15/08/2022, Available at: http://www.iiep.unesco.org/en/publication/isochrone-based-catchment-areas-educational-planning
42. Tewari, V. K. and S. Jena (1987). High school location decision making in rural India and location-allocation models. In Spatial analysis and location-allocation models, eds. A. Ghosh and G. Rushton . New York: Van Nostrand Reinhold.
43. The World Bank (2021), GDP growth (annual %) - Sub-Saharan Africa, Accessed: 18/07/2022, Available at: https://data.worldbank.org/indicator/NY.GDP.MKTP.KD.ZG?locations=ZG
44. The World Bank, (2021), School enrollment, preprimary (% gross) - Sub-Saharan Africa, Accessed 17/07/2022, Available at: https://data.worldbank.org/indicator/SE.PRE.ENRR?locations=ET
45. The World Bank (2022), School enrollment, secondary (% gross) - Sub-Saharan Africa, Accessed 17/07/2022, Available at: https://data.worldbank.org/indicator/SE.SEC.ENRR?locations=ZG
46. The World Bank (2022), School enrollment, primary (% gross) - Sub-Saharan Africa, Accessed 17/07/2022, Available at: https://data.worldbank.org/indicator/SE.PRM.ENRR?locations=ZG
47. The World Bank (2021), Rural population (% of total population) - Sub-Saharan Africa, Accessed 1/08/2022, Available at: https://data.worldbank.org/indicator/SP.RUR.TOTL.ZS?locations=ZG