Space and aerial architectures to expand global connectivity

by

Iñigo del Portillo Barrios

Ingeniero Superior de Telecomunicaciones, UPC BarcelonaTech (2014)
Ingeniero Superior Industrial, UPC BarcelonaTech (2014)
Ingeniero Superior en Electronica, UPC BarcelonaTech (2015)

M.Sc., Massachusetts Institute of Technology (2016)

Submitted to the Department of Aeronautics and Astronautics
in partial fulfillment of the requirements for the degree of
Doctor of Philosophy in Aeronautics and Astronautics
at the
MASSACHUSETTS INSTITUTE OF TECHNOLOGY
~~December 2019~~ February 2020

Signature redacted

Author ....
Department of Aeronautics and Astronautics
December 19, 2019

Signature redacted

Certified by .....
Prof. Edward F. Crawley
Professor, Aeronautics and Astronautics
Thesis Supervisor

Signature redacted

Certified by ....
Prof. Vincent Chan
Professor, Electrical Engineering and Computer Science
Member, Thesis Committee

Signature redacted

Certified by .....
Dr. Hamid Hemmati
Director of Engineering, Facebook
Member, Thesis Committee

Signature redacted

Certified by ....
Dr. Bruce G. Cameron
Director, MIT Systems Architecture Group
Member, Thesis Committee

Accepted by .....
Sertac Karaman
Associate Professor, Aeronautics and Astronautics
Chair, Graduate Program Committee

Space and Aerial Architectures to Expand Global Connectivity

by

Iñigo del Portillo Barrios

Submitted to the Department of Aeronautics and Astronautics
on January 10, 2020, in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy in Aeronautics and Astronautics

Abstract

Currently, 46.4% of the world's population does not have access to the Internet. Bringing the more than 3.5 billion individuals still unconnected online is the primary goal for multiple international organizations, including the ITU and the UN Broadband Commission. In the last ten years, there has been steady growth in the number of Internet users (around 200 - 300 million per year), but this has been considered insufficient to meet the target of having 60% of the world's population be connected by the end of 2020 (as set in *Resolution 200 - Connect 2020 Agenda for Global Telecommunication/ICT Development*). Besides, even more ambitious targets (75% of the world's population connected) have been proposed for 2025.

Two important barriers that restrict connectivity are the lack of infrastructure and affordability. To address these barriers, several novel concepts that involve spaceborne and airborne platforms have been proposed to provide connectivity at a lower cost (improve affordability) to a wider reach of people (extend infrastructure). This thesis explores the tradespace of architectures for space and aerial communication network concepts to extend global connectivity. In particular, constellations of geostationary satellites, large constellations of MEO and LEO satellites, and high- and low-altitude aerial platforms are studied. For each of these concepts, I develop end-to-end system models that include the RF propagation, atmospheric channel, power- and mass-sizing, system dynamics, and costs. Different frequency bands are considered, including current state-of-the-art Ku- and Ka-bands and future scenarios with extremely high-frequency bands (V/Q, E, and optical).

The potential of each of these concepts is then analyzed from a techno-economic perspective. Given the large scale of the problem (global connectivity), the different spatial and temporal scales on which each of the concepts operate, plus the large tradespace of potential architectures, evaluating the potential impact requires the development of large simulation models to compute realistic estimates for performance

and cost, as well as to identify trade-offs among concepts.

However, due to limited computing resources, an exhaustive evaluation of all design configuration is impractical and often not possible; consequently, the resources devoted to concept exploration need to be carefully allotted, balancing exploration and exploitation within the tradespace. To that end, this thesis presents a Bayesian optimization approach tailored for System Architecture problems, to explore tradespaces efficiently when there is a tightly-constrained budget for objective function evaluations.

Thesis Supervisor: Prof. Edward F. Crawley
Title: Professor, Aeronautics and Astronautics

Acknowledgments

First and foremost, I would like to thank my advisor, Prof. Ed Crawley, for taking me as a student and allowing me the opportunity to pursue my research interests with complete freedom. Ed has always been supportive of my work, and his advice and guidance have been paramount to the completion of this thesis.

The other members of my committee also deserve special mention. Prof. Vincent Chan has been an incredible source of knowledge and wisdom on communication architectures, and many of the ideas I was exposed to in his classes eventually found their way into this thesis. Dr. Hamid Hemmati has always supported my work during my internships at Facebook and also here at MIT. Finally, Dr. Bruce Cameron has always been my go-to for honest and pragmatic advice, which was vital to navigate the numerous hurdles and challenges of this highly-unpredictable journey. I would also like to show my appreciation for my two thesis readers, Prof. Joel Schindall and Prof. Daniel Selva. Prof. Selva played a crucial role in all of this, being the one who first introduced me to the System Architecture field when I arrived at MIT 6 years ago, and I cannot think of a more knowledgeable person on satellite constellations than Prof. Schindall. Thank you both for lending your expertise to this work.

I would also like to thank my former colleagues from Facebook's SoCal office, especially Dr. Kevin Birnbaum, my manager for the two internships I did at Woodland Hills during the summers of 2016 and 2017, and Dr. Slaven Moro, who continued to support my research afterward. In addition to the knowledge and resources my teammates have freely shared with me, I am eternally indebted to Facebook for providing access to data crucial to this research, as well as research funding in the last three years. The Fundacion Obra Social de LaCaixa also provided funding for my first two years of graduate school, and I would like to show my gratitude to them and acknowledge the exceptional job that they do in supporting young people in pursuing their dreams.

I feel incredibly thankful to have been able to share these past 6 years with a group of amazing people who make my day-to-day ever enjoyable and intellectually stimulating. To everyone in 33-409 (and the AeroAstro department at large), thank you all for being there every time I needed you. To Dani, Ana, Marc, Peter, Morgan, and Patricia, thanks for welcoming me into the MIT community when I first arrived, and for getting me started in this world called "research". Dr. Marc Sanchez deserves a special mention for being my personal mentor during the four years we shared in the office. To the more recent members of the SAL: Markus, Juanjo, Nils, Ruben, Sydney, and Skylar, thanks for seeing me through the last stages of the PhD. Also, to the rest of the people in AeroAstro: Raichelle, Johannes, Matthew, Alex, and Anne, thank you for your support and friendship. To Beth Marois and Ping Lee, thank you for being such champs when it comes to dealing with bureaucracy on my behalf.

I am very fortunate also to have shared these past years with some fantastic friends who have helped me make the best of this American experience. Among them, my flatmates Adria, Adrian, Alex, Ferran, Pablo, and Ricardo have always been there for me. Thank you for sharing in my highs and lows, and especially for tolerating me on my worst of days. I am also extremely grateful to Clara, Jordi, Laura, Will, Jordina, Joan, Juanito, David, Maite, Noel, Quique, Enrique, Ines, Ximo, Helena, Mikel, Xaide, Pablo, Gabriela, and the rest of people from Spain@MIT for making me feel a little bit closer to home here in Boston. I will miss you all lots!

I cannot forgo this opportunity to thank my friends back in Pamplona and Barcelona, who have been utterly supportive of my endeavors. Eric, Gemma, and Pablo deserve special shout-outs, as they have always taken the time to call and keep in touch with the ongoings in my life. And of course, to my dearest Silvia, thank you for our annual trips that kept me motivated and going strong!

Last but definitely not least, I would like to thank my family, who has supported me no matter what. To my partner (and unofficial thesis reader) Tai Boon, thank you for sharing in this PhD journey with me, and for being a constant source of joy and positivity. Finally, this dissertation is dedicated to my parents Emilio and Mila, who gave me everything without asking for anything in return, and to whom I will always look to for the most important lessons in life.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Context

Telecommunication services have significantly impacted technological progress, business, and social lives in the last century. In recent years, the widespread adoption of the Internet and computers has resulted in a *new economy*, where the focus has shifted from manufacturing and commodity-based businesses towards heavily technology-based companies, and where innovation in the form of new products and services occurs at an unprecedented pace.

However, not everyone is benefiting equally from this *new economy*. To date, 46.4% of the global population is still offline [1], and even though most of them live in areas covered by 3G and 4G connectivity, approximately 11% of the population live in areas with no broadband coverage. The majority of those in areas with no broadband connectivity are located in the least developed countries in Africa, South America, and Southeast Asia, being the two main reasons for the absence of connectivity the lack of infrastructure and affordability issues. In particular, due to the sparse distribution of the population and their relatively low incomes, traditional communications systems are not well-suited for extending connectivity into these regions since they are expensive to deploy and have limited coverage. Thus, new concepts such as satellite constellations, networks of interconnected balloons or unmanned aerial vehicles (UAVs), and tethered blimps have been proposed in the last few years.

This dissertation studies the impact that these space and aerial concepts can have in terms of bringing new users online, and analyzes which concepts and architectures have the highest potential in achieving this goal. To accomplish these objectives, there are two main elements that this work develops. First, accurate models for each of the concepts, capturing the different spatial and temporal scales of as well as the various design options, are constructed, which allow for realistic comparisons of their performance, costs, and impact. Second, due to the computational complexity of

the models developed, a methodology to conduct tradespace exploration when only a limited number of architectures can be evaluated is proposed.

1.2 Motivation

1.2.1 Connecting the other half

A recent study by Katz [2] showed that fixed and mobile broadband have had a significant impact on the global economy. An increase of 1% in fixed-broadband connectivity yields an increase of 0.08% in GDP, whereas a similar increase in mobile-broadband penetration results in an increase of 0.15% in GDP. Moreover, the same study concluded that digitalization improvements enabled by enhanced connectivity yield additional increases in GDP (~0.13% for every 1 per cent increase in the digital ecosystem development index) and productivity (~0.23%). Thus, the economic benefits of extended Internet connectivity (both direct and indirect) are evident.

However, by the end of 2019, 53.6% of the world's population is expected to be connected to the Internet [1]; bringing online the other half is a stated priority of various international organizations, such as the International Telecommunications Union (ITU) *Resolution 200 - Connect 2020 Agenda for Global Telecommunication/ICT Development* [3] and the UN Broadband commission [4]. The main goals of these two groups are to find scalable and replicable solutions to connect large rural offline populations at minimal costs and to find effective strategies for narrowing the usage gap across all regions.

To this end, each of these organizations has published a manifesto with objectives to be met by the years 2020 and 2025, respectively. Even though in the last ten years there has been a steady growth of Internet users (around 200 - 300 million per year [5], as shown in Figure 1-1a), and some continents have almost reached saturation with more than 80 % of the population connected (see Figure 1-1b), the growth trends exhibited, in addition to projections from studies [6], indicate that the objectives set in the Connect 2020 Agenda (60% of the population connected by 2020) and by the UN Broadband Commission (75 % of the population connected by 2025) are unlikely to be met.

Several studies have been conducted to determine the main factors which influence the number of people connected to the Internet in a given country [8, 9, 10]. These studies identified the degree of infrastructure roll-out (i.e., communications infrastructure, road infrastructure, power infrastructure) and affordability (i.e., device cost, service cost, electricity cost) as the two most important factors to be solved through technological innovations, while relevance (i.e., access to content in their primary language, perceived value of Internet use) and readiness (i.e., information technology skills, know-how, basic literacy) also play an essential role. In this dissertation, only

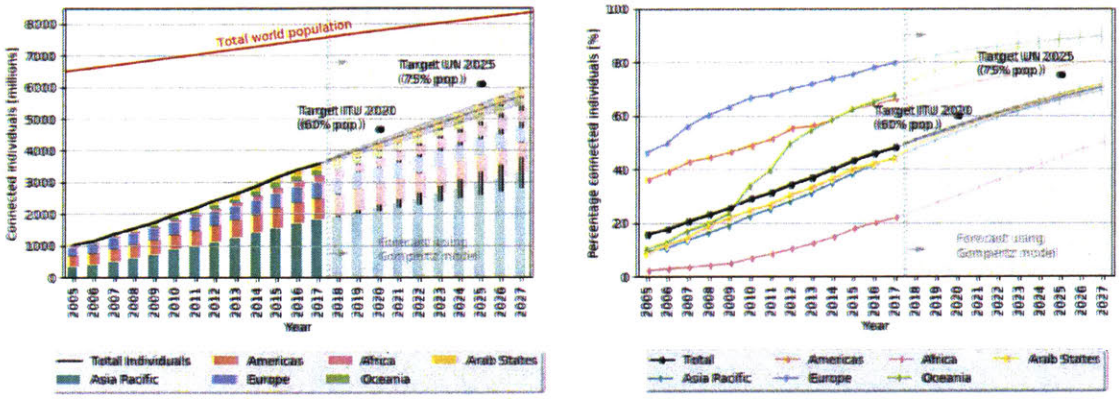


Figure 1-1: Internet users by year: (a) absolute numbers of individuals connected, (b) percentage figures. Forecasts for individuals connected for years 2018 - 2027 were made using the Gompertz model of technology diffusion [7]. Source data: ITU and Worldbank.

the techno-economic factors (infrastructure and affordability) are analyzed, leaving out the policy factors (relevance and readiness).

Affordability and infrastructure roll-out are closely related. Telecommunications operators do not deploy infrastructure in areas where no business opportunities are viable, and low-income individuals have few incentives to subscribe to broadband services. The current consensus among studies and organizations [10, 9, 11] is that for telecom services and devices to be considered "affordable", they should cost less than 2% of an individual's monthly income. And although historically the price of handheld devices has been a significant barrier to access, low-cost feature phones and smartphones, along with robust secondary smartphone markets, are rapidly lowering these costs. However, service costs remain an issue. A recent study by Thanki showed that while the first billion individuals to be connected had a monthly disposable income apportioned for communication services of at least $82.04 (according to the ITU 2% definition), the same measure for the last billion is merely $0.90 [12].

Deploying new infrastructure becomes increasingly complicated when the geographical distribution of those unconnected and their income levels is taken into account. 45.7% of the world's population lives in rural areas, and it is estimated that the poverty rate in these regions is three times higher than in urban areas [13]. Therefore, bridging the connectivity gap will require technologies that allow for sparse populations to be connected in cost-efficient manners.

New low-cost and wide-coverage solutions are an active area of research in both academia and industry, and several concepts have been proposed, developed, and tested within the last few years. Most of these new concepts differ from traditional broadband connectivity solutions (i.e., terrestrial systems like fiber, cell towers, or mm-wave links) in that they make use of space and aerial platforms to provide connectivity. In particular, the proposed concepts include different types of low-altitude

tethered platforms (tethered balloons and blimps), high altitude platforms (unmanned aerial vehicles and balloons flying in the stratosphere), and large constellations of satellites (both in low Earth orbit (LEO), medium Earth orbit (MEO), and geostationary orbit (GEO)). For example, SpaceX, Telesat, and OneWeb have all proposed large constellation of LEO satellites, whereas in 2021, SES will launch seven MEO satellites and Viasat a 1 Tbps (the highest throughput ever) GEO satellite. Moreover, firms like Airbus, Aurora, and HAPs Mobile have all proposed UAV-based aerial networks, and Google is proposing a network of stratospheric balloons.

A techno-economic analysis is required to quantify the impact (in terms of coverage capabilities and cost) of these newly-proposed concepts. This techno-economic analysis is part of a larger phase in the field of Systems Engineering, usually referred to as "Concept evaluation and concept selection", which takes place in the *conceptual design* phase. Conducting these techno-economic analyses requires the development of performance, cost, and financial models, which, given the global nature of the problem and the different scales of the concepts involved, are computationally expensive to evaluate. As an example, to assess the total cost and performance of a satellite system, the number of assets involved needs to be estimated. For extremely high frequencies, determining the number of ground stations requires consideration of the atmospheric attenuation at such frequencies and conducting optimization to determine their locations, since total uplink and downlink capacities will be highly conditioned by the location and number of ground stations. In addition, one will need to consider where and when the satellites might be saturated and estimate the number of unconnected individuals who will subscribe to the new service. Similarly, if tethered balloons were to be used, line-of-sight (LoS) conditions need to be evaluated from the balloons to thousands of population settlements, and the number of assets to be deployed needs to be determined using a placement algorithm.

The first and primary objective of this dissertation is to analyze the potential impact of these novel space and aerial concepts from a techno-economic perspective. Given the complexity of the models involved in these techno-economic analyses, traditional tradespace exploration methods based on the evaluation of a large number of design configurations are not practical. Instead, the decision regarding which design configurations to evaluate plays a crucial role when assessing the viability of different concepts. The next section motivates the analysis of this problem from a "System Architecture" standpoint.

1.2.2 A system architecture perspective

System architecture has emerged as the discipline at the forefront of the design of complex systems, products, and services. It is centrally concerned with the early-stage technical decisions that will determine most of a system's final performance and cost. One of the most crucial steps in the design process is concept selection, which is the process of evaluating, comparing, and finally deciding between competing

concepts that have been proposed to satisfy the system requirements. A poor decision at this point may result in significant cost and schedule overruns, as changes become increasingly difficult and expensive to implement in later stages.

From a theoretical perspective, a *concept* is a high-level intellectual construct that maps function to form and describes "how the system works". From an engineering perspective, a concept is an idea, sufficiently developed, such that a parametric model can be built to assess the performance and cost of one of the design configurations that emerge from that concept [14]. Each concept defines a design space, which is the set of decisions and variables that need to be specified to produce a design configuration. As an example, if one were to design a mechanism so that cars can cross from one side of a river to the other, potential concepts would include a bridge, an underwater tunnel, or a ferry service between both shores. Each of the concepts defines a different design space: for the bridge concept, the type of bridge, materials to be used, number of supports, all need to be considered; for the tunnel, the depth and number of lanes are relevant; and for the ferry service, the number of boats their capacity and frequency are crucial factors. If one assigns values to each of the variables in the design space, a design configuration is created, and performance and cost can then be evaluated.

Even though different methods for concept selection have been developed, most were conceived to compare just a handful of point-designs. Given that complex systems typically present large design spaces with trade-offs among different configurations, comparisons are frequently made between sub-optimal designs. In recent years, there has been increasing interest in exploration-based design methodologies that perform an extensive evaluation of designs in the tradespace. This has led to new methods for concept selection, which leverage large-tradespace exploration [15, 16, 17] and compare the Pareto fronts for different concepts instead of comparing point-designs.

Obtaining the Pareto fronts frequently requires the evaluation of thousands of design configurations, and therefore simplified computational models must be used to assess the performance and cost of such designs. Still, for systems of large-enough scale and complexity, even simplified models could require a considerable amount of computational resources, as shown in Table 1.1. Examples of such systems include multi-disciplinary systems (e.g., heterogeneous system-of-systems with different domain subsystems), systems where a critical subsystem needs to be optimized (e.g., a suborbital vehicle), systems that require processing large amounts of data (e.g., machine learning algorithms), large-scale systems (e.g., communication networks on a country-level), and systems subject to considerable uncertainty, which need numerous simulation runs to obtain statistically significant results. In all of these cases, one cannot simplify the models to reduce the computational burden, but need to evaluate any proposed design using models of certain computational complexity.

The main approach in tradespace exploration where large design spaces are involved has been to use model-free evolutionary algorithms (such as genetic algorithms or particle swarm optimization), which apply meta-heuristic rules to find the Pareto fronts. The main reason to use these types of algorithms is that by construction,

Table 1.1: Runtime of computationally-expensive models used for System Architecture in different domains. The element marked with a star (*) corresponds to one of the problems in this work.

Model name	Runtime	Application	Ref
FLOPS/ENGGEN ALCCA	5 min	Size aircraft and propulsion system, perform uncertainty analysis over the system.	[18]
LLVM	20 min	Compilation of the LLVM test suite.	[19]
FPGA synthesis	min - hours	Hardware implementations of a sorting network for 256 inputs. Each design is characterized by d = 3 parameters.	[20]
LEOsat	1 - 5 hours	Evaluate the performance of a constellation of satellites in Low Earth Orbit.	*
ASIC 2 design	2 hours - days	Generating energy-efficient MPSoC Network on Chips implementations of the AXI specification.	[21]
Crashworthiness	98 hours	Evaluate the capability of a vehicle to protect its occupants from the effects of an accident.	[22]
AlexNet	days	Train convolutional neural network for computer vision object recognition.	[23]

they are well suited to deal with discrete and categorical variables, which are relatively common in System Architecture problems (SAP) (see Section 1.4.2). However, as mentioned before, these algorithms rely on evaluating a large population of individuals iteratively, removing the least desirable designs and introducing random changes on surviving designs with each iteration, which becomes infeasible when the computational burden of each evaluation is heavy.

Therefore, there is a need for tradespace exploration methods that are capable of exploring large tradespaces and handling the types of decisions in SAPs, even when dealing with expensive computational models. Bayesian optimization (BO) is a technique commonly-employed to conduct optimization for black-box functions where a limited budget of function evaluations is possible, which has been used to optimize designs (mostly involving continuous variables) within the fields of aeronautics [24], sensor networks [25, 26], robotics [27, 28], and microprocessor architecture design [29]. The main idea underpinning BO is to build a surrogate model for the objective functions (usually performance and cost), which is then used to quickly estimate the values of the objective functions of thousands of designs and decide which design to evaluate next (using the computationally expensive model).

The second objective of this dissertation is to develop a method (tailored specially for System Architecture activities and design-decisions patterns) for performing tradespace exploration for computationally-expensive functions using Bayesian optimization. This method will then be applied to explore the tradespace of different space and aerial concepts that have been proposed to extend Internet connectivity globally.

1.3 Research questions

The main goal of this dissertation is to understand and identify space and aerial systems that can be used to expand broadband globally. This will require the analysis of both technical and economic factors, as well as comparing these systems to current terrestrial systems. In particular, the main research question that this thesis aims to answer is:

Research Question 1

What kind of novel space and aerial systems can complement existing infrastructure and contribute towards expanding global connectivity at an affordable cost? What is the potential impact of such systems in terms of connecting additional populations?

Furthermore, given that answering the questions above requires complex modeling and simulation efforts, this thesis aims to determine if Bayesian optimization is well suited for the tradespace exploration of SAPs, thereby answering the following research question:

Research Question 2

In the context of System Architecture, how should Bayesian optimization be used to conduct tradespace exploration when computationally expensive models are required and the number of point evaluations is limited?

1.4 Background

The objective of this section is to provide a brief introduction to "Connectivity technologies" and "System Architecture" (including "Tradespace Exploration"), as this will be recurrent topics throughout the dissertation. For readers who are already familiar with these topics, this section can be skipped.

1.4.1 Technologies to provide connectivity

As social animals, humans have always devised ways to communicate with each other. From ancient techniques such as smoke signals, drums, or carrier pigeons, to modern electric-based telecommunications, our methods of communication grow more complicated with the technological advancements of each generation.

In that vein, the invention of the telegraph at the beginning of the 19th century resulted in thousands of kilometers of copper wires being laid out, connecting selected

locations in all continents but Antarctica. Following the invention of the telephone by Bell in 1878, hundreds of millions of kilometers of copper wires were deployed to homes all over the world. The invention of the radio by Marconi at the beginning of the 20th century ushered in the era of wireless communications and resulted in the deployment of thousands of broadcast antennas, which proliferated even more with the invention of television. In the '60s, the first geostationary communications satellites were launched, making intercontinental high data rate transfer a possibility. At that time, satellite technology was envisioned as *the* technology that would dominate high-speed data-transfer across continents; however, since the '80s (as the mass-Internet grew more and more popular) fiber optic has instead become the dominant technology for high-speed data-networks, with cables spanning the world and connecting all continents. Finally, the advent of personal wireless communications for data connectivity (such as WiFi) and mobile networks has resulted in millions of cellphone antennas and wireless access points being deployed in the last 20 years.

This brief overview of the evolution of telecommunications in the last 150 years highlights the wide variety of technologies that have been used to provide connectivity. From point-to-point copper wires to large antennas and satellites for TV and radio broadcast, these technologies were the foundation that sustained economic growth in the so-called "Digital Revolution". The rest of this section is devoted to describing the most common technologies currently used to provide data connectivity around the world. Technologies are grouped into three broad groups (as shown in Figure 1-2), based on the portion of the network they support (access, backhaul, or backbone).

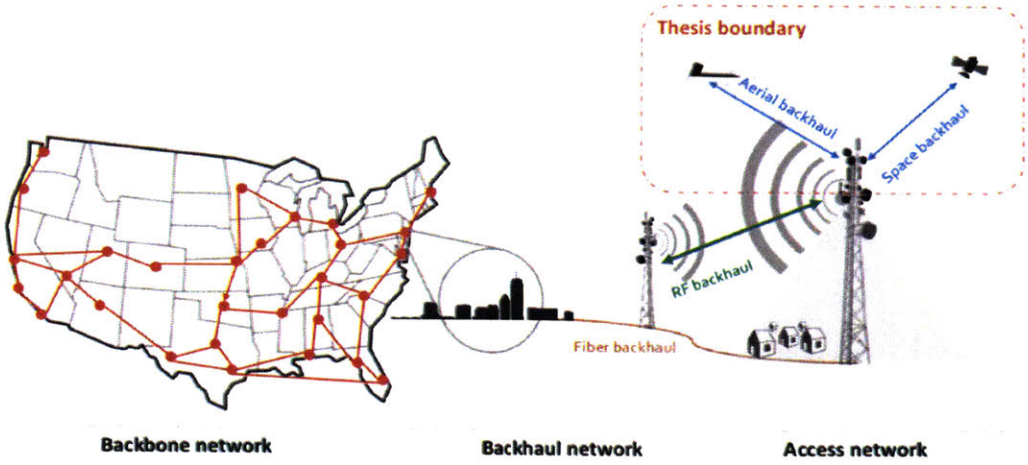


Figure 1-2: Types of networks. The red box represents the system boundary for this thesis.

Access network

The access network is the portion of the network that connects subscribers to their service provider. Technologies for access networks can be divided into two categories: wired and wireless. Examples of wired technologies include fiber optic and copper cables (Ethernet, coaxial, twisted pair), whereas examples of wireless technologies include WiFi, cellphone (LTE, UMTS, GSM), and satellite. Since the network equipment in an access network is only used when the subscribers are connected and actively using the Internet, wired technologies are normally underutilized, resulting in cost-ineffective deployments. In contrast, wireless technologies have higher utilization since they allow for traffic multiplexing sharing of the equipment; thus, they are more economical to deploy. Because of that, wireless technologies have grown in popularity in the last decade, especially in developing countries. It is estimated that wireless traffic currently represents 52% of the global access traffic, and will surpass 70% by 2022. Among wireless technologies, WiFi accounts for 54% of the traffic (59% in 2022), and LTE for 33% (29% in 2022). [30].

Throughout this dissertation, it is assumed that, due to its lower costs, only wireless technologies are used in the access network. In some cases, it is assumed that some customer premise equipment (CPE) with WiFi or LTE is deployed in the population settlements as part of the system, whereas in other cases, users connect directly to the concept under study (e.g., aerial systems beaming connectivity directly to user handsets).

Backbone network

The backbone network, which connects different high-level service providers (normally Tier 1[1] and Tier 2[2]), is the core of the network, and it comprises underground, aerial, and undersea fiber cables spanning all five oceans and six continents. Given the volume of traffic in the backbone network, most of the backbone links are wired – fiber optic being the dominant technology – and only a small fraction of the traffic is routed via satellite networks (currently 99% of international data traffic is routed through fibers whereas the remaining 1% is routed through satellite networks [31]). This is because the capacity of fiber is much larger than that of satellite links: a single fiber optic cable can transmit tens of Tbps, whereas the highest total capacity of a satellite is currently still below 500 Gbps.

Given this difference in throughput, the reach of the current fiber infrastructure, and the significant investments required to make new deployments and upgrade net-

[1]Tier 1 providers are those that can exchange traffic with any other network of the Internet via peering agreement (i.e., without paying fees). Normally these providers have global networks spanning multiple continents. Examples include AT&T, Deutsche Telekom, and CenturyLink.

[2]Tier 2 providers are those that peer for free with some networks, but still pay for transit IP to other networks. Generally, these are country-level providers like Vodafone, British Telecom, and China Telecom

work equipment, there are no prospective changes to the type of technology used for backbone networks in the coming years. Because of this, backbone network technologies will not be discussed further in this dissertation.

Backhaul network

Finally, the backhaul network comprises the intermediate links between the access network and the backbone network. An example of a backhaul network would be the link between a cellphone tower and the backbone network (normally accessed through an IXP or a point of presence (PoP)). As in the case of the access network, the technologies used for backhaul networks can be divided into wired and wireless. Wired technologies include fiber, and sometimes Ethernet links. Although wired technologies are generally preferred, the high costs of deploying such networks in low population density areas or rugged terrain have resulted in a significant portion of the backhaul links in developing countries being built with wireless technologies. These include point-to-point terrestrial RF relays (normally using millimeter-wave frequencies), satellite, and in very rare occasions free-space optics and aerial systems (mostly for disaster relief).

In the last few years, however, there have been several proposals in which space and aerial systems play a crucial role in backhaul networks. Since these aerospace systems can provide coverage to large regions, it has been argued that they could be good solutions to expand connectivity into medium-density and low-density areas. Understanding the impact that these systems could have on bringing new population to the Internet is thus the main objective of this dissertation.

1.4.2 System Architecture

System Architecture originated as a formal discipline in the late '80s, when system architects and engineers developed a theory to deal with the challenges of designing and building large-scale, complex systems [32]. The field encompasses all the activities of the early design stage in the systems engineering process, and studies the tools and methodologies used to define, document, communicate, certify, maintain, and improve complex systems throughout their life cycle [33]. Crawley defines the architecture of a system as "the embodiment of a concept: The allocation of physical/informational functions to elements of form, and the definitions of interfaces among them and with the surrounding context." [34]. In this definition, Crawley recognizes that the development of the architecture of a system is the process of materializing a concept, which he further defines as "the system vision, idea, notion, or mental image that maps function to form. It is a scheme for the system and how it works."[34]. This materialization or embodiment process is precisely a mapping of a system's function (i.e., the system's activities and operations that contribute to value creation) to particular elements of form (i.e., the physical entities that constitute the system).

Thus, it follows that a critical part of successfully engineering complex systems is choosing a good concept. Making comparisons among concepts is a highly-difficult task, since a large set of attributes (both quantitative and qualitative) needs to be weighted under initial uncertainty. Poor decisions at this stage will have a significant impact on final cost and schedule, since redressing mistakes at a later design stage tends to be highly expensive. For a given concept, thousands of design configurations (architectures) are possible; therefore, one should avoid reducing the design space prematurely, as this might take away crucial information from the decision-maker. Instead, tradespace exploration must be conducted before the comparison, to ensure that good designs for the different concepts are considered.

This dissertation is particularly interested in the methods and tools employed to select good concepts when the system's architect is limited by the number of configurations that can be evaluated. This limitation could be a result of finite resources, time, and cost; for example, in domains like CFD or communications network design, running computationally-expensive models or simulations is required to assess a design; in disciplines like chemistry, experiments need to be performed; in areas like drug design and advertising, human testing is required, which is both costly and time-consuming.

The main difference between the methods proposed in this thesis and previous efforts is that I present an explicit tradespace exploration method for situations in which *expensive functions* must be used (and thus a limited budget for function evaluations is available), whereas, in the past, methods developed to explore large design spaces where multiple concepts are considered ([14, 15, 35]) made use of simplified models which were not computationally-expensive to evaluate (hence the number of function evaluations was mostly unconstrained). Note that I do not renounce exploring, as broadly as possible, the conceptual design tradespace, but merely present a method to make this exploration more efficient. This method, however, incurs greater computational overhead, and therefore might not be suitable for cases where function evaluations can be performed quickly (less than 30 seconds per evaluation).

System architecture synthesis process

According to Crawley et al., the System Architecture Synthesis process can be decomposed into four principal steps [34]:

1. Identify and characterize the stakeholders (including beneficiaries, environment, regulatory bodies, etc.), determine their interrelations, and establish their needs.

2. Translate stakeholder needs to system goals and requirements. State the solution neutral statement (i.e., what the system must do).

3. Generate concepts by identifying functional elements that can satisfy the goals of the system.

4. Allocate the elements of function identified in Step 3 to elements of form and describe their interrelationships. In addition, the process flow, internal operands, supporting processes, and interfaces must be defined in this step too.

Successful application of these steps for a given problem produces a system architecture for each of the concepts. The system architecture synthesis process establishes the vector of design variables and operating parameters that need to be determined to produce a design configuration. By evaluating a design configuration (using a system model), one can assess its performance and verify if it is capable of satisfying the requirements for the system.

The process of mapping function to form for a particular concept (Step 4) can be characterized as a decision-making process; the system architect has to make a set of decisions regarding which elements of form will perform which functions. This set of decisions (and its potential alternatives) forms the design space for that particular concept. Note that each concept will have a different design space, since a different function-to-form mapping is made for each concept, requiring a different set of decisions.

The design space (decisions and options) has historically been represented using design structure matrices, decision trees, Markov processes, and meta-languages [36], although the most common method is the morphological matrix, as introduced by Zwicky in 1962 [37]. A morphological matrix contains one row for each decision, with the values in the columns representing alternative options for the decision. A design configuration is obtained by assigning a value/option for each decision, although some combinations of variables might be forbidden if they render infeasible architectures. By systematically assigning a value to each of the decisions, the architect can then enumerate all the potential design configurations. Several methods to generate an exhaustive enumeration of all the possible combinations of decisions, taking into account constraints, have been developed [35, 36].

Finally, even though the traditional approach has been to use a decision-options characterization to model decisions, Selva noted that there are decisions within the System Architecture domain that are better characterized if they are represented as a combinatorial problem, subject to certain constraints [38]. To that end, he defined a set of 6 different types of decision patterns that allow for some SAPs to be solved more efficiently by using tools tailored to the specifics of each decision pattern. Selva argued that these patterns constituted a domain-independent basis for the types of decisions, and thus when using this characterization for the system architecting process, the system designer could benefit from using tools originally developed for combinatorial optimization problems.

Tradespace Exploration

The tradespace exploration (TSE) paradigm has been the predominant systems engineering and concept selection tool in the last decade and has been widely adopted both in the industrial and academic contexts. As opposed to more traditional methods that largely rely on expert opinion and detailed analysis of a few point designs (known as "expert design"), tradespace exploration advocates for the evaluation of a multitude of designs corresponding to different concepts [39]. This allows the decision-maker to have a good understanding of the big picture, not just of the Pareto front (the set of non-dominated designs), but also how sub-optimal architectures rank. The Pareto front illustrates the key cost-benefit trade-off among design options, whereas analysis of the sub-optimal points (through main effects or sensitivity computations) provides insight on how certain decision-options affect the cost and performance of the system. In that sense, one could say that tradespace exploration – as its name suggests – focuses on the exploration of the design space, as opposed to traditional methods, whose main objective is to uncover design choices that are optimal or near-optimal [40] or to optimize the expert-selected single-point designs.

In practical terms, although one would ideally like to explore the full tradespace (i.e., evaluate all the possible design options) to obtain the best overview of the big picture, full exploration is infeasible in most cases. Whether this is due to the tradespace size being too large for full enumeration, the evaluation (or testing) routines being too computationally (or economically) expensive, or timeline constraints, different techniques have been proposed to conduct tradespace exploration, including sampling (when the focus is on exploration), tradespace optimization (when the focus is on finding the Pareto frontier), and user-guided exploration (when user feedback guides the exploration towards *interesting regions* of the design space.

The main benefit of TSE over other methods is a better understanding of the trade-offs and relationships between variables in the design space [41], which makes the technique very relevant during the conceptual design phase. These trade-offs and relationships are normally depicted as a 2D scatter plot, where one dimension represents the benefit of the architecture, and the other dimension represents cost.

1.5 Thesis Structure

The remainder of this dissertation is structured as follows: Chapter 2 reviews the literature in three fields that are directly related to this thesis, namely space and aerial systems for communications, concept selection methods, and tradespace exploration (including Bayesian optimization). Chapter 3 presents a method to apply Bayesian optimization to tradespace exploration in System architecture problems. The performance of this method when applied to SAPs is assessed by comparing its results to those obtained by other state-of-the-art tradespace exploration methods using a

set of test problems in Chapters 4 and 5. Then, Chapter 6 provides an overview of the current status of global connectivity and the main barriers to connectivity. Chapters 7, 8, and 9 are devoted to describing and validating the models, analyzing the tradespace, and presenting the results for space, aerial, and terrestrial systems, respectively. Chapter 10 compares the impact of each of the concepts in a competitive market. Finally, Chapter 11 summarizes the work conducted in this dissertation, restates the main contributions, and delineates potential areas of future research.

Chapter 2

Literature Review

2.1 Space and aerial systems to expand connectivity

2.1.1 Space systems

Even though initial communication satellite systems were launched in the '70s and '80s, it was not until the '90s when constellations were first proposed as a means to provide global broadband. Within that decade, multiple projects consisting of large constellations of LEO satellites were developed. This generated a broad set of research in technical topics such as constellation design [42, 43], inter-satellite links technology [44, 45, 46], and in-space routing devices and protocols [47].

In particular, significant effort was made to determine both the technical and economic feasibility of such systems, as well as to evaluate competing architectures. Kelic and Gumbert proposed the "*cost per T1 link-per-minute*" as a metric to assess the financial viability of broadband satellite systems, and used it to study 5 GEO and LEO systems [48, 49]. Shaw developed GINA, a generalized framework for evaluating information networks, and used it to assess the designs proposed by Cyberstarr, Spaceway, and Celestri [42]. Jilla extended GINA and integrated it with an MDO framework, which he used to conduct tradespace exploration of LEO, MEO, and GEO satellite systems [43]. Finally, de Weck proposed a methodology for tradespace analysis of LEO personal communication systems [50], that combined technical, economic, and policy aspects with which he analyzed how staged deployments can be used to reduce risks associated with the uncertainty on demand [51].

Much of the latest research in geostationary systems is focused on achieving Terabit/s satellites [52, 53] by using higher frequency feeder links (Q/V-band [54, 55], advanced multi-beam antennas with beam-forming capabilities, and optical technology [56, 57]), and by the use of advanced MODCODs. Finally, a second wave of

proposals to use large non-geostationary constellations as a means to provide global coverage has surfaced in the last three years. Several companies have developed ambitious plans to cover the skies with thousands of satellites [58, 59, 60] and use the proximity of low Earth orbit to provide global low-latency connectivity across the world; the related research in these areas is focused in ground-segment design [61], dynamic resource management (dynamic beam pointing, power, and bandwidth allocation, interference avoidance) [62, 63], and network control [64, 65]. However, the literature evaluating the technical and financial merits of these newly-proposed non-geostationary satellite orbit (NGSO) systems is scarce.

2.1.2 Aerial systems

Aerial systems have long been considered as a candidate solution to extend the reach of communication networks since they can cover larger areas compared to terrestrial systems. In this document, I shall use the term *aerial systems* to refer to both high altitude platform stations and low altitude platform stations. This section provides a literature review of the previous work done in these two fields.

On the one hand, a high altitude platform (HAP) is defined as "[...] telecommunications station located at an altitude of 20 to 50 km and at a specified fixed point relative to the Earth" [66]. HAPs are seen as the middle ground between terrestrial and space systems, which exploit the advantages of both: providing large coverage areas (100 - 10,000 km^2), requiring a small number of base stations, having low interference due to buildings and terrain obstructions, low latency, rapid, incremental, and localized deployments, ease of maintenance, and potential for reconfiguration [67].

Research in the use of HAPs as a platform for communication services boomed in the 2000s after ITU's WRC-97, where the spectrum allocation for HAPs services was adopted by over 50 countries. Within that decade, several projects for the research and development of HAPs were started, some being unsuccessful at the initial research stage (SkyStation [68], HALO [69], and StratSat [70]) while others reached the development phase. Among the latter group, some examples are aerostats (HiSentinel80 [71], JAXA's SPF [72], Stratospheric Airship [73, 74], HALE-D, Stratobus [75]), solar-powered unmanned aircraft (Helios [76, 77], Zephyr [78], Vulture [79]), and manned aircraft (U-2, WB-57, M-55 and G520 [80, 81]). A detailed overview of each of these systems can be found in [82].

In addition, various system analysis studies were conducted to identify key design parameters and challenges in deploying HAPs systems, as well as to provide different alternative architectures for HAPs [67, 83, 84, 85, 86]. The role of HAPs within the global connectivity landscape and their interactions with existing infrastructure (terrestrial and satellite) was the focus of [87, 88, 89]; other parts of the literature analyzed the coverage problem (i.e., the number of HAPs required to provide coverage to a particular region) for selected countries [84, 90, 91].

With regard to the enabling-technologies, studies were conducted from both the communications payload and platform perspectives. Research into the communications payload focused on the development of antennas [92], MIMO technology [93], cognitive radio [94], and free-space optics [95] tailored to HAPs; on the platform end, extensive research was conducted on system design and optimization of long-endurance platforms, both for solar-powered unmanned aircraft [96, 97, 98] and airships [99, 100, 101].

On the other hand, a low altitude platform (LAP) is defined as "small aerial telecommunication stations that fly at altitudes below 1 km at specified fixed points relative to Earth". LAPs include tethered aerostats (commonly balloons and blimps) and station-keeping enabled UAVs. Even though the principal use cases for LAPs have been to facilitate rapid recovery of damaged terrestrial wireless infrastructure due to natural disasters [102, 103], to augment wireless networks to cover foreseen massive crowd movements [104, 105], and for temporary military deployments [106], in recent years several commercial companies have proposed using LAPs to provide long-term communication infrastructure in rural and low-population density areas [107].

Since LAPs fly at lower altitudes, communications are affected by similar effects as mobile cellphone networks. Given this, research in LAPs has focused on characterizing the link channel [108, 109, 110, 111], and developing algorithms to determine the optimal altitudes and spacing between stations such that performance is maximized while interference with existing terrestrial networks is minimized [112, 113, 114]. With the increasing popularity of low-altitude UAVs, some cellphone companies have studied the use of "swarms" of UAVs to form low-altitude networks that extend the existing terrestrial networks [115, 116]. Finally, the literature in buoyancy-lift platforms, such as tethered blimps and balloons, is not nearly as extensive and is generally focused on platform development [117].

2.1.3 Summary of technologies to expand connectivity

Table 2.1 shows a summary of previous research in both space and aerial networks. As can be seen, most of the works considered end-to-end models of the system, and some identified the most important architectural decisions.

However, there is a lack of analysis involving tradespace exploration while taking into account the economic feasibility of the different systems. This is mainly due to the lack of reliable economic models for these systems, which are in a very incipient stage of development.

Table 2.1: Summary of technologies to expand connectivity

		Reference	Models end-to-end system	Identifies architectural decisions	Conducts tradespace exploration	Evaluates economic feasibility
Satellite networks		[42] Shaw 1999	Yes	Yes	Partially	Yes
		[43] Jilla 2002	Yes	Yes	Yes	Yes
		[50] de Weck 2002	Yes	Yes	Yes	No
		[48] Gumbert 1996	Yes	No	No	Yes
		[49] Kelic 1998	Yes	No	No	Yes
		[51] de Weck 2004	Yes	Yes	Yes	Yes
Aerial networks	High altitude platforms	[67] Djuknic 1997	Yes	Yes	No	No
		[83] Karapantazis 2005	Partially	Yes	No	No
		[85] Grace 2001	Yes	Partially	No	No
		[86] Pace 2004	Partially	Partially	No	No
		[84] Tozer 2001	Partially	Yes	No	No
		[90] Milas 2003	No	Partially	No	No
		[91] Miura 2001	Yes	No	No	No
	Low altitude platforms	[108] Al-Hourani 2014	No	Yes	No	No
		[111] Amorim 2017	Partially	Yes	No	No
		[112] Al-Hourani 2014	Partially	Partially	No	No
		[113] Al-Hourani 2015	Partially	Partially	No	No

2.2 Concept selection methods

Concept selection plays a crucial role in the engineering design process since the outcome of this step will heavily influence the quality, cost, and performance of the final system or product. More precisely, concept selection refers to the conceptual design stage in which, after a *diverging* phase where different concept alternatives are generated, these concepts are evaluated, compared, and the best alternatives chosen for further study in a *converging* phase [118].

The importance of this step stems from the fact that a bad decision made at this juncture can rarely be corrected in later stages without incurring prohibitive high costs and causing great delays. Because of this, a significant amount of research has gone into concept section within the last 40 years [119]. This section provides an overview of the different concept selection methodss (CSM)text that have been developed, which can be broadly classified into qualitative and quantitative methods.

2.2.1 Qualitative methods

Qualitative CSMs are those that shun mathematical evaluations as part of the comparison, using instead subjective preferences to inform decisions. Some authors have argued that at the initial conceptual stage of design, mathematics is of little or no utility, perhaps even harmful to the decision-making process [120].

According to Mattson and Messac, qualitative methods are based on three different principles [121]:

1. Unstructured selection methods based on individual or collective preferences (e.g., voting).

2. Structured selection methods based on feasibility judgment, intuition, go/no-go screening, pros and cons, and technology-readiness assessment.

3. Decisions-matrix selection methods.

CSMs that fall within categories 1 and 2 are characterized by their subjectivity and lack of structure. They usually do not consider a fixed set of attributes, and hence, there is no clear criterion that specifies how comparisons between concepts should be made. In contrast, methods that fall within category 3 present a clearly-defined methodology to compare concepts. The most popular decision-matrix method in engineering is Pugh's method for concept selection, introduced in 1991 [122]. This method sets a concept as the baseline (datum) and evaluates all other concepts against it. Each decision criteria is compared against the datum, and a '+', '-', or 's' is assigned if the concept outperforms, under-performs, or has similar performance to the datum (respectively). The number of '-'s is subtracted from the number of '+'s, and the lowest-scoring concepts are removed.

A more advanced design-matrix-based CSM is Quality Function Deployment (QFD) [123], which was developed in Japan in the '60s. The most popular version of this method is the House of Quality (HoQ), which uses a ⌂-shaped relational matrix to map requirements to the engineering characteristics of each concept [124]. The main advantage of HoQ over the Pugh Matrix is that HoQ allows for the weighting of metrics and explicit representation of coupled decisions.

2.2.2 Quantitative methods

In contrast, quantitative methods evaluate the performance of each concept against a set of decision criteria *numerically*. These methods have gained popularity in recent years following a movement towards employing more formal frameworks in the decision-making process. Among the advantages of using numerical methods to perform concept selection are having a explicit quantification of the difference in "goodness" between concepts (i.e., how much better is one concept over another), the possibility of establishing weights to evaluate the relative importance of each decision criteria, and also the possibility of incorporating measures of uncertainty into the decision-making process.

Some of the most widely-used classes of quantitative CSMs are those based on utility theory, which try to maximize the expected value of a utility function [125]. The utility function represents the values of a given concept according to a set of criteria, which have each an associated weight representing its importance. Given a concept, the designer assigns a score to each of the *criteria* (i.e., the utility along that dimension), and the final outcome of the utility function is the dot product between

the weights and scores. In the systems engineering context, this approach was first proposed to be applied in design by Pahl in 1984 [126], and several extensions, such as multi-attribute utility analysis, have been developed subsequently [127, 128]. The major criticism of utility theory is the difficulty in determining *a priori* the shape of the utility functions and coming up with the corresponding weight values for the criteria.

Another popular method is analytic hierarchy process (AHP), developed in 1980 by Saaty [129, 130]. This method involves creating a hierarchy of criteria and comparing the pairwise-performance of all the alternatives for each criterion. The method was later adapted by Marsh in 1993 for design decision-making [131], and an AHP-based method tailored for concurrent-design principles and multiple-person decision-making processes was further developed by Mullens in 1995 [132].

The three main drawbacks of the two quantitative CSMs mentioned above are that 1) they do not take into account the uncertainty in criteria and imprecise information inherent to early decision-making stages, 2) they are typically applied to cases where a single design is considered for each of the concepts compared, and 3) they scale poorly when large tradespaces are considered. Different CSMs developed in the last few years have tried to overcome these limitations specifically.

To account for uncertainty in the early decision stages, CSMs that explicitly consider uncertainty modeling have been developed. We can classify these methods into three broad categories. First, methods that use a probabilistic approach such as those by Otto and Wood [133], and Takai and Ishii [134]; these add probabilistic analyses and sensitivity methods to Pugh's matrix to quantify the probability of satisfying each criterion (these probability values are used in-place of the '+'s and '-'s). Second, non-classical mathematical methods based on triangular fuzzy numbers have been used to extend AHP [135, 136, 137], which allow for the computation of the weight vectors ranking alternatives to take into account uncertainties. Finally, fuzzy clustering methods such as product concept generation and selection (PCGS) have also been proposed for consumer products concept selection [138]. PCGS groups concepts together with the Fuzzy C-means algorithm after capturing expert knowledge assessments of three different domains for each concept: functional, commercial, and marketing.

To enable more thorough tradespace exploration, methods based on evolutionary computation have been proposed. Buonanno and Mavris argued that in most instances CSMs select only a handful of designs from the billions of possible designs on which to perform detailed analysis and (qualitatively or quantitatively) evaluate "concept goodness" [15]; in order to overcome this limitation, they came up with a genetic algorithm-based CSM that allows for the exploration of large design spaces. A genetic algorithm is a meta-heuristic optimization algorithm that emulates the "survival of the fittest" principle by applying bio-inspired operators to a set of candidate design configurations [139]. Buonanno and Mavris proposed a dual-stage multi-fidelity optimization algorithm, where a low-fidelity fast-computation algorithm is used first to

solve a set of two-criteria (2-dimensional) multi-objective problems in parallel, and then the best designs from the first step are evaluated using higher-fidelity models to determine the final designs. The method keeps iterating between the first and second stages until convergence is achieved.

Later on, Mattson and Messac proposed a Pareto-frontier CSM [14, 16], based on the argument that the best design configurations will be those which appear on the global Pareto front of the multi-objective optimization problem. In particular, they denoted the s-Pareto front (*set of concepts* Pareto front) as the set of non-dominated architectures across all concepts, which forms a hyper-surface in the N-dimensional space (N being the number of criteria considered); the "goodness" of each concept is then measured by its contribution to the area of that hyper-surface. In the case where the s-Pareto front is generated using a method that ensures an even distribution of architectures, the "goodness" of a particular concept can be approximated as being proportional to the number of architectures on the s-Pareto front [121] derived from that concept.

Finally, Moshaiov and Avigad have worked on the multiple-concept, multi-objective problem [17, 140]. Their work includes tailoring evolutionary computing algorithms such as ϵ-MOEA [141] and NSGA-II [142] to deal with the multi-concept approach, as well as new methods that take into account delayed-decisions [143], tradespace exploration under run-time limitations [144], and human-in-the-loop interactive approaches [145, 146].

2.2.3 Summary of concept selection methods

Table 2.2 presents a comparison of the main CSMs in terms of whether they are suitable for tradespace exploration and expensive function evaluation, and whether they consider uncertainty measures. Note how the current methods that are suitable for tradespace exploration are not valid in cases where expensive functions need to be evaluated, as they rely on optimization methods that require a large number of function evaluations. Conversely, methods that are currently used when expensive functions are involved are not suitable for tradespace exploration, as only a handful of pre-selected designs can be compared.

2.3 Tradespace exploration

As illustrated in Figure 2-1, techniques for conducting tradespace exploration (TSE) can be grouped into three broad categories: full enumeration, sampling, and optimization.

Full enumeration consists of evaluating all the configurations from the design space, and is the preferred method when the size of the search space is small, as this provides

Table 2.2: Comparison of concept selection methods.

	Reference	CSM name	Considers Uncertainty	Suitable for Tradespace Exploration	Suitable for expensive functions
Qualitative	[122] Pugh 1990	Pugh matrix	No	No	Yes
	[123] Coelho 2005	QFD	No	No	Yes
Quantitative	[126] Pahl 1984	Utility functions	No	Partially	Partially
	[125] Otto 1993	Utility functions	No	Partially	Partially
	[129] Saaty 1980	AHP	No	No	Yes
	[133] Otto 1995	Probabilistic Pugh	Yes	No	Yes
	[136] Chang 1996	Fuzzy AHP	Yes	No	No
	[138] Yan 2006	PCGS	Yes	No	No
	[15] Buonnano 2004	Variable fidelity GA	No	Yes	No
	[16] Mattson 2002	s-Pareto	No	Yes	No
	[17] Moshaiov 2007	EC for multi-concept	No	Yes	No

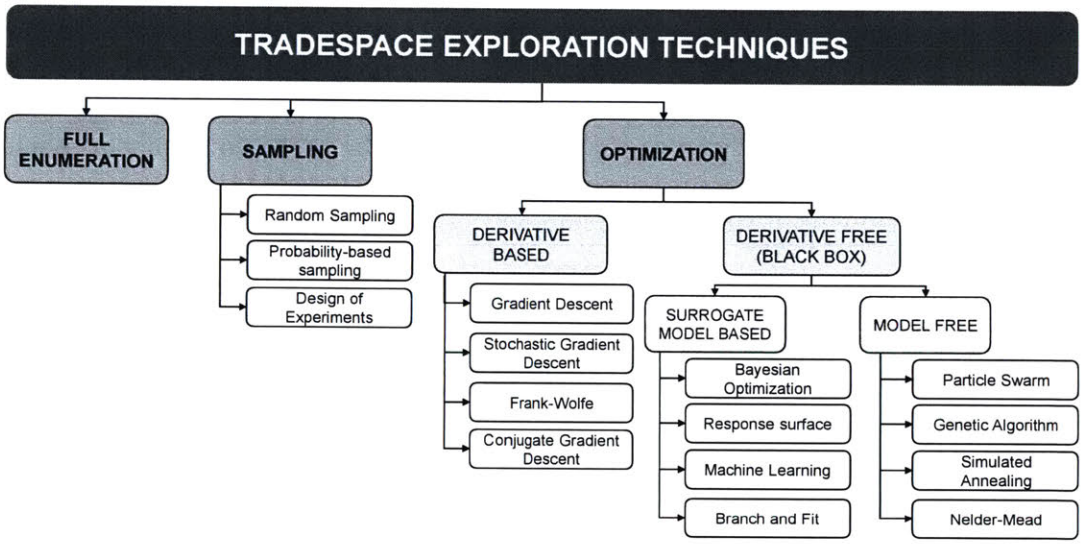


Figure 2-1: Taxonomy of tradespace exploration techniques.

the most information regarding the tradespace and ensures that the optimal architectures are found. In sampling techniques, only a subset of the configurations in the search space is evaluated. This subset is chosen using a set of rules (whose objective is to have good coverage of the design space) before evaluation. Finally, in tradespace optimization techniques, too, only evaluate a subset of configurations, but the points to evaluate are dynamically chosen (trying to find optimal and semi-optimal designs) based on the values obtained from previously-evaluated points. Note that one of the key differences between these techniques resides in the ultimate goal of the search process: whereas for the sampling and full enumeration techniques the objective is to gain information of the structure of the whole tradespace, for the tradespace optimization techniques the objective is to find the *best* architectures.

Tradespace optimization techniques can be further sub-divided into two categories: derivative-based optimization and black-box functions optimization (or derivative-free optimization). On the one hand, in derivative-based optimization, the user knows the gradients of the function and thus can employ numerical methods such as steepest descent or the conjugate gradient method for optimization. However, knowing the gradients is not possible in most SAP, as the decision variables are usually categorical, and the complexity of the problem and the interrelations between elements make the dependencies of the system hard to model analytically (in particular, none of the SAPs analyzed in this dissertation belongs to this category). Because of this, derivative-based optimization methods will not be covered further in this section (a thorough review is provided in [147]). On the other hand, black-box functions are those where, given an input, it is possible to obtain the corresponding output, but additional information about the gradients or the characteristics of the underlying functions is not available. These types of functions are common in System Architecture, and the rest of this section is devoted to reviewing past work on the optimization of such functions.

2.3.1 Sampling

Sampling is seldom used as the only technique during the tradespace exploration phase. Nevertheless, it is common to have an initial subset of points be selected via sampling, evaluated, and later used to guide the search of the tradespace either by constructing a meta-model [148] or by providing general insight to the user. In addition to random sampling, in which configurations are selected without careful choice, structured methods such as Design of Experiments (DoE) (also referred to as experimental design), and probability-based sampling are also used.

The most common sampling technique is DoE, where the objective is to obtain subsets of architectures with good space-filling properties (i.e., they uniformly cover the design space). DoE comprises a broad set of techniques, with the most popular being Latin hypercube sampling [149, 150], orthogonal arrays [151], and Hammersley designs [152]. These methods are mainly used when nothing can be assumed about the relationship between inputs and outputs of a model.

In contrast to DoE, in probability-based sampling, the user attaches probabilities to some areas of the design space being sampled, thus evaluating more (or fewer) configurations from specific regions. This allows the user to incorporate prior knowledge and/or preference in the process; by focusing the exploration on a subspace of the design space, they increase the likelihood of a particular set of architectures being selected [153, 154]. Examples of probability-based sampling techniques include random sampling, stratified random sampling, systematic random sampling, and cluster random sampling [155].

2.3.2 Black-box functions optimization

Among the techniques for optimizing black-box functions, one can distinguish between model-free methods and model-based methods. Model-free methods are based on heuristics that guide the search of the design space using the outputs already evaluated, whereas model-based methods use the input-output pairs resulting from function evaluations to construct a surrogate model, and use the model to guide the optimization process.

Model-free methods

Model-free methods use meta-heuristics that mimic natural processes such as evolution (e.g., evolutionary strategies[156], genetic algorithm (GA)[157]) or swarm intelligence (e.g., particle swarm optimization (PSO)[158, 159], ant colony optimization (ACO)[160]) to obtain solutions for the optimization problem.

GAs and evolutionary strategies have been applied to solve SAPs where categorical and combinatorial parameters play a crucial role, as they offer a natural way of encoding these types of variables. In particular, GAs have been used to architect systems in the domains of space [161], communications [162, 163], suborbital spacecraft [164], aeronautics [15], and integrated circuits [165]. GAs are based on the iterative application of three basic operations to a population of individuals: selection of good candidate solutions, crossover between pairs of solutions, and mutation of solutions. The selection operation picks a subset of the population individuals as the "parents" of the next generation; crossover combines pairs of good solutions to produce new candidates; whereas mutation applies random changes to individuals to increase diversity and avoid getting stuck in local optima.

PSO strategies are predominantly used in situations where the parameters of the design space are continuous, while ACO is mainly used in graph-type problems. These two types of techniques are based on swarm intelligence: PSO mimics the behavior of flocks of birds in flight, whereas ACO imitates the procedure that ants follow when gathering food. The general idea behind these methods is that there is a population of agents who act according to a set of simple rules, which leads to the emergence of "intelligent" global behavior. For example, in ACO each agent (ant) lays down "pheromones" at levels of intensity proportional to the quality of the solution they have found; the rest of the ants then, adapt their behavior to move towards the regions of the space in which the "pheromones" are more intense. Analogously, in PSO, each of the "birds" communicates with a set of neighboring birds and individually adapts its speed to get closer to the birds that represent the best solutions within its communication-group.

All these methods are population-based methods, meaning that multiple agents search through the solution space cooperatively and globally. These methods rely on randomness and repeated evaluation of the objective function to solve the optimiza-

tion problem, which makes them poorly suited to situations where there is a tight constraint on function evaluations.

Model-based methods

Model-based methods use the pairs of inputs-outputs from evaluations of the black-box function to build a surrogate model, which is later used to guide the optimization process. It is common to update the surrogate model as new points of the black-box are evaluated. Bayesian optimization itself is a model-based method, with the particularity that it constitutes a probabilistic framework in which prior information about the function to optimize can be incorporated. Since Bayesian optimization is central to this thesis, a literature review for the topic will be presented independently in the next subsection.

In terms of surrogate modeling techniques, there are three methods that are most prevalent in the literature: response surfaces, machine learning techniques, and Kriging. A review of these methods and their applications in system engineering and system architecture can be found in references [166, 167]. The remaining of this section provides a brief overview of each of the techniques.

Response surface methodology (RSM) was firstly proposed in 1951 by Box and Wilson [168]. It comprises a collection of statistical techniques, to fit a lower order model to a set of observation points, in order to predict a set of outcomes in a computationally-inexpensive manner. Conventional approaches employ second-order polynomial models that are fitted to minimize the mean square error [169]. RSM has been used extensively in engineering to perform optimization [170, 171, 172], design exploration [173], and trade-off analysis for the effects of different design variables on the output. In relation to system architecture analysis, RSM has been used in a wide variety of contexts, such as the generation of Pareto frontiers with evenly distributed points [174], architecting multiprocessor systems-on-chip [175], and aircraft design [176, 177].

Machine learning techniques include methods such as artificial neural networks (ANNs) and decision trees. ANNs are computational systems inspired by biological neural networks; an ANN is composed of a set of "neurons" in different interconnected layers [178]. Each neuron performs a nonlinear transformation of the sum of its inputs. The connections mimic synapses in biological neurons, with weights determining the relative strengths of the signals from different neurons. The values of the weights are determined when fitting the ANN model to the existing data, a process commonly referred to as "training". ANNs are best suited for the approximation of deterministic functions in regression-type applications [179] and should be used when a large number of configurations can be evaluated, as training them requires a large number of input-output pairs. A decision tree is an inductive learning method that derives a set of rules (based on the inputs) that partitions the data into discrete categories, to which a value is assigned. Decision trees are best suited for problems where

the inputs are discrete values, although they can also be adapted to deal with continuous inputs. Both decision trees and ANNs have been extensively used as surrogate models in optimization in multiple disciplines [180, 181, 182, 183]

Kriging is an alternative name used to refer to Gaussian process regression, in honor of Daniel Krige, who invented the method for its application in geological applications [184]). These techniques were not widely used in engineering design applications before the '90s since simple computer implementations were not readily available. This changed in the 2000s when the first Kriging software toolboxes became widespread [185, 166]. The rising popularity of Kriging in recent years is due to the fact that it is an extremely flexible interpolation method capable of handling deterministic data.

Finally, in the last ten years, multi-fidelity models have gained increasing popularity. The central idea of these methods is that expensive high-fidelity models are combined with inexpensive (but less accurate) low-fidelity models to derive the optimal solution. A reduction of fidelity can be obtained by simplifying the physical equations that govern the system (e.g., relaxing the constraints), as well as using reduced-order models, a coarser discretization of the space (in finite element simulation), or even historical experimental data instead of mathematical models. Most of the work with multi-fidelity models has focused on optimization under uncertainty, with the most common fields being fluid and solid mechanics. Theoretical papers in these fields have focused on methods to select the surrogate models for the low-fidelity models and methods to combine the results of these different fidelity models [186, 187].

2.3.3 Bayesian optimization

Bayesian optimization is another model-based method for black-box optimization. Given that an essential part of this dissertation is studying the feasibility of using Bayesian optimization for tradespace exploration in SAPs when evaluating the performance and cost of a single design configuration is computationally expensive, an individual section has been devoted to discussing recent advances in the field.

Bayesian optimization in single-objective problems

In 1998, Jones published his seminal paper in which he described a Bayesian method to optimize expensive black-box functions [188] (see Section 3.1). In his method, a stochastic process which acts as a surrogate model is fitted to a set of function evaluations ("observations"), and the surrogate model, together with an acquisition function, is used to determine the point that is the most promising candidate for the next evaluation. In particular, Jones made use of Gaussian processes to obtain the fitted surface, and the decision of the next point to evaluate was based on the *expected*

improvement criterion, which considers both the predicted value for each point by the response surface and the confidence in that prediction.

Since then, considerable work has been done in the field of Bayesian optimization to speedup convergence and reduce the number of function evaluations needed. A significant body of research has focused on developing acquisition functions based on probabilistic frameworks [189, 190], optimistic policies [191, 192], and information theory policies [193, 194, 195]. Brochu proposed an approach in which he used a portfolio of different acquisition functions (instead of a single function), and a multi-armed bandit approach to determine the next point to evaluate [196]. Similarly, several authors have proposed methods to adapt the Bayesian optimization framework to multi-task settings, where the data from one task provides information about another task. [197, 198, 199]

In most of the earlier Bayesian optimization algorithms, the next evaluation point is chosen in a greedy manner (looking at the immediate next step), which has been deemed as a sub-optimal strategy [200]. This happens because making predictions with a multi-step lookahead strategy involves computing nested expectations, which can quickly become computationally infeasible. However, an analytical expression for the two-step expected improvement acquisition function [200], a multi-step Bayesian optimization [25], and a dynamic programming approach [201] have been developed.

Finally, from an application standpoint, several software packages have been developed to optimize single-objective problems using Bayesian optimization, with the Sequential Model for Algorithms Configuration (SMAC) [202], Spearmint [203], and Hyperopt [204] being the most popular ones. SMAC uses random forests as its surrogate model, whereas Hyperopt uses tree-structured Parzen density estimators. The main use case for these packages is hyper-parameter tuning in machine learning tasks and automatic algorithm configuration.

Bayesian optimization in multi-objective problems

Unlike in single-objective problems, for multi-objective problems there is not a singular optimal solution, but rather a set of non-dominated points which present trade-offs. To handle these trade-offs, numerous Bayesian optimization algorithms for multi-objective problems have been proposed in the last ten years, the most relevant being ParEGO [205], SMS-EGO [206], ϵ-PAL[207, 208], SUR [209], and PESMO [210]. The rest of this section provides an overview of each of these algorithms.

ParEGO uses scalarization of the metrics vector by introducing different coefficients in each iteration, thereby transforming the multi-objective problem into a single-objective problem. A Gaussian process model is fitted to the observations, and the original algorithm developed by Jones [188] is used to determine the new evaluation point. SMS-EGO uses the hypervolume improvement of an optimistic prediction for each design point as its acquisition function. An extension of SMS-EGO has been

proposed by Emmerich, using the expected hypervolume improvement (EHI) instead of the *hypervolume improvement for an optimistic prediction* [211]. This metric constitutes the natural extension of the approach used by Jones [188] for single-objective problems to multi-objective problems. However, the main limitation with this approach is that there is no closed-form expression for the EHI and thus its value is computed by numerical integration or Monte Carlo sampling. Despite the latest efforts to speed up the computation of the EHI, this is still a slow process that can only be applied to 2 or 3-objective problems.

ϵ-Pareto Active Learning (ϵ-PAL)[208] uses the Gaussian process predictions at each iteration to establish an uncertainty region around each of the points. The uncertainty region is a hyper-rectangle of size $2 \cdot (k\sigma(x)_k + \epsilon)$ along each dimension of the space, and each point is classified after each iteration as non-dominated, dominated, or undetermined, depending on whether all of the points in its uncertainty region are non-dominated by any other uncertainty region, dominated by at least one other uncertainty region, or overlap with other uncertainty regions, respectively. ϵ-PAL is best applied towards problems where all possible design configurations can be enumerated exhaustively (i.e., finite discrete domains), and its extension to continuous domains is work-in-progress. ϵ-PAL has been used during conceptual studies for FPGA-based systems, to find optimal configurations for the LLVM compiler flags, as well as for drug discovery research.

Sequential uncertainty reduction (SUR) [209] considers the expected decrease in the hyper-volume metric over the whole domain of objectives, which is equivalent to the one-step optimal policy in terms of reducing the uncertainty on the objective function. This is a slow operation, computed by dividing the objective space into cells and then integrating the probability of improvement over each cell.

Finally, predictive entropy search for multi-objective optimization (PESMO) [210] uses a decoupled acquisition function (with respect to the multiple objectives), which measures the expected reduction in entropy of the posterior distribution of the Pareto set. Because objectives can be decoupled, one of the advantages of this method is that there is no need to evaluate each design-configuration against all the metric functions, which leads to important speedups. A comparison of PESMO against the rest of the methods proved its superior performance [210]. PESMO has been thus used for system architecture studies to design neural network hardware accelerators, where in addition to the network performance, power is to be minimized [212].

Bayesian optimization for combinatorial problems

Bayesian optimization for combinatorial problems is very much an active area of ongoing research. Initial approaches by Moraglio generalized the Bayesian optimization framework to combinatorial spaces by replacing Euclidean distance measures used in the surrogate models with distance measures more suited to the combinatorial problem representation (e.g., Hamming distance (HD) or problem-tailored edit distances).

In particular, Moraglio used an radial basis function network (RBFN) as the surrogate model, together with the edit distance between binary strings, on the NK-Landscape problem and outperformed model-free evolutionary algorithm approaches. Other authors have since expanded on the work done by Moraglio by adapting the method for permutation problems [213, 214] and genetic programming problems [215].

Zaefferer [216] used a similar method based on edit distances (in particular the swap distance and the interchange distance) to transform the discrete domain of the combinatorial problem into a continuous domain, and then used the *Efficient Global Optimization* algorithm proposed by Jones [188]. A comparison of this approach against random sampling, genetic algorithms, and surrogate-based optimization, as applied to several benchmark problems, demonstrated the superior performance of Zaefferer's approach.

Baptista [217] proposed the use of a second-order sparse Bayesian linear regression model (i.e., a model that contains first-order effects as well as interactions of up to two categorical variables) as the surrogate model. In this surrogate model, parameters (those that need to be adjusted using the already evaluated data-points) are treated as random variables. This approach was shown to outperform other approaches such as SMAC [202] and simulated annealing in binary quadratic programming, sparsification of Issing model, contamination control, and aero-structural problems [217].

Finally, Oh presented COMBO [218], a Bayesian optimization algorithm that used a graph to represent the combinatorial design space, with each configuration being represented as a vertex in the graph. Their method decomposed the design space graph into the Cartesian product of subgraphs and used diffusion kernels and Gaussian processes within the common Bayesian optimization framework. Through four different problems, they compared the performance of COMBO against other Bayesian optimization methods and showed that their method is on-par with or outperforms the state-of-the-art.

Summary of Bayesian optimization research

Table 2.3 contains a comparison of the principal work in the Bayesian optimization literature to date. This comparison focuses on both theoretical and practical contributions. Note that even though there exists broad literature in Bayesian optimization methods for continuous problems (with both single and multiple optimization objectives), there is a research gap on combinatorial algorithms for multi-objective problems. Moreover, Bayesian optimization has not been applied to all types of combinatorial problems and has focused mainly on combining and permuting problems.

Table 2.3: Comparison of Bayesian optimization previous work

	Reference	Number objectives	Purpose	Comment
Continuous single-objective	[189] Mockus 1975	Single	Theoretical	Acquisition function: EI
	[191] Srinivas 2009	Single	Theoretical	Acquisition function: GP-UCB
	[193] Villemonteix 2009	Single	Theoretical	Acquisition function: IG
	[195] Henning 2012	Single	Theoretical	Acquisition function: PES
	[202] Hutter 2011	Single	Practical	Software package: SMAC
	[203] Snoek 2012	Single	Practical	Software package: Spearmint
	Bergstra 2013	Single	Practical	Software package: Hyperopt
Continuous multi-objective	[205] Knowles 2006	Multiple	Theoretical	Adapt SMS-EGO for MO
	[208] Zuluaga 2016	Multiple	Practical	Software package: e-PAL
	[209] Picheny 2013	Multiple	Theoretical	Acquistion function: SUR
	[210] Hernandez-Lobato 2016	Multiple	Theoretical Practical	Acquisiton function: PES for MO
Combinatorial	[213] Kim 2014	Single	Theoretical	Edit distance to transform combinatorial problem in continuous
	[216] Zaefferer 2014	Single	Theoretical	
	[218] Oh 2019	Single	Practical	Software package: COMBO
	[217] Baptista 2018	Single	Theoretical	Sparse Bayesian linear regression 2^{nd} order model.

2.4 Research opportunities

The following key findings summarize the insight gained from the literature review:

- Although space and aerial concepts have been proposed in the past as concepts to expand connectivity into remote regions, most of the analyses have focused on end-to-end modeling of the system, and there is little work on a) tradespace exploration, and b) economic factors. More precisely, the impact that space and aerial systems might have as a cost-effective solution (versus terrestrial systems) to connect unconnected and under-served communities has not been analyzed yet.

- Even though there have been several CSMs that are adapted to tradespace exploration, there are no methods for concept selection in situations where a large tradespace of options exists, and where there is a tight budget on the number of architectures that can be evaluated.

- Most of the methods currently used for black-box functions optimization rely on a large number of functions evaluations. This is true for both model-free methods and model-based methods.

- Bayesian optimization is the state-of-the-art technique for black-box functions optimization under limited budgets, and significant work has been conducted to apply Bayesian optimization to continuous problems, from both theoretical and practical standpoints.

Based on these key findings, the following research gaps were identified:

- There is a need for comprehensive techno-economic analyses that compare different space and aerial concepts to expand connectivity, as well as for rigorous methods to determine the most cost-efficient architectures for each of the concepts.
- There are no methods for tradespace exploration in System Architecture for situations involving large tradespaces and requiring computationally-expensive functions to be evaluated for each design configuration. This shortcoming arises because most current methods employ model-free population-based evolutionary algorithms to explore the tradespace, which requires a large number of function evaluations for convergence.
- Bayesian optimization for combinatorial problems being an incipient field of research, the initial approaches taken to adapt the traditional continuous-space method to combinatorial spaces (by using edit-distances and methods which use inherently discrete surrogate models) should be assessed for their applicability to SAPs.

2.5 Thesis Statement

Thus, given the research gap identified in this literature review, and to address the research questions posed in Section 1.3, this dissertation's problem statement is:

Thesis statement

To identify the most effective space and aerial communication concepts to complement existing infrastructure and expand global connectivity **by**:

i Developing realistic technical and financial models for each of these concepts,

ii Conducting techno-economic analyses for each of the concepts proposed,

iii Comparing the impact of these concepts in terms of expanding connectivity to uncovered and under-served regions,

using Bayesian optimization as a method to efficiently explore tradespaces when there is a tightly-constrained budget for evaluating objective functions.

Part I

Bayesian optimization for System Architecture problems

Chapter 3

Bayesian optimization fundamentals

Bayesian optimization (BO) is a set of optimization techniques used to find the extrema of black-box functions that are expensive or infeasible to evaluate repeatedly. Examples of such functions include cases where the function is computationally expensive (e.g., Monte-Carlo simulations), has an associated monetary cost (financial investments) or moral cost (drug trials, experiments with humans or animals), and where physical experimentation (destructive tests, prototyping) or human intervention (user testing, A/B testing) may be required. In these instances, Bayesian optimization is very effective in minimizing the number of function evaluations required to locate the extrema.

In particular, this Chapter explores the following two research questions:

Research Question 3.1

Can Bayesian optimization be used in the combinatorial problems common in System Architecture problems?

Research Question 3.2

How can Bayesian optimization be adapted to be applicable to the different types of System Architecture problems?

3.1 Introduction to Bayesian optimization

Historically, Bayesian optimization has been used extensively for problems with the following properties:

- The input domain is $\mathbb{R}^d$, where d, the number of dimension of the problem, is

not large, normally smaller than 15.

- The function domain is compact; that is, it contains a set of limit points and is bounded, such as $\{x \in \mathbb{R}^d : a_i \leq x_i \leq b_i\}$.
- The objective functions are continuous and present a certain degree of smoothness.
- One or more of the objective functions are "expensive or infeasible to evaluate repeatedly", either due to computational, monetary, or moral constraints.
- The objective functions are black-box functions. That is to say: when evaluating a candidate solution point x, one only observes $f(x)$ but does not have access to its derivatives, which prevents us from using numerical optimization methods such as gradient descent, or algorithms based on Newton's method.

Let us denote $f(x)$ as the particular black-box function which is to be optimized. One can try to model the unknown function $f(x)$ using a cheaper-to-evaluate model-function $\hat{f}(x)$, and then use $\hat{f}(x)$ to guide the optimization process within the search space by balancing exploration and exploitation. The main idea behind Bayesian optimization is to incorporate prior belief about the space of possible functions $f(x)$ when making the choice of $\hat{f}(x)$; even though the precise function $f(x)$ is unknown, some of its properties can be assumed (such as smoothness or continuity), making some model-functions $\hat{f}(x)$ more likely than others. As its name suggests, Bayesian optimization uses *Bayes theorem* to obtain the posterior probability of the model $\hat{f}(x)$ given a set of data observations. In this context, data observations are recorded as the set of pairs $\mathcal{D} = \{(x, f(x))\}_{j=1}^{n}$ that have been evaluated so far. Hence, Bayesian optimization computes a probability distribution for a model-function $\hat{f}(x)$ given the data $\mathcal{D}$ as

$$\Pr(\hat{f}|\mathcal{D}) \propto \Pr(\mathcal{D}|\hat{f})P(\hat{f}) \tag{3.1}$$

Importantly, Bayesian optimization does not merely consist of adjusting a surrogate model to the observed data and using it for optimization purposes, but also uses prior beliefs about the properties of the objective function (and therefore the model) to compute a posterior probability distribution over the space of surrogate model functions. Furthermore, the posterior distribution is used to determine the next evaluation point following some criterion which maximizes the utility of the evaluation. The criterion is commonly given by an *acquisition function* which explicitly captures the balance between exploration (evaluate new points in unexplored regions) and exploitation (evaluate new points in the proximity of known good solutions).

Maximizing the acquisition function (also referred to as solving the auxiliary problem) is much cheaper than evaluating the objective function, and so the computational overhead is reduced. The point resulting from the optimization of the acquisition function is evaluated using the original function $f(x)$, and is then added to the set

of observations $\mathcal{D}$, which is used to update the probability distribution of the model $\hat{f}(x)$. Algorithm 1 presents the typical Bayesian optimization procedure.

Algorithm 1 Bayesian optimization

1: $\mathcal{D} = \{\}$ ▷ Initialize observations set
2: Evaluate $f(x)$ in k points (either randomly selected or using an space-filling design) and fit initial regression model $\hat{f}(x)$ ▷ Optional step
3: **for** $t = 0$ to N **do**
4: $x^*_{t+1} = \arg\max_x \alpha(x|\mathcal{D}_t)$ ▷ Find next evaluation point x_{t+1} by optimizing the acquisition function (α) over $\hat{f}$
5: $y_{t+1} = f(x^*_{t+1})$ ▷ Evaluate the objective function at x^*_{t+1}
6: $\mathcal{D}_{t+1} = \mathcal{D}_t \cup \{(x^*_{t+1}, y_{t+1})\}$ ▷ Augment the observation set
7: $\Pr(\hat{f}|\mathcal{D}_t) \propto \Pr(\mathcal{D}_t|\hat{f})P(\hat{f})$. ▷ Update the posterior of the model $\hat{f}$
8: **end for**

Figure 3-1 illustrates several iterations of Algorithm 1, where a single-objective function $f(x)$ (depicted as a red line) is to be minimized. The black dashed line corresponds to the mean function of the model $\hat{f}(x)$, whereas the gray shaded area represents the uncertainty around the mean. The blue line is the value of the acquisition function (to be maximized). The model is initially fitted using two randomly chosen points, and a third point (shown in Fig. 3-1a) is chosen for evaluation such that it maximizes the acquisition function (marked with an inverted red triangle in the Utility graph). Once the new point is evaluated, the model is updated and using the new posterior, the next point for evaluation is chosen (Fig. 3-1b). This process is then repeated again.

Notice how the acquisition function switches from exploration of regions where the model-uncertainty is high (e.g., in iterations 0-5) to exploitation of good points (e.g., from iteration 10 onward). This trade-off is fundamental to guarantee convergence of the algorithm towards the optimal value.

Although most of the discussion up to this point has focused on single-objective problems, the Bayesian optimization framework can also be expanded to multi-objective problems. In this case, the common practice is to fit a surrogate model to each of the objective functions and use specialized acquisition functions for multi-dimensional cases. An exhaustive comparison of proposed multi-objective acquisition functions can be found in [219].

Bayesian optimization has become a hot research topic in the past five years, and the latest developments have turned it into the primary method for optimizing expensive black-box functions and hyper-parameter tuning in areas like machine learning [202, 203], algorithm configuration [220], and experimental design [221]. Section 2.3.3 contains a literature review of the latest advances in Bayesian optimization for single-objective problems, multi-objective problems, and combinatorial problems.

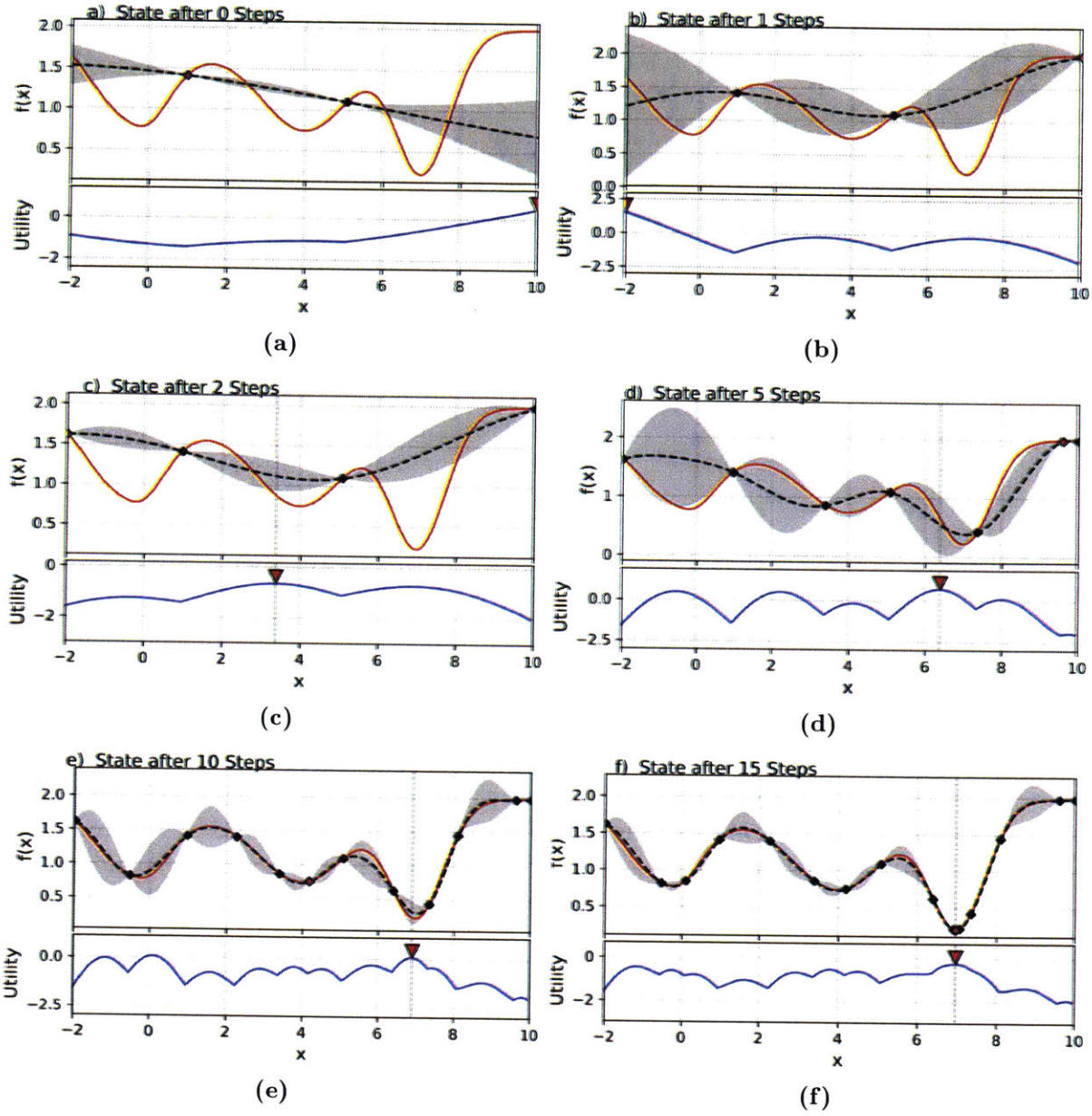


Figure 3-1: Illustration of Bayesian optimization for a one-dimensional problem. State after a) 3, b) 4, c) 6, d) 10, e) 15, and f) 20 iterations is depicted. Upper pictures show the objective and model functions, lower pictures show the acquisition function.

3.2 Mathematical formulation of Bayesian Optimization using Gaussian Processes

The previous section has introduced the three main components of Bayesian optimization: computing predictions of new values using the surrogate model, choosing the next point to evaluate using the acquisition function, and computing the poste-

rior distribution of the model once the new point has been evaluated. This section presents a mathematical formulation for each of those steps using a Gaussian process (GP) as the surrogate model. Even though other non-parametric priors can also be used, GPs are the most popular choice due to their adaptability in continuous domain problems and well-studied mathematical formulation.

3.2.1 The Gaussian process: A Prior distribution over the space of functions

A Gaussian process (GP) is a collection of random variables (a stochastic process), any finite number of which have a joint Gaussian distribution [222]. Just as a Gaussian distribution is a distribution over a random variable, defined entirely by its mean and variance, a GP $\hat{f}(x)$ describes a probability distribution over *functions*, and is defined entirely by its *mean function* ($\mu(x)$) and its *covariance function* ($k(x, x')$). A GP is denoted as:

$$\hat{f}(x) \sim \mathcal{GP}(\mu(x), k(x, x')) \tag{3.2}$$

with $\mu(x)$ and $k(x, x)$ defined as,

$$\mu(x) = \mathbb{E}[f(x)] \tag{3.3}$$

$$k(x, x') = \mathbb{E}[(f(x) - \mu(x))(f(x') - \mu(x'))] \tag{3.4}$$

Any finite collection of random variables $\{\hat{f}(x_1)...\hat{f}(x_m)\}$ drawn from the stochastic process has the following joint distribution:

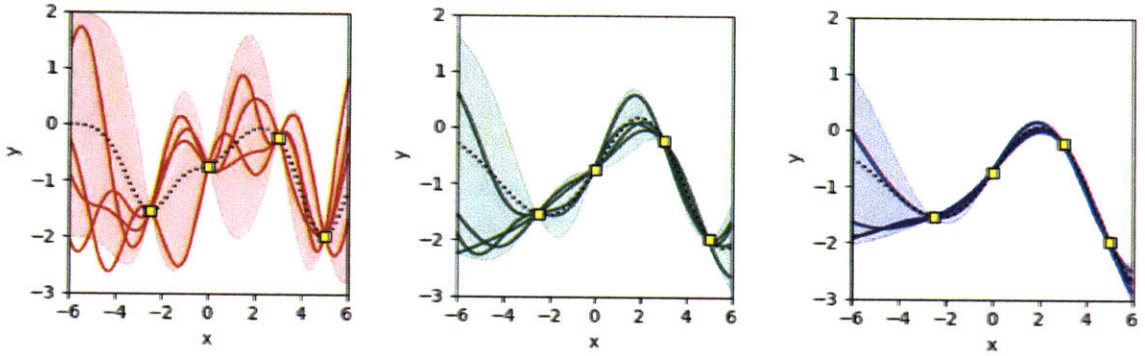
$$\begin{bmatrix} \hat{f}(x_1) \\ \vdots \\ \hat{f}(x_m) \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} \mu(\mathbf{x}) \\ \vdots \\ \mu(x_m) \end{bmatrix}, \begin{bmatrix} k(x_1, x_1) & \dots & k_(x_1, x_m) \\ \vdots & \ddots & \vdots \\ k(x_m, x_1) & \dots & k(x_m, x_m) \end{bmatrix} \right) \tag{3.5}$$

A GP can be thought of as a function that, when evaluated at a new point x, returns not a scalar but the mean and variance of a normal distribution that characterizes the probability of occurrence of the possible values of $f(\mathrm{x})$ at point x. Note that a GP also fulfills the marginalization (or consistency) property, which states that the univariate marginal distribution of one component of the Gaussian multivariate distribution is also Gaussian. Therefore, a GP is a collection of random variables jointly Gaussian, (with mean $\mu(x)$ and covariance function $k(x, x')$)), and each of those random variables is itself a Gaussian random variable.

The covariance function $k(x, x')$) (also called the kernel) gives the covariance between pairs of random variables. A popular choice for such functions is the radial basis function (RBF), which is given by:

$$k(x, x') = \exp\left(-\lambda \frac{\|\mathbf{x} - \mathbf{x}'\|^2}{2} \right). \tag{3.6}$$

with λ being the scale parameter. It can be observed that the covariance between two points that are very close is almost 1, whereas the further away the points, the closer to 0 their covariance will be. The specification of the kernel implies specifying a distribution over functions (and thus prior information regarding the function to be optimized), as the kernel defines the covariance matrix of a multivariate Gaussian random variable. When sampling from such distribution, different curves can be obtained. For example, Figure 3-2 shows the effect of the parameter λ on the functions drawn from the posterior (i.e., the RBF prior conditioned on the sample points) after fitting the GP to four sample points (yellow squares). It can be observed that the smaller the value of λ, the larger the variability in the sample functions obtained (when fitting the Gaussian process to a set of 4 sample points).



(a) $\lambda = 1$ **(b)** $\lambda = \frac{1}{2}$ **(c)** $\lambda = \frac{1}{3}$

Figure 3-2: Effect of the scale parameter λ on the sample functions obtained from a Gaussian process (GP) that uses the RBF kernel and has been fitted to four samples points (yellow squares).

3.2.2 Acquisition functions

The other important piece of the Bayesian optimization procedure is the acquisition function, $\alpha(x)$. The acquisition function guides the search for optima, balancing exploration of new points with exploitation of already known areas. To select the next point to evaluate, one has to solve the subproblem of maximizing the acquisition function. Expressed mathematically, the next point to evaluate for the expensive function $f(x)$ is $x^*_{t+1} = \arg\max_x \alpha(x|\mathcal{D}_t)$. Since this subproblem uses the surrogate model, evaluations of the acquisition function $\alpha(x)$ are cheaper compared to evaluations of the original function $f(x)$, and therefore the optimization of the acquisition function is computationally tractable.

The main idea behind the acquisition function is that it serves as the lever controlling the exploration-exploitation trade-off. As with any global optimization problem, one is presented with the following disjunctive: is it better to exploit regions of the space that are already known to render good solutions (small value for $\hat{f}(x)$), or

to explore new areas where there is a large uncertainty (large variance $\sigma(x)$ in the posterior distribution) regarding the values of the function?

To address this question, several acquisition functions have been proposed in the literature both for single and multi-objective problems. Below is a summary of these functions:

Single objective acquisition functions

- **Upper confidence bound (UCB)**: The upper confidence bound acquisition function uses an optimistic approach, establishing a fixed probability of improvement for each point (that is, an upper bound). Analytically, the value of the acquisition function is given by

$$\alpha(x;\mathcal{D}) = -\mu(x) + \beta_n \sigma(x) \tag{3.7}$$

  where the optimal value of β_n can be determined for each of the iterations of the Bayesian optimization algorithm [223, 191].

  Note that in this acquisition function the value of parameter β_n balances the importance of the mean function of the model ($\mu(x)$) and the standard deviation or uncertainty function ($\sigma(x)$) in selecting the next evaluation point. If the value of β_n is very high, the acquisition function will have a maximum in regions where the value of $\sigma(x)$ is large, thus favoring exploration of regions where there is large uncertainty. In contrast, if the value of β_n is small, the acquisition function will have a maximum where $\mu(x)$ is small, thus exploiting those regions of the input space where good solutions (assuming that we are trying to find the minimum the black-box function) have been found. Thus, when using this acquisition function there is an explicit trade-off between exploration and exploitation.

- **Probability of improvement (PI)**: The PI criteria calculates the probability that when evaluating a new point x it would render an improvement upon a threshold τ [224]. When using a Gaussian process as the model, it is computed analytically as

$$\alpha(x;\mathcal{D}) = \Pr(f(x) > \tau) = \Phi\left(\frac{\mu(x) - \tau}{\sigma(x)}\right) \tag{3.8}$$

  where Φ is the cumulative distribution of the standard normal distribution.

- **Expected improvement (EI):** The EI is similar to the probability of improvement, but it incorporates a measure of the amount of improvement [189].

For Gaussian processes it is computed as:

$$\alpha(x;\mathcal{D}) = \mathbb{E}[\max(f(x)-\tau,0)] = \\ = (\mu(x)-\tau)\Phi\left(\frac{\mu(x)-\tau}{\sigma(x)}\right) + \sigma\phi\left(\frac{\mu(x)-\tau}{\sigma(x)}\right). \tag{3.9}$$

where Φ is the cumulative distribution of the standard normal distribution, ϕ is the probability distribution function of the standard normal distribution, and τ is a parameter that determines the amount of exploration done during the optimization (the larger the value of τ the more exploration).

- **Predictive entroypy search (PES)**: Predictive entropy search is an information-theory based acquisition function which selects the point that maximizes the information-gain. Computing the entropy search analytically is prohibitive for continuous spaces, and therefore it is commonly evaluated numerically using Monte-Carlo or Thomson sampling [225]. In particular, the expression to be approximated is:

$$\alpha(x;\mathcal{D}) = H(x^*|\mathcal{D},x) - \mathbb{E}_{p(y|\mathcal{D},x)}\left[H(x^*|\mathcal{D}\cup(x,y))\right]. \tag{3.10}$$

where $H(x) = -\int p(x)\log p(x)$ is the differential entropy, x^* represent the unknown global optimum, and the expectation is taken with respect to the posterior predictive distribution of y given x.

Multiple objective acquisition functions

- **Expected hypervolume improvement (EHI) [219, 226]**: the EHI is the analogous metric to the EI criteria in one-dimensional problems. The hypervolume is a scalar measure of the space dominated (with respect to a reference point) by the Pareto front. The hypervolume of the Pareto front, with respect to a reference point $P = \{p_1, ..., p_D\} \in M$, is defined as the Lebesgue measure of the intersection of the globally-dominated set[1], and the hypervolume defined by $\{0, p_1] \times [0, p_2] \times ... \times [0, p_D]\}$, where p_i are the coordinates of a reference point, and a D-dimensional space is considered.

  The idea of the hypervolume measure is better understood if shown pictorially, as illustrated in Figure 3-3a. In the figure, the plain-gray shaded area is the hypervolume measure of the Pareto front, represented by the black points; the white dots represent dominated points, and the line pattern shaded area

[1] The dominated set of a point μ^* (D_{μ^*}) is the set of points that are dominated by that point: $D_{\mu^*} = \{\mu \in \mathcal{M} : \mu \prec \mu^*\}$. The globally-dominated set ($D_{\mathcal{M}}$), is the union of the dominated sets of all the elements of the Pareto front.

$$D_{\mathcal{M}} = \bigcup_{\mu^* \in PF(\mathcal{M})} D_{\mu^*} = \bigcup_{\mu^* \in PF(\mathcal{M})} \{\mu \in \mathcal{M} : \mu \prec \mu^*\} \tag{3.11}$$

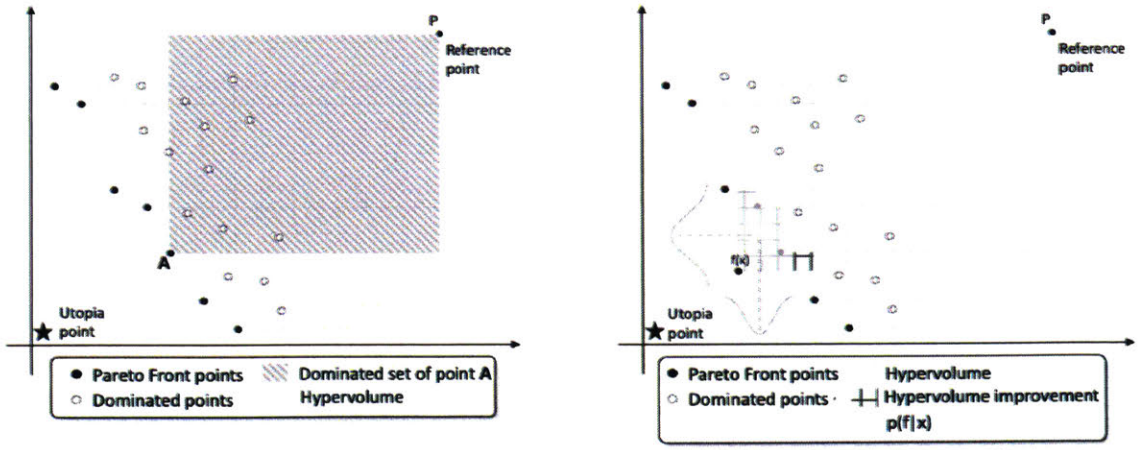
represents the *dominated set* by point A. The hypervolume measure is the hypervolume of the union of the *dominated sets* of each of the points on the Pareto front.

The EHI represents the expected improvement in the hypervolume measure, relative to the Pareto front of the points that were previously evaluated. The hypervolume improvement is simply the difference between the hypervolume measure with and without the new point, which will be > 0 if and only if the new point is not dominated by the current Pareto front. Since the surrogate model provides a probability distribution for the value of the new point, the value of the hypervolume improvement needs to be weighted according to the PDF of the new point, and therefore, one would compute its expected value as

$$\alpha_t(x;\mathcal{D}_t) = \mathbb{E}\left[\mathrm{HVI}(x)\right] = \int_{\hat{f}(x)} \Big[\mathrm{HV}(\mathrm{PF}(\mathcal{D}_t) \cup \hat{f}(x)) - \mathrm{HV}(\mathrm{PF}(\mathcal{D}_t))\Big]\, p(\hat{f}|x)df = \\ = \int_{\hat{f}(x)} \mathrm{HV}(\mathrm{PF}(\mathcal{D}_t) \cup \hat{f}(x))p(\hat{f}|x)df - \mathrm{HV}(\mathrm{PF}(\mathcal{D}_t)), \tag{3.12}$$

where $\mathcal{D}_t$ are the previously evaluated data-points, and $\mathrm{PF}(\cdot)$ and $\mathrm{HV}(\cdot)$ are functions to compute the Pareto front and hypervolume measure respectively.

This same concept is illustrated in Figure 3-3b, where the checked pattern area represents the hypervolume improvement of point $\hat{f}(x)$, and the yellow-shaded area is the probability distribution of the value of the new point.



(a) Conceptual representation of the hypervolume metric.

(b) Conceptual representation of the expected hypervolume improvement.

Figure 3-3: Conceptual representations of the a) hypervolume measure and b) expected hypervolume improvement.

- **Predictive entropy search (PES)**: PES can be applied to multi-objective problems by maximizing the expected reduction in the entropy of the posterior

distribution over the Pareto set ($\mathcal{X}^*$) [210]. The formulation is similar to Eq. 3.10:

$$\alpha_n(x) = H\left(\mathcal{X}^* | D_n\right) - \mathbb{E}\left[H\left[\mathcal{X}^*(D_n \cup \{(x,y)\})\right]\right] \tag{3.13}$$

Since this expression is impractical to evaluate exactly, its value can be approximated using expectation propagation [210].

- **Sequential uncertainty reduction (SUR) [209]**: SUR chooses as the next evaluation point the one which maximizes the reduction of the volume of the excursion sets below the current Pareto set. This excursion set below the current Pareto set measures performance with regard to the optimization problem; the volume of the set is large if the optimum is not well located, and small if little is to be gained from extra evaluations). The SUR metric is computed via numerical integration of a tessellation of the space using hyper-rectangular cells, and is best suited for problems with 2 or 3 output variables [209].

3.2.3 Posterior distribution computation

Throughout the optimization process, we are interested in deriving the posterior distribution of a new observation. This involves using existing observations of the black-box function to fit a GP model and then use the model to compute the posterior of new points. With a given set of n pairs of observations $\mathcal{D} = (x, f(x))^n$, the objective is to predict the outcome of the black-box function at a new point x_{n+1}. Then, the prediction ($\hat{f}_{n+1}$) would be jointly Gaussian with the points already evaluated ($\hat{\mathbf{f}}$) in $\mathcal{D}$, and their joint distribution according to the prior is given by:

$$\begin{bmatrix} \hat{\mathbf{f}} \\ \hat{f}_{n+1} \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} \mu(\mathbf{x}) \\ \mu(x_{n+1}) \end{bmatrix}, \begin{bmatrix} \mathbf{K} & \mathbf{k} \\ \mathbf{k}^T & k(x_{n+1}, x_{n+1}) \end{bmatrix}\right) \tag{3.14}$$

where $\mathbf{K}$ is the $n \times n$ covariance matrix, and $\mathbf{k}$ is the $1 \times n$ vector of covariance values between x_{n+1} and $\mathbf{x}$, as defined below:

$$\mathbf{x} = \begin{bmatrix} x_1 & x_2 & \dots & x_n \end{bmatrix} \tag{3.15}$$

$$\hat{\mathbf{f}} = \begin{bmatrix} \hat{f}(x_1) & \hat{f}(x_2) & \dots & \hat{f}(x_n)) \end{bmatrix} \tag{3.16}$$

$$\mathbf{k} = \begin{bmatrix} k(x_{n+1}, x_1) & k(x_{n+1}, x_2) & \dots & k(x_{n+1}, x_n) \end{bmatrix} \tag{3.17}$$

$$\mathbf{K} = \begin{bmatrix} k(x_1, x_1) & \dots & k(x_1, x_n) \\ \vdots & \ddots & \vdots \\ k(x_n, x_1) & \dots & k(x_n, x_n) \end{bmatrix} \tag{3.18}$$

Matrix $\mathbf{K}$ must be semi-definite positive (and thus invertible) to be a valid covariance

matrix corresponding to a multivariate Gaussian distribution. This is guaranteed as long as the covariance function $k(x, x')$, normally denoted as the *kernel*, fulfills Mercer's theorem [227].

The posterior over the space of functions for $\hat{f}_{n+1}$ is obtained by conditioning the joint Gaussian prior distribution on the observations, which results in:

$$\hat{f}(x_{n+1}) = \mu(x_{n+1}) + \mathbf{k}^T\mathbf{K}^{-1}(\mathbf{f} - \mu(\mathbf{x})) \tag{3.19}$$

whereas the variance of the predictive distribution is given by:

$$\sigma^2_{n+1} = k(x_{n+1}, x_{n+1}) - \mathbf{k}^T\mathbf{K}^{-1}\mathbf{k}. \tag{3.20}$$

Thus, one can compute the conditional distribution of a new point x_* by substituting x_{n+1} with x_* and evaluating Eq. 3.19 and Eq. 3.20. However, making predictions using the Gaussian process is expensive, as it involves inverting the covariance matrix $\mathbf{K}$, an $\mathcal{O}(n^3)$ operation, where n is the number of observations. To reduce the computational burden, given that $\mathbf{K}$ is a semi-definite matrix, one can compute its Cholesky factorization (as denoted in Eq. 3.21) once, store it in memory, and then use it to predict new values, which lowers the cost to $\mathcal{O}(n^2)$.

$$\mathbf{L} = \text{cholesky}(\mathbf{K}) \tag{3.21}$$

Once L is obtained, the posterior values of x_* can be computed as follows. Let $\mathbf{k}_* = \begin{bmatrix} k(x_*, x_1) & k(x_*, x_2) & \dots & k(x_*, x_n) \end{bmatrix}$, then the auxiliary variables α and $\mathbf{v}$ are given by:

$$\alpha = \mathbf{L}^T\backslash(\mathbf{L}\backslash y), \tag{3.22}$$

$$\mathbf{v} = \mathbf{L}\backslash\mathbf{k}_*, \tag{3.23}$$

and can then be used to compute the value of the posterior distribution on a point x_* as

$$\hat{f}(x_*) = \mathbf{k}_*^T\alpha \tag{3.24}$$

$$\sigma(x_*) = k(x_*, x_*) - \mathbf{v}^T\mathbf{v} \tag{3.25}$$

Compared to the direct inversion of the covariance matrix, this approach is not only faster but also more stable numerically.

3.3 Types of decision patterns in System Architecture problems

Before discussing how to adapt the BO formulation presented above to SAPs, it is important to understand the nature of the problems faced when conducting System Architecture tasks.

Most of the System Architecture frameworks developed in the last years revolve around the decision-making process during system conception. Many of the tasks that a system architect needs to conduct would benefit from being formulated as a decision-making problem in a mathematically rigorous manner, as this enables the designer to: a) better understand the problem at hand, since all the tradespace options can be enumerated systematically, and b) conduct optimization over the tradespace of options efficiently by using well-established methods and tools, thereby saving time and resources. Instances of such tasks include allocating goals to solution-neutral functions, decomposing or aggregating elements of form, mapping elements of function to elements of form, specializing elements and form, or connecting function to form [38].

With this goal in mind, Selva introduced in 2012 a set of 6 different patterns which can be employed to gain insight into the decision-problem. These patterns are the combining pattern, the assigning pattern, the downselecting pattern, the partitioning pattern, the permuting pattern, and the connecting pattern. The rest of this section is devoted to describing each of the patterns, assuming (without loss of generality) that there is a single-objective function to be minimized.

Combining Pattern

The **combining pattern** can be described as follows: given a set of decisions $\mathcal{D} = \{d_1, d_2, ..., d_N\}$ each with a set of possible options $\mathcal{O}_i = \{o_{i1}, o_{i2}, ..., o_{ik_i}\}$, select one option for each decision such that the objective function is minimized. Mathematically, it is formulated as:

$$x^* = \arg\min_{\mathbf{x}} f(\mathbf{x}), \tag{3.26}$$
$$\mathbf{x} = \begin{bmatrix} x_1 & x_2 & ... & x_N \end{bmatrix}, \; x_i \in \mathcal{O}_i$$

where $\mathbf{x}$ is an assignment of options to decisions (an *architecture*), and $f(\mathbf{x})$ is the objective function. Figure 3-4 contains a visual representation of the combining pattern, as well as three example architectures and a possible encoding of these. The encoding is performed using a vector of N positions where the number on the i-th position corresponds to the option chosen for the i-th decision.

The combining pattern is the most popular pattern for SAPs, appearing in the

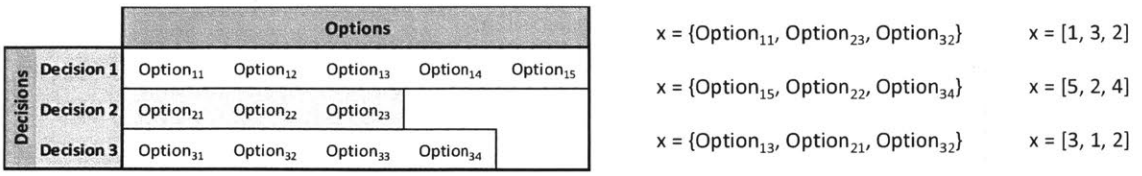


Figure 3-4: Visual representation of the combining pattern (left), three example architectures (center), and a possible encoding of the architectures (right).

"mapping elements of function to elements of form" and "specializing elements of form" tasks, as well as on experimental design, decision trees, and morphological matrices.

Assigning Pattern

The **assigning pattern** can be described as follows: given two sets of entities $\mathcal{E} = \{e_1, e_2, ..., e_N\}$ and $\mathcal{E}' = \{e'_1, e'_2, ..., e'_M\}$, assign entities from set $\mathcal{E}$ to entities in set $\mathcal{E}'$ such that the objective function is minimized. Note that an entity in $\mathcal{E}$ can be assigned to none or multiple entities in $\mathcal{E}'$, and that an entity in $\mathcal{E}'$ can have none or multiple entities from $\mathcal{E}$ assigned. An alternative formulation of the problem is to assign a subset of $\mathcal{E}$ to each entity in $\mathcal{E}'$. Mathematically, it is formulated as

$$x^* = \arg\min_{\mathbf{x}} f(\mathbf{x}), \tag{3.27}$$
$$\mathbf{x} = \begin{bmatrix} S_1 & S_2 & ... & S_M \end{bmatrix}, \; S_i \subseteq \mathcal{E},$$

where $\mathbf{x}$ is an assignment of entities in $\mathcal{E}$ to entities in $\mathcal{E}'$, and $f(\mathbf{x})$ is the objective function. Figure 3-5 contains a visual representation of the assigning pattern, as well as two example architectures and a possible encoding of these. The encoding is done using a vector of M positions (dimension of $\mathcal{E}'$), where the element in the i-th position is a set (represented using a variable size vector) containing the index of the elements from $\mathcal{E}$ assigned to e'_i.

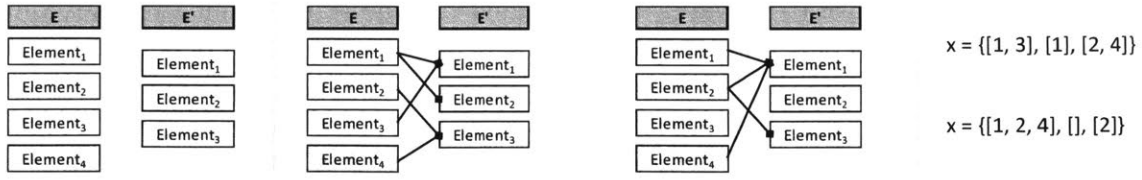


Figure 3-5: Visual representation of the assigning pattern (left), two example architectures (center), and the encoding of the architectures (right).

Downselecting Pattern

The **downselecting pattern** can be described as follows: given a set of elements $\mathcal{E} = \{e_1, e_2, ..., e_N\}$, select a subset of elements from $\mathcal{E}$ such that the objective function

is minimized. Mathematically, it is formulated as

$$x^* = \arg\min_{\mathbf{x}} f(\mathbf{x}), \tag{3.28}$$
$$\mathbf{x} = \begin{bmatrix} x_1 & x_2 & \dots & x_N \end{bmatrix}, \; x_i \in [0, 1],$$

where $\mathbf{x}$, is a vector encoding the elements that have been downselected (1 indicates selection, 0 not selection), and $f(\mathbf{x})$ is the objective function. Figure 3-6 contains a visual representation of the downselecting pattern, as well as three example architectures and a possible encoding of these. The encoding is performed using a binary vector of N positions, where a 1 in the i-th position denotes that element e_i has been selected.

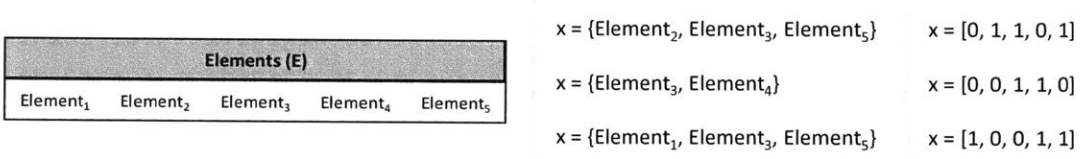


Figure 3-6: Visual representation of the downselecting pattern (left), three example architectures (center), and a possible encoding of the architectures (right).

Downselecting patterns are common when one has to choose a subset of objectives or assets, such as when defining the scope of the system, or when picking a set of ground stations from a pool of candidate locations. Among combinatorial optimization problems, the 0-1 knapsack problem is best represented using the downselecting pattern.

Partitioning Pattern

The **partitioning pattern** can be described as follows: given a set of elements $\mathcal{E} = \{e_1, e_2, ..., e_N\}$, divide the elements in $\mathcal{E}$ into mutually-exclusive, collectively-exhaustive (MECE) subsets such that the objective function is minimized. In other words, for an architecture to be valid, each element of $\mathcal{E}$ must be in one and only one subset, each subset must have at least one element, and the union of all subsets must equal $\mathcal{E}$. Mathematically it is formulated as

$$x^* = \arg\min_{\mathcal{S}} f(\mathcal{S}), \tag{3.29}$$
$$\mathcal{S} = \begin{bmatrix} S_1 & S_2 & \dots & S_m \end{bmatrix}, \; S_i \subseteq \mathcal{E}, \; S_i \neq \{\emptyset\}$$
$$s.t.$$
$$\bigcup_{i=1}^{m} S_i = \mathcal{E} \tag{3.30}$$
$$\bigcap_{i=1}^{m} S_i = \emptyset,$$

where $\mathcal{S}$ is the collection of MECE subsets (a partition of $\mathcal{E}$), and $f(\mathcal{S})$ is the objective function. Figure 3-7 contains a visual representation of the partitioning pattern, as well as three example architectures and a possible encoding of these. The encoding is done using a variable-length vector, where each element is a subset (also represented by a different variable size vector) containing the index of the elements from $\mathcal{E}$ assigned to a particular subset.

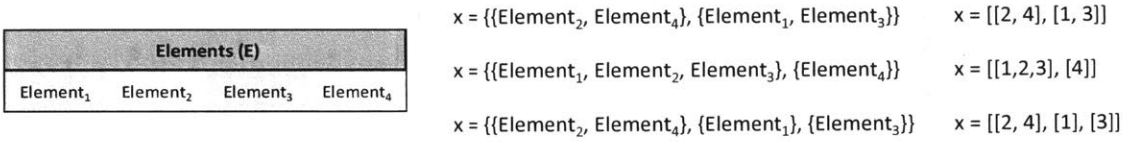


Figure 3-7: Visual representation of the partitioning pattern (left), three example architectures (center), and a possible encoding of the architectures (right).

Permuting Pattern

The **permuting pattern** can be described as follows: given a set of elements $\mathcal{E} = \{e_1, e_2, ..., e_N\}$ and n positions, assign each of the elements in $\mathcal{E}$ to one position such that the objective function is minimized. Note that this is equivalent to finding the sequence (or ordering) of elements in $\mathcal{E}$ that minimizes the objective function. Mathematically, it is formulated as

$$x^* = \arg\min_{\mathbf{x}\in\Pi} f(\mathbf{x}), \tag{3.31}$$

where $\mathbf{x}$ is a permutation of the elements in $\mathcal{E}$, Π is the set of all permutations, and $f(\mathbf{x})$ is the objective function. Figure 3-8 contains a visual representation of the permuting pattern, as well as three example architectures and a possible encoding of these. The encoding is performed using a vector of N positions where the number i-th position corresponds to the index (j) of the element (e_j) that occupies that position.

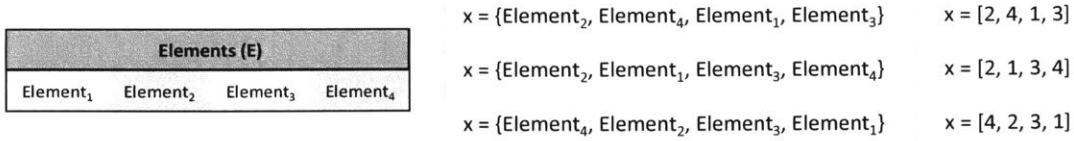


Figure 3-8: Visual representation of the permuting pattern (left), three example architectures (center), and a possible encoding of the architectures (right).

Permuting patterns are common in SAPs that involve the determination of a sequence of actions, such as in scheduling or DSM functional sequencing. Moreover, several classical combinatorial optimization problems are best represented by this pattern; these include the traveling salesman problem (TSP), quadratic assignment problem (QAP), linear ordering problem, and flow shop problem (FSP).

Connecting Pattern

The **connecting pattern** can be described as follows: given a set of nodes $\mathcal{N} = \{n_1, n_2, ..., n_N\}$, determine the connections among nodes such that the objective function is minimized. In other words, given the set of nodes, assign a set of edges $e = (n_i, n_j),\ i \neq j$ that minimizes the objective function. Since the connecting pattern is rather infrequent in SAPs, and when it does occur, the number of potential edges for the graph tends to be very large, these types of problems are not well suited for Bayesian optimization (which requires low input dimensionality), and therefore will not be considered further in this dissertation.

3.4 Adapting Bayesian Optimization to System Architecture problems

Section 3.1 introduced the fundamentals of Bayesian optimization and the conditions under which BO works best. Among these conditions were the continuity of the function domain and the function image; however, these rarely occur in System Architecture problems (SAP). Specifically, the following issues are commonly present in SAPs:

- Most of the decision variables (input space) are categorical. This makes it hard to use traditional regression methods (such as Gaussian processes) in which the variables are assumed to be continuous.
- Moreover, the discrete decision variables often correspond to decisions involving combinatorial problems, which have their own characteristic structure (i.e., a set of permutations, a set of partitions).
- It is not uncommon for certain combinations of decision-values to render infeasible architectures (in that they do not satisfy the system requirements), which might not be known until after the expensive function has been evaluated. Hence, even though the function domain is a countable set, it might be impossible to define valid architectures pre-evaluation.

While the first two issues can be overcome by adapting the BO formulation to combinatorial problems (as described below), the third is more problematic. If the problem at hand is such that invalid architectures occur frequently, then BO is in general a bad option, since there is no mechanism to introduce information regarding infeasible architectures into the model.

As explained in Section 3.2.1, the most common approach for Bayesian optimization is to use a Gaussian Process with a radial-basis function (RBF) covariance function (kernel) which uses the Euclidean distance between points. The RBF kernel is

defined as

$$k(x, x') = \exp\left(-\lambda \frac{\|\mathbf{x} - \mathbf{x}'\|^2}{2}\right). \tag{3.32}$$

Thus, taking this into account, there are three main strategies to adapt Bayesian optimization to combinatorial problems:

1. **Use a different model**, better tailored for SAPs than the Gaussian process. Rather than using a continuous surrogate model, one can use models better suited for combinatorial problems, as long as they provide a measure of uncertainty in the prediction. Examples of such models are random forest or tree-structured Parzen estimators.

2. **Use a different covariance function**, well suited for combinatorial problems. For some decisions patterns (e.g., combining or permuting patterns), it is possible to use tailored covariance functions.

3. **Use a different distance function** that adapts to the combinatorial space. In this approach, the combinatorial nature of the problem is abstracted by defining a distance function between points of the input space. This distance function is specific for each type of decision pattern, and is later used in-place of the Euclidean distance in a RBF kernel, which becomes $k(x, x') = \exp\left(-\lambda \frac{d(x,x')^2}{2}\right)$.

Combinations of the above strategies are also possible. For example, different kernels can be used with different distance functions.

The rest of this section details these three alternative approaches, which will be explored when assessing the suitability of BO as an optimization method for SAPs involving expensive black-box function evaluations.

3.4.1 Using other models better suited for System Architecture problems

Bayesian optimization can be used with other types of probabilistic models. As long as a model predicts both the expected value and the variance of any new point, it can be adapted for use in the Bayesian optimization framework. Examples of such models that have been studied in the literature include random forests (using the mean and variance of the predictions by each of the decision trees in the model), and tree-structured Parzen density estimator (TPE).

Random forests are an ensemble learning method for regression which use a multitude of decision trees. Each decision tree is trained individually using a different

set of training data-points, usually chosen by repeated sampling of the data (with replacement). In general, decision trees that are grown very deep tend to overfit to the model, thus producing low-bias but high-variance estimates [228]. Random forests reduce the variance by averaging the predictions of multiple independent trees. In particular, assuming that T trees are used in the random forest algorithm, the predicted value and variance for a particular point x' are:

$$\mu(x') = \frac{1}{T}\sum_{t=1}^{B} f_t(x') \tag{3.33}$$

$$\sigma(x') = \sqrt{\frac{\sum_{t=1}^{T}(f_t(x') - \mu(x'))^2}{T-1}} \tag{3.34}$$

where $f_t(x)$ denotes each of the trees.

A tree-structured Parzen density estimator is a model well suited for tree-structured input spaces (where the valid options of some decisions depend on the values of other prior-decisions). The model splits the observations into two groups: the first group is comprised of the best performing γ observations (where γ is a fraction of the total number of observations), and the second group contains the other observations. Then, the likelihood probabilities of a new point coming from each group (i.e., $p(x \in \text{group } 1|y)$, $p(x \in \text{group } 2|y)$) is modeled using two different Parzen estimators. Then, the value of $p(y|x)$ can be computed using Bayes rule since $p(y) = \gamma$

When these models are employed, the mathematical expression for the posterior distribution and the acquisition function, must (in most cases) be adapted accordingly to the characteristics of the model. If interested, the reader can find the formulations for the cases of random forests and TPE in references [202] and [204].

Besides, one can derive alternative models using ensemble methods, which are composed of multiple regression models trained with different subsets of the training data (similar to cross-validation). However, this approach has some disadvantages, as the estimated uncertainty can be non-zero at the already evaluated samples (thus introducing a non-existent noise at those points) and longer training times are typically required.

3.4.2 Tailoring kernels to System Architecture problems

The kernel of a Gaussian process refers to the functions used as the covariance functions. For $k(x, x')$ to be a valid covariance function, it must be positive semi-definite, as stated in Mercer's theorem. Multiple stationary kernels have been proposed for continuous domains, the most common being the squared exponential kernel (or RBF), the γ-exponential kernel, the rational quadratic kernel, and the family of Matérn kernels.

In this section, I will focus on the definition of kernels that are suited to combinatorial problems similar to those present in SAPs, as presented in Section 3.3.

Permutation Kernels

Given two permutations σ, and σ', let us define $n_d(\sigma, \sigma')$ and $n_c(\sigma, \sigma')$ as the number of discordant and concordant pairs respectively between the two permutations:

$$n_d(\sigma, \sigma') = \sum_{i<j} \left(\mathbb{1}_{\sigma(i)<\sigma(j)} \mathbb{1}_{\sigma'(i)>\sigma'(j)} + \mathbb{1}_{\sigma(i)>\sigma(j)} \mathbb{1}_{\sigma'(i)<\sigma'(j)} \right) \tag{3.35}$$

$$n_c(\sigma, \sigma') = \binom{n}{2} - n_d(\sigma, \sigma') = \sum_{i<j} \left(\mathbb{1}_{\sigma(i)<\sigma(j)} \mathbb{1}_{\sigma'(i)<\sigma'(j)} + \mathbb{1}_{\sigma(i)>\sigma(j)} \mathbb{1}_{\sigma'(i)>\sigma'(j)} \right) \tag{3.36}$$

These two quantities can be further used to define two kernels that measure the similarity between σ and σ':

1. The **Mallows kernel** is defined for $\lambda \geq 0$ as

$$k_M^{\lambda}(\sigma, \sigma') = e^{-\lambda n_d(\sigma, \sigma')} \tag{3.37}$$

2. The **Kendall kernel** is defined as

$$k_\tau(\sigma, \sigma') = \frac{n_c(\sigma, \sigma') - n_d(\sigma, \sigma')}{\binom{n}{2}} \tag{3.38}$$

3.4.3 Tailoring distance functions to System Architecture problems

This approach exploits the mathematical encoding for each of the decision patterns developed by Selva to compute distances between configurations in the input space, thus transforming the combinatorial domain into a continuous domain. These distances can then be used together with distance-based kernels in the traditional Bayesian optimization framework.

As an example, a very commonly used kernel is the RBF kernel, defined as:

$$k(\mathbf{x}, \mathbf{x}') = \exp\left(-\lambda \frac{\|\mathbf{x} - \mathbf{x}'\|^p}{2} \right) = \exp\left(-\lambda \frac{d_E(\mathbf{x}, \mathbf{x})^p}{2} \right), \tag{3.39}$$

with p and λ being two parameters of the model. In the traditional formulation, the

Euclidean distance ($||\mathbf{x} - \mathbf{x}'||$) between points is used; in this approach, one would use the same kernel but with a new distance function tailored to each specific SAP.

$$k(\mathbf{x}, \mathbf{x}') = \exp\left(-\frac{d_{\text{SAP}}(\mathbf{x}, \mathbf{x})^p}{2\sigma^2}\right) \tag{3.40}$$

Combining and Downselecting distance functions

A natural choice for a distance function for the combining and downselecting problems is the Hamming distance. Given two vectors (σ, σ') containing the same number of elements (n), the Hamming distance is equal to the number of elements that are different:

$$d_H(\sigma, \sigma') = \sum_{i=0}^{n} \mathcal{I}(\sigma_i, \sigma_i') \qquad \text{where} \qquad \mathcal{I}(x, y) = \begin{cases} 0 & \text{if} \quad x = y \\ 1 & \text{if} \quad x \neq y \end{cases} \tag{3.41}$$

In addition, the weighted Hamming distance ($d_{HW}(\sigma, \sigma')$) is obtained by multiplying each of the terms in Eq. 3.41 by a coefficient w_i, as shown in Eq. 3.42. These coefficients are useful in situations where the decisions (or elements) have different magnitude contributions to the objective function. When this distance function is used, the kernel is called an anisotropic radial-basis function (ARBF), and the values of the weights are adjusted during the model fitting process.

$$d_{HW}(\sigma, \sigma') = \sum_{i=0}^{n} w_i \mathcal{I}(\sigma_i, \sigma_i') \tag{3.42}$$

Partitioning distance functions

Similar to the combining and downselecting cases, a kernel for the partitioning problem can be constructed using an alternative distance function in the RBF kernel.

An appropriate distance measure for the partitioning pattern is the Damerau–Levenshtein distance ($d_L(\sigma, \sigma')$), which, given two partitions σ and σ', measures the minimum number of operations that would be needed to transform σ in σ'. An operation is defined as moving an element from a subset to another; for example, given the partitions $\sigma = \{\{e_1, e_2, e_3\}, \{e_4, e_5\}, \{e_6\}\}$ and $\sigma' = \{\{e_1, e_3\}, \{e_2, e_5\}, \{e_4\}, \{e_6\}\}$, the Damerau–Levenshtein distance equals 2, as σ can become σ' by moving e_2 to the second subset, and then moving e_4 into a new subset.

Permuting distance functions

The permuting pattern also allows for the Hamming distance, which in this case accounts for the number of elements in different positions of the permutation π. In addition, as already introduced in Section 3.4.2, the number of discordant pairs can be used as a distance function.

3.4.4 Numerical conditioning of the covariance matrix

Using the Gaussian process to predict the value of new points requires the inversion of the covariance matrix. For this inversion to be possible, the determinant of the covariance matrix must be non-zero. Moreover, if the covariance matrix is positive definite, its inverse can be computed efficiently using the Cholesky decomposition, as detailed in Section 3.2.1. However, due to the substitution of the Euclidean distance with a custom distance function, the covariance function might lose its positive semi-definiteness, and the covariance Matrix cannot be inverted. Moreover, the numerical conditioning of the covariance matrix might cause errors in the Cholesky decomposition due to machine precision (i.e., when some of the eigenvalues of the covariance matrix are close to zero). This often happens when two input points used to compute an element of the covariance matrix are very close to each other in the input-space.

To avoid this, a multiple of the identity matrix is added to the covariance matrix, using the procedure described in Algorithm 2, such that it becomes positive definite. First, parameters β and τ_0 are initialized to values equal to the Frobenius norm and the minimum (positive) element of the diagonal of the covariance matrix, respectively. If no positive element in the diagonal can be found, τ_0 is simply initialized to $\beta/2$. Then, the algorithm tries to compute the Cholesky decomposition of $K + \tau_k I$. If successful, the algorithm terminates; else, the value of τ_k is updated to be the maximum between $2\tau_k$ and $\beta/2$.

Due to its simplicity, this strategy is preferable to other methods in which advanced modified factorization of the covariance matrix is required. However, there are two disadvantages inherent to this method: first, the Cholesky decomposition of the matrix needs to be computed multiple times, which makes the algorithm slow; second, the final value of τ might be quite large, and thus, the resulting matrix could be very different from the original matrix.

The final value of τ can be interpreted as a random noise contribution to the modified covariance matrix $K' = K + \tau I$. If this noise is very large, it might overpower the information contained in K. However, this has not been an issue in the experiments in Chapters 4 and 5, where there has seldom been need for numerical conditioning of the covariance matrix. When the procedure described in Algorithm 2 *was* run, the values of τ have been small relative to the rest of the eigenvalues in K.

Algorithm 2 Cholesky with multiple of identity

$\beta \leftarrow ||K||_F$
if $\min_i k_{ii} > 0$ **then** ▷ Assign initial value for τ_0.
    $\tau_0 \leftarrow 0$
else
    $\tau_0 \leftarrow \beta/2$
end if

for $k = 0, 1, 2, ...$ **do** ▷ Iterate until a successful factorization is possible.

    Try to compute the Cholesky factorization (L) of $K + \tau_k I$
    $$LL^T = K + \tau_k I$$
    if factorization was successfully **then**
        return L, τ_k
    else
        $\tau_{k+1} \leftarrow \max(2\tau_k, \beta/2)$
    end if
end for

3.5 Chapter conclusions

This chapter introduced the concept of Bayesian optimization and studied how one may adapt the traditional formulation, appropriate for continuous domain problems, to the most common combinatorial problems that practitioners face in the field of System Architecture. The main conclusions of this chapter are:

What is Bayesian optimization? Bayesian optimization is an effective way of exploring the tradespace when one is limited by the number of function evaluations that can be carried out. The approach uses a model (which is cheaper to evaluate) together with an acquisition function to guide the optimization process. The Bayesian optimization algorithm has four steps: evaluate a new point using the expensive function, update the data-set of points evaluated, fit (or update) a model using the data-set of points evaluated, then find the next point to evaluate by maximizing the acquisition function obtained from the model.

Research Question 3.1

Can Bayesian optimization be used in the combinatorial problems common in System Architecture problems?

Even though most of the research on Bayesian optimization addresses continuous domains, it is possible to adapt the formulation and algorithms to combinatorial domains as long as the dimensionality of the problem is small ($d < 20$) and infea-

sible architectures can be detected without the need to evaluate them explicitly (or alternatively, if valid architectures can be systematically enumerated).

Research Question 3.2

How can Bayesian optimization be adapted to be applicable to the different types of System Architecture problems?

Bayesian optimization can be adapted to combinatorial problems by using different models (e.g., random forest, tree Parzen estimator), different kernels (e.g., Matérn, Kendall, Mallows), and different distance functions (e.g., Hamming distance, Damerau-Levenshtein distance). The appropriate technique(s) would depend on the type of combinatorial pattern involved. When using these techniques to adapt Bayesian optimization, the covariance matrix might become ill-conditioned or lose its positive-semi-definiteness. This challenge can be overcome by the use of numerical re-conditioning methods on the covariance matrix.

Chapter 4

Bayesian Optimization for Single-Objective SAPs

The adapted BO procedure from Chapter 3 is first validated against scenarios involving single-objective SAPs. Although this is not the norm, given that SAPs typically present trade-offs among multiple objectives, evaluating the proposed method as applied to single-objective problems will help us understand its strengths and limitations before considering expanding its applications to multi-objective problems.

Hence, the core research questions of this chapter are:

Research Question 4.1

Is Bayesian optimization a suitable method to solve one-dimensional System Architecture problems?

Research Question 4.2

Which of the methods to adapt Bayesian optimization to combinatorial problems is most effective for each of the decision patterns common in SAPs?

Research Question 4.3

What is the improvement achieved by using Bayesian optimization, compared to other methods currently employed in the literature?

4.1 Methodology overview

In this chapter, a set of artificial test problems are used to demonstrate that BO can indeed offer improvements, vis-a-vis model-free optimization algorithms and other methods in the literature (when applied to combinatorial patterns). Specifically, the following methodology is used:

1. Define a library of test problems for five of the six decision patterns described in Section 3.3. The connecting pattern is not analyzed for two reasons: first, it is a rather uncommon decision pattern in SAP; second, it is normally a very high-dimensional problem and thus not well suited to Bayesian optimization methods.

2. Define the set of optimization methods to study. Some of these methods are Bayesian optimization methods (i.e., the three approaches described in Section 3.4), whereas others are existing optimization methods commonly used in the literature.

3. Run each optimization method (when applicable) on each test problem, imposing a limited budget of N_f function evaluations (where N_f depends on the size of the design space of the function being optimized). To obtain statistically significant results, 20 runs with different random seeds are computed for each method.

4. Finally, the these results are analyzed to rank the different optimization methods and conduct time-savings studies.

Note that this methodology results in a large number of optimization runs since there are 5 patterns, each with 8 - 12 test problems, 6 - 13 optimization methods being considered, and 20 runs conducted per combination (resulting in more than 10,000 individual optimization runs). The rest of this chapter discusses each of the points above in detail.

4.2 Library of Test Problems for System Architecture problems

The performance of the Bayesian optimization framework is compared against other optimization methods using a library of test problems for each decision pattern common in SAPs. The test problems selected for the library include both real problems from the literature as well as synthetically generated functions where the input-dimension and the degree of interaction between inputs can be tuned. In total, 12

different classes of problems are considered, and 65 different single-objective test problems are included in the library. The same objective holds across all test problems, which is to minimize a given function, as described further below.

4.2.1 Assigning Pattern

Instances of the generalized assignment problem (GAP) are used to test the effectiveness of Bayesian optimization when applied to assigning patterns. In the GAP, there are N jobs to be assigned to M agents [229]. The i-th agent incurs a cost $c(j, i)$ when doing the j-th job, and the objective is to minimize the total cost incurred by all agents:

$$f(x) = \sum_{i}^{M} \sum_{j}^{N} c(j, i) \tag{4.1}$$

Notice that each job can be assigned to a unique agent, but there is no limit to the number of jobs that can be performed by each agent.

In addition, a modified *NK problem* (which will be described in depth in Section 4.2.3) where each element in the solution can take L different values, is also used as a test problem for the assigning pattern. Briefly, the problem statement gives "given L elements in set $\mathcal{E}$, assign them to N slots in set $\mathcal{E}'$ (allowing repetition and empty assignments)". Table 4.1 contains a summary of the test instances used for the assigning pattern.

Table 4.1: Test instances used for the assigning pattern. d is the number of dimensions, l is the number of levels, S is the size of the input space, and O is the optimal value. Values with the '*' symbol have been computed using a genetic algorithm with 100,000 function evaluations.

Assigning pattern					
Problem	**d**	**l**	**S**	***O***	**Comments**
GAP('gap515-1')	15	5	$3 \cdot 10^{10}$	240	General assignment problem
GAP('gap515-4')	15	5	$3 \cdot 10^{10}$	251	General assignment problem
GAP('gap520-2')	20	5	$9.5 \cdot 10^{13}$	248	General assignment problem
GAP('gap824-3')	24	8	$4 \cdot 10^{21}$	372	General assignment problem
NK(N=15, K=1, L=4)	15	4	$1.1 \cdot 10^{9}$	0.14*	No interactions
NK(N=15, K=2, L=4)	15	4	$1.1 \cdot 10^{9}$	0.14*	Second order interactions
NK(N=15, K=4, L=4)	15	4	$1.1 \cdot 10^{9}$	0.13*	Fourth order interactions
NK(N=10, K=1, L=5)	10	5	$1 \cdot 10^{6}$	0.10*	No interactions
NK(N=10, K=2, L=5)	10	5	$1 \cdot 10^{6}$	0.11*	Second order interactions
NK(N=10, K=5, L=5)	10	5	$1 \cdot 10^{6}$	0.13*	Fifith order interactions

4.2.2 Combining Pattern

Four "real-life" test problems are used in the combining pattern.

- The first data set, SNW, is taken from Zuluaga [230]. The design space consists of 206 different hardware implementations of a sorting network, where each design is characterized by d = 3 parameters. The objective functions are the area and throughput of the network when synthesized for a field-programmable gate array (FPGA) platform. This synthesis is very costly and can take up to many hours.

- The second data set, NoC, is taken from Almer [21]. The design space consists of 259 different implementations of a tree-based network-on-chip to be implemented in application specific integrated circuits (ASICs) and multi-processor system-on-chip designs. Each design is defined by d = 4 parameters, and the objective functions are the energy consumption and runtime for the synthesized designs.

- The third data set, SW-LLVM, is taken from Siegmund [19]. The design space consists of 767 different compiler settings for the LLVM compiler framework, where each setting is specified by d = 11 binary flags. The objective functions are the performance and memory footprint for a given suite of software programs when compiled with these settings.

- The fourth (and final) data-set is the BO4CO data-set [231], which contains throughput and latency measurements for 3 different Big data stream benchmark applications under different configurations. Each application has between 3-6 decisions with 3-18 possible values for each.

Table 4.2 contains a summary of the dimensions, size of the input space, and optimal values for each of the test problems considered for the combining pattern. Since only 1D problems are considered in this chapter, each output of the combinatorial problem is optimized individually.

4.2.3 Downselecting Pattern

NK-fitness landscapes are stochastic fitness functions on binary strings initially proposed by Kauffman [232], where each of N bits interacts with K other bits. These function can be generated synthetically and the value of the function is computed as the sum of the contributions from each bit (which include the coupling effect):

$$f(\mathbf{x}) = \frac{1}{N}\sum_{i}^{N} g_i(x_i; x_{i1}, x_{i2}, ..., x_{iK}), \tag{4.2}$$

Table 4.2: Test instances used for the combining problem. d is the number of dimensions, l is the number of levels, S is the size of the input space, and O is the optimal value.

Combining pattern					
Problem	**f(x)**	**d**	**l**	**S**	O
SNW	area	3	3-14	206	-16.24
SNW	throughput	3	3-14	206	-14.71
NoC	energy	4	2-7	259	9.96
NoC	runtime	4	2-7	259	4.3
SW-LLVM	performance	11	2	767	-270.4
SW-LLVM	memory-footprint	11	2	767	29
rs-6d-c3	throughput	6	2-8	3840	-232000
rs-6d-c3	latency	6	2-8	3840	1.9
sol-6d-c2	throughput	6	2-6	2866	-115000.0
sol-6d-c2	latency	6	2-6	2866	1.2
wc+sol-3d-c4	throughput	3	4-7	196	-63734
wc+sol-3d-c4	latency	3	4-7	196	2.18
wc-5d-c5	throughput	5	3-6	1080	-20591
wc-5d-c5	latency	5	3-6	1080	47.4
wc-c1-3d-c1	throughput	3	6-18	1343	-23075
wc-c1-3d-c1	latency	3	6-18	1343	148.8
wc-c3-3d-c1	throughput	3	6-18	1512	-11930
wc-c3-3d-c1	latency	3	6-18	1512	247.45

where x_i is the i-th bit of the binary string $\mathbf{x}$, $x_{i1}, ..., x_{iK}$ are the K bits that influence bit x_i, and g_i is the function that describes the contribution of the i-th bit. Note that g_i can take at most $2^{(K+1)}$ values, and therefore the function is usually evaluated using N lookup tables (one for each g_i) of $2^{(K+1)}$ values.

NK landscape functions are a good test case for downselecting problems as their binary-nature is a natural way of encoding the downselecting pattern; a "1" indicates that an element has been selected, whereas a "0" indicates the converse.

Table 4.2 contains a summary of the values used for the test problem. Test instances were selected to have a good combination between dimensions (10-25), size of the search space (thousands of points to tens of millions), and interactions among the input variables (no interactions to fifth-order interactions).

4.2.4 Partitioning Pattern

A modified version of the NK-landscape problem is used to evaluate partitioning patterns (which will be denoted as PAR(N,K)). The partitioning pattern consists of dividing a set of elements into subsets, and the main idea is that two elements

Table 4.3: Test instances used for the downselecting problem. d is the number of dimensions, l is the number of levels, S is the size of the input space, and O is the optimal value. Values with the '*' symbol have been computed using a genetic algorithm with 100,000 function evaluations.

Downselecting pattern					
Problem	**d**	**l**	**S**	O	**Comments**
NK(N=10, K=1)	10	2	1024	0.33	No interactions
NK(N=10, K=2)	10	2	1024	0.36	Second order interactions
NK(N=10, K=5)	10	2	1024	0.24	Fifth order interactions
NK(N=15, K=1)	15	2	$3.3 \cdot 10^4$	0.27	No interactions
NK(N=15, K=2)	15	2	$3.3 \cdot 10^4$	0.23	Second order interactions
NK(N=15, K=5)	15	2	$3.3 \cdot 10^4$	0.22	Fifth order interactions
NK(N=20, K=1)	20	2	$1 \cdot 10^6$	0.28*	No interactions
NK(N=20, K=2)	20	2	$1 \cdot 10^6$	0.25*	Second order interactions
NK(N=20, K=5)	20	2	$1 \cdot 10^6$	0.20*	Fifth order interactions
NK(N=25, K=1)	25	2	$3.3 \cdot 10^7$	0.24*	No interactions
NK(N=25, K=2)	25	2	$3.3 \cdot 10^7$	0.26*	Second order interactions
NK(N=25, K=5)	25	2	$3.3 \cdot 10^7$	0.21*	Fifth order interactions

in the same subset might interact positively, negatively, or not at all. Thus, the objective functions of the PAR(N, K) problem is constructed as follows: first, generate a symmetrical square matrix (known as the *interaction score* matrix), whose entries contain a positive value if two elements in the same subset interact positively, a negative value if they interact negatively, and zero if they do not interact (the value for each entry is randomly generated between -1 and 1); then, for each row zero all entries with exception of the K highest positive values (if there are more than K) and the K lowest negative values (if there are more than K), so that each element interacts at most with 2K other elements (with at most K positively and at most K negatively). Finally, to obtain the value for a given partition, sum the *interaction score* for all pairs of elements in the same subset, as shown in Eq 4.3:

$$f(\mathbf{x}) = -\sum_{s \in \mathcal{S}} \sum_{i \in s} \sum_{j \in s} \iota(i,j) \qquad \forall i > j, \tag{4.3}$$

where $\iota(i,j)$ is the interaction score between elements i and j (which belong to the same subset s). Since the objective of all test functions presented so far is to minimize the function value (and since positive interactions were assigned positive values to), a negative sign is added in front of Eq. 4.3.

Table 4.4 contains the values for each of the test instances considered. Again, values were selected to strike a good balance between dimension and degree of interaction.

Table 4.4: Test instances used for the partitioning problem. d is the number of dimensions, S is the size of the input space, and O is the optimal value. Values with the '*' symbol have been computed using a genetic algorithm with 100,000 function evaluations.

Partitioning pattern				
Problem	**d**	**S**	O	**Comments**
PAR(N=7, K=1)	7	877	-1.26	Interaction 2 elements
PAR(N=7, K=2)	7	877	-2.29	Interactions 4 elements
PAR(N=10, K=1)	10	$1.1 \cdot 10^5$	-1.70*	Interactions 2 elements
PAR(N=10, K=2)	10	$1.1 \cdot 10^5$	-2.39*	Interactions 4 elements
PAR(N=10, K=4)	10	$1.1 \cdot 10^5$	-4.02*	Interactions 8 elements
PAR(N=15, K=2)	15	$1.3 \cdot 10^9$	-6.69*	Interactions 4 elements
PAR(N=15, K=4)	15	$1.3 \cdot 10^9$	-8.02*	Interactions 8 elements

4.2.5 Permuting Pattern

Different test instances from OR-Library [233], QAPLIB [234], and TSPLIB [235] are used to test the performance of BO in the permuting pattern. OR-Library is a collection of test data sets for a variety of Operations Research (OR) problems, QAPLibe is a unified testbed which contains data instances for the Quadratic Assignment Problem, and TSPLIB is a library of sample instances for the TSP (among other related problems).

In particular, the test instances considered (also summarized in Table 4.5) are:

- The quadratic assignment problem (QAP): in the QAP, a set of n facilities needs to be assigned to n locations. A *flow* of supplies needs to be transported between each pair of facilities, and a matrix with the *distances* between locations is provided. The objective is to assign each facility to a location such that the product of *flows* and *distances* is minimized. The solution can be represented as a permutation sequence of facilities, where the first facility in the ordered sequence is assigned to the first location, the second facility to the second location, and so on. Mathematically, the objective function of the problem is:

$$f(x) = \sum_{i=1}^{n} \sum_{j=1}^{n} f_{i,j} d_{x(i),x(j)} \tag{4.4}$$

  where x is the permutation denoting the solution, $f_{i,j}$ is the flow from facility i to facility j, $d_{i,j}$ is the distance between locations i and j, and n is the number of facilities and locations.

- The traveling salesman problem (TSP): in the TSP, a set of n locations and a matrix with the *distances* between locations are provided. The objective is

to find a sequence that visits each location exactly once (and returns to the starting point), while minimizing the total distance traversed.

$$f(x) = \sum_{i=1}^{n-1} d_{x(i),x(i+1)} + d_{x(n),x(0)} \tag{4.5}$$

- The flow shop problem (FSP): in the FSP the order in which n jobs are executed needs to be scheduled. Each job is composed of m tasks, to be performed sequentially in m machines. The processing time of each task in each machine is different among jobs, and machines can only process one job at a time. Thus, all tasks must be executed sequentially. The goal is to determine the job sequence that minimizes the makespan (i.e., the time at which the last job finishes).

- Finally, in the unimodal-permutation problem, there is a fixed permutation of size N (e.g., $\pi_r = \{1, 2, 3, ...N\}$) and the objective function is simply the Hamming distance between a solution and the reference permutation. Therefore, the objective of the unimodal-permutation problem is simply to find a given permutation.

Table 4.5: Test instances used for the permuting pattern. d is the number of dimensions, S is the size of the input space, and O is the optimal value.

Permuting pattern				
Problem	**d**	**S**	O	**Comments**
UNI(10)	10	$3.6 \cdot 10^{06}$	0	Random permutation
UNI(15)	15	$1.3 \cdot 10^{12}$	0	Random permutation
UNI(20)	20	$2.4 \cdot 10^{18}$	0	Random permutation
UNI(30)	30	$2.7 \cdot 10^{32}$	0	Random permutation
TSP(ulysses16)	16	$2.1 \cdot 10^{13}$	6859	Traveling salesman problem
TSP(fri26)	26	$4.0 \cdot 10^{26}$	937	Traveling salesman problem
TSP(gr21)	21	$5.1 \cdot 10^{19}$	2707	Traveling salesman problem
TSP(gr17)	17	$3.6 \cdot 10^{14}$	2085	Traveling salesman problem
TSP(bays29)	29	$8.8 \cdot 10^{30}$	2020	Traveling salesman problem
FSP(reC27_30_15)	27	$1.1 \cdot 10^{28}$	2373	Flow shop problem
FSP(hel2_20_10)	20	$2.4 \cdot 10^{18}$	135	Flow shop problem
FSP(car2_13_4)	13	$6.2 \cdot 10^{09}$	7166	Flow shop problem
FSP(car7_7_7)	7	$5.0 \cdot 10^{03}$	6590	Flow shop problem
QAP(nug16a)	16	$2.1 \cdot 10^{13}$	1610	Quadratic assignment problem
QAP(nug21)	21	$5.1 \cdot 10^{19}$	2438	Quadratic assignment problem
QAP(nug27)	27	$1.1 \cdot 10^{28}$	5234	Quadratic assignment problem
QAP(tho30)	30	$2.7 \cdot 10^{32}$	149936	Quadratic assignment problem
QAP(tai35a)	35	$1.0 \cdot 10^{40}$	2422002	Quadratic assignment problem

4.3 Optimization methods considered

In this section, Bayesian optimization is compared against other methods proposed in the literature which are and commonly used for SAPs. These methods include: random search (RS), genetic algorithms (simple (GA-EAS); μ, λ (GA-MCL); $\mu + \lambda$ (GA-MPL)), and existing Bayesian optimization optimizers for (some) combinatorial problems such as SMAC, Hyperopt, and BOCS.

The parameters for each of the alternative methods are described below:

- In random sampling (RS), N_f points are randomly generated and evaluated.
- In genetic algorithm evolutionary-algorithm simple (GA-EAS), a population size of 20 is used. The crossover operation is "uniform crossover" or "partially-matched crossover", depending on the the type of pattern, with a crossover rate equal to 0.95. The mutation operation is "uniform mutation" or "shuffle mutation", with mutation rate set to $1/N$, where N is the dimension of the problem. Selection is conducted using tournament selection of size 3, and the algorithm terminates when N_f function evaluations have been performed.
- In genetic algorithm, mu comma lambda (GA-μ, λ) and genetic algorithm mu plus lambda (GA-$\mu + \lambda$), $\mu = 30$ % and $\lambda = 70$ % of the population size, respectively. The rest of the settings are similar to those used in GA-EAS.
- SMAC is the algorithm presented in [202] for algorithm configuration. The open source python implementation[1] with its default parameters is used for all applicable problems. However, one of the shortcomings of SMAC is that it cannot deal with permuting and partitioning problems.
- Hyperopt is the algorithm presented in [204], which uses a tree-structured Parzen density estimator as its surrogate model[2]. As with SMAC, Hyperopt is limited in the type of problems it can deal with.
- BOCS refers to the BOCS-SA[3] algorithm presented in [217], with a value of $\lambda = 10^{-4}$. BOCS is only applicable to combining and downselecting problems.

With regard to the Bayesian optimization algorithms, different surrogate models, types of kernels, and distance functions are used (as applicable to each test instance).

- **Different models:**
  - In BO-TREE, a random forest composed of 50 decision trees is used as the surrogate model within the Bayesian optimization framework.

[1] Available at https://github.com/automl/SMAC3
[2] Available at https://github.com/hyperopt/hyperopt
[3] Available at https://github.com/baptistar/BOCS

- In BO-TREE-FT, a random forest of 50 decision trees is also used, but the inputs to the tree are a set of features derived for each of the problems. For example, for the combining and downselecting problems, the features include the first and second-order interactions, whereas for the partitioning problem, the features indicate whether two elements are in the same subset or not.
- In BO-RBFN, a set of radial basis functions (which only depend on the distance of points to the center) is used to build a surrogate model for the function to be optimized. In this case, Gaussian functions are used as radial basis functions:

$$\hat{\mathbf{y}}(\mathbf{x}) = w_0 + \sum_{i=1}^{N} w_i \phi(x, c_i) \tag{4.6}$$

$$\phi(||x - c_i||) = \exp(-\beta d(x, c_i)) \tag{4.7}$$

where w_0, w_i are weights to be determined, $\phi(||x - c_i||)$ is a radial function that depends on the distance between x and c_i, and d_H is a distance function between x and c_i. The value of the weights is computed as:

$$w_0 = \frac{1}{N} \sum_{i=1}^{N} y_i \tag{4.8}$$

$$w = (G^{\cdot} G)^{-1} G^{\cdot} y \tag{4.9}$$

- **Different kernels**:
  - In BO-MT32, a Matérn kernel with $\nu = 3/2$ is used in combination with the distance function appropriate to the problem instance under analysis.
  - In BO-MT52, a Matérn kernel with $\nu = 5/2$ is used in combination with the distance function appropriate to the problem instance under analysis.
  - In BO-Kendall, the Kendall kernel for permutations described in Section 3.4.2 is used.
  - In BO-Mallows, the Mallows kernel for permutations described in Section 3.4.2 is used.
- **Different distance functions**:
  - In BO-RBF, the radial basis function kernel is used in combination with the distance function appropriate to the problem instance under analysis, as described in Section 3.4.2.
  - In BO-ARBF, an *anisotropic* radial basis function kernel is used with the appropriate distance function. In the ARBF kernel, the value of λ is different for each dimension of the input.

In all the Bayesian optimization methods, a genetic algorithm is used to maximize the acquisition function. The genetic algorithm runs for 1,000 generations using a population size of 500, with crossover probability equal to 0.95 and mutation probability equal to 0.05. For the initial population, 3/4 of the points are generated randomly whereas the remaining 1/4 is extracted from the "hall of fame". The selection mechanism is tournament selection (with a tournament size of 3) for the single-objective case, and NSGA-II in the multi-objective case.

4.4 Results

This section presents the results of the Bayesian optimization framework when applied to a library of one-dimensional SAPs.

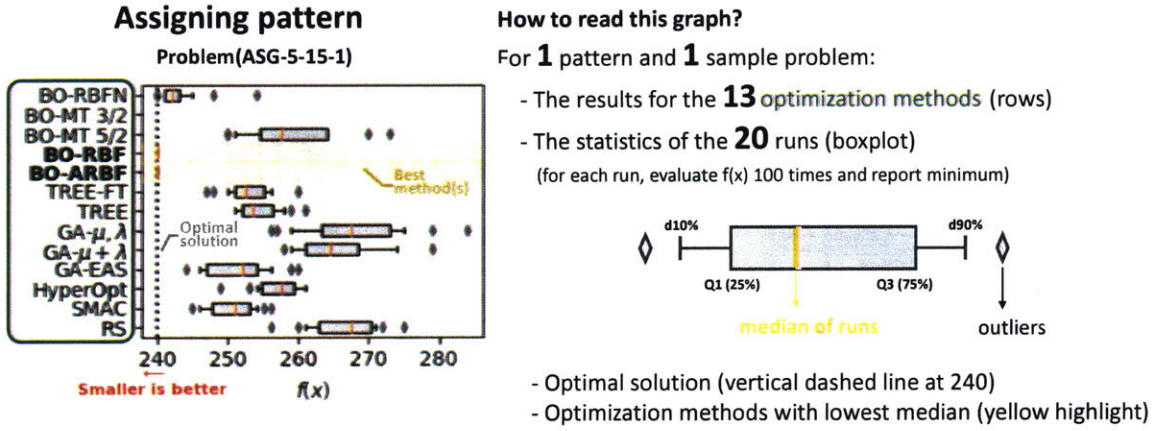


Figure 4-1: Example box-plot depicting the results for a particular problem. In this instance, the results for the problem ASG-5-15-1 of the assigning pattern are depicted. The dashed vertical line shows the optimal value of the functions, whereas the bloxplots illustrate the statistics of 20 runs for each method.

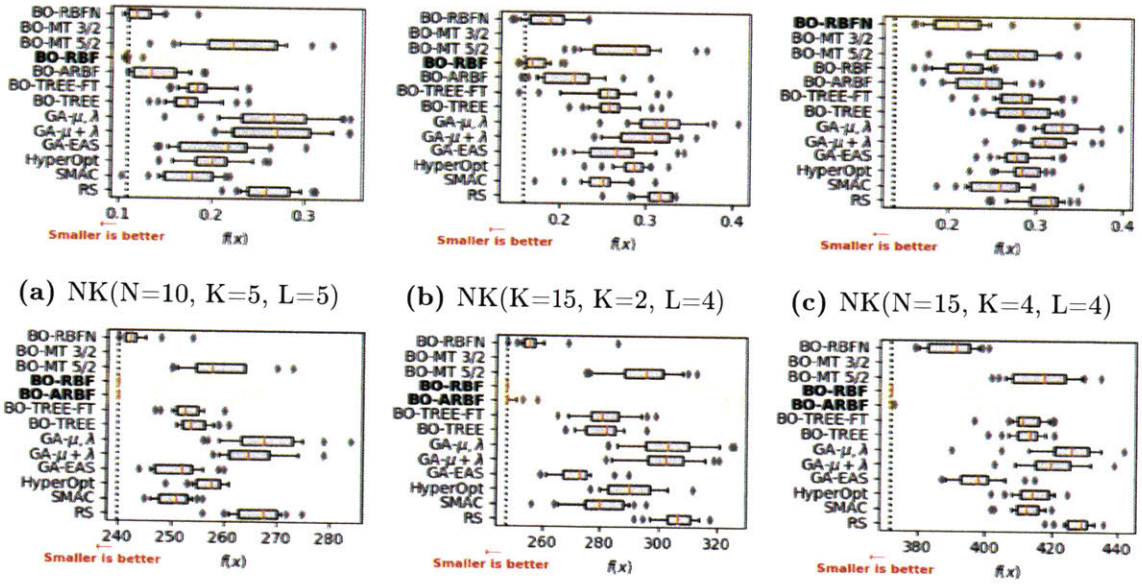
Figure 4-1 shows an example of the graphs used to represent the results for a particular problem (ASG-5-15-1) of the assigning pattern, with each row corresponding to a different optimization method (as described in Section 4.3). The boxplots depict the statistics (of the 20 runs) for the best function values (i.e., the minimum) obtained after N_f function evaluations; the orange line represents the median of the minimum function values (i.e., the optimization result) obtained across the 20 runs; whereas the rectangular region contains the values between the first and third quartiles (i.e., half of the obtained values fall within the rectangular region), and the whiskers extend from the first to ninth decile. The dots outside this range are outliers. Where available, the optimal value of the function is depicted as a dashed vertical line, whereas the method(s) that provided the best solution (using the median of runs as the metric) are highlighted with a yellow band and printed in bold type.

4.4.1 Overview of the different decision patterns

First, the visual analysis is presented for each of the patterns. Numerical values for all the decision patterns, test problems, and optimization methods are included in Appendix B.

Assigning Pattern

Figure 4-2 shows the results for six illustrative assigning problems. It can be observed that Bayesian optimization methods (in particular BO-RBF and BO-ARBF) are superior to other methods in all cases but for NK(N=15, K=4, L=4). This happens because in NK(N=15, K=4, L=4), there are a lot of interactions between decisions[4], which make the problem harder to model, and consequently, the surrogate Gaussian Process model cannot fully capture the structure of the problem. Finally, note how BO-RBF presents a very small variance across different runs in all the different test instances analyzed, which makes it a robust optimization method.



(a) NK(N=10, K=5, L=5) **(b)** NK(K=15, K=2, L=4) **(c)** NK(N=15, K=4, L=4)

(d) GAP(515-1) **(e)** GAP(520-2) **(f)** GAP(824-3)

Figure 4-2: Selected results for the single-objective assigning problem. Smaller values are better.

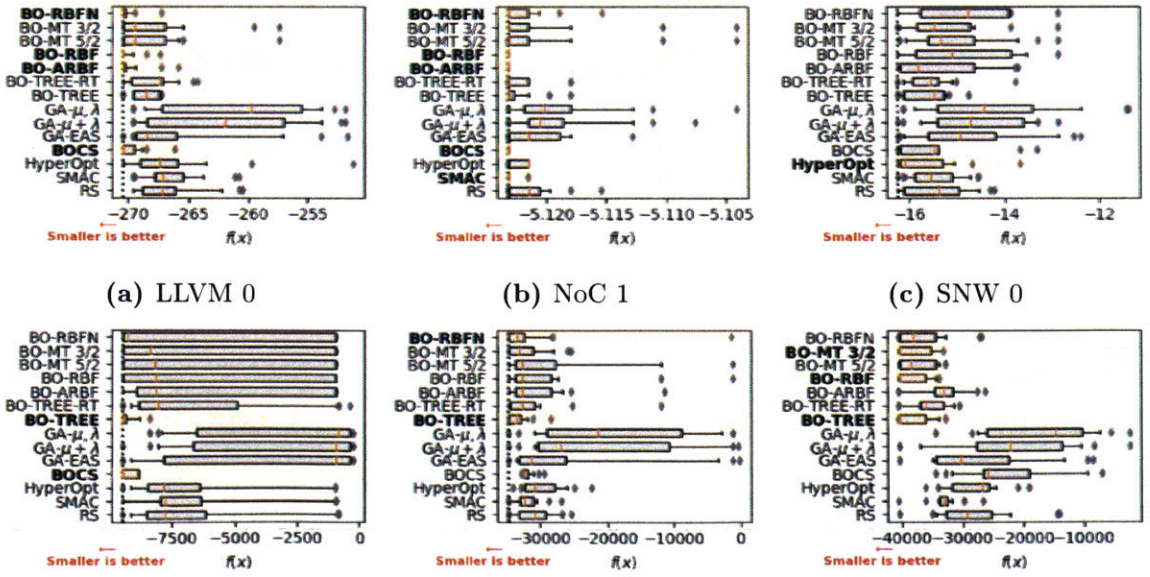
In addition, it can be observed that RS performs the worst in most cases (as expected). Surprisingly, both GA-μ, λ and GA-$\mu + \lambda$ algorithms also perform very

[4]These situations are not common in real-life problems, and when they do occur, there is no other alternative than evaluating a large number of points of the tradespace, as the lack of exploitable structure of the problem makes all optimization methods ineffective.

poorly, with results comparable to those obtained by random sampling. This is because both genetic-algorithm variants rely on the evaluation of large populations, and are therefore sub-optimal when working with a limited budget for function evaluations. Finally, among non-Bayesian optimization methods, SMAC and GA-EAS offer the best performances.

Combining Pattern

Figure 4-3 shows the results for six test problems of the combining pattern. It can be seen that there is no clear trend across problems: in LLVM0 and NoC1, BOCS, BO-RBF and BO-RBFN are capable of reaching the optimal solution in most of the runs; in SNW 0 it is HyperOpt the best performing algorithms; and in the problems from the BO4CO dataset (depicted in the lower row of Figure 4-3) BO-TREE is the best performer. Despite of this variance, it can be observed that the Bayesian optimization methods outperform random selection and genetic algorithm-based methods in all cases, even though there is great variability in terms of the specific BO method that performed the best: BO-RBFN, BO-RBF, and BO-TREE reached the best solution in three out of the six problems depicted. In addition, these three methods show a lower variance in the test problems where they perform the best, as compared to other approaches, such as GA-based methods or BOCS which is indicative of a robust performance for those particular problems. Overall, there seems to be no clear winner for all cases.



(a) LLVM 0 **(b)** NoC 1 **(c)** SNW 0

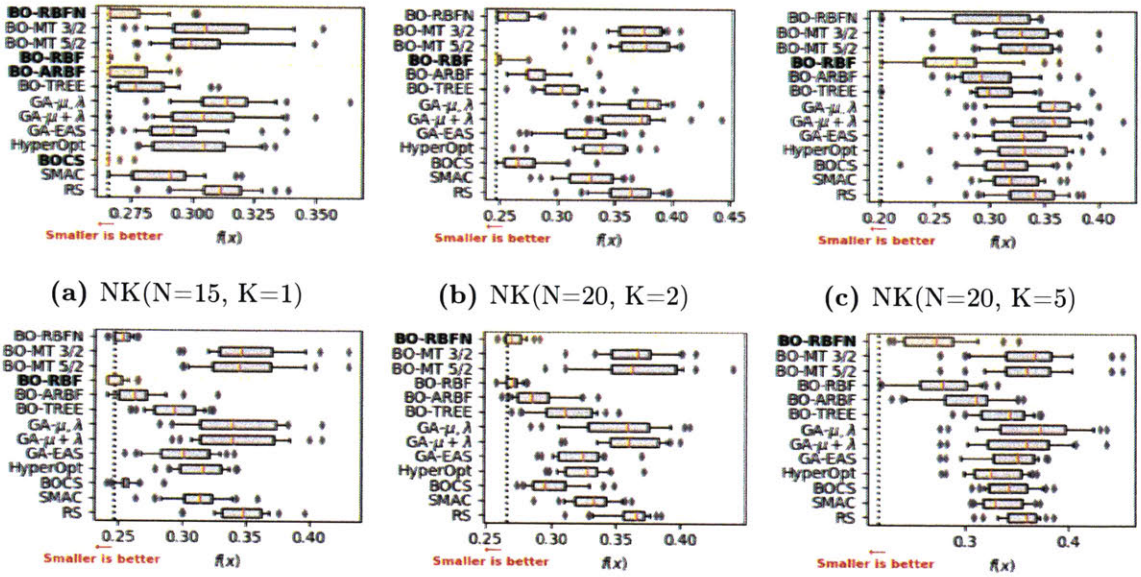
(d) wc-c1-3d-c1 1 **(e)** rs-6d-c3 1 **(f)** sol-6d-c2 1

Figure 4-3: Selected results for the single-objective combining pattern. Smaller values are better.

Downselecting Pattern

Figure 4-4 shows the results for six selected instances of the downselecting pattern. The Hamming distance has been used in all BO methods. In this case, the BO-RBF method offers the best solutions for three out of the six problems considered, whereas BO-RBFN offers the best solutions for two problems, with both methods producing comparable results for the remaining problem. In all cases, the performance of these two Bayesian optimization methods is superior to other optimization methods. Among the other BO methods, those based on the Matérn kernel (BO-MT 3/2 and BO-MT 5/2) produced very poor results, with a performance comparable to that of RS. Also note how BOCS, which was designed to capture first and second-order interactions, works reasonably well for the K=1 and K=2 cases, but performs similarly to RS when the number of interactions is high (K = 5).

The fact that BO-RBF and BO-RBFN were superior to other methods in all problems shows that these two BO methods can be effective for situations where there is minimal interaction between the different elements of the downselecting problem (e.g., the K=1, 2 cases), as well as in cases where there are numerous interactions among inputs (e.g., the K=5 case).



(a) NK(N=15, K=1) **(b)** NK(N=20, K=2) **(c)** NK(N=20, K=5)

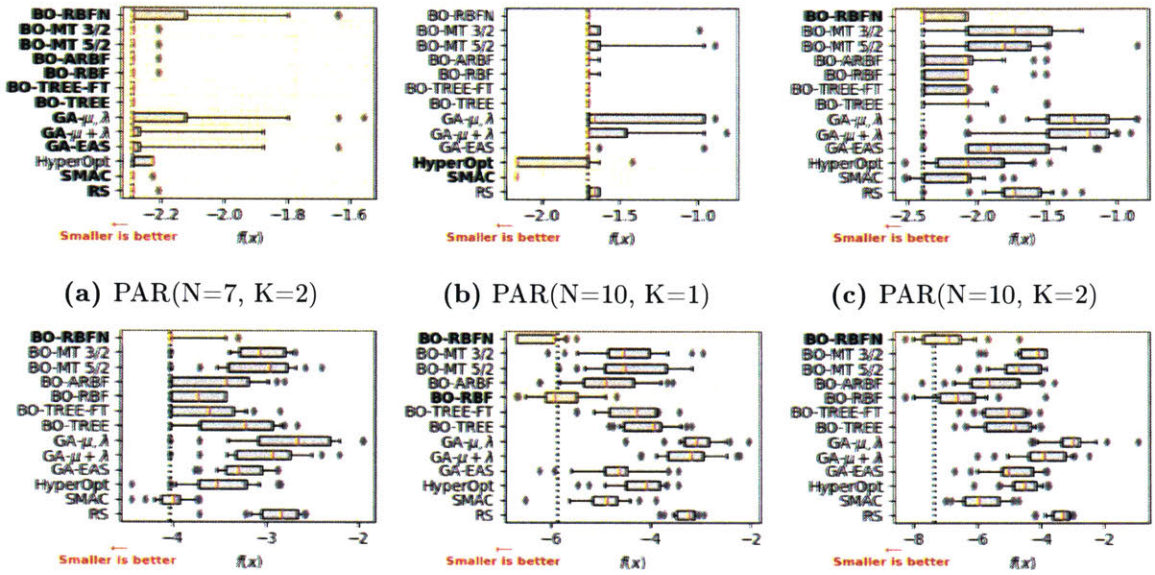
(d) NK(N=25, K=1) **(e)** NK(N=25, K=2) **(f)** NK(N=25, K=5)

Figure 4-4: Selected results for the single-objective downselecting pattern. Smaller values are better.

Partitioning Pattern

Figure 4-5 shows the results for six of the partitioning pattern test instances. It can be seen that for this type of problem, there is larger variability in the results:

for the small dimensional problems PAR(N=7,K=2) and PAR(N=10, K=2), most methods achieve the optimal solution for most of the runs; for the large dimensional problems PAR(N=15, K=2) and PAR(N=15, K=4), both BO-RBF and BO-RBF offer better performance than other methods. Overall, BO-RBFN is the preferred method in five out of the six problems. Note that in this pattern, the differences between optimization methods are subtle, and the variance within each method is larger than in previous patterns. In addition to BO-RBFN, SMAC also shows good performance in these test instances, especially on problems with large dimensionality ($N \geq 10$), where it paces second and third best (after BO-RBFN and BO-RBF).



(a) PAR(N=7, K=2) **(b)** PAR(N=10, K=1) **(c)** PAR(N=10, K=2)

(d) PAR(N=10, K=4) **(e)** PAR(N=15, K=2) **(f)** PAR(N=15, K=4)

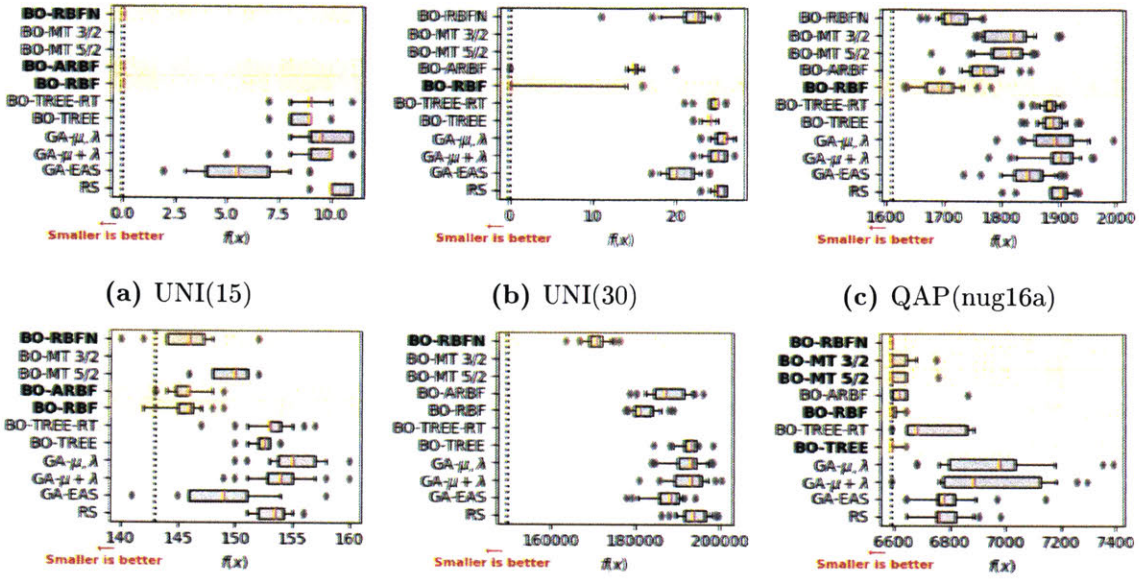
Figure 4-5: Selected results for the single-objective partitioning problem. Smaller values are better.

Permuting Pattern

Figure 4-6 shows the results for six selected test instances for the single-objective permuting pattern. In five of them, BO-RBF produced the best results, whereas BO-RBFN was the best option for four out of the six cases. For the UNI problems, the difference between BO-RBF and the rest of the methods is outstanding.

For the permuting pattern, tree-based methods did not produce great results (comparable to random sampling), and the genetic algorithm-based algorithms were also sub-optimal. This is not surprising, given that these methods lack the capabilities to capture the structure of the relationships between permutations (i.e., it is difficult to encode the solutions in such a way that the algorithms can exploit the similarities or differences between different permutations). In addition, the approaches that used the Matérn kernel (both 3/2 and 5/2) were not able to complete the runs for all

problems, since the covariance matrix quickly lost its positive semidefiniteness; the procedure described in Section 3.4.4 to solve these kind of situations was not able to fix the in all cases, and the algorithms crashed for some of the problems.



(a) UNI(15) (b) UNI(30) (c) QAP(nug16a)

(d) FSP(hel2_20_10) (e) QAP(tho30) (f) FSP(car7_7_7)

Figure 4-6: Selected results for the single-objective permuting problem. Smaller values are better.

4.4.2 Ranking of methods

This section recommends the best method(s) for each decision pattern. However, because it is rare for a single method to outperform all others in all test instances, recommendations will be based on rigorous, quantitative rank-metrics. This is the same approach taken in the literature in the context of algorithm selection for classification tasks [236], where the performance of classification algorithms is evaluated over multiple datasets. The same three ranking metrics proposed in reference [236] are considered:

- The **average ranking (AR)** is computed by averaging the rank of each method in each problem. Let r_m^p be the rank of method m in problem p, then the average ranking of method m is $AR_m = \frac{1}{|\mathcal{P}|}\sum_{p\in\mathcal{P}} r_m^p$, with $\mathcal{P}$ being the set of problems evaluated. Lower values of the AR metric indicate higher rank.
- The **success rate ratio (SRR)** ranking uses pairwise comparisons between methods to compute the ranking. In this case, the error rate proposed in reference [236] is replaced by the percentage difference between the method-obtained value and the optimum value for each test instance. Higher values of the SRR metric indicate higher rank.

- The **significant wins (SW)** ranking uses the number of methods that perform significantly worse than a given method as its main metric. The statistical significance is computed using a paired t-test (with significance level of 5%) over the difference in performance between the algorithms. Higher values of the SRR metric indicate higher rank.

Appendix C contains a more complete description and the mathematical formulation of these rank-metrics.

Tables 4.6 - 4.10 show the value of the AR, SRR, and SW metrics for each of the patterns analyzed. The number in brackets represents the ranking of each method according to that particular metric. It can be seen that BO-RBF is the highest-ranked method for the assigning, downselecting, and partitioning patterns, whereas BO-RBFN is best for the combining and permuting patterns. Note that all metrics provide consistent results (i.e., they always place the selected algorithms in either the first or second place). Furthermore, it is worth noting that for the assigning, combining, downselecting, and permuting patterns, the Bayesian optimization algorithms clearly outperform other optimization methods (GA-based and other optimizers such as HyperOpt and SMAC).

Table 4.6: Ranking for assigning pattern algorithms.

	BO-RBF	BO-ARBF	BO-RBFN	SMAC	BO-TREE-FT	BO-TREE	GA-EAS	BO-MT 5/2	HyperOpt	GA-$\mu + \lambda$	RS	GA-μ, λ	BO-MT 3/2
AR	1.30 [1]	2.20 [2]	2.50 [3]	5.00 [4]	5.50 [5]	6.00 [6]	6.10 [7]	7.60 [8]	8.00 [9]	10.40 [10]	11.00 [11]	11.40 [12]	13.00 [13]
SRR	46.35 [1]	40.40 [3]	45.52 [2]	24.70 [5]	23.38 [6]	25.54 [4]	11.56 [9]	16.48 [7]	15.59 [8]	0.68 [11]	1.46 [10]	0.61 [12]	nan [13]
SW	0.85 [1]	0.69 [2]	0.69 [2]	0.38 [4]	0.31 [5]	0.31 [5]	0.23 [7]	0.23 [7]	0.23 [7]	0.08 [10]	0.00 [11]	0.00 [11]	0.00 [11]

Table 4.7: Ranking for combining pattern algorithms.

	BO-RBFN	BO-TREE	BO-RBF	BO-MT 3/2	BO-ARBF	BO-MT 5/2	BOCS	BO-TREE-RT	SMAC	HyperOpt	RS	GA-EAS	GA-$\mu + \lambda$	GA-μ, λ
AR	3.06 [1]	3.39 [2]	3.39 [2]	3.67 [4]	4.00 [5]	4.17 [6]	5.00 [7]	5.06 [8]	5.89 [9]	6.56 [10]	8.89 [11]	9.28 [12]	11.44 [13]	11.72 [14]
SRR	1.17 [2]	1.18 [1]	1.15 [4]	1.15 [3]	1.15 [6]	1.14 [7]	1.15 [5]	1.14 [8]	1.12 [9]	1.12 [10]	1.08 [11]	0.93 [12]	0.87 [13]	0.84 [14]
SW	0.43 [1]	0.43 [1]	0.43 [1]	0.36 [4]	0.29 [5]	0.29 [5]	0.14 [10]	0.21 [7]	0.21 [7]	0.21 [7]	0.14 [10]	0.14 [10]	0.07 [13]	0.00 [14]

Table 4.8: Ranking for downselecting pattern algorithms.

	BO-RBF	BO-RBFN	BO-ARBF	BOCS	BO-TREE	SMAC	GA-EAS	HyperOpt	BO-MT 3/2	BO-MT 5/2	GA-$\mu + \lambda$	RS	GA-μ, λ
AR	1.42 [1]	1.92 [2]	2.50 [3]	3.83 [4]	4.25 [5]	5.92 [6]	7.17 [7]	8.00 [8]	9.33 [9]	9.58 [10]	10.08 [11]	11.00 [12]	12.08 [13]
SRR	1.23 [1]	1.19 [2]	1.16 [3]	1.13 [4]	1.10 [5]	1.03 [6]	1.00 [7]	0.99 [8]	0.93 [9]	0.92 [10]	0.87 [12]	0.88 [11]	0.83 [13]
SW	0.85 [1]	0.69 [2]	0.69 [2]	0.62 [4]	0.62 [4]	0.46 [6]	0.38 [7]	0.38 [7]	0.08 [9]	0.08 [9]	0.08 [9]	0.08 [9]	0.00 [13]

Table 4.9: Ranking for partitioning pattern algorithms.

	BO-RBF	SMAC	BO-ARBF	BO-RBFN	BO-TREE-FT	GA-EAS	BO-TREE	HyperOpt	BO-MT 5/2	BO-MT 3/2	RS	GA-$\mu + \lambda$	GA-μ, λ
AR	1.86 [1]	2.00 [2]	2.86 [3]	3.00 [4]	3.57 [5]	4.29 [6]	5.29 [7]	5.57 [8]	5.57 [8]	5.71 [10]	8.43 [11]	9.00 [12]	10.86 [13]
SRR	1.38 [1]	1.37 [2]	1.29 [3]	1.09 [10]	1.26 [5]	1.24 [6]	1.22 [7]	1.26 [4]	1.20 [8]	1.18 [9]	1.09 [11]	1.02 [12]	0.98 [13]
SW	0.54 [2]	0.69 [1]	0.46 [3]	0.00 [11]	0.23 [5]	0.31 [4]	0.23 [5]	0.23 [5]	0.15 [8]	0.15 [8]	0.08 [10]	0.00 [11]	0.00 [11]

Table 4.10: Ranking for permuting pattern algorithms.

	BO-RBFN	BO-RBF	BO-ARBF	GA-EAS	BO-TREE	BO-TREE-RT	BO-MT 3/2	BO-MT 5/2	GA-$\mu + \lambda$	RS	GA-μ, λ
AR	1.44 [1]	2.22 [2]	3.06 [3]	4.39 [4]	6.17 [5]	7.17 [6]	7.17 [6]	7.44 [8]	7.78 [9]	8.17 [10]	8.39 [11]
SRR	39.44 [1]	5.16 [2]	2.61 [3]	1.52 [4]	0.93 [7]	0.92 [8]	1.04 [5]	1.03 [6]	0.92 [10]	0.92 [9]	0.90 [11]
SW	0.64 [1]	0.55 [2]	0.55 [2]	0.45 [4]	0.00 [6]	0.27 [5]	0.00 [6]	0.00 [6]	0.00 [6]	0.00 [6]	0.00 [6]

4.4.3 Computational speedup

In this section, the computational speedup is analyzed. First, an iso-performance analysis is carried out to determine the number of *extra function evaluations* that other algorithms need to carry out with respect to the recommended algorithm identified in the previous section (i.e., how many extra function evaluations would each of the other methods require to achieve a similar performance as the preferred algorithm?). Second, the time speedup is analyzed as a function of the time required to evaluate the black-box function (i.e., how much time is saved by using Bayesian optimization?).

Figure 4-7 shows the ratio between the average number of function evaluations required by the recommended algorithm and the average number of function evaluations required by other methods to achieve an equal performance (the one achieved by the recommended algorithm after N_f function evaluations). There is a limit of 5,000 function evaluations for population-based and random search algorithms (i.e., if the RS or GA algorithm cannot achieve similar performance as the reference algorithm even with 50× more function evaluations, the algorithm is terminated), whereas a hard limit of 2,000 (20×) function evaluations is used for SMAC and HyperOpt optimizers.

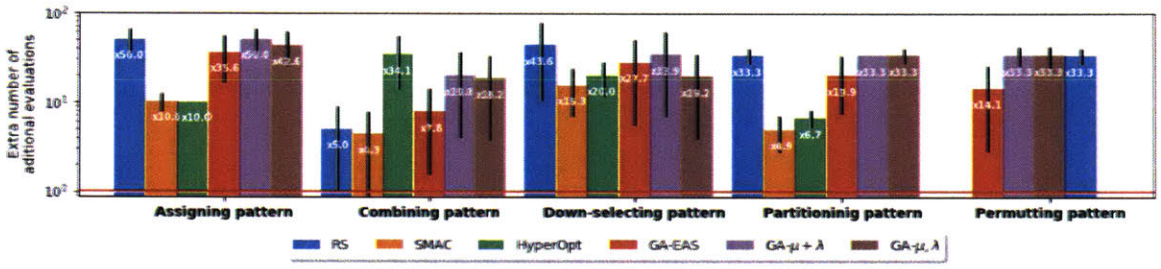
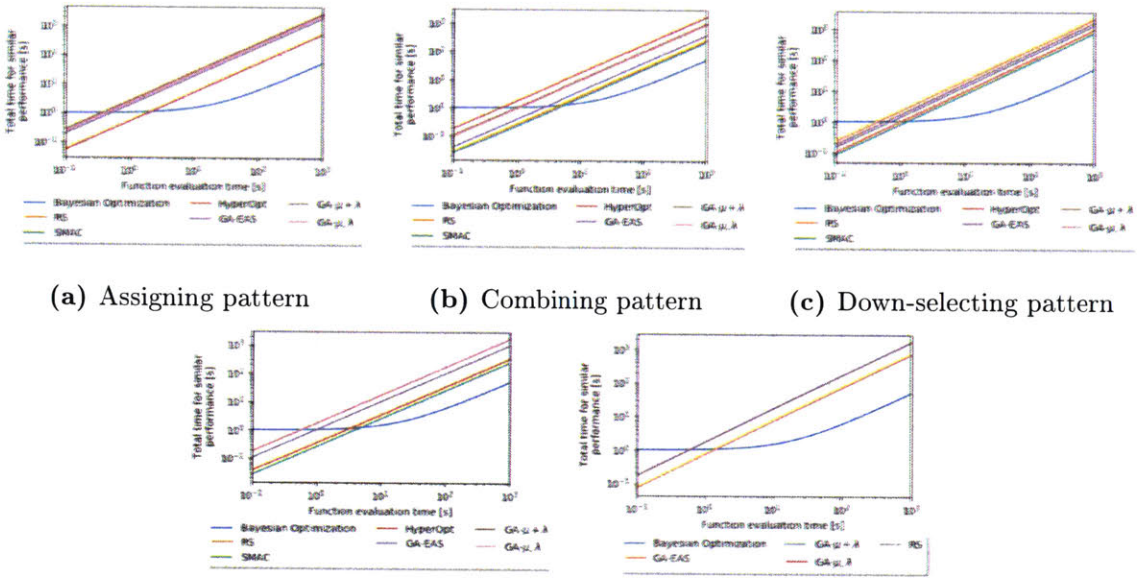


Figure 4-7: Computational speedups of the recommended method when compared against other optimization methods for single-objective problems. The computational speedup is measures as the number of additional function evaluations required by other methods to achieve a similar performance to the recommended method.

It can be observed that the speedup factor (in terms of extra function evaluations) ranges from 4.3× in the partitioning pattern (achieved using SMAC) to more than 10× in the assigning pattern (when using SMAC or HyperOpt). Note that SMAC is the fastest algorithm among "traditional" algorithms, for the patterns where it is applicable. This was expected as SMAC was designed to operate in mixed-integer single-objective problems. Also, it is interesting to note that random sampling works relatively well for the combining pattern (with performance superior to GA-based algorithms and HyperOpt). This is because given the very limited number of function evaluations, the genetic algorithm is unable to explore the space extensively, and instead converges earlier to sub-optimal solutions. This phenomenon has been observed in previous literature, when one of the dimensions of the input space dominates the

function output, for example, in the field of hyper-parameter tuning [237].

Finally, Figure 4-8 shows the computational speedup achieved for each of the five patterns analyzed as a function of the function evaluation time. For single-objective problems, Bayesian optimization has an overhead (fitting the model, maximizing the acquisition function) of 5-20 seconds, depending on the dimension of the search space. Therefore, if the function evaluation time is very small, it does not pay off to use Bayesian optimization. As shown in the graph, in general there are no benefits from using Bayesian optimization if the function evaluation time is shorter than 5 seconds. For longer evaluation times, however, Bayesian optimization allows for potential reductions in computational time. Asymptotically (i.e., when the function evaluation time is much higher than the overhead), the benefit is proportional to the speedups obtained in Figure 4-7, which ranged from 4.3× in the combining pattern to 15.3× for the downselecting pattern.



(a) Assigning pattern **(b)** Combining pattern **(c)** Down-selecting pattern

(d) Partitioning pattern **(e)** Permuting pattern

Figure 4-8: Computational speedups of Bayesian optimization algorithms when applied to multi-objective problems.

4.5 Chapter conclusions

The objective of this chapter was to answer the three research questions below.

Research Question 4.1

Is Bayesian optimization a suitable method to solve one-dimensional System Architecture problems?

The analysis conducted in this chapter shows that Bayesian optimization can be successfully applied to different one-dimensional SAPs. To do so, one can adapt the classical BO formulation by using a tailored model, kernel, or distance function (or a combination of them). The results show that when compared against classical optimization methods used in the literature for these types of problems (including genetic algorithms, and classical optimizers), the BO adaptation is the preferred method for the five patterns analyzed.

Research Question 4.2

Which of the methods to adapt Bayesian optimization to combinatorial problems is most effective for each of the decision patterns common in SAPs?

For the assigning and downselecting patterns, the Gaussian process model using the radial basis function (BO-RBF) kernel, together with the Hamming distance, was the preferred method; for the partitioning pattern, the same model and kernel, together with the Damerau-Levenshtein distance, ranked first. For the combining and permuting patterns, the radial basis function network (BO-RBFN) model was superior, and the BO-RBF method with the Hamming distance ranked second with comparable performance.

Research Question 4.3

What is the improvement achieved by using Bayesian optimization, compared to other methods currently employed in the literature?

Bayesian optimization allowed the number of expensive function evaluations to be reduced by a factor of 4.3× - 15.3× (depending on the problem) when compared to other optimization methods in the literature. However, since the Bayesian optimization procedure has significant overhead in fitting the model and optimizing the acquisition function, this reduction in the number of function calls translates to time savings only if the time required to perform a function call is higher than 5 seconds. For faster function calls, the overhead of the Bayesian optimization method dominates the runtime, whereas if a function call takes several minutes (or several hours, as is the case of the simulations in the second part of this dissertation), the gains of using Bayesian optimization are clear.

Finally, it is important to note that Bayesian optimization is only recommended for problems of small or moderate dimensionality ($d < 20$), and that the higher the number of interactions among decisions, the poorer the performance of BO. Table 4.11 summarizes the recommendations for each of the five patterns:

Table 4.11: Summary of recommendations for single-objective combinatorial patterns.

	Method	Distance function	Speedup	Comments
Assigning	BO-RBF	Hamming	>10x	Works well for all problems considered
Combining	BO-RBFN	Hamming	4.3x	Best for small num. of interactions
Downselecting	BO-RBF	Hamming	15.3x	Best for small num. of interactions
Partitioning	BO-RBF	Damerau-Levenshtein	4.9x	Best for small num. of interactions
Permuting	BO-RBFN	Hamming	14.1x	Works well for all problems considered

Chapter 5

Bayesian Optimization for Multi-Objective SAPs

This chapter aims to answer similar research questions like the ones posed in the previous chapter, but in the context of multi-objective SAPs.

Research Question 5.1

What method to adapt Bayesian optimization to combinatorial problems works best for each of the decision patterns common in SAPs

Research Question 5.2

What is the improvement achieved by using Bayesian optimization, compared other methods used in the literature?

The Bayesian optimization approach is extended to multi-objective problems by fitting a different surrogate model for each of the objective functions, and adapting the acquisition function to handle the multi-objective space. Section 3.2.2 discussed various acquisition functions for multi-objective BO commonly used in the literature. In this work, the expected hypervolume improvement (EHI) is used as the acquisition function.

5.1 Methodology overview

The methodology used to evaluate the performance of different optimization algorithms within a multi-objective setting is similar to that used in the single-objective case, but the normalized hypervolume of the Pareto front is used instead as the metric

to compare the performance of the different optimization methods. To compute this metric, the hypervolume of the solutions obtained by each optimization method is divided by the "true" hypervolume, which is computed by evaluating 300,000 points using a genetic algorithm and serves as a reference for normalization.

As in the single-objective case, the optimization methods compared against BO include random sampling (RS) and genetic algorithms (simple, NSGA-II, SPEA2 as described below)). SMAC, Hyperopt, and BOCS cannot be used as they do not accommodate for multi-objective problems. In addition, all the Bayesian optimization methods with different surrogate models, types of kernels, and distance functions presented in the previous section are used where applicable (BO-TREE, BO-TREE-FT, BO-MT3/2, BO-MT5/2, BO-RBF, and BO-ARBF).

- NSGA-II is a fast and elitist multi-objective genetic algorithm described in [238]. In NSGA-II, the selection stage is carried out as follows:

  i points with lower Pareto ranking are selected first

  ii among points with similar Pareto ranking, those with lower crowding distances are given preference

  Thus, this method selects the best points (i.e., the Pareto front) while preserving population diversity, since it prioritizes solutions in less crowded areas of the Pareto front.

- SPEA2 is the Strength Pareto Evolutionary Algorithm described in [239]. In a nutshell, the algorithm assigns a fitness value and a strength value to each point in the population and the hall-of-fame (an archive of the Pareto front obtained up to the current step). The strength value of a point x, $S(x)$, represents the number of population members that are dominated by or equal to x, whereas the fitness value $F(x)$ is the sum of the strength values of all members in the hall-of-fame that dominate x. The lower the fitness value, the higher the probability of a point being selected for the mating and mutation phase of the GA.

5.2 Test Instances for Multi-Objective Problems

Two types of problems are used to test multi-objective SAPs: problems which are inherently multi-objective, or alternatively, synthetically-generated multi-objective problems. The latter are created using different test problems for the same decision pattern for each objective (or the same test problem with different seeds). For example, a multi-objective permuting problem can be created by using a TSP as the first objective function, and an FSP (of the same dimension) as the second objective. The only condition is that all the problems used for each objective function have the same input dimensions.

Table 5.1 contains information about the collection of test instances used for each decision pattern. Overall, 37 different test problems were considered for the five decision patterns. Considerations regarding dimension, size of the tradespace, and order of interactions were taken into account when choosing the problems' parameters.

Table 5.1: Test instances used for multi-objective problems. $f_1(x)$ and $f_2(x)$ refer to the first and second objective of the problem. If the same type of problem is used for both functions, different seeds are used for each objective. d is the number of dimensions in the input, l is the number of options on each input dimension (when applicable), and S is the size of the input space.

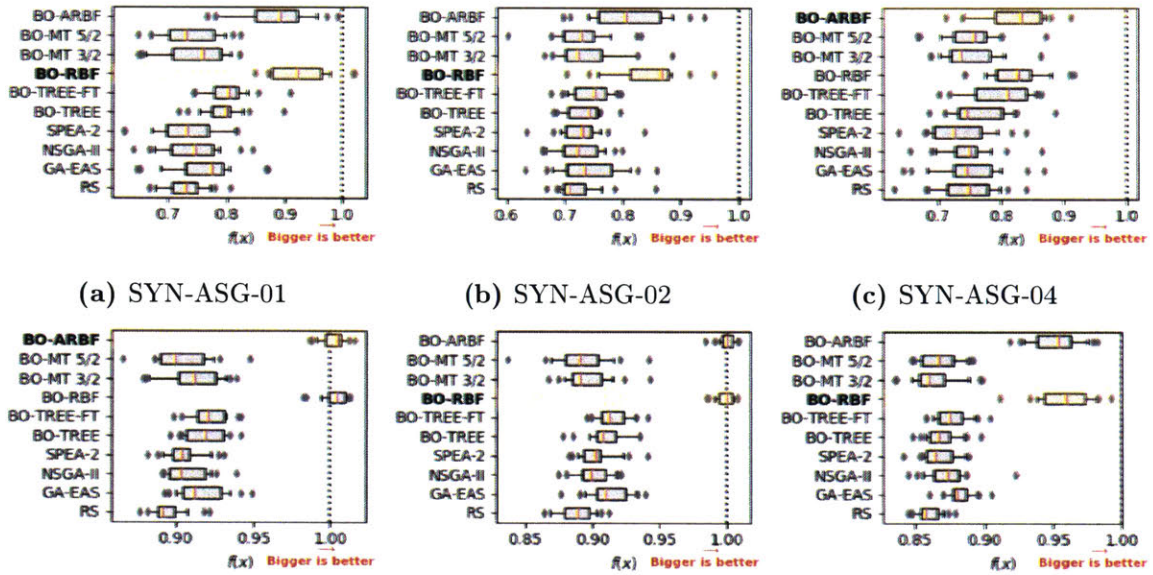
Assigning pattern						
Problem	$f_1(x)$	$f_2(x)$	**d**	**l**	**S**	**Comments**
SYN-ASG-01	NK(N=15, K=1, L=4)		15	4	$1.1 \cdot 10^9$	No interactions
SYN-ASG-02	NK(N=15, K=2, L=4)		15	4	$1.1 \cdot 10^9$	Second order interactions
SYN-ASG-03	NK(N=10, K=1, L=4)		10	4	$1 \cdot 10^6$	No interactions
SYN-ASG-04	NK(N=10, K=2, L=4)		10	4	$1 \cdot 10^6$	Second order interactions
SYN-ASG-05	gap515-1	gap515-2	15	5	$3 \cdot 10^{10}$	
SYN-ASG-06	gap515-3	gap515-4	15	5	$3 \cdot 10^{10}$	
SYN-ASG-07	gap520-2	gap520-3	20	5	$9.5 \cdot 10^{13}$	
SYN-ASG-08	gap824-1	gap824-2	24	8	$4 \cdot 10^{21}$	
Combining pattern						
Problem	$f_1(x)$	$f_2(x)$	**d**	**l**	**S**	**Comments**
SNW	area	throughput	3	3-14	206	ASIC design
NoC	energy	runtime	4	2-7	259	ASIC design
SW-LLVM	performance	memory	11	2	767	Software compiler
rs-6d-c3	throughput	latency	6	2-8	3840	Big data system configuration
sol-6d-c2	throughput	latency	6	2-6	2866	Big data system configuration
wc+sol-3d-c4	throughput	latency	3	4-7	196	Big data system configuration
wc-5d-c5	throughput	latency	5	3-6	1080	Big data system configuration
wc-c1-3d-c1	throughput	latency	3	6-18	1343	Big data system configuration
wc-c3-3d-c1	throughput	latency	3	6-18	1512	Big data system configuration
Downselecting pattern						
Problem	$f_1(x)$	$f_2(x)$	**d**	**l**	**S**	**Comments**
SYN-DWN-01	NK(N=10, K=1)		10	2	1024	No interactions
SYN-DWN-02	NK(N=10, K=2)		10	2	1024	Second order interactions
SYN-DWN-03	NK(N=15, K=1)		15	2	$3.3 \cdot 10^4$	No interactions
SYN-DWN-04	NK(N=15, K=2)		15	2	$3.3 \cdot 10^4$	Second order interactions
SYN-DWN-05	NK(N=15, K=5)		15	2	$3.3 \cdot 10^4$	Fifth order interactions
SYN-DWN-06	NK(N=25, K=2)		25	2	$3.3 \cdot 10^7$	Second order interactions
SYN-DWN-07	NK(N=25, K=3)		25	2	$3.3 \cdot 10^7$	Third order interactions
Partitioning pattern						
Problem	$f_1(x)$	$f_2(x)$	**d**		**S**	**Comments**
SYN-PAR-01	PAR(N=7, K=1)		7		877	Interactions 2 elements
SYN-PAR-02	PAR(N=7, K=2)		7		877	Interactions 4 elements
SYN-PAR-03	PAR(N=10, K=1)		10		$1.1 \cdot 10^5$	Interactions 2 elements
SYN-PAR-04	PAR(N=10, K=2)		10		$1.1 \cdot 10^5$	Interactions 4 elements
SYN-PAR-05	PAR(N=10, K=4)		10		$1.1 \cdot 10^5$	Interactions 8 elements
SYN-PAR-06	PAR(N=15, K=2)		15		$1.3 \cdot 10^9$	Interactions 4 elements
SYN-PAR-07	PAR(N=15, K=4)		15		$1.3 \cdot 10^9$	Interactions 8 elements
Permuting pattern						
Problem	$f_1(x)$	$f_2(x)$	**d**	**l**	**S**	**Comments**
SYN-PER-01	UNIHD(10)		10		$3 \cdot 10^6$	All elements have same weight
SYN-PER-02	UNIHD(15)		15		$1.3 \cdot 10^{12}$	All elements have same weight
SYN-PER-03	UNIHD(30)		30		$2.7 \cdot 10^{32}$	All elements have same weight
SYN-PER-04	nug12	chr12a	12		$4.8 \cdot 10^8$	Elements have different weight
SYN-PER-05	ulysses16	nug16a	16		$2.1 \cdot 10^{13}$	Elements have different weight
SYN-PER-06	tho30	nug30	30		$2.7 \cdot 10^{32}$	Elements have different weight

5.3 Results

This section presents the results of the evaluation of the Bayesian optimization framework as applied to a library of multi-objective SAPs. The interpretation of the boxplot graphs in the next subsection is similar to the ones in the previous chapter (see Section 4.4), with the only difference being that the objective is now to *maximize* the value of the function, since the normalized hypervolume metric (i.e., volume enclosed by the Pareto front and a reference point) is being used to quantify the quality of the solutions, and higher values for this metric represent higher-quality solutions.

5.3.1 Overview of the different decision patterns

Assigning Pattern



(a) SYN-ASG-01 (b) SYN-ASG-02 (c) SYN-ASG-04

(d) SYN-ASG-05 (e) SYN-ASG-06 (f) SYN-ASG-08

Figure 5-1: Selected results for the multi-objective assigning pattern. Bigger values are better.

Figure 5-1 shows the results for six problems for the assigning pattern. As was the case for the single-objective assigning pattern in Section 4.4.1, it can be observed that BO-RBF and BO-ARBF are superior to other methods for all the multi-objective problems displayed. Moreover, both of them achieved very similar performance across all problems, being able to produce solutions with 10-20% larger hypervolume than the rest of the algorithms. In some instances (SYN-ASG-05 and SYN-ASG-06), the resulting Pareto front is superior to the "true" Pareto front, which is possible since

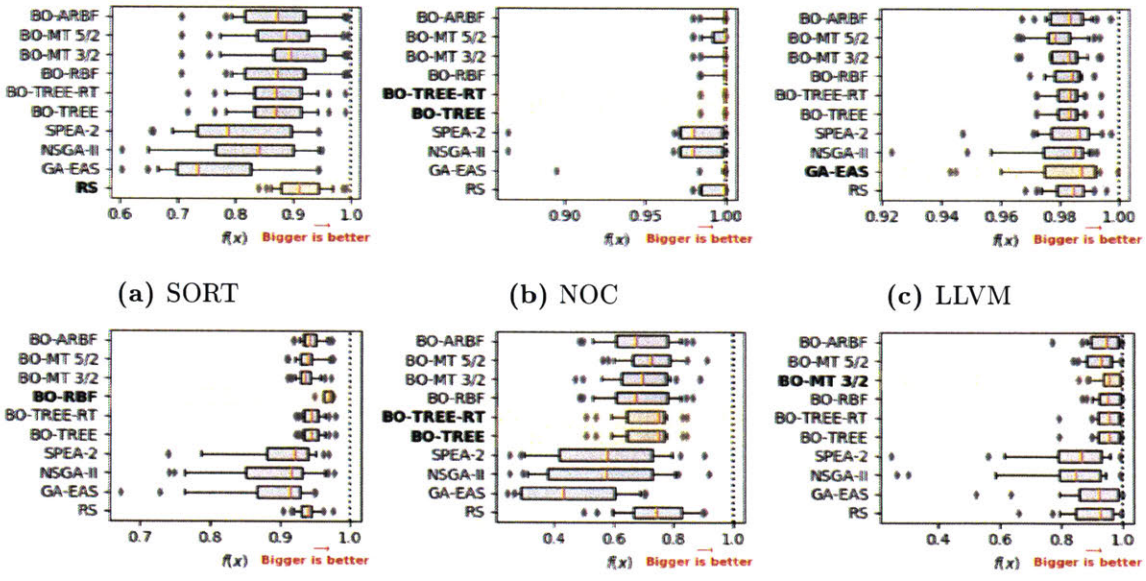
the "true" Pareto front was computing by *just* evaluating 300,000 points using a genetic algorithm. Thus, this only indicates that for the problems herein considered, evaluating 100 points with BO-RBF produces better results than those obtained by evaluating 300,000 points with a GA.

Combining pattern

Figure 5-2 shows the results for six of the combining problems analyzed. It can be seen that the genetic algorithms exhibit the worst performance and the largest variance in the results, which shows the subpar performance of these methods in scenarios where a limited number of function evaluations are allowed. Regarding the best methods, it can be observed that:

i The performance of all the BO-methods and RS is comparable across problems

ii The best method greatly varies from problem to problem.



(a) SORT (b) NOC (c) LLVM

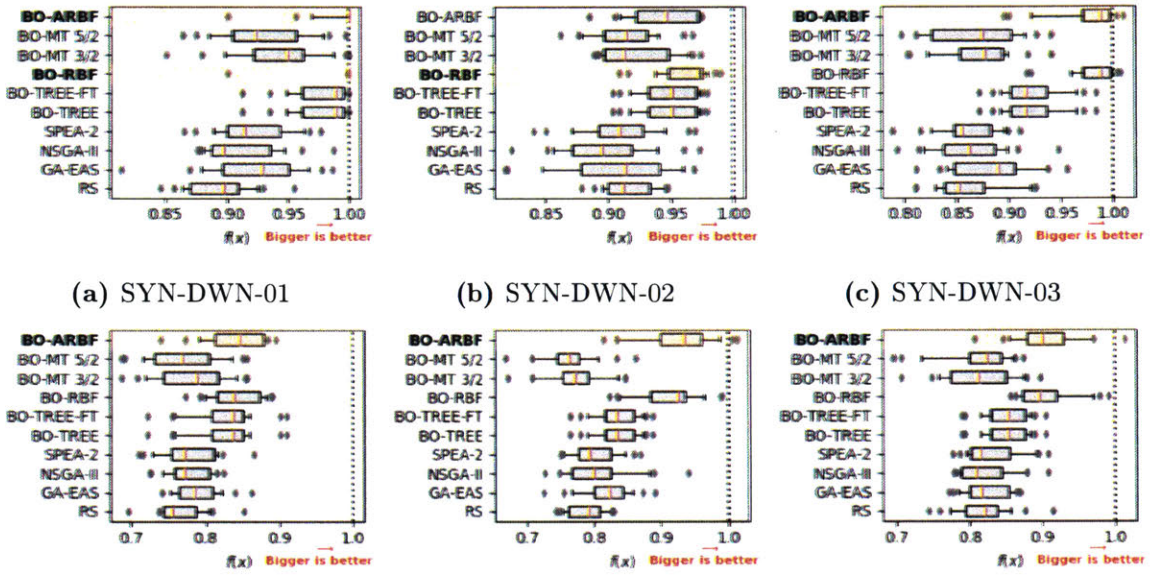
(d) WC-5d-c5 (e) WC-c3-3d-c1 (f) WC+sol-3d-c4

Figure 5-2: Selected results for the multi-objective combining pattern. Bigger values are better.

Downselecting Pattern

Figure 5-3 shows the results for six selected instances of the multi-objective downselecting pattern. The BO-ARBF method (which uses the Hamming distance) offers the best results for five out of the six problems depicted, followed by BO-RBF which

ranks first or second in all problems. In all instances but SYN-DWN-02, these two Bayesian optimization method offer significantly better results (5% to 15% superior) than random sampling and GA-based methods. Moreover, their variance is smaller than other methods for SYN-DWN-01 and SYN-DWN-03, which is a sign of robustness. Finally, not that for the BO algorithms that use the Matérn-kernel performance was worse than or equal to random sampling and the genetic algorithm based methods.



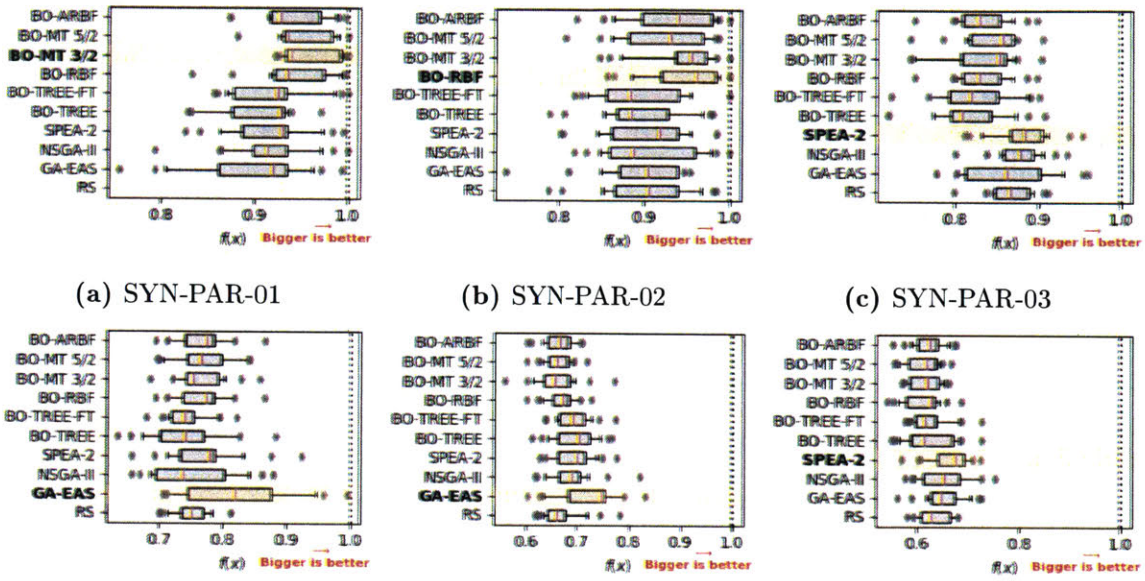
(a) SYN-DWN-01 **(b)** SYN-DWN-02 **(c)** SYN-DWN-03

(d) SYN-DWN-04 **(e)** SYN-DWN-06 **(f)** SYN-DWN-07

Figure 5-3: Selected results for the multi-objective downselecting pattern. Bigger values are better.

Partitioning Pattern

Figure 5-4 shows the results for the multi-objective partitioning problems considered. For this pattern, different methods produced the best performance for each of the problems. For small-dimensional problems with reduced input spaces (SYS-PAR-01 and SYN-PAR-02, both with input dimension of 7), BO methods are superior to the other methods; whereas for large-dimensional problems, all of the methods offer comparable performance, with genetic algorithm-based methods having just a slight edge. Thus, one can conclude that there is no single optimization method for the partitioning pattern that clearly dominates other.



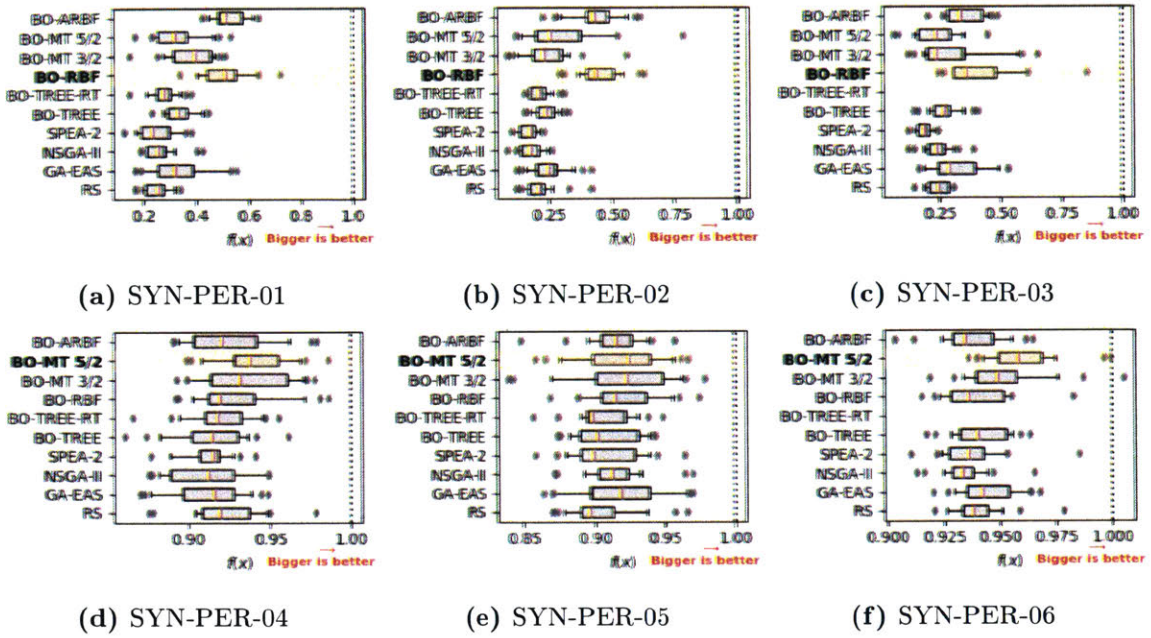
(a) SYN-PAR-01 (b) SYN-PAR-02 (c) SYN-PAR-03

(d) SYN-PAR-04 (e) SYN-PAR-06 (f) SYN-PAR-07

Figure 5-4: Selected results for the multi-objective partitioning pattern. Bigger values are better.

Permuting Pattern

Figure 5-5 shows the results for six selected test instances of the multi-objective permuting pattern. In three of them, BO-RBF produced the best results, and BO-MT52 was the best option for the other three the cases, even though for some problems (SYN-PER-04, SYN-PER-05) the performance of all methods was comparable, and in fact, BO-MT 5/2 showed a rather large variance. For some of the problems (SYN-PER-01, SYN-PER-2), the Bayesian optimization methods produced results which were better (by a factor of $\sim 2\times$) than the other methods, which obtained results very far from the "true" Pareto front (with normalized hypervolume values below 0.3); whereas for the rest of the problems, the improvements were less significant, given the large variance of the solutions.



(a) SYN-PER-01 (b) SYN-PER-02 (c) SYN-PER-03

(d) SYN-PER-04 (e) SYN-PER-05 (f) SYN-PER-06

Figure 5-5: Selected results for the multi-objective permuting pattern. Bigger values are better.

5.3.2 Ranking of methods

A methodology similar to the one described in Section 4.4.2 is used to rank the different optimization methods in the multi-objective scenario. Tables 5.2 - 5.6 show the ranking as assigned by the average ranking (AR), success rate ratio (SRR), and significant wins (SW) metrics for each of the patterns analyzed.

It can be observed that Bayesian optimization with a radial basis function kernel (BO-RBF) is the preferred method for multi-objective assigning and downselecting patterns, being the winner based on all three metrics. BO-ARBF presents a comparable performance, being just slightly inferior in the AR metric but ranking with very similar scores in the SRR and SW metrics.

For combining patterns, the tree-based method offers the best results, even though their performance is not far from random sampling and other BO-based methods. All BO-methods and RS show superior performance to GA-based methods, but there is no clear method that outperforms the others.

In the partitioning pattern, the GA-based methods (SPEA2 and GA-EAS) produced better results than the Bayesian optimization methods, although there is no single method that is clearly superior to the others (according to the significant win ratio, all methods perform comparably). As was already discussed in Section 5.3.1, Bayesian optimization methods work best for small input spaces, whereas GA-based methods tend to produce better performance in problems with a large number of

decisions.

Finally, for the permuting problem, Bayesian optimization with the 5/2 Matérn kernel is the highest-ranked method according to the average ranking metric, whereas the BO-RBF and BO-ARBF rank first and second (respectively) according to the SRR and SW metrics. Given the more mathematical approach of these two methods, and the fact that they account for the statistical significance of the differences between algorithms (as described in-depth in Appendix C), BO-RBF is preferred.

Table 5.2: Ranking for multi-objective assigning pattern algorithms.

	BO-RBF	BO-ARBF	BO-TREE-FT	BO-TREE	GA-EAS	BO-MT 5/2	NSGA-II	BO-MT 3/2	SPEA-2	RS
AR	1.25 [1]	1.62 [2]	3.38 [3]	4.75 [4]	5.12 [5]	6.88 [6]	7.12 [7]	7.38 [8]	7.88 [9]	9.12 [10]
SRR	1.13 [1]	1.11 [2]	1.01 [3]	0.99 [4]	0.99 [5]	0.97 [6]	0.97 [8]	0.97 [7]	0.96 [9]	0.95 [10]
SW	0.90 [1]	0.80 [2]	0.60 [3]	0.50 [4]	0.50 [4]	0.10 [6]	0.10 [6]	0.10 [6]	0.00 [9]	0.00 [9]

Table 5.3: Ranking for multi-objective combining pattern algorithms.

	BO-TREE	BO-TREE-RT	RS	BO-RBF	BO-MT 3/2	BO-ARBF	BO-MT 5/2	SPEA-2	GA-EAS	NSGA-II
AR	3.44 [1]	3.56 [2]	3.67 [3]	3.89 [4]	4.56 [5]	5.00 [6]	5.33 [7]	7.22 [8]	7.56 [9]	7.67 [10]
SRR	1.06 [2]	1.06 [2]	1.06 [1]	1.05 [5]	1.06 [4]	1.05 [7]	1.05 [6]	0.96 [8]	0.84 [10]	0.95 [9]
SW	0.30 [2]	0.30 [2]	0.40 [1]	0.30 [2]	0.30 [2]	0.30 [2]	0.30 [2]	0.00 [8]	0.00 [8]	0.00 [8]

Table 5.4: Ranking for multi-objective downselecting pattern algorithms.

	BO-RBF	BO-ARBF	BO-TREE	BO-TREE-FT	GA-EAS	BO-MT 5/2	BO-MT 3/2	NSGA-II	RS	SPEA-2
AR	1.43 [1]	1.57 [2]	2.86 [3]	2.86 [3]	5.71 [5]	6.86 [6]	7.00 [7]	8.14 [8]	8.43 [9]	8.43 [9]
SRR	1.08 [1]	1.08 [2]	1.03 [3]	1.03 [3]	0.98 [5]	0.97 [7]	0.97 [6]	0.96 [9]	0.96 [10]	0.96 [8]
SW	0.80 [1]	0.80 [1]	0.60 [3]	0.60 [3]	0.30 [5]	0.00 [6]	0.00 [6]	0.00 [6]	0.00 [6]	0.00 [6]

Table 5.5: Ranking for multi-objective partitioning pattern algorithms.

	SPEA-2	GA-EAS	BO-RBF	BO-ARBF	BO-MT 3/2	BO-MT 5/2	RS	NSGA-II	BO-TREE	BO-TREE-FT
AR	2.86 [1]	3.43 [2]	4.29 [3]	4.57 [4]	5.71 [5]	5.86 [6]	6.00 [7]	6.00 [7]	7.29 [9]	8.14 [10]
SRR	1.03 [2]	1.04 [1]	1.00 [3]	1.00 [4]	0.99 [6]	0.99 [7]	0.99 [8]	1.00 [5]	0.98 [9]	0.98 [10]
SW	0.50 [1]	0.30 [2]	0.00 [3]	0.00 [3]	0.00 [3]	0.00 [3]	0.00 [3]	0.00 [3]	0.00 [3]	0.00 [3]

Table 5.6: Ranking for multi-objective permuting pattern algorithms.

	BO-MT 5/2	BO-MT 3/2	BO-RBF	BO-ARBF	GA-EAS	BO-TREE	RS	BO-TREE-RT	NSGA-II	SPEA-2
AR	2.50 [1]	2.83 [2]	3.17 [3]	3.50 [4]	4.17 [5]	5.83 [6]	7.00 [7]	8.00 [8]	8.17 [9]	8.67 [10]
SRR	1.03 [4]	1.13 [3]	1.49 [1]	1.42 [2]	1.03 [5]	0.98 [6]	0.89 [8]	0.91 [7]	0.89 [9]	0.85 [10]
SW	0.20 [4]	0.10 [5]	0.60 [1]	0.40 [2]	0.30 [3]	0.00 [6]	0.00 [6]	0.00 [6]	0.00 [6]	0.00 [6]

5.3.3 Computational speedup

As in the previous chapter, for each decision pattern, I assess the extra number of function calls required to achieve similar performance to the one achieved by the

recommended optimization method. Figure 5-6 shows the ratio between the average number of function evaluations required by required by different methods and the average number of function evaluations of the reference algorithm.

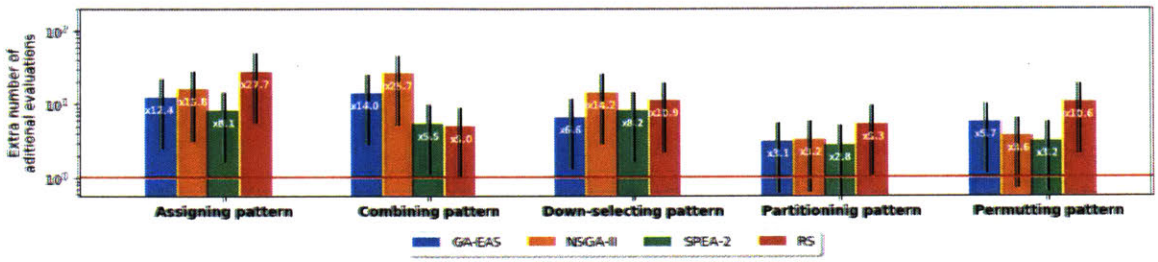
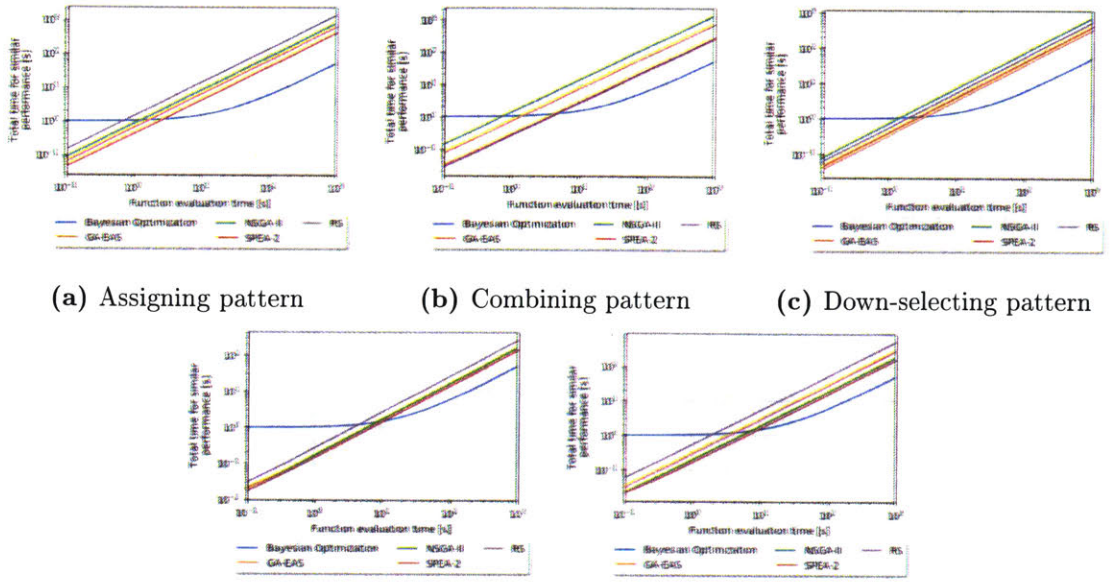


Figure 5-6: Computational speedups of the recommended method when compared against other optimization methods for multi-objective problems. The computational speedup is measures as the number of additional function evaluations required by other methods to achieve a similar performance to the recommended method.

It can be observed that the speedup factor (in terms of extra function evaluations) ranges from 2.8× for the partitioning pattern to 8.1× for the assigning pattern. As one could expect, random sampling performs the worst for all patterns but for the combining pattern, where it is indeed the best performing of the non-BO methods. This behavior was also observed in the one-dimensional problems, as the tight budget of function evaluations limits the genetic algorithm's ability to explore the space extensively, which causes early convergence leading to sub-optimal solutions. This phenomena has been noted in previous literature whenever one of the dimensions of the input space dominates the function output, for example, in the field of hyper-parameter tuning [237].

Finally, Figure 5-7 shows the computational speedup achieved for each of the five patterns analyzed, in terms of the function evaluation time. For multi-objective problems, Bayesian optimization has an overhead (adjusting the model and maximizing the acquisition function) of 20-120 seconds. Thus, if the function evaluation time is very short, it might not make sense to use Bayesian optimization. As shown in the graph, in general there are no benefits to using Bayesian optimization if the function evaluation time is shorter than 30 seconds. For longer evaluation times, however, Bayesian optimization will help achieve an overall reduction in the computation time. Similar to the single-objective case, the asymptotical benefit (i.e., when the function evaluation time is much higher than the overhead) is proportional to the speedups obtained in Figure 5-6.



(a) Assigning pattern (b) Combining pattern (c) Down-selecting pattern

(d) Partitioning pattern (e) Permuting pattern

Figure 5-7: Computational speedups of Bayesian optimization algorithms when applied to multi-objective problems.

5.4 Chapter conclusions

In this Chapter, the performance of Bayesian optimization when applied to multi-objective System Architecture problems was discussed. The conclusions are as follows:

Research Question 5.1

What method to adapt Bayesian optimization to combinatorial problems works best for each of the decision patterns common in SAPs?

Unlike for single-objective problems, there is not a one-size-fits-all BO method (or family of methods) that can be applied to all decision patterns. For the assigning and downselecting patterns, BO-RBF, used together with the Hamming distance, proved to be clearly superior to other methods. For the permuting pattern, the AR metric revealed BO-MT 5/2 as the best method, but the SRR and SW metrics ranked BO-RBF first. Given that the latter two metrics provide a more rigorous mathematical comparison (as explained in depth in Appendix C), BO-RBF is preferred. For the partitioning pattern GA-based methods were preferred over the others; In particular, SPEA-2 showed the best performance. Finally, for the combining pattern, the BO methods and RS were superior to GA-based methods, but there was no clear winner among them.

Research Question 5.2

What are the time savings achieved by using Bayesian optimization in comparison to other methods used in the literature?

As in the previous chapter, Bayesian optimization only produces time savings if the objective functions are expensive to evaluate. For multi-objective SAPs, since 1) there is an individual model fitted to each of the function objectives, and 2) the acquisition functions are more expensive to evaluate (as the EHI needs to be computed), the average overhead of BO is higher (between 30 to 120 seconds depending on the pattern). Thus, Bayesian optimization is only recommended for scenarios where the function evaluation time is greater than 30 seconds. In those cases, time savings range between 2.3× (for the partitioning pattern) and 8.1× (for the assignment pattern) in an iso-performance setting.

Table 5.7 summarizes these recommendations.

Table 5.7: Summary of recommendations for multi-objective combinatorial patterns.

	Method	Distance function	Speedup
Assigning	BO-RBF	Hamming	8.1x
Combining	BO-TREE, BO-RBF, RS	Hamming	5x
Downselecting	BO-RBF	Hamming	6.6x
Partitioning	SPEA2, GA-EAS	-	2.8x
Permuting	BO-RBF, BO-MT3/2	Hamming	3.2x

Part II

Expanding connectivity using novel space and aerial concepts

Chapter 6

The connectivity problem

The second part of this thesis is devoted to studying the impact that space and aerial concepts can have in bringing people online, as stated in Research Question 1.

Research Question 1

What kind of novel space and aerial systems can complement existing infrastructure and contribute towards expanding global connectivity at an affordable cost? What is the potential impact of such systems in terms of connecting additional populations?

To consider novel solutions, it is imperative to first have a basic understanding of the current state of connectivity as well as the main barriers to adoption. Despite the remarkable advancements made in terms of connectivity within the last ten years (driven mainly by the widespread adoption of cellular communications), almost half of the world's population still lacks connectivity. Termed as the digital divide, this is considered as one of the most pressing issues that need to be resolved by the international community in order to guarantee equal access to opportunity. Increasing access to connectivity infrastructure and basic technical tools is deemed an essential step in accomplishing the UN Sustainable Development goals, since greater Internet access has a direct impact in effecting better health care and education, higher economic development, and greater gender equality outcomes[1].

To raise awareness of the importance of expanding broadband connectivity and

[1]Studies [2] have shown that for low-income countries, every 10% increase in *mobile* broadband penetration results in a 2% increase in GDP. Interestingly, for low-income countries, the studies did not find direct correlations between higher *fixed* broadband penetration and GDP increases. Even more surprising, for high-income countries, the effect is exactly opposite, with increases in *mobile* broadband penetration having no correlation with GDP increases, but increases in *fixed* broadband penetration having an impact (every 10% increase in fixed-broadband penetration renders a 1.4% increase in GDP).

boost the issue's prominence within the international policy agenda, the ITU and UNESCO set up the *Broadband Commission for Digital Development* in 2010, whose main tasks include advocating at the international level for the expansion of broadband, defining target connectivity values to be met, and conducting progress monitoring, evaluation, and reporting activities. The latest targets defined in 2018 [11] are as follows:

Broadband Commission for Digital Development targets for 2025

1. By 2025, all countries should have a funded national broadband plan or strategy, or include broadband in their universal access and services definition.
2. By 2025, entry-level broadband services should be made affordable in developing countries, at less than 2% of monthly GNI per capita.
3. By 2025 broadband-Internet user penetration should reach:
   - 75% worldwide
   - 65% in developing countries
   - 35% in least developed countries
4. By 2025, 60% of youth and adults should have achieved at least a minimum level of proficiency in sustainable digital skills.
5. By 2025, 40% of the world's population should be using digital financial services.
6. By 2025, unconnectedness of Micro-, Small- and Medium-sized Enterprises should be reduced by 50%, by sector.
7. By 2025, gender equality should be achieved across all targets.

This dissertation focuses on technical solutions to tackle the issues related to the second and third targets. Notably, the second target sets a threshold that defines what is considered *affordable* broadband service. The specific value of the threshold has changed over time; for the targets set prior to 2018, the threshold was 5% of the monthly GNI per capita, but that was determined to be too high for widespread adoption to become a reality.

The next two sections provide an overview of the current state of connectivity and the main barriers which prevent widespread adoption of broadband Internet connectivity, whereas the last part of this chapter is devoted to introducing the methodology to be used in future chapters to evaluate the impact of space and aerial networks in expanding global connectivity.

6.1 Current state of connectivity

In the past decade, the number of Internet users[2] has more than doubled, reaching 4.1 billion people (~53.6% of the world's population) by the end of 2019. This growth has been primarily fueled by the Asia-Pacific region, which accounted for 70% of new Internet users in the last five years. However, reports suggest that the rate of growth has begun to slow down [240, 241, 242], which makes it unlikely that target 3 of the UN Broadband Commission will be met. Out of the 4.1 billion connected users, 1.1 billion are **fixed** broadband users, whereas 3.5 billion are **mobile** broadband users [241]. Particularly in low- and middle-income countries, 57% of the population access the Internet exclusively using cellphone wireless connections [243]; in Southeast Asia, this figure is as high as 75%, and in countries like Myanmar 94% [243].

Among those not connected to the Internet (i.e., 46.4% of the population), ~70% live in areas where mobile broadband connectivity (3G/4G technology) is available. Figure 6-1 shows the distribution of offline users by region and country[3]. The area of each wedge is proportional to the absolute number of people unconnected, whereas colors indicate the percentage of the population with broadband coverage on each country. The image highlights that the usage gap (i.e., those living in areas covered by mobile broadband networks but are not connected) is several times larger than the coverage gap (i.e., those living outside of areas covered by mobile broadband networks), which suggests that there are other factors other than the lack of network infrastructure preventing people from becoming online (as discussed in the next section).

6.2 A summary of the barriers to connectivity

Pursuing global connectivity requires that we understand the main reason why people remain offline. Numerous studies [8, 9, 244, 245] have concluded that barriers to adoption can be classified into four groups: lack of infrastructure, affordability, readiness, and relevance. These barriers rarely exist in isolation, and overcoming them would require both technical- and policy-oriented actions to be carried out collaboratively by all the stakeholders involved (industry, governments, NGOs, and non-profits).

In this section, each of the four kinds of barriers mentioned is briefly analyzed. The objective is to provide an overview that allows the reader to understand and contextualize the rest of this dissertation.

[2]The number of Internet users includes anyone who used the Internet from any location for any purpose in the last months, as reported by the ITU.

[3]This plot was inspired by the one in https://www.bloomberg.com/news/features/2019-06-07/the-next-big-phones-could-bring-a-billion-people-online (Accessed on September 2019)

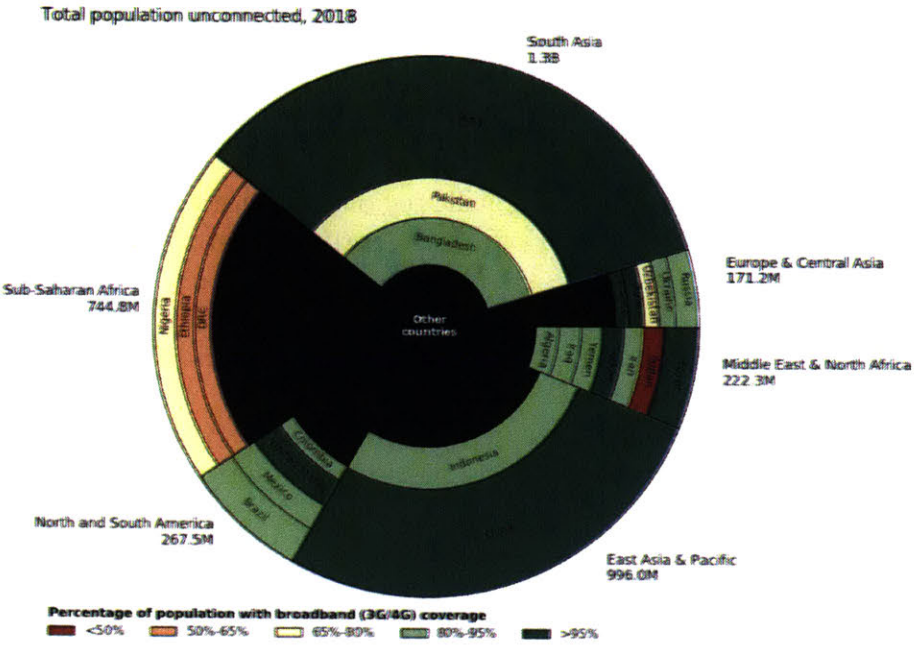


Figure 6-1: Distribution of offline users vs. percentage of broadband coverage at the end of 2018. The area of each of the wedges is proportional to the population unconnected. Data sources: ITU (population unconnected) and GSMA Connectivity index (% of broadband coverage).[2]

6.2.1 Lack of infrastructure

Lack of infrastructure is one of the major blockers preventing ubiquitous mobile connectivity access. Although infrastructure is usually associated with coverage, other obstacles prevent the rollout of new infrastructure, such as the lack of electricity, lack of road access, and underdeveloped (or outdated) backhaul and backbone networks. These obstacles are particularly prevalent in remote areas because of the challenging terrain, large capital expenditures, higher operating costs, and lower average revenue per user (ARPU).

The rest of this section is devoted to exploring how various infrastructure issues affect connectivity.

Mobile coverage

To determine the mobile coverage over inhabited areas, data from the Gridded Population of the World (GPWv4) was combined with the cellular tower database from OpenCellId[4]. The GPWv4 dataset contains estimated population counts in 2020 over a 30 arc-seconds by 30 arc-seconds grid; whereas the OpenCellId database contains

[4]Available at www.opencellid.org. Accessed on 11 July 2018

the location, tower ID, and technology used (GSM, CDMA, UMTS, or LTE) by more than 42 million cellphone towers in 219 countries.

Figure 6-2 shows the *most common* mobile technology serving inhabited regions of the world (as of 2018). This map shows several differences between regions in terms of technology used: 4G is the most common technology only in very developed nations and in high population-density areas, whereas 3G is the most common technology worldwide. Furthermore, in low and mid-income countries, there exist large inhabited areas without connectivity, and where 2G connectivity is available, it is the most commonly-used technology.

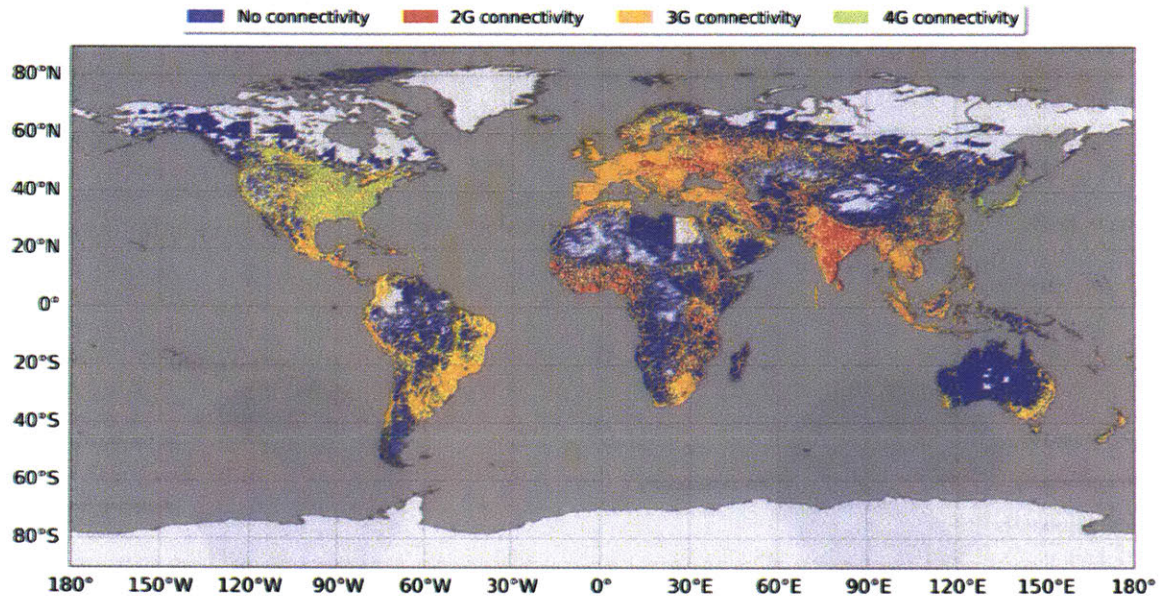


Figure 6-2: World map of the most common generation of mobile technology by region. (Only inhabited areas are colored).

Figure 6-3 shows the *best* technology available over inhabited regions of the world. As with Figure 6-2, we see how most of the world's inhabited area is covered by mobile broadband technology (i.e., 3G or 4G), which leads us to consider other barriers to connectivity beyond the lack of infrastructure. This realization is one of the more important takeaways from this section. Although most space and aerial companies have as their main missions broad goals about closing the digital divide, the data shows that close to 90% of the world's population is already covered by broadband networks (3G/4G).

The relationship between 3G and 4G technology coverage and population density is better understood when looking at Figure 6-4a and Figure 6-4b, which show the percentage of the population covered by each technology and the percentage of inhabited areas covered by each technology respectively. Although 78.2% of the population (5.7 billion people) live in an area where there is 4G technology available, the area they occupy is only 28.2% of the total inhabited area. In contrast, those who

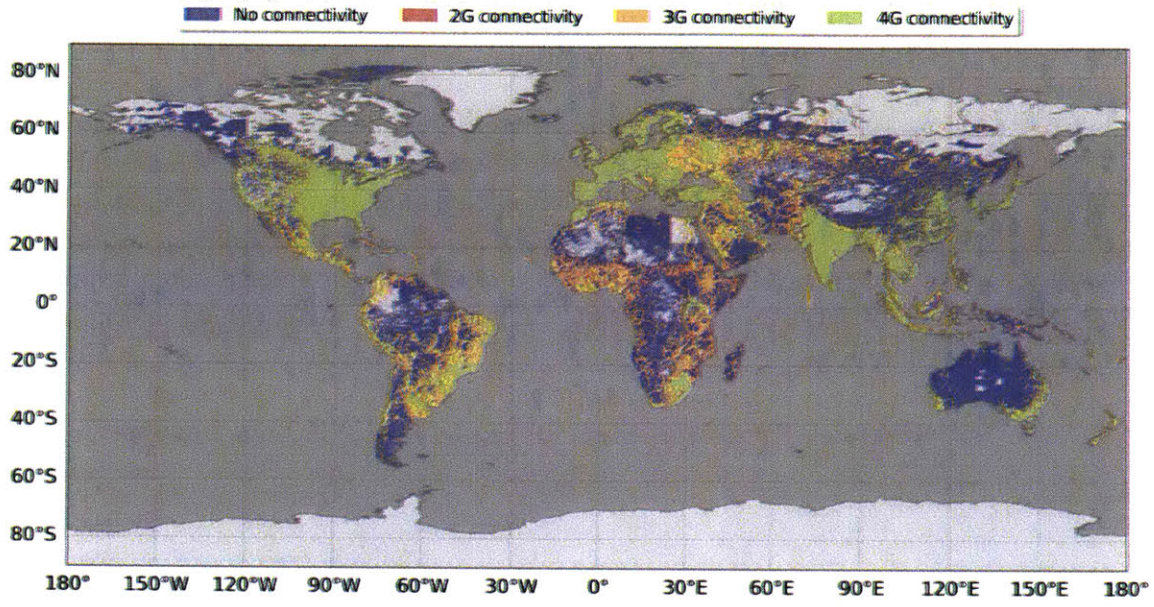


Figure 6-3: World map of most best mobile technology available at a worldwide level over populated areas.

(a) Population

(b) Inhabited area

Figure 6-4: Fraction of a) population and b) inhabited area by best mobile technology available.

live in areas with no mobile coverage only represent 10.6% of the population (750 million people), and yet it would be necessary to cover 54.6% of the inhabited land area to provide service to them. Therefore, the challenge presented by such sparsely distributed populations is evident.

Electrical grid infrastructure

Electricity is an essential requirement for connectivity, from both the deployment and user perspectives, since base stations need to be powered and users require electricity to charge their handset terminals or CPE.

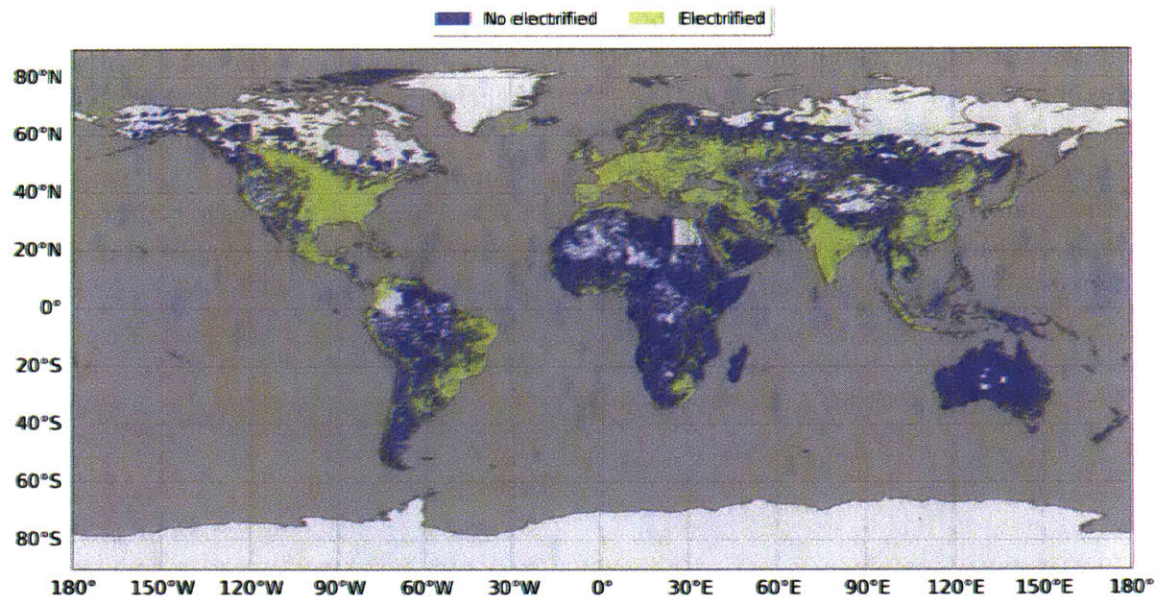


Figure 6-5: Presence of electricity infrastructure at a worldwide level over populated areas.

Figure 6-5 shows a map indicating availability (or lack thereof) of lack of electrical infrastructure, estimated using NOAA's Global DMSP-OLS Nighttime Lights Time Series[5] [246] and VIIRS Cloud Mask - Nighttime Lights[6] [247] datasets. It can be observed that the lack of electricity is most pronounced in sub-Saharan Africa, where, according to WorldBank data, more than half of the population (or about 570 million people), do not have regular access to electrical power. Another 170 million people, or almost a tenth of the population of South Asia, also live off the grid. Where electrical infrastructure is in place, the cost of access can still be a barrier to connectivity. Overall, according to the International Energy Agency, in 2017, there were 840 million people without access to electricity [248].

Road infrastructure

The lack of road access is another common challenge which prevents infrastructure deployment. In inaccessible regions, material transportation needs to be done by hand, or by aerial means, which increases costs and deployment timeframes signifi-

[5]This dataset consists of cloud-free grid-composites (at 30 arc-seconds × 30 arc-seconds resolution) made using all the available images taken by the OLS instrument in the US Air Force DMSP program in a given calendar year. Each pixel of the dataset represents the average radiance at night of that point in a 0-63 range. A threshold value of 2 was used as the criteria to determine the presence or lack of electricity

[6]The VIIRS Cloud Mask - Nighttime Lights is a cloud-free grid-composite (at 15 arc-seconds × 15 arc-seconds resolution) made using the available images taken by the VIIRS instrument in a given calendar year. Each pixel of the dataset represents the average radiance at night of that point in $nW/cm^2/sr$. The `vcm-orm-ntl` dataset, in which outliers have been removed and background noise has been set to zero, was used to determine the presence or lack of electricity.

cantly. Figure 6-6 charts the regions on Earth with existing road infrastructure. This image was produced using the Open Street Maps dataset, which contains worldwide geographical road information. On the Figure, each pixel corresponds to a 0.05 degrees by 0.05 degrees tile (approximately 5.5 km $\times$ 5.5 km at the Equator), and is classified as either having or lacking road infrastructure depending on whether there are any roads within the tile. When compared to Figure 6-5, it can be seen that the lack of road infrastructure is less prevalent than the lack of electricity, but still a challenge in regions of sub-Saharan Africa and South America.

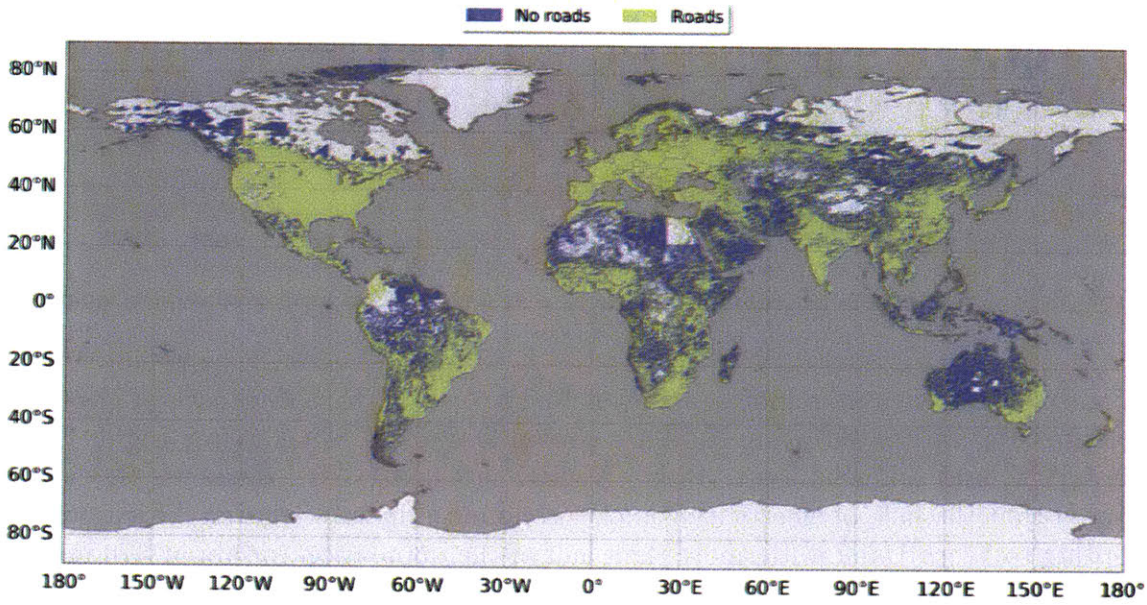


Figure 6-6: Presence of road infrastructure at the worldwide level over populated areas.

6.2.2 Affordability

Affordability is defined as the cost of connectivity relative to income. The severity of this barrier is driven by the users' disposable incomes and the costs of cellphone devices and data package services. Historically, the device costs have been a significant barrier to access for lower-income groups, but the proliferation of low-cost terminals (with purchase prices as low as $20 in some developing markets[7]) within the last few years has progressively reduced this barrier. The prices of the data services, on the other hand, remain the main barrier to access in most of sub-Saharan Africa.

As stated in the *Broadband Commission* targets, a connectivity service is considered affordable if its price is below 2% of monthly GNI per capita. Therefore, it is critical to take into account the income distribution across and within countries when estimating the potential impact of new connectivity technologies. Figure 6-7 shows the distribution of unconnected populations by country, with colors mapping

[7] https://thenextweb.com/plugged/2019/02/26/a-20-phone-for-africa-is-mwcs-unlikeliest-hero/

to levels of GNI per capita. As can be observed, most of unconnected populations live in countries with GNI less than $5,000; in sub-Saharan Africa and South Asia in particular, GNIs per capita are below $2,000, which, based on the ITU threshold implies a maximum price of $3.30 per month for connectivity services to be considered affordable.

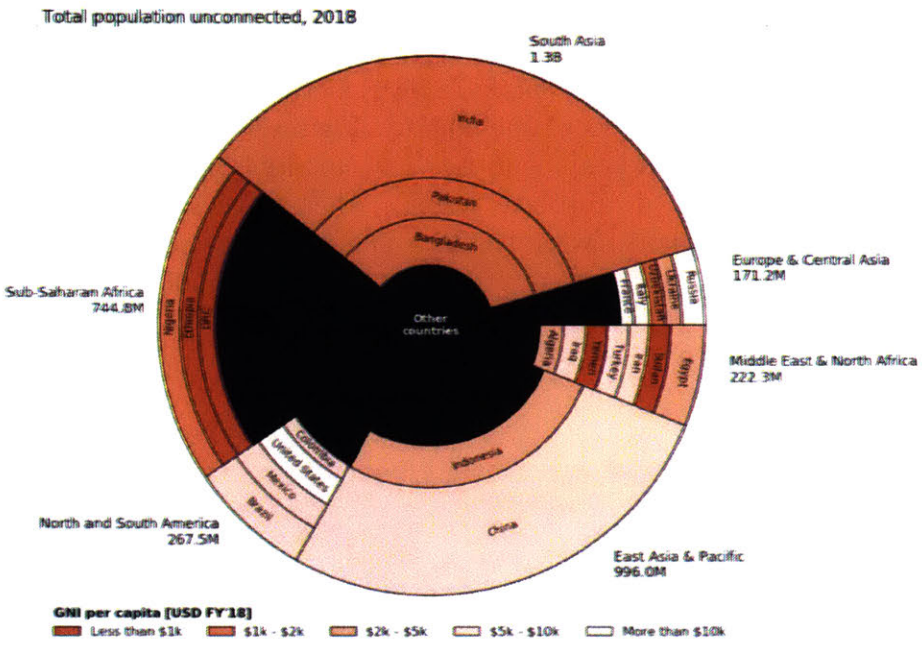


Figure 6-7: Distribution of offline users vs. GNI per capita. The area of each wedge is proportional to the population unconnected. Data sources: ITU (population unconnected) and WorldBank (GNI per capita).[2]

Broadly speaking, there are two types of broadband networks: fixed-broadband and mobile-broadband. Fixed-broadband refers to high-speed data transmission networks serving residential and business venues (i.e., fixed locations) by any means of technology. Commonly, DSL, fiber, cable-modem, fixed wireless, and satellite technologies are used. In contrast, mobile-broadband refers to high-speed data transmission networks that provide service to portable devices, normally cellular handsets. In this case, the technologies used are categorized by the release generation. In that sense, the second generation of technologies (2G) comprises GSM (including GPRS and EDGE); the third generation (3G) is mainly composed of UMTS and CDMA, and the fourth generation (4G) uses HSPA+, LTE (including LTE-Advanced), and WiMax technologies.

The costs of mobile- and fixed-broadband data plans are analyzed next.

The cost of mobile connectivity

This section analyzes the prices of basic monthly mobile data plans, in both absolute terms ($/month) and as a percentage of the GNI per capita for each country. For the rest of this thesis, a basic mobile plan is assumed to provide 1GB/month, and is considered affordable if it is cheaper than 2% of the average income threshold.

To estimate prices of basic mobile data plans in different countries, information on 6,000 data plans was extracted from three specialized websites: Prepaid Data SIM Card Wiki"[8], Cable.co.uk[9], and the Alliance for Affordable Internet (A4AI) index[10]. All data sources were fused into a single table that estimates the monthly price of a basic plan for 189 countries (see Appendix A). Figure 6-8 shows the average price of a 1 GB data plan for different countries. Countries colored in dark-blue have the cheapest mobile plans (in absolute terms), whereas those colored in red have relatively expensive data plans. Notably, there is great variation across countries: in India, Sudan, and Russia, the average price of a basic 1GB/month data plan can be less than $1, whereas, in countries like the United States, Chad, and Finland, the average price of a similar plan is more than $30/month.

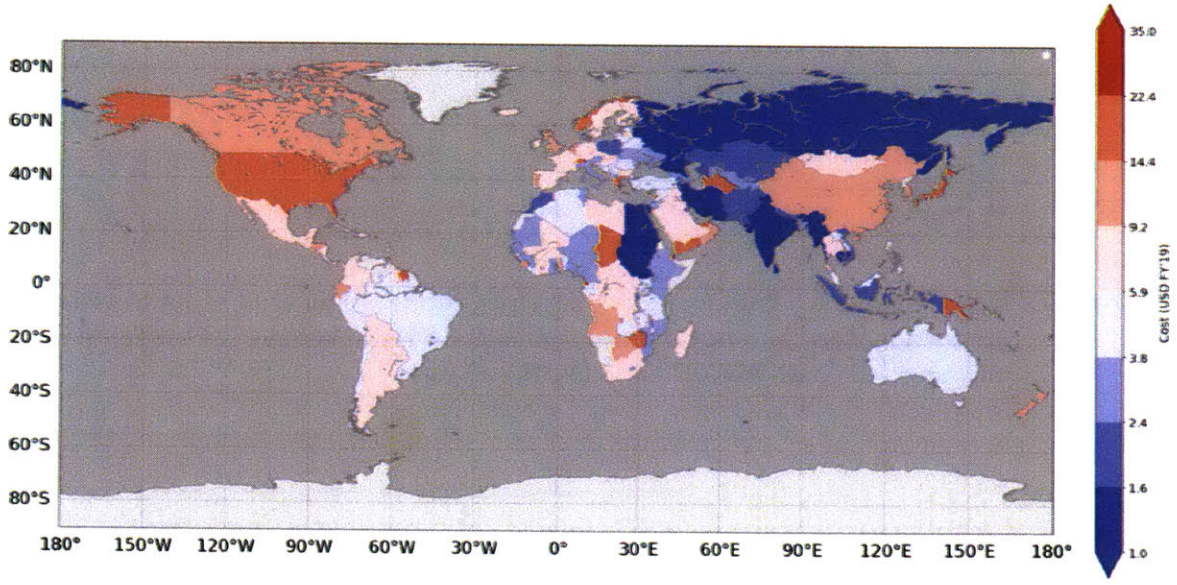


Figure 6-8: Price of a 1GB/month data plan

More tellingly, Figure 6-9 shows the price of a 1GB/month data plan as a percentage of the monthly GNI per capita. In blue-colored countries, the price of a 1GB monthly plan is below the 2% GNI per capita threshold, whereas in red-colored countries, the price is above this value. It can be observed that the price of a monthly plan in most of sub-Saharan Africa is well over 2% of GNI, whereas as a group, OECD

[8] https://prepaid-data-sim-card.fandom.com/

[9] http://cable.co.uk

[10] https://a4ai.org/extra/mobile_broadband_pricing_usd-2019Q2

countries have the most affordable data plans (with an average cost of just 0.98% of the monthly GNI per capita). It is also noteworthy how in many developing countries in South America, while the prices of data plans have dropped significantly in the last years (making them well under the 2% threshold), due to large income inequalities between urban and rural regions within these countries, there are still broad areas where connectivity remains out of reach for the majority of the population.

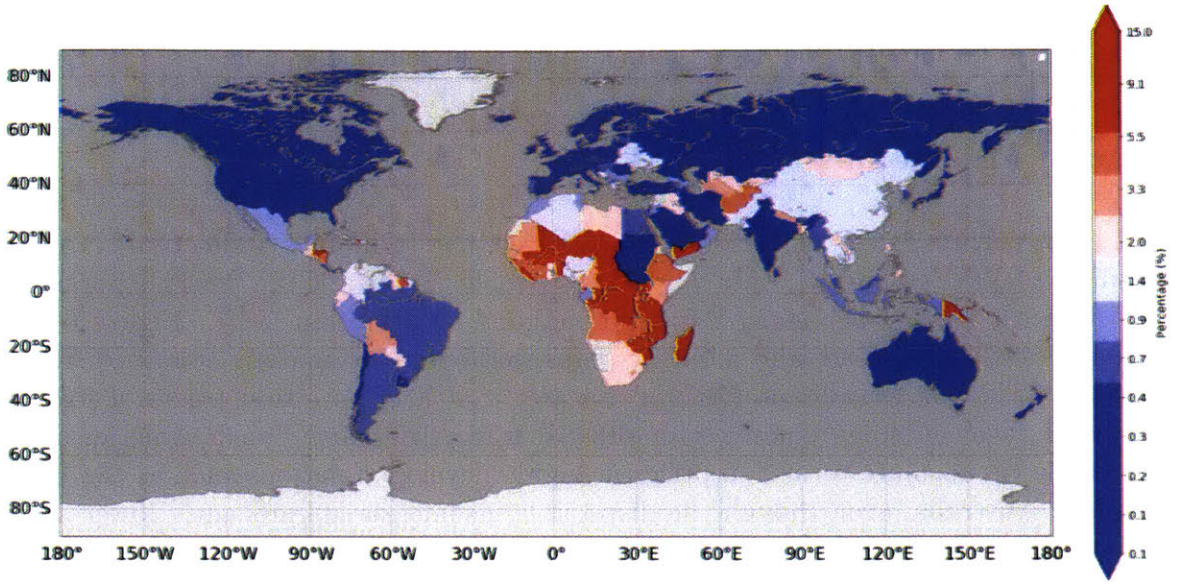


Figure 6-9: Price of a 1GB/month data plan as a percentage of the monthly GNI per capita.

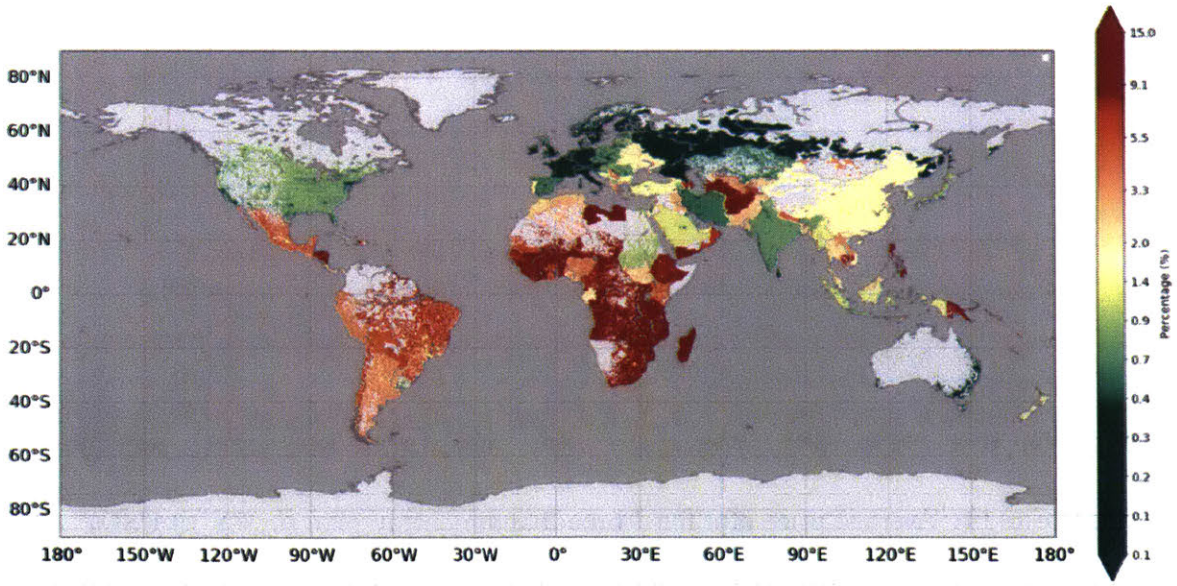


Figure 6-10: Price of a 1GB/month data plan as a percentage of the monthly income.

Figure 6-10 shows the price of a 1GB/month data plan as the percentage of the average income, taking into account the geospatial income distribution *within* each

country. The large differentials between rural and urban populations become evident in this image; in most countries in South America, where the average data plan prices are well below the 2% threshold, the price of a basic monthly data plan is above the affordability index is still unaffordable for populations in large regions of the country.

A similar trend is observed in many parts of Southeast Asia and sub-Saharan Africa, where connectivity is affordable for urban populations but less so for those living in rural areas. Unsurprisingly, these low-income rural regions correlate with regions where no mobile broadband connectivity is currently available.

The cost of fixed connectivity

Fixed connectivity user penetration is much lower than of mobile connectivity, with, according to ITU estimates, only 1.1 billion people (~14% of the world's population) currently having access to a fixed-broadband subscription [241]. Figure 6-11 shows prices of fixed-broadband monthly plans worldwide, using data extracted from Cable's "Worldwide Broadband Pricing" dataset[11]. It is of note that prices of fixed-broadband plans in some countries in sub-Saharan Africa and South America are comparable (in absolute terms) to prices in highly-developed nations such as the US, Sweden or Australia, even though the disposable incomes are significantly lower and the connection speeds offered also trend to be inferior.

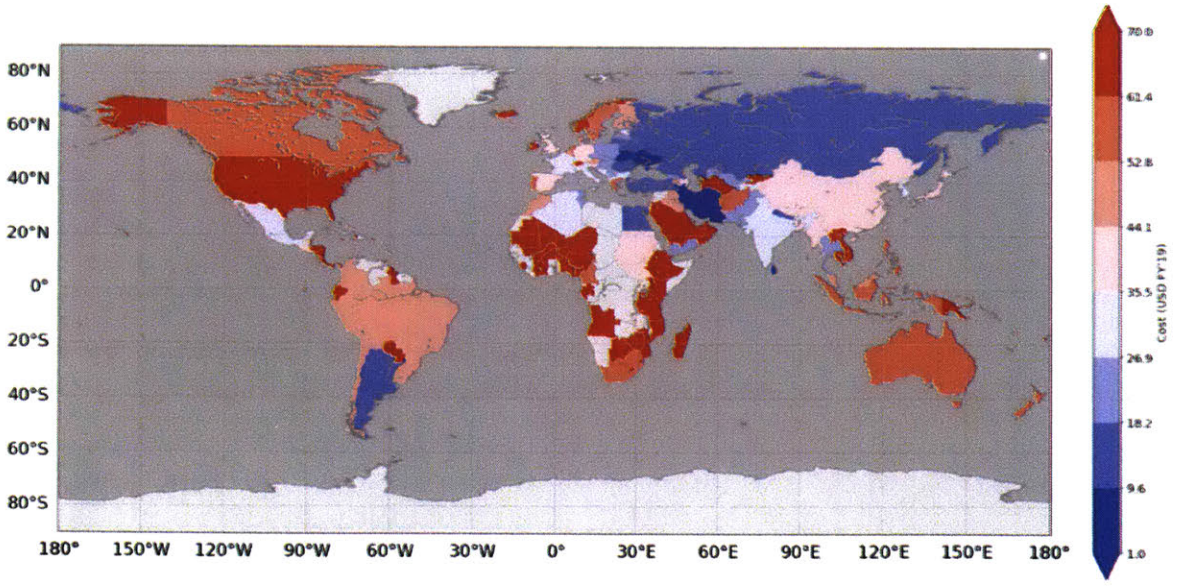


Figure 6-11: Average price of a fixed-broadband monthly plan (in FY'19 USD)

Figure 6-12 shows the price as a percentage of the monthly GNI per capita for each country, with blue-colored countries being those where a fixed-broadband plan costs

[11] See https://www.cable.co.uk/broadband/pricing/worldwide-comparison/ for raw data and a description of the methodology used to compute the values.

less than the ITU's 2% threshold, and red-colored countries being where the price is above the threshold. Here the difference is starker; fixed-broadband is only affordable in highly-developed countries and largely unaffordable in most of Southeast Asia and Africa, where costs can represent more than 20% of the average monthly income. In countries like Ethiopia, Tanzania, Mozambique, and Niger, costs can be as high as more than twice the monthly average income (>200%)

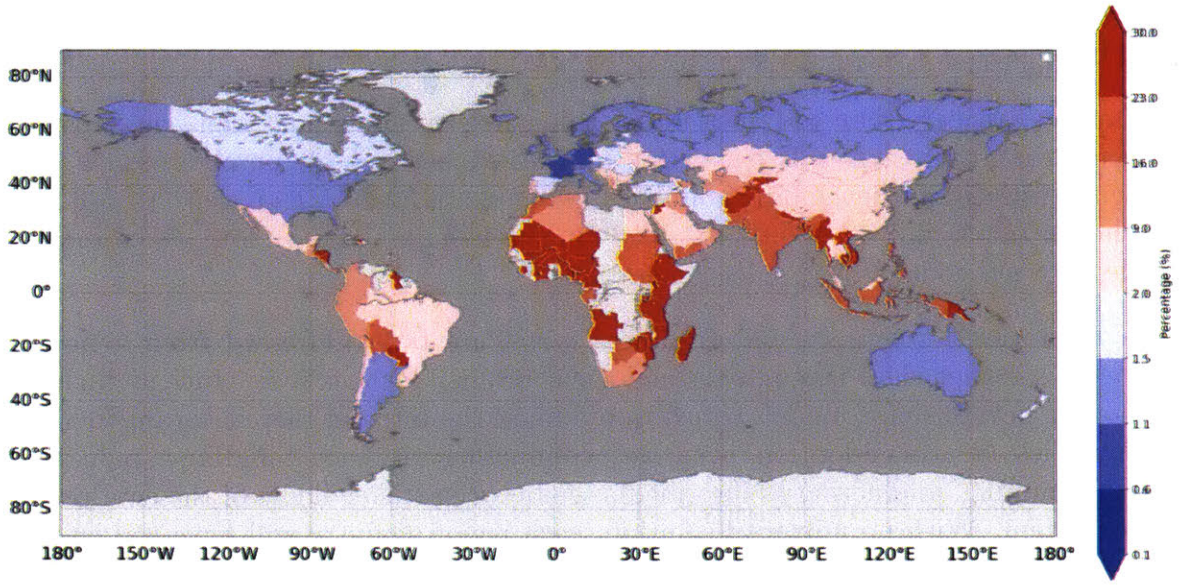


Figure 6-12: Average price of a fixed-broadband monthly plan, as a percentage of the monthly GNI per capita.

6.2.3 Relevance and readiness

Relevance barriers are those that arise from a perceived lack of usefulness (of the Internet) or because of accessibility issues. The most significant barrier in this category is the lack of content and services in local languages, which profoundly affects accessibility. Even though more than 7,000 languages are currently spoken in the world, only 10 (representing the first language of ~3 billion people) account for 89% of the Internet content [9]. That said, the advances in automatic translation technology and the rapid growth of user-generated content platforms in recent years, such as social media, instant messaging applications, and collaborative websites, can drive content creation across a greater diversity of languages.

In contrast, readiness issues are related to the lack of digital abilities; for example, a lack of knowledge on how to operate a computer or a mobile device, or a lack of awareness on what the Internet is and how it may be used. In this regard, illiteracy is a major barrier, given that there are an estimated 1 billion people that currently cannot read or write. Furthermore, previous studies have shown that a large part of those unconnected to the Internet are not *aware of the Internet*: either they have

never heard of the concept, or they have heard of the "Internet" but do not know what it is. The fact is, one cannot connect to something they does not know about. Finally, even when people are literate, have the skills to use a cellphone, and there exists relevant content on the Internet in their language, surveys have shown that many do not see any value or benefit from using the Internet, or worse, have security concerns (especially in South America) regarding the Internet [243]. Because the utility of the Internet (measured as the value that service brings to the users) is best realized first-hand (i.e., experience drives utility), changing personal beliefs might be a big challenge.

Another critical group of issues under this category are gender and age gaps, in the sense that, there currently exists large disparities between the number of men and women connected in developing countries (women are 50-60% less likely to use the Internet in India, and up to 40% less likely in sub-Saharan Africa), and also a digital divide between younger and older generations [241].

In most cases, relevance and readiness issues are best addressed from a policy perspective rather than through technological means. Prospective solutions include developing new business models which would make the Internet more relevant to local communities, creating adoption incentives, conducting advocacy, building capabilities through training, and acting as intermediaries between stakeholders of the connectivity ecosystem. With no universal solution, each region/country will have to identify the strategies best-suited to address their particular challenges; however, given that the focus of this dissertation is on technical solutions (i.e., novel concepts to expand connectivity) to address infrastructural and affordability concerns, policy-based solutions will not be discussed further in the rest of this dissertation.

6.3 Methodology overview

Given the complexity of the problem at hand, and the interplay between infrastructural and affordability issues (as discussed in earlier sections of this chapter), evaluating the impact of novel concepts requires a techno-economic approach that considers not only the technical performance of the systems studied, but also their economic viability and potential adoption rates in a competitive market.

In this dissertation, I propose a techno-economic methodology that can be applied to both space and aerial concepts, to evaluate the viability of each concept in bringing connectivity to uncovered and under-served settlements and regions. The main difference with respect to other methods in the literature is that it provides a systematic procedure to analyze and compare dissimilar concepts that operate at different spatial and temporal scales.

The methodology comprises five steps, namely: 1) defining the scope, 2) developing the technical and economic models, 3) evaluating for technical performance,

4) evaluating financials, and 5) analyzing the potential impact within a competitive market. A tabular overview of the methodology is depicted in Figure 6-13, where each of the five steps is represented by a column. For each step, the first row contains a summary of the tasks associated with each of the steps, and the second and third rows contain the inputs required and outputs produced by these tasks, respectively. The first four steps are applied to each architecture of each concept proposed, and the last step takes as inputs the outputs of all previously analyzed architectures. A more detailed explanation of each of the five steps is given in the remainder of this section.

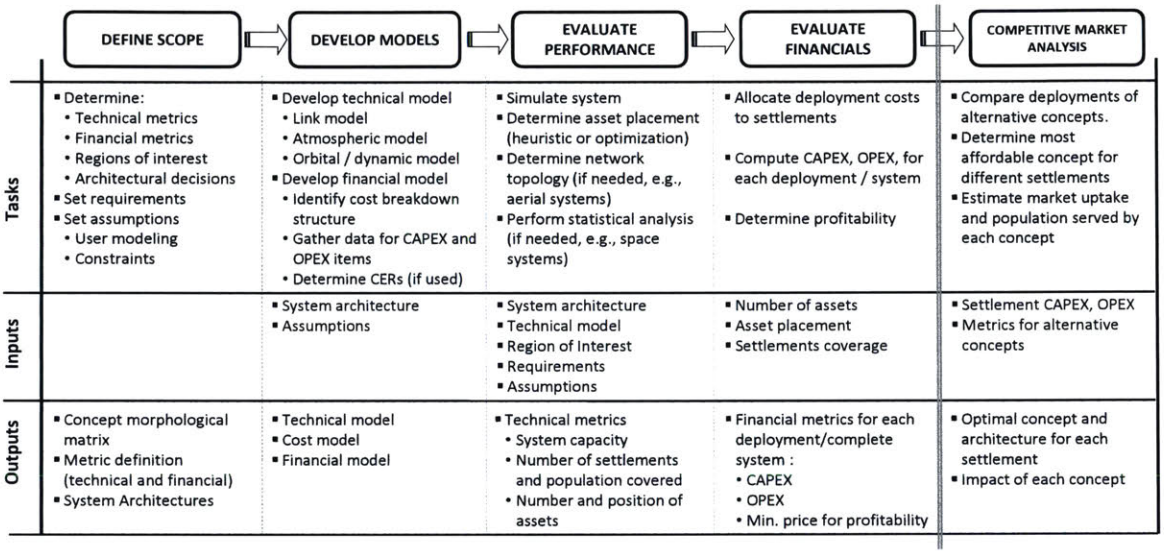


Figure 6-13: Overview of the techno-economic methodology used to evaluate the impact of new concepts proposed to expand global connectivity through space, aerial, and terrestrial networks.

6.3.1 Definition of the scope

The scope definition step consists of the following tasks: determining the regions of interest, defining the technical and financial metrics, establishing the architectural decisions under consideration, as well as characterizing the service and user requirements.

Since the focus of this thesis is on space and aerial architectures as means of providing backhaul to wireless access networks (i.e., cellphone or WiFi-based networks) in uncovered and under-served regions, and it is assumed that space and aerial architectures will not be competitive to terrestrial infrastructure in those regions where there is currently broadband connectivity, the regions of interest are limited to places where there is currently **no connectivity at all (i.e., uncovered) or only 2G connectivity (i.e., under-served)**, which represent ~60% of the Earth's inhabited surface and contain 12.9% of the population (and approximately 30% of those unconnected).

It is important to emphasize that this thesis does not seek to evaluate the impact of space and aerial concepts in closing the usage gap (i.e., addressing populations within areas where broadband is available but are not connected to the Internet), as it is assumed that affordability, relevance, and readiness issues are the main barriers in those instances, and that leveraging existing networks would be more cost-effective in expanding connectivity than introducing new space and aerial concepts.

The two main technical metrics of interest are the **total sellable capacity** (i.e., the capacity that could be realistically sold to users) and the **number and locations of assets** required. On the financial side, the focus is on estimating the **capital expenditure (CapEx)** and **operational expenditure (OpEx)** for each of the architectures (or deployments), together with the **minimum selling price per Mbps/month** that would achieve an internal rate of return (IRR) of 15%.

The architectural decisions to be considered include the types and altitudes of the space and aerial platforms, the network topology, and the operational frequencies of the payloads, given that these are the main drivers of performance and costs for the systems. The specific decisions and corresponding morphological matrices for space and aerial networks are described in Sections 7.5.1 and 8.4.1, respectively. Since terrestrial concepts for backhauling are well established and leave little potential for disruptive innovations to take place, tradespace exploration will not be conducted for these types of networks. Instead, the models for fiber and wireless backhaul networks will be based on the performance and costs of current technologies.

Finally, in terms of service requirements, it is assumed that a minimum service availability of 99% is required for all the concepts and that an entry-level broadband service providing 1 GB/month per user for those living in uncovered regions and 3 GB/month per user for those in under-served regions (with a conservative compound annual growth rate of 9% [249]) would be provided.

6.3.2 Technical and economic models

The technical and economic models are specific to each of the concepts considered, and therefore they will be described in greater detail in Chapters 7, 8, and 9 for space, aerial, and terrestrial networks, respectively. However, it is important to make a brief remark here about a chief difference between the models used to evaluate space networks and those used for aerial and terrestrial networks.

Due to the rotation of the Earth, satellite constellations in non-geostationary orbit (i.e., LEO and MEO constellations) can provide coverage over (almost) the entirety of the Earth once launched[12]. Therefore, the models used for their evaluation need to

[12]Only circular-orbit constellations are considered in this thesis, and since with these types of orbits it is not possible to selectively cover only some regions of the Earth (partial or staged deployments are not considered either), once the constellations are launched, complete coverage within a latitude band is assumed.

consider the demographics and state of connectivity on a global scale. At the same time, the economic viabilities of these architectures would count on capturing customers across all segments of the market (unconnected, under-served, and covered), and thus, the economic models also need to take into account the geospatial income distribution at a global scale. At first glance, these might appear to be difficult; however, since satellite beam footprints are relatively large (in the range of *hundreds of kilometers*) and LoS conditions can be usually be taken for granted, low resolution (i.e., 0.1°× 0.1°grid) population and income models can be used for the analysis.

In contrast, both aerial and terrestrial networks can be deployed to target specific regions where, given the number of users projected to be covered, ARPU, and proximity to existing infrastructure (among other factors), their economic viabilities are guaranteed. However, given their smaller coverage footprints and tighter LoS restrictions (mainly due to terrain occlusions), evaluating these types of networks requires detailed data of the locations, states of connectivity, and income distributions of uncovered and under-served settlements. For this reason, the analysis for aerial and terrestrial concepts focuses on just 37 countries[13] (shown in Table 6.1) in Africa, South America, and Southeast Asia for which income and population data were readily available.

Table 6.1: Countries considered in the comparative analysis of technologies.

ISO2	Country name	ISO2	Country name	ISO2	Country name
AF	Afghanistan	ID	Indonesia	PG	Papua New Guinea
AR	Argentina	IN	India	PH	Philippines
BD	Bangladesh	KE	Kenya	PK	Pakistan
BF	Burkina Faso	LK	Sri Lanka	RW	Rwanda
BR	Brazil	LS	Lesotho	SD	Sudan
CD	Congo (DRC)	MM	Myanmar	TH	Thailand
CI	Cote d'Ivore	MW	Malawi	TR	Turkey
CM	Cameroon	MX	Mexico	TZ	Tanzania
DZ	Algeria	MY	Malaysia	UA	Ukraine
EG	Egypt	MZ	Mozambique	UG	Uganda
ET	Ethiopia	NG	Nigeria	ZA	South Africa
GH	Ghana	PE	Peru	ZM	Zambia
HT	Haiti				

6.3.3 Performance and financial evaluation

Evaluating the performance and financial viability of the different concepts is done by running technical and economic models tailored to the specifics of each concept. Due to the large scales and complexities of these models, evaluating a single architecture

[13]The total uncovered and underserved population in the 37 countries analyzed represents 16% of the global uncovered and underserved population.

may take anywhere from several minutes to several hours, and therefore careful deliberation on which architectures to evaluate is paramount. It is at this stage where the Bayesian optimization formulation discussed in the first part of this thesis comes into play. In particular, the tradespace exploration of space and aerial concepts is framed as a combining pattern problem, and BO-TREE, BO-RBF, and random sampling (RS) (as described in Section 4.1), the best performing methods for this kind of problems, are used to conduct the exploration and their performance compared.

In addition to the tradespace exploration results, the main effects of the architectural decisions considered will also be examined (i.e., what impact did each decision and option have on the final metrics). Furthermore, the sensitivity of the results to the uncertainties associated with the most important parameters of the models employed will also be analyzed.

6.3.4 Competitive market analysis

The last step of the techno-economic methodology determines which of the concepts studied is the most affordable for each region of interest. Besides space and aerial concepts, terrestrial networks are also included in the comparative analysis as baseline systems.

In addition to the estimates for sellable capacity and the minimum selling price that would achieve profitability, this step requires two more pieces of data: the locations of the uncovered and unconnected settlements, and population and income estimates for each of them. This allows us to assess the relative impact of each of the concepts (in terms of population covered that can afford the service and in terms of the number of new users brought online) if they were to operate within a competitive market with other backhaul options available.

Chapter 10 is devoted in its entirety to this last step.

6.4 Chapter conclusions

The main conclusions drawn from this chapter are:

1. Although there has been remarkable growth in the number of people with Internet access in the last five years, with 1.3 billion new users being brought online, the rate of growth has begun to slow down. At the current growth rates, it is unlikely that the targets set by the UN Broadband Commission (75% of the worldwide population connected) would be met by 2025.

2. While more than 6.5 billion people (87.1% of the population) live in areas currently covered by 3G and 4G networks, only 3.5 billion are actually connected

to mobile broadband. This statistic reveals that the usage gap (i.e., the number of people who live in covered areas but are unconnected) is, in absolute terms, 4 times larger than the coverage gap (i.e., the number of people who live in areas with poor or no connectivity), which indicates that addressing other factors such as affordability, relevance, and readiness issues is just as (if not more) important in closing the connectivity gap.

3. The 10.6% of the population that have no connectivity (i.e., are uncovered) and spread out across more than 50% of the inhabited land area on Earth, making it extremely challenging to provide connectivity to them using terrestrial infrastructure. To this end, space and aerial concepts lend themselves better to providing connectivity.

4. When evaluating the impact of novel space and aerial concepts both technical and economic considerations need to be taken into account. This dissertation proposes a five-step techno-economic methodology which can be applied to both space and aerial concepts, to evaluate the viability of each concept in bringing connectivity to uncovered and under-served settlements and regions.

Chapter 7

Satellite networks

The visionary idea of using satellites for communication purposes was originally proposed in 1945 by Arthur C. Clarke described how three "communications stations" placed in geostationary orbit could allow for worldwide RF communications [250, 251, 252]. It was not until almost two decades later, in 1964, that his ideas would materialize, when the first geostationary communications satellite, Syncrom 3, was built and launched by Hughes. Syncron 3 was preceded by several other satellites, themselves each a milestone in the history of satellite communications: project SCORE (1958) was the first broadcasting system in space; Echo-1 (1960) was the first passive reflector communication satellite; Telstar-1 (1962), built by AT&T, was the first real-time communications relay satellite and the predecessor of modern communication satellites, and Syncrom 2 (1963) was the first geosynchronous communications satellite [253].

If the '60s was the decade of the first communication satellites, the '70s was characterized by the commercialization of satcoms, where services such as TV broadcast (cable-TV and direct broadcast), telephony and data services, and mobile satellite services became widespread. During this time, satellites were characterized by the use of low-frequency bands (VHF, C-band, S-band, L-band) and low dry masses (less than 1,000 kg). Besides, advanced technologies such as three-axis stabilization, position-keeping propulsion systems, navigation services, Ku-band transceivers, and solid-state power amplifiers were developed during these years, which were crucial for scaling the performance of communication satellites. In the '80s, the liberalization of the satellite market resulted in the creation of multiple private ventures (such as Echostar, SES, Viasat, or Thaicom), together with the development of very-small aperture terminals (VSATs), contributed to further growth of the satellite communications market [253]. On the technological front, the transition towards Ka-band, along with the use of new medium-access protocols and digital communications [254], were some of the main advances in satellite communications in this decade.

In the '90s, satellite systems started being deployed as a means to provide broad-

band data connections to remote areas. During this decade there was a boom of proposals for launching large LEO constellations, including Iridium, Orbcomm, Globalstar, Teledesic, Skybrdge, and Astrolink [255]. Most of the proposals were eventually canceled due to concerns over the viability of their business models [256], and those that were actually built and deployed (Iridium, Orbcomm, and Globalstar) filed for bankruptcy shortly after launch [257]. After the failure of these large LEO constellations projects, research efforts in the 2000s focused on the development of geostationary high throughput satellite (HTS), which (nowadays) have capacities in excess of 300 Gbps. This increase in throughput was driven by higher frequency reuse factors, enabled by the use of multi-beam antennas, larger reflectors, and higher spectral efficiency modulation and coding schemes (MODCODs), along with advances in power-amplifier technology and the development of larger buses for GEO satellites [258].

In the last five years, the space communications sector has seen even more disruptive proposals, including HTS with throughputs in the Tbps range, mega-constellations in LEO comprising thousands of satellites and with a total throughput of several Tbps, and MEO systems with tens of thousands of fully-flexible and configurable spot beams. These proposals, if brought to fruition, will dramatically decrease the cost of providing connectivity, and it has been argued that they will be vital in bridging the digital divide and providing connectivity to any remaining uncovered and under-served regions.

This chapter presents the models used to evaluate satellite networks, which are validated using current and past constellation designs, as well as the results of several analyses regarding the space architectures best-suited for expanding connectivity into uncovered and under-served regions.

In particular, the following three research questions (which derive from Research Question 1 in Chapter 1) are explored:

Research Question 7.1

What are the characteristics of dominant space architectures?

Research Question 7.2

Taking into account both technical and economic factors, what is the potential impact of space systems in terms of connecting additional populations?

Research Question 7.3

Which technical decisions play the most important role in the design of space systems? Which technologies, if further developed, would yield the most benefit for such systems?

7.1 Model overview

This section presents the model used to evaluate a satellite system based on performance and economic viability. Figure 7-1 shows an overview of the models, with sub-models depicted as gray-shaded boxes, and inputs and outputs as white boxes. As already mentioned in Chapter 6, the primary performance output metric is the total system sellable capacity, whereas the economic metrics of interest are the system cost, the architecture's financial feasibility, and the minimum price per Mbps/month which would be required to meet a stipulated minimum internal rate of return.

System performance is determined using a two-step process. First, the optimal locations and number of feeder gateways are computed using a genetic algorithm. Second, the results locations are combined with atmospheric models, link budget models, and orbital dynamic models to determine statistically the total system throughput (sellable capacity) through Monte Carlo analysis.

To evaluate the economic metrics, first, the masses of the satellites are estimated using a sizing model. Next, the masses are used inputs for the launch model, to determine the number of rockets required and the launch costs for the constellation. Using the masses, launch cost, and ground segment design as inputs, the cost model computes the total system cost. Finally, these parameters are fed into a financial model which determines the project viability and the minimum price per Mbps/month required to achieve profitability.

The next two sections describe sub-models for the performance and economic models, respectively.

7.2 Performance model

The models herein presented are similar to those in [259], by the same author. For the clarity and completeness of this dissertation, part of that work is reproduced herein. Occasionally, some sections have been adopted verbatim from that publication. Overall, the text was significantly rewritten and adapted before its inclusion in this dissertation.

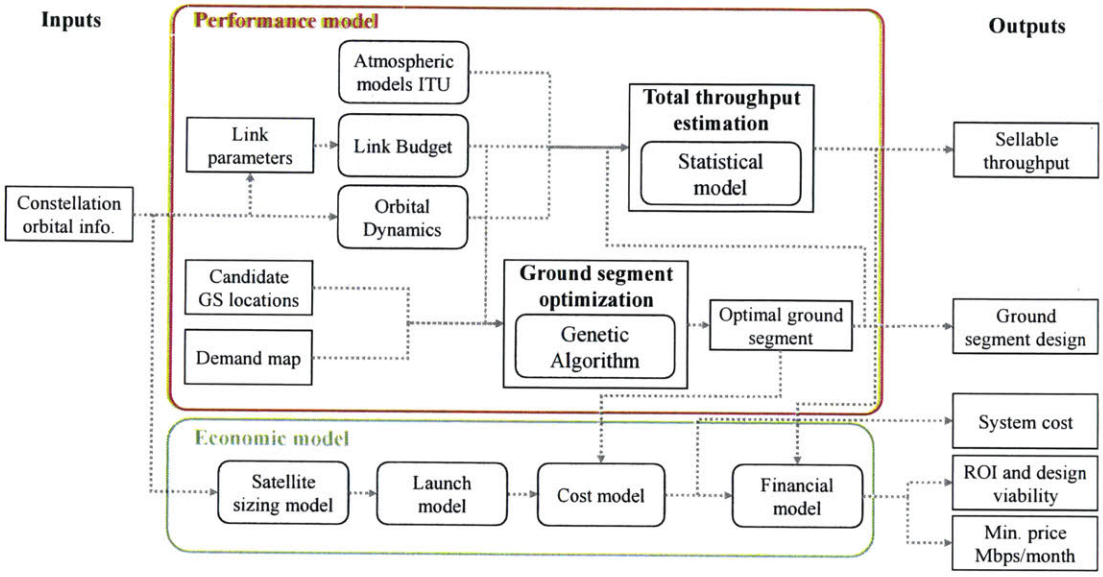


Figure 7-1: Overview of the methodology to determine the ground segment characteristics and estimate total system throughput.

7.2.1 Atmospheric Attenuation model

Atmospheric attenuation is the main external factor that affects the performance of communications links. At Ka-frequencies, extremely high frequency (EHF), and optical band frequencies, its effects cause link outages for non-negligible periods of time, since up to several tenths of dBs of attenuation may be introduced. In addition to the use of adaptive coding and modulation strategies, site diversity is necessary to mitigate the effects of atmospheric attenuation. A satellite should have several ground stations within LoS so that if weather conditions at one particular site cause a link outage, data can still be transmitted through the other ground stations.

The atmospheric attenuation for each frequency band is computed using the guidelines provided in recommendation ITU-P R.618-12 [260], which considers gaseous, cloud and fog attenuation, as well as tropospheric scintillation and rain impairments. In particular, recommendations ITU-R P.676-10 and ITU-R P.840-6 are used to compute gaseous and cloud attenuation, respectively; while the maps in recommendations ITU-R P.837-5, ITU-R P.838-3, and ITU-R P.839-4 are used to estimate rainfall-rates, rain specific attenuation, and rain heights, respectively.

The method described in ITU-P R.618-12 to compute rain attenuation is valid only for percentages of time smaller than 5%. In order to overcome this limitation, it is assumed the rain attenuation is zero for percentages of time greater than the rain probability (P_0, which is computed using the method described in ITU-R P.837-5) and that the transition from the rain attenuation value at 5% to P_0 is linear.

For each ground station, the total atmospheric attenuation at different percentages of time ($\{25, 50, 75, 90, 95, 97, 99, 99.5, 99.7\}\%$) is computed, and for different elevation angle values ($\{10, 30, 50, 70, 90\}$ degrees), creating a set of cumulative distribution function (CDF) curves for the total atmospheric attenuation. Figure 7-2 shows the CDFs for a ground station operating in Ka-band located in Los Angeles, for different elevation angles.

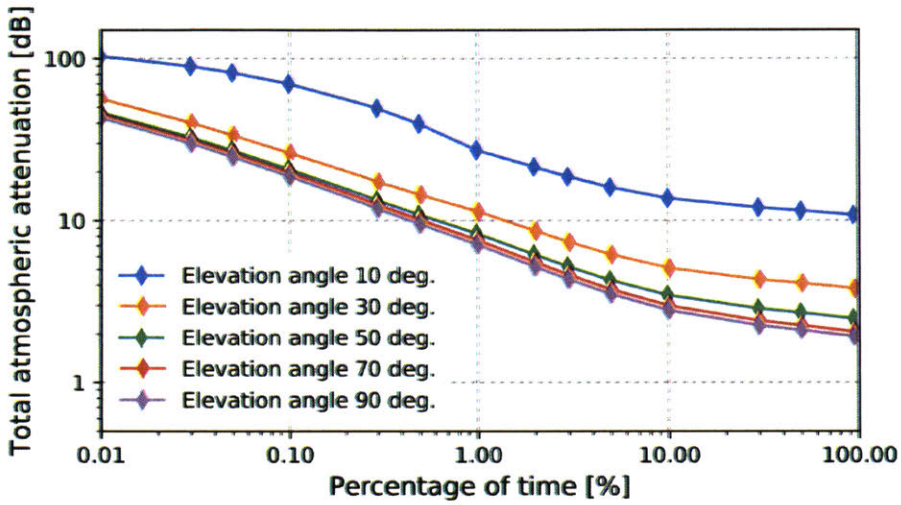


Figure 7-2: Atmospheric attenuation for a ground station operating in Ka-band located in Los Angeles, for different elevation angles.

Whenever present, rain attenuation is the most significant contributor to atmospheric attenuation; in adapting existing models, an important caveat needs to be added. The rain attenuation model for slant paths in recommendation ITU-P R.618-12 is only recommended for frequencies below 55 GHz, as no measurement of atmospheric effects at higher bands were used to create the model (in fact, no measures of atmospheric attenuation at higher frequency bands exist). This issue also applies to other rain attenuation computation techniques described in recommendation ITU-P R.618-12, such as the frequency-scaling methods. Using the rain attenuation model outside of the recommended range might incur errors, a limitation widely recognized in the literature [54]. However, pending further experimental data on space-to-Earth links in EHF bands, this recommendation shall be used, since it is the best available model.

7.2.2 Link budget model

The link budget model is used to determine both the satellite-to-user and satellite-to-gateway capacities. Figure 7-3 shows an overview of the link-model, as well as the parameters considered by each of the blocks. As can be seen, in the link budget computation, the effects of the full RF chain, from digital signal modulation to demodulation, including power amplification and LNB considerations, are all taken

into account.

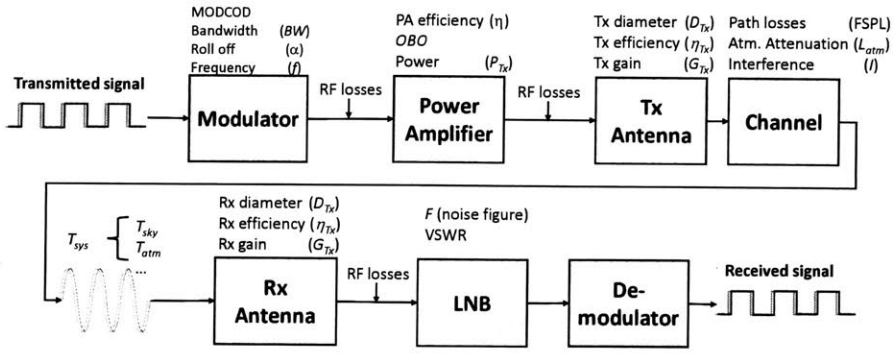


Figure 7-3: Link-budget model block diagram

The rest of this section presents the equations used to compute the link budget for a single beam. First, the link C/N_0 is computed as:

$$\frac{C}{N} = P_{T_x} - \text{OBO} + G_{T_x} + G_{R_x} - L - 10\log_{10}(k\, T_{sys}) - 10\log_{10}(BW) \quad \text{[dB]} \quad (7.1)$$

$$L = \text{FSPL} + L_{atm} + L_{RF_{T_x}} + L_{RF_{R_x}} \quad \text{[dB]} \quad (7.2)$$

where P_{Tx} is the transmitted power (dB), OBO is the power-amplifier output back-off, G_{T_x} and G_{R_x} are the transmitting and receiving antenna gains, respectively (dB), T_{sys} is the system temperature (K), BW is the link bandwidth (Hz), and L represents the different losses considered (dB). In particular, free-space path losses (FSPL), atmospheric losses (L_{atm}), and losses in the transmitting and receiving RF chains ($L_{RF_{T_x}}$ and $L_{RF_{R_x}}$ respectively) are all considered.

The system temperature is computed using Friis transmission equation, as shown in Eq. 7.3,

$$T_{sys} = T_{ant} \cdot 10^{-(L_{RF}/10)} + T_{atm} \cdot 10^{-(A_t + L_{RF})/10} + T_w \cdot (1 - 10^{-(L_{RF}/10)}) \quad \text{[K]} \quad (7.3)$$

where T_{ant} is the antenna temperature (K), T_{atm} is the atmospheric temperature (K), and T_w is the waveguide temperature (K). A_t are the total atmospheric losses (dB), and L_{RF} are the RF losses in reception (dB).

Next, the link $E_b/(N+I)$ is computed as

$$\frac{C}{N+I} = \left(\frac{1}{\text{CABI}} + \frac{1}{\text{CASI}} + \frac{1}{\text{CXPI}} + \frac{1}{\text{C3IM}} + \frac{1}{C/N}\right)^{-1} \quad (7.4)$$

$$\frac{E_b}{N_0 + I_0} = \frac{C}{N+I} \cdot \frac{BW}{R_b} \quad (7.5)$$

where R_b is the link data rate (see below) (bps), and BW is the bandwidth allocated

to that beam (Hz). Notice how our link budget equation considers four different types of interference (CABI, CASI, CXPI, and C3IM). In Eqs. 7.4 - 7.5, all terms are in linear scale.

Finally, the beam data rate is computed as

$$R_b = \frac{\text{BW}}{1+\alpha_r} \cdot \Gamma\left(\frac{E_b}{N_0+I_0}\right) \qquad \text{[bps]}, \tag{7.6}$$

where α_r is the roll-off factor, and Γ is the spectral efficiency of the modulation and coding scheme (MODCOD) (bps/Hz), which depends on $E_b/N+I$, as described below.

It is assumed here that adaptive coding and modulation (ACM) strategies are used, and therefore the MODCOD used for each link is the one that provides the maximum spectral efficiency and satisfies the condition:

$$\left.\frac{E_b}{N_0}\right|_{\text{th}} \geq \frac{E_b}{N_0+I_0} + \gamma \qquad \text{[dB]}, \tag{7.7}$$

where $\left.\frac{E_b}{N_0}\right|_{\text{th}}$ is the MODCOD threshold "energy per bit to noise spectral density" (dB), $\frac{E_b}{N_0+I_0}$ is the actual link energy per bit to noise plus interference spectral density ratio (dB), given by in Eq. 7.5, and γ is the desired link margin (dB). Note that due to the cyclic dependency within Eqs. 7.4 - 7.7, in order to carry out the link budget computations, one needs to assume *a priori* that a given MODCOD scheme is used, compute Eqs. 7.1 - 7.6, and *then* verify whether the condition in Eq. 7.7 is satisfied.

The link budget module is next combined with the atmospheric models to compute the data rates achievable for the uplink and downlink communications under different atmospheric conditions. The code implementation for the link budget is parametric and designed to allow for fast computation of the optimal MODCOD scheme for each combination of ground stations and operating conditions. Moreover, it is designed to handle both bent-pipe architectures, where a frequency translation occurs between uplink and downlink, as well as regenerative architectures, where the uplink and downlink use different MODCOD schemes.

In the performance estimation model, it is assumed that the modulation-coding schemes prescribed in the standard DVB-S2X [261], developed by the Digital Video Broadcast Project in 2014, are used since DVB-S2X is the predominant standard for broadcasting, broadband satellite communication, and interactive services. The standard defines the framing structure, channel coding, and a set of modulation schemes. In particular, more than 60 MODCODs are included, with modulations ranging from BPSK to 256-APSK and coding rates from 1/4 to 9/10.

Finally, to estimate the output-backoff (OBO) for each of the MODCODs, a synthetic sequence of 100,000 symbols is generated and it is assumed that the OBO

equals the peak-to-average power ratio of such a sequence (computed as the ratio between the 99.9th percentile power and the average power). Note that this is an over-estimation of the required OBO, as in a real scenario, one could optimize the OBO by first simulating the channel and the RF chains in transmission and reception, and then reducing it by using pre-distortion techniques to push the amplifier closer to saturation.

The rest of the parameters in the link budgets include the diameters, efficiencies, and noise temperatures of the transmitter and receiver antennas, as well as the values for the different losses over the RF chain and the carrier-to-interference values. Table 7.1 shows an example link budget for the satellite-user link and different frequency bands for a MEO satellite used in the simulations (see Section 7.5).

Table 7.1: Example of values for the link budget for the satellite-user link for a MEO satellite, for different frequency bands.

Parameter	Ku	Ka	Q	E	Units
Frequency	13.5	22.5	39	73.5	GHz
Bandwidth	2	2.1	3.5	5	GHz
Tx antenna diameter	3	3	3	3	m
EIRP	64.2	68.6	73.4	82.3	dBW
MODCOD name	16APSK 13/18	16APSK 25/36	16APSK 23/36	8PSK 2/3	-
Roll-off factor	0.1	0.1	0.1	0.1	-
Spectral efficiency	2.6	2.5	2.3	1.8	bps/Hz
Path distance	9946.2	9946.2	9946.2	9946.2	km
Elevation Angle	31.9	31.9	31.9	31.9	°
Free Space Path Loss	195.01	199.44	204.22	209.73	dB
Atmospheric loss	0.79	3.51	5.61	13.41	dB
Rx antenna gain	36.4	40.8	45.6	50.9	dBi
System temperature	236.4	416.6	557.5	726.8	K
G/T	12.7	14.7	18.2	22.3	dB/K
Rx C/N0	14.0	13.1	12.3	10.5	dB
Rx C/ACI	27	27	27	27	dB
Rx C/ASI	27	27	27	27	dB
Rx C/XPI	25	25	25	25	dB
HPA C/3IM	30	30	30	30	dB
Rx Eb/(N0 + I0)	9.06	8.43	8.09	7.55	dB
MODCOD req. Eb/N0	5.57	5.30	4.77	4.07	dB
Link Margin	3.49	3.13	3.31	3.48	dB
Achieved data rate (single pol.)	5193.1	5241.9	8033.3	9002.9	Mbps
Dist. to Shannon limit	2.34	2.26	2.39	2.87	dB

Finally, when using the link budget to compute the maximum satellite-to-user data rate ($R_b^{\max}$), one needs to take into account the number of channels, frequency reuse factor, and number of polarizations, as shown in Eq. 7.8:

$$R_{b,\text{us}}^{\max} = N_{\text{pol}} \cdot R_b^{lb} \frac{n_{\text{beams}}}{k} \tag{7.8}$$

where R_b^{lb} is the data rate reported by the link budget, N_{pol} is the number of polarizations used (assumed to be 2), n_{beams} is the number of beams of the satellite, and k is the frequency reuse factor (i.e., how many different bandwidth chunks is the spectrum allocation divided into). In contrast, when computing the satellite-to-gateway data rate, the value is just taken as $R_{b,\text{gw}}^{\max} = N_{\text{pol}} \cdot R_b^{lb}$.

7.2.3 Demand model

Dimensioning communication networks is commonly done by estimating the average throughput on a per user basis, considering how the patterns of data consumption vary by time. Typically, the computation starts with estimating the per user data tonnage allowance per month, which considers the mix of traffic for different services required by a "representative" user. After applying a set of correcting factors, the data tonnage is transformed into an average throughput per user required to guarantee a given quality of service during the busy hour (i.e., the sliding 60-minute period during which occurs the maximum total traffic load in a given 24-hour period). In particular, the average (busy-hour) throughput per user ($R_b^{(bh)}$) is given by:

$$R_b^{(bh)} = \underbrace{\frac{D_t}{30}}_{\substack{\text{data-volume} \\ \text{per day}}} \cdot \overbrace{\frac{B}{24}}^{\substack{\text{busy hour} \\ \text{ratio}}} \cdot \underbrace{\frac{1}{60 \cdot 60}}_{\substack{\text{conversion} \\ \text{to bps}}} \cdot \overbrace{\frac{(1+\alpha)}{\eta}}^{\substack{\text{QoS and} \\ \text{utilization ratio}}} \tag{7.9}$$

where D_t is the monthly data tonnage (in bits), B is the busy-hour ration (equal to 2.5-3 for backhauling systems), α is a security factor to guarantee a minimum QoS during unexpected peaks (assumed to be equal to 10%), and η is the utilization factor (assumed equal to 0.85). The total demand generated by a set of users under the footprint of a beam or satellite is computed simply by multiplying the average throughput per user and the total number of subscribers, which in turn can be estimated by multiplying the number of people within the footprint with an assumed adoption rate.

For the purpose of this thesis, the data tonnage is segmented based on the level of existing connectivity[1]. In that sense, for an uncovered person, a basic 1GB/month entry plan is to be provided; whereas for under-served and covered individuals a 3GB/month plan and a 10GB/month plan are to be provided, respectively. Furthermore, it is assumed that the compound annual growth rate (CAGR) is 9%, which will approximately triplicate these values in a 10-year time period. When these values are

[1]Ideally segmentation would be done by market segment (e.g., maritime, aeronautical, trunking, backhaul, broadband), and by current connectivity status (connected vs. unconnected), but this is very hard to estimate at a granular level

applied to Eq. 7.9, the resulting values of $R_b^{(bh)}$ in the first year are approximately 10 kbps, 30 kbps and 100 kbps for the uncovered, under-served, and covered segments respectively, which will go up to 30 kbps, 90 kbps, and 300 kbps after 10 years. Finally, it is worth noting that this demand model intentionally focuses on backhaul infrastructure as used to expand cellphone networks, as opposed to satisfying the demands of other markets (such as end-user, military, in-flight, marine, off-shore connectivity, etc.).

For the technical model, however, the interest is in providing an upper bound to the total sellable capacity for each architecture, and therefore it is assumed that everyone is provided a 30 GB/month service along with an exaggerated adoption rate of 10%. Making this assumption will render the results of the technical model less sensitive to the particular market adoption assumptions; during the discussion of the results of this chapter, this assumption will be revisited and scenarios with more realistic parameters considered for each of the market segments.

The procedure to generate the demand at different orbital positions is as follows. For a given orbital altitude, a gridded map (of resolution 0.1°×0.1°in latitude and longitude) which determines the number of people covered by the beams of a satellite located in a particular orbital position is generated, using population data from the GPWv4[2] [262]. The minimum elevation angle constraints imposed by each of the satellites are also taken into account. Furthermore, it is assumed that users within a region are evenly distributed across all the satellites within their LoS. Finally, the demand is capped at the maximum data rate per satellite (see Eq. 7.12), as shown in Eq. 7.10, where n_{FOV} is the number of satellites within LoS of a ground location:

$$d_{\text{sat}} = \min\left(n_{\text{pop}} \cdot 10\% \cdot 300\text{kbps}\frac{1}{n_{\text{FOV}}}, Rb_{\text{sat}}^{\max}\right) \tag{7.10}$$

Figure 7-4 shows the demand data rate for a LEO, 18-plane polar constellation with 40 satellites per plane (maximum data rate per satellite is 8.8 Gbps) similar to OneWeb's initial design. The orbital positions with higher demand are displayed in brighter tones, while the orbital positions with lower demand are in darker tones. Orbital positions where demand is zero are not colored.

7.2.4 Ground segment optimization

A procedure similar to the one described in [263] is used to determine the optimal gateway locations. This procedure consists of running an optimization algorithm to

[2]This dataset estimates population counts for the year 2020 over a 30-arc-second resolution grid based on census data

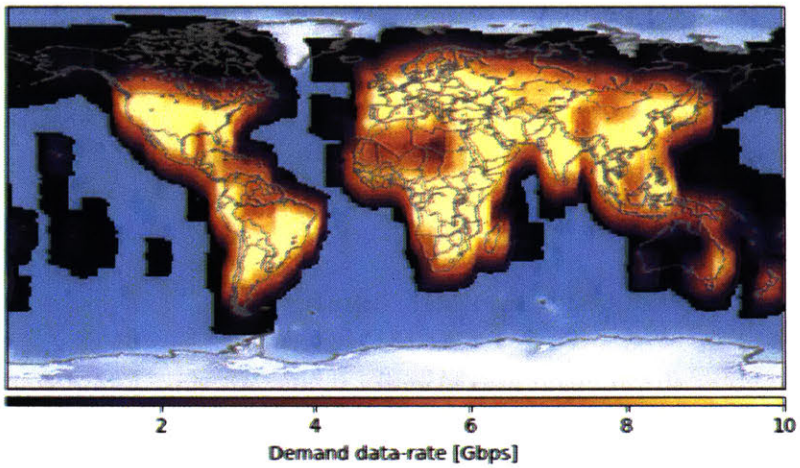


Figure 7-4: Demand data rate for different orbital positions of a sample LEO satellite constellation, with 720 satellites distributed in 18 planes. The maximum data rate per satellite is 8.8 Gbps.

maximize the objective function,

$$O = \frac{1}{2}\text{cov}_{95} + \frac{1}{2}\text{cov}_{99.5} \tag{7.11}$$

while minimizing the number of ground stations required. In Eq. 7.11, cov_{95} and $\text{cov}_{99.5}$ represent the percentage of orbital positions that are covered by a gateway with a throughput of at least 50% of the maximum gateway throughput, under atmospheric conditions present in less than 5% and 0.5% of the time respectively. It is assumed that the minimum elevation angle for ground stations to communicate with a satellite is 10°.

Mathematically, this optimization problem can be framed as a down-selecting problem, where one needs to pick the N ground stations that offer the best performance. A pool of 160 different locations spread across the world is considered, which results in a search space of $2^{168} \sim 3.8 \cdot 10^{49}$ points, making full enumeration and evaluation close-to-impossible. Therefore the use of optimization algorithms is called for. Even though a single function evaluation takes between 60 - 120 seconds, given the large dimensionality of the problem (d = 160), genetic algorithms are better suited than Bayesian optimization for this problem. In particular, the method used is the *genetic algorithm-evolutionary algorithm simple* (GA-EAS), as described in Section 4.3.

Furthermore, the geographical structure of the problem is exploited to speed up the convergence of the optimization algorithm. Given that the selection of ground stations in one region has little impact on which ground stations are selected in another region, the optimization process is divided into two phases. In phase A, the optimal ground segment architectures for each of the six regions considered (Africa, Asia, Europe, North America, Oceania, and South America) are determined using

the GA algorithm described above ($N_{\text{pop}} = 200$, $N_{\text{gen}} = 200$). Then, in phase B, the GA is applied globally, but instead of initiating with a random population (step 1), the best architectures from phase A are used as the generating components for the initial population. In other words, a ground segment architecture for phase B is generated by choosing a good solution from each of the regions in phase A. This new population serves as the initial population for the phase B NSGA-II algorithm ($N_{\text{pop}} = 200$, $N_{\text{gen}} = 80$).

7.2.5 Total system throughput estimation

To evaluate the system throughput statistically a computational model that calculates an upper bound for the maximum sellable capacity for each of the mega-constellations was developed. The need for this statistical model is due to the fact that 1) the system dynamics make the number of customers and gateways within LoS of each satellite vary over time, and 2) the atmospheric conditions that introduce varying attenuation fading and thus varying data rates, are also stochastic by nature.

The procedure to determine the total system throughput is as follows. First, the orbits of the satellites on the constellation are propagated for a day, using 60 seconds time-steps. Then, for each orbital configuration, 10,000 atmospheric attenuation samples for each ground station are drawn, assuming that the samples are statistically independent and distributed according to the probability distribution curves computed with the atmospheric model. These samples were then used as inputs to the link budget model to estimate the achievable link data rates for each of the ground stations. Finally, the total system throughput is computed in two different ways, depending on whether the satellites have inter-satellite links (ISLs) or not.

If the constellation does not have ISLs, the throughput of each satellite is computed according to Eq. 7.12, where d_{sat} is the user-demand, $\sum_{i=0}^{N} R_{b,\text{gw}_{\text{sat}}}$ represents the sum of the data rate of the N best performing ground stations, and $R_{b,\text{us}}^{\max}$ is the maximum satellite-to-user data rate. This is done for each orbital position and set of atmospheric conditions, resulting in 14.4 million samples. The total system forward capacity for each of the scenarios (where a scenario is a combination of orbital positions + atmospheric conditions) is computed by adding the throughput of all satellites.

$$TH_{\text{sat}} = \min\left(d_{\text{sat}}, \sum_{i=0}^{N} R_{b,\text{gw}_{\text{sat}}}, R_{b,\text{us}}^{\max}\right) \tag{7.12}$$

On the other hand, if ISLs are present, the following five-step procedure is used to compute the total system throughput:

1. Compute the total system forward capacity that could be potentially transmitted using all feeder gateways available.
2. Compute the CDF of the total system forward capacity.

3. Select a subset of 1,000 scenarios, evenly spaced on the CDF curve to conduct further analysis with the ISLs taken into account.

4. For each of the selected scenarios:

   - Construct a network graph where the satellites, users, and the ground stations are the nodes, and the RF and optical links are the edges.
   - Set the cost of the ISLs to 1 and the cost of the rest of the links to 0. The capacity of each edge is determined by:
     - the demand captured by the satellite in the case of users-satellite links,
     - the ISL data rate in the case of satellite-satellite links, and
     - the gateway-link data rate in the case of gateway-satellite links.
   - Solve the "minimum-cost, maximum-flow" problem and determine the flow from each satellite to the gateways.
   - Compute the total system throughput by adding the flows from all the satellite-to-gateway links.

5. The total system forward capacity is the 5% percentile of the total system throughputs obtained for the 1,000 scenarios evaluated.

7.2.6 Summary of other assumptions

Some other assumptions made in the estimation of the total system throughput are also worth highlighting:

- User demand is concentrated within land areas and is proportional to the population within the reach of a satellite. Maritime and aeronautical demands are not considered.
- Customers with multiple satellites within LoS select one at random to communicate with, therefore the demand is evenly distributed among satellites within LoS.
- adaptive coding and modulation (ACM) is used on the satellite-gateway links, thus, for any orbital position and atmospheric conditions, the MODCOD that maximizes throughput is selected.
- Satellites produce enough power to communicate at maximum EIRP whenever required.
- User terminals are not a limiting factor, as they are capable of tracking satellites continuously and communicating at the required data rates.
- There are no outages caused by foliage, building obstruction, or other factors that might affect the user links at any elevation angle.

- Performance degradation due to interference between LEO satellites from different constellations is not considered.
- Ground stations can be located over *any* land area, i.e., there are no political, legal, landing rights, or geographical constraints to their placements.
- ISLs can be used to route excess demand to other satellites, but only satellites in the same orbital set (both in-plane and cross-plane) can communicate through ISLs.
- There is no maximum number of hops that data can traverse through ISLs, even though latency shall be minimized.

7.3 Economic model

The goal of the economic model is to obtain the total cost of the system, assess its economic feasibility, and compute the minimum average selling price (in Mbps/month) that the operator would need to set to achieve the desired internal rate of return (IRR) of 15%. To that end, the economic model comprises four sub-models: the satellite-sizing model, the launch model, the cost model, and the financial model. The sub-sections that follow describe each of these models in turn.

7.3.1 Satellite-sizing model

The satellite-sizing model is based on the subsystems models described in Space Mission Analysis and Design (SMAD) [264]. The main inputs to the algorithm are the payload parameters (mass, average data rate, average and peak power), which are estimated using a regression model built using data from communication satellite payloads launched in the last 15 years.

Since the masses of some of the subsystems depend on both the payload parameters and the total satellite dry-mass (sat_{dm}) (thus creating a cyclic dependency, as the satellite dry-mass is the sum of the subsystems' masses), an iterative algorithm is used. The algorithm starts by assuming a value for the satellite dry-mass and computes the masses of each of the subsystems. The new satellite dry-mass is then computed by adding up the masses of each one of its subsystems. If the difference between the newly-computed satellite dry-mass and the assumed one is below 0.1%, the algorithm is taken to have converged; otherwise, the algorithm is rerun using the newly-computed satellite dry-mass as the initial guess.

The rest of this section prescribes the equations used to compute the power and mass required by each of the subsystems. In the equations, quantities in bold correspond to constants and technological parameters. Table 7.2 contains the values used for all constants; whereas the values for the different variables involved in the

satellite-sizing model are shown in Table 7.3 using a OneWeb satellite as a canonical example.

Payload

The payload model was built by fitting a nonlinear regression to a database with the masses of 65 GEO payloads, which includes both traditional GEO and HTS designs. The main driver of the model is the number of beams, which is closely related to the number of transponders, power amplifiers, mixers, and the other RF components of the payload. Given the increasing digitalization of the payloads and advancements in component integration, the payload weight exhibits a very non-linear relationship with the number of beams; while, there are some components (like the antenna, local oscillators, IF equipment) that add a fixed weight, the miniaturization of components and transition towards phased-array and multi-feeder architectures have drastically changed the outlook for the weight of payloads with a large number of beams. Thus, to capture this non-linear dependency, a model of the form $\alpha \cdot (1 - e^{-(\beta \log(n_{\text{beams}})^p)}$ was chosen.

After fitting the model to the 65 data points, the following curve was obtained:

$$m_{\text{GEO}} = 1096 \cdot (1 - e^{-\left(0.25 \log(n_{\text{beams}})^2.05\right)} \qquad [kg] \tag{7.13}$$

Since for LEO and MEO constellations the number of data points available was insufficient to fit a regression model (7 and 2 data points respectively), curves presenting similar trends like the one shown in Eq. 7.13 and that fit well to the data points available were chosen. In other words, the value of α was adjusted (α_nMEO = 638 and α_nLEO = 285) such that the lower weights associated with MEO and LEO orbits were captured. Figure 7-5 shows these curves for GEO, MEO, and LEO payloads.

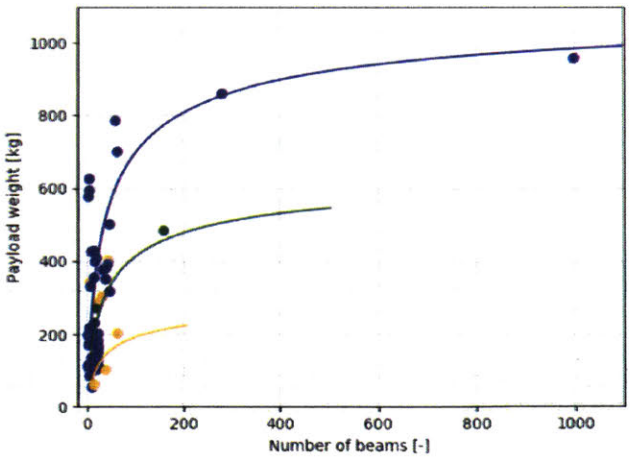


Figure 7-5: Payload model curves for GEO, MEO, and LEO constellations.

In addition to the mass estimate obtained from Eq. 7.13 (and the analogous equations for MEO and LEO payloads), for cross-linked architectures the ISL terminal masses are estimated to be 8 kg [265] and 15 kg when optical and RF technologies are used, respectively. These values are then multiplied by 4, since that is the number of ISLs per satellite assumed for cross-linked architectures.

Finally, the average payload RF-power and data rate are derived directly from the link budget model described in Section 7.2.2 and Eq. 7.8, respectively. The peak payload power is assumed to be 50% higher than the average payload power, and to obtain the DC-power of the payload, different efficiencies are assumed based on the frequency band of the link: 55%, 40%, 30%, and 20% for the Ku, Ka, Q/V, and E bands, respectively.

Propulsion Subsystem

The propulsion system is sized first by computing the total ΔV amount required over the lifetime of a satellite

$$\Delta V = \Delta V_{\text{drag}} + \Delta V_{\text{ADCS}} + \Delta V_{\text{deorbit}} + \Delta V_{\text{inj}} \tag{7.14}$$

$$\Delta V_{\text{injection}} = \sqrt{\frac{\boldsymbol{\mu_E}}{sa_{\text{orbit}}}} - \sqrt{\boldsymbol{\mu_E}\left(\frac{2}{sa_{\text{orbit}}} - \frac{1}{sa_{\text{transfer}}}\right)} \tag{7.15}$$

$$\Delta V_{\text{drag}} = \begin{cases} 40 \cdot l & \text{if orbital altitude } < 400 \text{ km} \\ 12 \cdot l & \text{if orbital altitude } < 500 \text{ km} \\ 5 \cdot l & \text{if orbital altitude } < 600 \text{ km} \\ 2 \cdot l & \text{if orbital altitude } < 1000 \text{ km} \\ 0 & \text{otherwise} \end{cases} \tag{7.16}$$

$$\Delta V_{\text{ADCS}} = \frac{(39.93 + 14.45 + 1.715) \cdot 1}{1000} \tag{7.17}$$

$$\Delta V_{\text{deorbit}} = 1 \tag{7.18}$$

where l is the lifetime of the mission (in years), μ_E is the standard gravitational parameter for the Earth, sa is the semi-major axis of the orbit, and the different ΔV values are computed following the recommendations of [264]. Then, the fuel mass is computed as the sum of the injection fuel and the ADCS fuel masses:

$$m_{\text{fuel-inj}} = \boldsymbol{\gamma_{\text{inj}}} \cdot \text{sat}_{\text{dm}} \left(e^{\frac{\Delta V_{\text{injection}} * 1000}{9.81 \cdot \boldsymbol{\text{Isp}}}} - 1\right) \tag{7.19}$$

$$m_{\text{fuel-ADCS}} = \boldsymbol{\gamma_{\text{ADCS}}} \cdot \text{sat}_{\text{dm}} \left(e^{\frac{\Delta V_{\text{ADCS}} * 1000}{9.81 \cdot \boldsymbol{\text{Isp}}}} - 1\right) \tag{7.20}$$

$$m_{\text{fuel}} = m_{\text{fuel-inj}} + m_{\text{fuel-ADCS}} \tag{7.21}$$

where γ_{inj} and γ_{ADCS} are margin coefficients, sat_{dm} is the dry-mass of the satellite, and Isp is the specific impulse of the fuel used.

Electric Power Subsystem

The electric power subsystem mass is estimated using the six-step procedure described below.

1. Compute the total power that needs to be produced by the solar arrays, assuming that the payload consumes 80% of the total satellite power, and considering both solar and eclipse conditions:

$$P_{\text{eclip}} = P_{\text{sun}} = \frac{P_{\text{payload}}}{0.8} \tag{7.22}$$

$$P_{\text{tot}} = \frac{P_{\text{eclip}} \cdot k_{\text{eclip}} + P_{\text{sun}} \cdot k_{\text{sun}}}{k_{\text{sun}}} \tag{7.23}$$

   where k_{eclip} and k_{sun} are the time fractions that the satellite spends in eclipse and non-eclipse conditions respectively, and P_{tot} is the total power that needs to be produced by the solar arrays.

2. Determine the beginning of life (BOL) and end of life (EOL) power densities of the solar array:

$$p_{\text{BOL}} = \boldsymbol{G_{SC}\eta_{SA}} \cdot \cos(\theta_{\text{sun}}) \tag{7.24}$$

$$p_{\text{EOL}} = P_{\text{BOL}} \left(1 - \boldsymbol{d_{SA}}\right)^{1} \tag{7.25}$$

   where G_{SC} is the solar constant, η_{SA} is the solar array cell efficiency, θ_{sun} is the sun worst angle, and d_{SA} is the yearly solar array annual degradation.

3. Estimate the solar array area A_{SA}:

$$\text{A}_{\text{SA}} = P_{\text{tot}}/p_{\text{EOL}} \tag{7.26}$$

4. Determine the solar array mass:

$$\text{m}_{\text{SA}} = p_{\text{BOL}}/\rho_{SA} \tag{7.27}$$

   where ρ_{SA} is the solar array specific power (in W/kg).

5. Determine the battery mass by computing the capacity of the battery:

$$C_{\text{batt}} = \frac{P_{\text{eclip}} \cdot k_{\text{eclip}} T_{\text{orbit}}}{3600 \cdot \boldsymbol{\eta_{\text{batt}}} \cdot \mathbf{DoD_{batt}}} \tag{7.28}$$

$$\text{m}_{\text{batt}} = \frac{C_{\text{batt}}}{\boldsymbol{E_{\text{batt}}}} \tag{7.29}$$

   where η_{batt} is the battery efficiency, DoD_{batt} is the battery depth-of-discharge, T_{orbit} is the orbital period, and E_{batt} is the battery specific energy (in W/kg).

6. Determine the total EPS subsystem mass by adding the masses of the solar arrays, batteries, and additional electronics:

$$m_{EPS} = m_{SA} + m_{batt} + 2\% \cdot sat_{dm} + 3.25\% P_{BOL} \tag{7.30}$$

Attitude Determination Control Subsystem

The ADCS subsystem is sized as follows:

- Compute the maximum torque that will need to be corrected by the reaction wheels.

$$\Gamma_{grav} = 1.5 \frac{\boldsymbol{\mu_E}}{sa_{orbit}^3} \cdot |I_y - I_z| \sin(2\theta_s) \tag{7.31}$$

$$\Gamma_{drag} = 0.5 \cdot \boldsymbol{C_{ADCS}} \cdot \boldsymbol{\rho_{atm}} \cdot v_{orbit}^2 \cdot L_x \cdot L_y C_{PA} \tag{7.32}$$

$$\Gamma_{solar} = \frac{F_s}{c} L_x \cdot L_y (1 + \boldsymbol{q}) C_{PS} \cos(\theta_w) \tag{7.33}$$

$$\Gamma_{magnetic} = \boldsymbol{D_r} \frac{2M}{sa_{orbit}^3} \tag{7.34}$$

$$\Gamma = \max(\Gamma_{grav}, \Gamma_{drag}, \Gamma_{solar}, \Gamma_{magnetic}) \tag{7.35}$$

  where I_y and I_z are the moments of inertia along the $\hat{y}$ and $\hat{z}$ axes, θ_s is the slew angle, C_{ADCS} is the drag coefficient, ρ_{atm} is the atmospheric density at the orbital altitude, v_{orbit} is the orbital velocity, L_x and L_y are the dimensions along the $\hat{x}$ and $\hat{y}$ axes, C_{PA} is the difference between the center of aerodynamic pressure and the center of gravity, F_s is the solar constant, c is the speed of light, C_{PS} is the difference between the center of solar pressure and the center of gravity, q is the reflectance factor, θ_w is the angle of incidence of the Sun, D_r is the residual dipole of the satellite, and M is the magnetic moment of the Earth.

- Compute the mass of the reaction wheels, which corresponds to the mass of the control subsystem.

$$\Gamma_{rw} = \frac{0.25}{\sqrt{2}} \Gamma \cdot T_{orbit} \tag{7.36}$$

$$m_{CS} = 1.5 \cdot (\Gamma_{rw})^{0.6} \tag{7.37}$$

  where T_{orbit} is the orbital period.

- Compute the attitude determination mass using a mass estimation ratio.

$$m_{AD} = 10 \cdot (\theta_{req})^{-0.316} \tag{7.38}$$

  where θ_{req} is the ADCS pointing requirement (in arc-seconds).

- Compute the total mass of the ADCS subsystem by adding up the masses of each of the components:

$$m_{\text{ADCS}} = 6 \cdot m_{\text{AD}} + 4 \cdot m_{\text{CS}} + \boldsymbol{m_{\text{GPS}}} + \boldsymbol{m_{\text{Earth-horizon}}} + \\ + 2 \cdot \boldsymbol{m_{\text{thruster}}} + 3 \cdot \boldsymbol{m_{\text{mag-torq}}} \tag{7.39}$$

Tracking, Telemetry, and Command Subsystem

The mass of the TT&C (and avionics) subsystem is computed as a fixed percentage of the satellite dry-mass (sat_{dm}):

$$\text{m}_{\text{TTC}} = 3.28\% \cdot \text{sat}_{dm} \tag{7.40}$$

Thermal Subsystem

The mass of the thermal subsystem is computed as a fixed percentage of the satellite dry-mass (sat_{dm}).

$$\text{m}_{\text{thermal}} = 2.8\% \cdot \text{sat}_{dm} \tag{7.41}$$

Structure Subsystem

The mass of the structure subsystem is computed as a fixed percentage of the satellite dry-mass (sat_{dm}):

$$\text{m}_{\text{structure}} = 20\% \cdot \text{sat}_{dm} \tag{7.42}$$
$$\text{m}_{\text{adapater}} = 1\% \cdot \text{sat}_{dm} \tag{7.43}$$

Total satellite mass

The final component of the satellite-sizing model to estimate the total satellite dry-, wet-, and launch-masses, as well as other dimensional variables such as the moments of inertia and lateral dimensions.

$$\text{sat}_{dm} = \text{m}_{\text{payload}} + \text{m}_{\text{EPS}} + \text{m}_{\text{ADCS}} + \text{m}_{\text{thermal}} + \text{m}_{\text{structure}} + \text{m}_{\text{TTC}} \tag{7.44}$$
$$\text{sat}_{wm} = \text{sat}_{dm} + m_{\text{fuel}} \tag{7.45}$$
$$\text{sat}_{lm} = \text{sat}_{wm} + \text{m}_{\text{adapater}} \tag{7.46}$$

To compute the satellite dimensions and moments of inertia

$$L_x, L_y, L_z = \left(\frac{\text{sat}_{dm}}{\rho_{\text{sat}}}\right)^{\frac{1}{3}} \tag{7.47}$$

$$I_x, I_y, I_z = 0.01 \cdot (\text{sat}_{dm})^{\frac{5}{3}} \tag{7.48}$$

Table 7.2: Constants and technical parameters for the satellite-sizing model

Param	Value	Units
Injection fuel Isp	310.0	s
Injection fuel security margin	1.20	
ADCS fuel Isp	230.0	s
ADCS fuel security margin	1.00	
Spacecraft average density	140.0	kg/m3
Residual dipole	0.50	A m2
Reflectance factor Q	0.60	
Magnetic torquer mass	5.00	kg
3-axis magnetometer mass	1.00	kg
GPS mass	10.0	kg
Earth Horizon sensor mass	1.00	kg
Thruster mass	1.00	kg
Slew Angle	0.04	rad
Solar constant	1386.0	W/m2
Eclipse SA efficiency	65%	
Day SA efficiency	85%	
SA cells efficiency	20%	
SA inherent degradation	77%	
Yearly SA degradation	2.75%	
SA specific power	60.0	W/kg
Efficiency batt-load	90%	
Batt specific energy	40.0	Wh/kg
Battery DoD	80%	

7.3.2 Launch model

Launch cost is computed as the cost per launch vehicle multiplied by the number of launches required. A database of currently-operational and in-development launch vehicles containing cost, mass-to-LEO, and mass-to-GEO launch capabilities, as well as fairing dimensions, was compiled (see Table D.1 in Appendix D). The fairing dimensions correspond to the measures of a simplified, notional fairing, composed of a cylindrical section joined to a conical frustum section, as shown in Figure 7-6. Data on size and dimensions were extracted from the user manuals of the respective launch vehicles, while the cost values were obtained from official websites and historical launch data from various sources [266, 267].

An algorithm to determine the number of satellites that could be fitted on each

Table 7.3: Example of the different values of the satellite-sizing algorithm for a OneWeb satellite.

Param	Value	Units
Orbit parameters		
Satellite orbit	LEO	-
Satellite orbit altitude	1200.00	km
Satellite orbit inclination	55.00	degree
Satellite orbit eccentricity	0.00	degree
Satellite orbit sa	7578.14	km
Satellite orbit apogee	7578.14	km
Design lifetime	5	years
Satellite orbit period	6565.30	s
Payload Subsystem		
Electronics mass	36.00	kg
Antenna mass	24.00	kg
Payload mass	60.00	kg
Payload frequency	25.00	GHz
Payload data rate	8.80	Gbps
Payload maximum power	262.50	W
Payload average power	210.00	W
EPS Subsystem		
Fraction sunlight	68.16	%
Power sunlight	220.5	W
Time sunlight	5165.1	s
Power eclipse	220.5	W
Time eclipse	2090.5	s
Power BOL	456.1	W
Power EOL	396.7	W
Solar array area	2.14	m2
Solar array mass	7.60	kg
Battery capacity	177.8	Wh
Battery mass	4.45	kg
EPS other mass	17.68	kg
Total EPS mass	29.72	kg

Param	Value	Units
Propulsion Subsystem		
Delta V apogee insertion	0.28	m/s
Delta V drag	0.0	m/s
Delta V ADCS	0.10	m/s
Delta V de-orbit	1.00	m/s
Total Delta V	1.38	m/s
Mass propellant Injection	16.21	kg
Mass propellant ADCS	6.46	kg
Mass propellant AKM	1.03	kg
ADCS Subsystem		
Gravity gradient torque	0.00	Nm
Aero. drag torque	0.00	Nm
Solar pressure torque	1.31	uNm
Magnetic torque	18.29	uNm
Maximum torque	18.29	uNm
Reaction wheels capacity	0.02	Nms
Att. control mass	0.15	kg
Att. determination mass	1.86	kg
ADCS propellant tank mass	0.63	kg
ADCS sensors mass	12.60	kg
ADCS structure mass	1.43	kg
ADCS total mass	14.65	kg
Subsystem		
Telemetry mass	4.68	kg
Thermal Subsystem		
Thermal mass	3.99	kg
Structure Subsystem		
Structure mass	28.53	kg
Structure adapter mass	1.43	kg
Total Satellite sizing		
Satellite dry-mass	142.61	kg
Satellite wet-mass	165.29	kg
Satellite launch-mass	166.71	kg

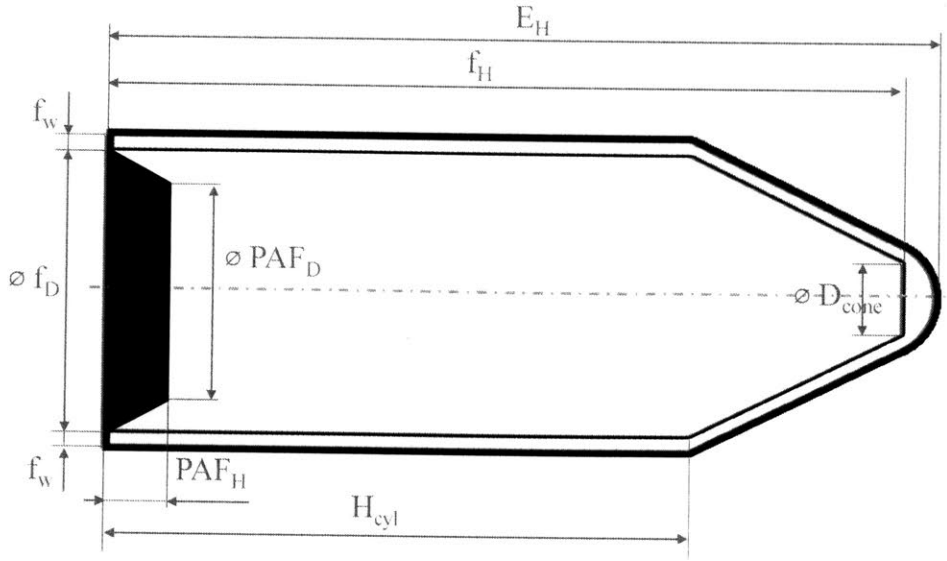


Figure 7-6: Schematic of a rocket's fairing dimensions.

launch was developed. This algorithm considers both volumetric and weight constraints, and consists of the following steps:

- Estimate the maximum number of satellites that can be launched, subject to weight constraints:

$$N_{mass} = \frac{0.9 \cdot \text{pay}_{\text{orbit}}}{\text{sat}_{lm}} \tag{7.49}$$

  where N_{mass} is the maximum number of satellites that can be launched, subject to mass constrains, $\text{pay}_{\text{orbit}}$ is the mass that the rocket can carry into orbit, and sat_{lm} is the satellite launch mass. Note that a margin of 10% of the total satellite weight is set aside for the adapter and launch structures.

- Determine the number of satellites that can be launched attending to volumetric constraints, following the procedure below:

  i Determine the number of satellites that can be fitted within the cylindrical section: first, for each satellite-orientation, determine the number of satellites that can be laid out across in a "layer"; then determine the number of "layers" that can be stacked. It is assumed that there is a minimum separation between satellites (Δ) of 10 cm. Finally, select the satellite orientation that allows for pack the largest number of satellites to be packed into the cylindrical section.

  ii Determine the number of satellites that can be fitted within the conical frustum section: first, divide the truncated cone section into layers of dimensions $\{L_x, L_y, L_z\} + \Delta$ (lateral dimensions of the satellite plus some margin); then, for each layer, determine the number of satellites that will

fit within a cylindrical section of diameter equal to the diameter of the conical frustum at the current layer upper limit. Iterate this process until it is no longer possible to divide the conical frustum into further layers.

iii Volumetrically, the total number of satellites that volumetrically could be fitted in the fairing (N_{vol}) is the sum of the results from steps i and ii.

- The number of satellites per vehicle is the minimum value of satellites among the number that can be fitted subject to mass constraints (N_{mass}), the number that can be fitted subject to volumetric constraints (N_{vol}), and the total number of satellite in the constellation (N_{const}):

$$N = \min\left(N_{mass},\ N_{vol},\ N_{const}\right) \tag{7.50}$$

7.3.3 Cost model

Figure 7-7 depicts the cost breakdown structure of the cost model, which is based on the reference-model presented in [264]. Non-recurring costs, such as funding for research and development and ground segment construction costs, are marked with a white triangle with letters "NR" at the lower right corner of each box. Recurring costs, i.e., those that are incurred per each unit built or are paid annually as operational expenses, are marked with a black triangle with letter "R".

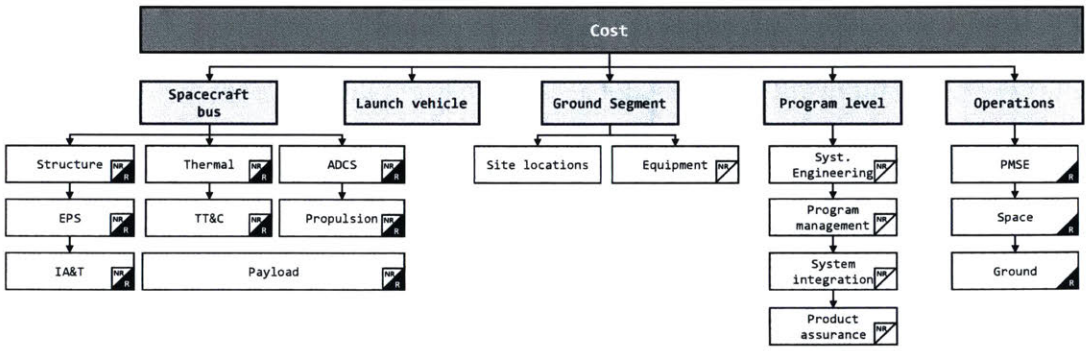


Figure 7-7: Work breakdown for the cost structure. White and black triangles indicate that the component is part of the non-recurring and recurring costs elements respectively. Adapted from [264].

The cost model is divided into five broad categories, namely spacecraft, launch, ground segment, program-level costs, and operational expenses. The spacecraft bus is split into different subsystem costs, whereas the ground segment costs consider both site-location conditioning (i.e., new building and infrastructure development) as well as the procurement of RF equipment and antenna-dishes. The rest of this section

will go into detail on the spacecraft and ground segment costs; the launch vehicle costs are simply computed by adding up the costs of the rockets required to put the satellites into orbit, described in the previous section. Finally, the program-level and operational costs are included as part of the satellite cost model (see Table 7.4).

Satellite cost model

Two different parametric cost models (shown in Table 7.4) are used to estimate the satellite cost: SSCM and USCM-10. SSCM, the *Small Satellites Cost Model*, was developed by the Aerospace Corporation to predict the development costs of modern, small satellites (under 1,000 kg.) [268]. USCM-10, the *Unmanned Spacecraft Cost Model*, was developed by the Air Force to estimate the costs associated with earth-orbiting, unmanned space vehicle programs [269].

These models are based on cost estimating relationships (CERs), which, given the mass (or other parameters) of a subsystem, provide estimates for its costs. Such models are better suited for trade studies during the conceptual design stage, as this phase is characterized by the lack of precise information to inform detailed cost estimates. USCM-10 provides estimates for both non-recurring costs (e.g., research and development costs, prototype unit costs, program costs) and recurring costs (e.g., manufacturing costs per unit), whereas SSCM provides combined estimate for non-recurring plus first unit costs (it is assumed that the non-recurring costs are 60% of this value and the first unit costs 40%). For satellites weighing over one tonne, the estimates from USCM-10 are used; otherwise, the SSCM model is employed. Cost figures from each model are adjusted to \$FY'18 values.

Finally, a learning curve is used to capture the savings from productivity improvements, cost reductions, and human learning obtained when multiple units are manufactured. Using this technique, the total cost of producing N satellites is

$$C_{total} = \text{FPU} \cdot N^{\left(1-\frac{\log \frac{1}{S}}{\log 2}\right)}, \tag{7.51}$$

where FPU is the cost of the first production unit, and S is the learning curve slope, whose value depends on the number of units manufactured (0.95 if less than 10 units are produced, 0.9 when between 10 and 50 units are built, and 0.85 if more than 50 units are fabricated).

Ground segment cost model

The ground segment cost is computed using the recommendations from the *DoD Facilities Pricing Guide* [270], which states that the cost of building (C^{GS}_{const}) and

Model		Aerospace Corp. SSCM [FY'10] recurring +non-recurring	USAF USCM-10 [FY'12] non-recurring	recurring
Structure	ST	$407 + 19.3X_1 \log_{10}(X_1) + 355 + 5.7X_2^2$	$23.83X_1 + 156.2X_2^{0.4338}$	$8.388X_1 + 11.63X_2^{0.5787}$
ADCS	ADCS	$1850 + 11.7X_3^2$	$68.4X_3^{0.4945}$	$68.4X_3^{0.4945}$
EPS	EPS	$1261 + 539X_4^{0.72}$	$54.87X_4^{0.7042}$	$34.92X_4^{0.8949}0.7647$
Telemetry	TT&C	$486 + 55.5X_5^{1.25}$	$613.8X_5^{0.4366}$	$197.4X_5^{0.8124}0.6625$
Bus cost	SC_b	ADCS + EPS + TT&X + ST	ADCS + EPS + TT&X + ST	
Payload	PL	$0.4 * SC_b$	$125.95 * X_6^{0.8027}$	
Satellite cost	SC	SC_b + PL	SC_b + PL	
IA&T	IA&T	$0.139SC_b$	$0.1627SC_b$	$0.13(SC_b$ + PL$)0.6603X_8^{0.1143}$
Program	P	$0.229SC_b$	0.2048(IA&T + SC_b + PL)	0.2095(IA&T+SC_b+PL)* 0.7621
Aerospace and ground	AGE	$0.066SC_b$	2583.36	-
Launch Ops	LOPS	$0.061SC_b$	-	$255.7X_9^{0.4431}$
		X_1: Structure mass [kg] X_2: Thermal mass [kg] X_3: ADCS mass [kg] X_4: EPS mass [kg] X_5: TT&C mass [kg]	X_1: Structure mass [lb] X_2: Thermal mass [lb] X_3: ADCS mass [kg] X_4: EPS mass [lb] X_5: TT&C mass [lb]	X_6: Antenna mass [kg] X_7: Burn time [s] X_8: Number of antennas X_9: Satellite dry-mass [lb]

Table 7.4: Comparison of the USCM-10 and SSCM parametric cost models for different satellite systems. The different variables used as inputs for each of the models are described at the bottom of the table.

maintaining (C^{GS}_{sust}) a new satellite communications facility can be computed as:

$$C^{GS}_{const} = \text{RUC} \cdot \text{ACF} \cdot \eta_{PD} \cdot \eta_{SIOH} \cdot \eta_{CF}, \quad \text{and} \tag{7.52}$$

$$C^{GS}_{sust} = \text{SUC} \cdot \text{SACF} \cdot I_t \tag{7.53}$$

where

- RUC is the *replacement unit cost*, which depends on the country where the ground station is located.
- ACF is the *area cost factor*, which also depends on the country where the ground station is located.
- η_{PD} is a factor that accounts for the planning and design of the facility, and is equal to 1.09.
- η_{SIOH} is a factor that accounts for "supervision, inspection, and overhead activities" associated with the construction of the facility, and is equal to 1.057.
- η_{CF} is a factor that accounts for construction contingencies, and is equal to 1.05.
- SUC is the *sustainment unit cost*
- SACF is the *sustainment area cost factor*, which also depends on the country where the ground station is located.
- I_t is an inflation factor that accounts for increases in maintenance costs in year t.

Finally, the cost per gateway antenna (including modem, electronics, etc.) is based on the model from [271] and is given by

$$C_{\text{GW-ant}} = 50 + 39 \left(\frac{D_{GW} \cdot N_p}{N_c} \right)^{2.7} \quad [k\$], \tag{7.54}$$

where D_{GW} is the antenna diameter, N_p is the number of polarizations of the systems, and N_c is the number of colors. This formula estimates the cost of a 2.4 m antenna to be \$113k, and the cost of a 3.5 m antenna to be \$226k.

In our models, it is further assumed that one ground station also doubles up as the control facility (located in the USA or Canada by default), while the rest are only satellite communications facilities.

7.3.4 Financial model

The financial model evaluates the economic feasibility of a given constellation using a cash flow analysis, which takes into account the revenue, CapEx, and operational expenditure (OpEx) items. The goal of the model is to determine the minimum average price per Mbps/month that a company would need to charge its users to achieve an IRR of at least 15%. Cash flow projections are computed assuming a 20-year horizon; in addition, the following assumptions were made:

Revenue generation

- Satellites are first launched in year 2, and service operations begin in year 3.
- Capacity is sold progressively, starting at 20% in the first year of operations, and stepping up to 40%, 50%, and 65% in years 2, 3, 4 respectively. From year 5 onward, 70% of the system's total system throughput is sold. The variable ρ is used to represent the percentage of capacity sold.
- The average price per Mbps/month is a variable chosen such that the IRR within a 20-year horizon equals 15%.
- The average price per Mbps/month decreases at a rate of 5 % annually.
- Finally the annual revenue is computed as:

$$\text{Revenue} = \text{System capacity} \times \rho \times \text{price Mbps} \tag{7.55}$$

Capital Expenditure

- The items that constitute CapEx are: satellite manufacturing costs, launch costs, insurance costs, ground segment costs, and R&D costs.
- The satellite manufacturing costs, launch costs, ground segment costs, and R&D costs are given by the cost models described in the previous two sections.
- The insurance costs are assumed to be 8% of the launch value (launch costs + total costs of the satellites on-board), following the recommendations of [264].
- Constellation replenishment costs are distributed uniformly across all years of operations.

Operational Expenditure

- The items that constitute OpEx are: ground segment sustainment costs, back-haul costs, and other operational expenses (including salaries, marketing, etc).

- The ground segment sustainment costs are given by the ground segment cost model.

- The backbone transit costs are assumed to be \$5 per Mbps/month, with an annual decrease rate of 5%.

- Other operational expenses (including salaries, sales, marketing, leases, power, etc.) are assumed to be 35% of revenue once the constellation starts offering service, and up to 10% of the CapEx in the years prior operations begin.

Once the revenue, CapEx and OpEx items have been taken into account, the cash flow for each year is simply given by:

$$\mathrm{CF} = \mathrm{Revenue} - \mathrm{CapEx} - \mathrm{OpEx} \tag{7.56}$$

Finally, the internal rate of return of (the discount rate that makes the difference between current investments and the future NPV equal to zero), is the solution to Equation7.57,

$$0 = \sum_{t=0}^{20} \frac{\mathrm{CF}_t}{(1+\mathrm{IRR})^t}, \tag{7.57}$$

where CF_t is the cash flow in year t.

7.4 Model validation

This section details the steps taken to validate the performance, satellite-sizing, and cost model estimates.

- The performance model is validated in Section 7.4.1 by comparing the total system throughput predicted against that of current systems (plus systems under design).

- The satellite-sizing model is validated in Section 7.4.2 by comparing the estimates for mass against the values in several historical missions, extracted from the TelAstra 2016 database.

- The cost models are validated in Section 7.4.3 by comparing the cost estimates against those of several legacy systems.

7.4.1 Performance model validation

The performance model validation is carried out by comparing the predicted performance of several systems to current and future (proposed) systems. In particular, the total system throughput (i.e., the sellable capacity) of OneWeb, Viasat-1, and Viasat-3 are used as reference values. The actual value for OneWeb's system (after accounting for atmospheric effects, the geographical distribution of the demand, and other factors) has been reported to be in the 1.4 Tbps - 1.6 Tbps range [272], whereas the throughputs of Viasat-1 and Viasat-3 are 140 Gbps and 1 Tbps respectively.

Table 7.5: Validation of the performance model

	Sellable capacity [Gbps]		
System	**Model**	**Truth**	**Error (%)**
Viasat-1	122.3	140	-12.8 %
Viasat-3	865	1,000	-13.5 %
OneWeb	1,506	1,400	7.57 %

Table 7.5 contains the comparison of the predicted and actual values for the two systems selected. It can be seen that our model underestimates the sellable throughput of Viasat-1 and Viasat-3 by approximately 13%; whereas for OneWeb, the predicted figure is within the range of the reported values. A possible reason for the lower values of Viasat's satellites is that the model estimates sellable throughput with the statistics for atmospheric attenuation taken into account, whereas values reported by the satellite operators are typically under clear-sky conditions. In any case, given the low relative errors, the model will suffice for the analyses carried out in this dissertation.

7.4.2 Satellite-sizing model validation

The satellite-sizing model validation is conducted by comparing the estimates produced by the model (Section 7.3.1) against actual values from the TelAstra 2016 database, which contains more than 100 records of past communications missions (both in LEO and GEO). The records in the database date back to as early as 1972 (ANIK A), while the latest record being from 2016 (Viasat-1).

LEO satellites

To validate the satellite-sizing model for LEO satellites, five representative communications satellite systems are used: Globalstar-1, Globalstar-2, Iridium, Teledesic, and Skybridge. The first three systems have been launched, whereas for Teledesic and Skybridge, the projects were canceled before launch, and thus the values reported correspond to design values.

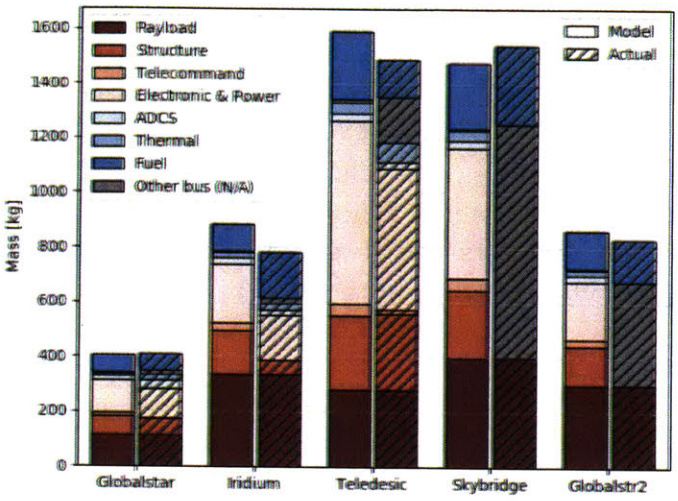


Figure 7-8: Comparison of real launch mass and model mass for 6 different LEO satellites

Figure 7-8 shows the relative mass of the subsystems, and Table 7.6 presents the actual values along with the relative errors of the model. For the five spacecraft considered, it is assumed that the payload mass was provided as an input to the sizing algorithm. It can be observed that the predicted launch mass is within ±15% of the actual launch mass for all satellites, whereas the predicted satellite dry mass is within ±8% of actual values for all cases but Iridium (for which our model overestimates the mass of the structure subsystem). In terms of relative errors by subsystem, the TT&C subsystem shows the largest deviations; however, since its mass is relatively small compared to the total mass of the spacecraft, it has a small impact on the final values for the dry and launch masses.

Table 7.6: Comparison of predicted and actual values for the subsystem and satellite masses for 5 LEO satellites

Satellite	Globalstar			Iridium			Teledesic		
	Model	Actual	% error	Model	Actual	% error	Model	Actual	% error
Payload	115.0	115.0	0.0	338.5	338.5	0.0	283.4	283.4	0.0
Structure	68.4	60.0	14.0	157.0	51.0	207.9	268.0	273.5	-2.0
Telecommand	11.2	1.0	1021.5	25.8	-	-	44.0	17.7	148.4
EPS	117.4	104.0	12.9	215.1	161.7	33.0	665.9	513.6	29.6
ADCS	17.4	32.0	-45.5	23.2	20.1	15.5	30.1	23.6	27.5
Thermal	9.6	23.0	-58.4	22.0	24.8	-11.4	37.5	72.8	-48.5
Dry-mass	341.9	350.0	-2.3	785.1	613.7	27.9	1340.3	1350.0	-0.7
Launch-mass	406.3	410.0	-0.9	884.0	780.0	13.3	1592.8	1490.0	6.9

Satellite	Skybridge			Globalstr2		
	Model	Actual	% error	Model	Actual	% error
Payload	400.0	400.0	0.0	300.0	300.0	0.0
Structure	247.3	-	-	145.4	-	-
Telecommand	40.6	-	-	23.8	-	-
EPS	474.4	-	-	208.5	-	-
ADCS	28.8	-	-	22.5	-	-
Thermal	34.6	-	-	20.4	-	-
Dry-mass	1236.7	1250.0	-1.1	726.8	676.0	7.5
Launch-mass	1476.4	1538.0	-4.0	864.7	832.0	3.9

GEO satellites

To validate the satellite-sizing model for GEO satellites, six different communications missions were selected (from both commercial and governmental organizations). These missions include Arabsat, TDRSS-J, Telstar-4, Viasat-1, Wildblue, and Koreasat, whose masses range from 1,000 kg (characteristic of missions from small operators) to more than 6 tonnes (characteristic of lastest generation HTS launched by well-established operators). Figure 7-9 shows the relative masses of the subsystems, and Table 7.7 presents the actual masses along with model errors. Again, it is assumed that the payload mass for each of the satellites was provided as an input to the algorithm. The results obtained for GEO satellites are similar to those obtained for LEO satellites; the launch-mass errors are within 10% error for all satellites but TDRSS-J, whereas the dry-mass errors are within ± 17% in all cases.

Table 7.7: Comparison of predicted and actual values for the subsystem and satellite masses for 6 GEO satellites

Satellite	Arabsat			TDRSS-J			Telstar-4		
	Model	Actual	% error	Model	Actual	% error	Model	Actual	% error
Payload	122.4	122.4	0.0	625.0	625.0	0.0	390.7	390.7	0.0
Structure	104.5	90.4	15.6	280.1	305.5	-8.3	316.5	176.4	79.4
Telecommand	17.1	29.2	-41.3	45.9	52.1	-11.8	51.9	78.5	-33.9
EPS	182.6	177.0	3.1	268.6	434.0	-38.1	625.1	566.7	10.3
ADCS	56.0	66.5	-15.9	74.4	138.9	-46.4	77.9	70.6	10.4
Thermal	14.6	30.2	-51.6	39.2	76.4	-48.7	44.3	90.7	-51.1
Dry-mass	522.3	573.1	-8.9	1400.6	1736.0	-19.3	1582.6	1621.0	-2.4
Launch-mass	990.4	1140.0	-13.1	2655.9	3196.0	-16.9	3000.9	3331.2	-9.9
Satellite	Viasat-1			Wildblue			Koreasat-1		
	Model	Actual	% error	Model	Actual	% error	Model	Actual	% error
Payload	794.1	794.1	0.0	500.0	500.0	0.0	202.0	202.0	0.0
Structure	698.5	532.3	31.2	416.5	-	-	127.6	-	-
Telecommand	114.6	159.8	-28.3	68.3	-	-	20.9	-	-
EPS	1507.7	824.0	83.0	851.9	-	-	180.1	-	-
ADCS	112.0	155.7	-28.0	87.3	-	-	58.6	-	-
Thermal	97.8	313.9	-68.8	58.3	-	-	17.9	-	-
Dry-mass	3492.8	3296.6	6.0	2082.5	1925.0	8.2	637.8	641.0	-0.5
Launch-mass	6623.2	6739.6	-1.7	3948.8	4451.0	-11.3	1209.4	1411.0	-14.3

Given the results from the validation process, it can be concluded that the satellite-sizing model is accurate enough for tradespace exploration in the context of space communications architectures. Moreover, the effects that the uncertainty (or errors) of the satellites-sizing model have on the output metric will be further studied through sensitivity analysis.

7.4.3 Satellite cost model validation

The cost model is validated by comparing the estimates for the 11 missions considered (5 LEO, 6 GEO) against the actual values reported within the TelAstra database. The model estimates correspond to the sum for the non-recurring and recurring costs

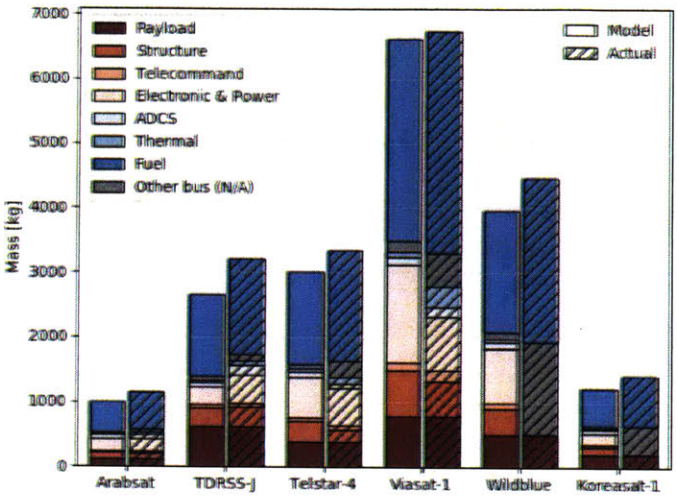


Figure 7-9: Comparison of real launch mass and model mass for 6 different GEO satellites

of each unit. For constellations where more than one unit was produced, the non-recurring costs are uniformly allocated to each of the satellites, whereas the recurring costs are the average recurring costs after taking into account all discounts associated with the learning rate (i.e., discounts from mass-manufacturing).

Table 7.8 reports the cost values estimates for the 6 GEO satellites, whereas Table 7.9 contains the estimate values for the 5 LEO satellites. All cost values (both estimates and actual values) are quoted in the year each satellite was manufactured.

Table 7.8: Comparison of model and actual cost for 6 GEO satellites

	Actual	# sats	Model	
	[$M]	-	Est	% error
Name				
Arabsat	45.0	3.0	38.2	-15.1
TDRSS-J	229.0	1.0	233.2	1.8
Telstar-4	70.0	3.0	93.3	33.3
Viasat-1	501.0	1.0	556.7	11.1
Wildblue	265.0	1.0	286.7	8.2
Koreasat-1	78.0	2.0	71.3	-8.5

Table 7.9: Comparison of model and actual cost for 5 LEO satellites

	Actual	# sats	Model	
	[$M]	-	Est	% error
Name				
Globalstar	12.0	56.0	10.9	-9.1
Iridium	24.0	66.0	19.7	-18.1
Teledesic	35.0	324.0	30.2	-13.8
Skybridge	36.0	80.0	37.5	4.2
Globalstr2	22.0	48.0	28.8	30.8

It can be observed that the cost model does provide reasonable estimates for both LEO and GEO satellites. Overall, relative cost errors are within 30% of the real

values. Although these errors might seem high at first glance, it is not uncommon for cost models to deviate by such magnitudes (in fact, a 30% error would be considered relatively modest in terms of cost estimation). Importantly, these cost estimates only need to be accurate enough to allow for comparison among architectures of the same concept (e.g., between multiple space connectivity systems) and between different concepts (e.g., satellite architectures vs. aerial systems), and thus, it is when interpreting and drawing conclusions from the tradespace exploration results that these considerations should be taken into account, for example, through sensitivity analysis.

7.5 Tradespace exploration for space networks

7.5.1 Design space description

The design space is defined as the set of decisions that need to be taken when designing a satellite constellation, together with the options for each decision. Table 7.10 shows the morphological matrix for space networks.

Table 7.10: Morphological matrix for space systems

Decision name	Decision options						
Orbital Altitude	GEO	MEO	LEO				
Number of planes	1	5	10	15	20	50	
Number of satellites / plane	3	5	10	20	30	50	70
Orbital inclination	0	30	45	60	90		
Number of payload beams	7	19	37	74	370	740	
Frequency User $\leftrightarrow$ Satellite	Ku	Ka	V/Q				
Frequency Satellite $\leftrightarrow$ Gateway	Ka	V/Q	E	Optical			
Frequency Inter-satellite Links	E	Optical	None				

In particular, the following decisions were considered when defining space networks.

- **Orbital altitude and inclination**: possible orbital altitudes include GEO (35,768 km), MEO (8,000 km), and LEO (1,200 km). The orbital inclination can take one of the values in {0°, 30°, 45°, 60°, 90°}, with the following constraints for GEO networks, only equatorial orbits (inclination = 0°) are allowed; for LEO networks, only non-equatorial orbits are allowed.
- **Number of planes and number of satellites per plane**: these decisions determine the overall constellation structure and the total number of satellites. The number of planes considered can take any value in {1, 5, 10, 15, 20, 50}, whereas the number of satellites per plane is {3, 5, 10, 20, 30, 50, 70}. The following constraints apply to the decision: for equatorial configurations (inc=0°), the number of planes is always 1; for non-equatorial configurations, it is always

greater than 1. Moreover, for MEO constellations, the maximum number of planes is 20 and the number of satellites per plane is any option between 5 to 30; finally, only GEO architectures can have 3 satellites per plane, since constellations in other orbits cannot possibly achieve global coverage with such a small number of planes.

- **Number of beams per satellite**: the number of beams per satellite can take any value in {7, 19, 37, 74, 370, 740}[3]. However, the following constraints apply: for LEO networks, the number of beams must be fewer than or equal to 74 beams (due to the assumed power limitations of the spacecraft), whereas GEO networks must have a number of beams greater than or equal to 37 (i.e., architectures comprised of "small-GEO" satellites are not considered). Finally, MEO networks with a large number of planes (10, 20, 50) must have fewer than 100 beams per satellite.

- **Operational frequencies**: these frequencies determine the bandwidth available (according to the ITU frequency allocation guidelines) and the achievable data rates. For the user links, frequencies {Ku, Ka, V/Q} are considered; for the feeder links: {Ka, V/Q, E, optical}, and for the ISLs: {None, E, optical}.

- **Number and locations of ground stations**: although not featured as a decision per se in the morphological matrix, the number and locations of the ground stations are optimized for each architecture, following the procedure described in Section 7.2.4.

Finally, it is worth mentioning that all the architectures are assumed to require a CPE to connect to users, as direct-to-user strategies are unaffordable for uncovered and under-served populations.

7.5.2 Tradespace exploration methodology

This section contains the tradespace exploration results for space networks. To preserve the flow of this part of the thesis, the performance of using Bayesian optimization for this tradespace exploration exercise is not discussed here but can instead be found in Appendix E.

[3]It is assumed that each beam has 1 GHz of usable bandwidth. Thus, the number of beams equals the total usable bandwidth in the satellite.

7.5.3 Tradespace exploration results

Overview of results

Figure 7-10 shows the tradespace exploration results for space networks. Each point corresponds to an architecture, that is, a unique combination of decision-options from Table 7.10. The horizontal axis represents the CapEx required to begin operations, whereas the vertical axis is the total sellable forward link (FWD) capacity. In addition to the architectures obtained from the morphological matrix, values for several of the currently-proposed systems have been included in the graph. The yellow dot at the top-left corner represents the *utopia point*, that is, the ideal (unattainable) point (where infinite capacity is available at close-to-zero CapEx costs).

The Pareto front has GEO, MEO, and LEO points, which indicates that there are good designs in all orbits. For low CapEx and low sellable capacities, GEO systems are predominant; for mid-capacities, MEO satellites are best. Among current proposals, mPower and Viasat-3 (the global constellation of 3 satellites) are closest to the Pareto front, whereas OneWeb is the furthest away.

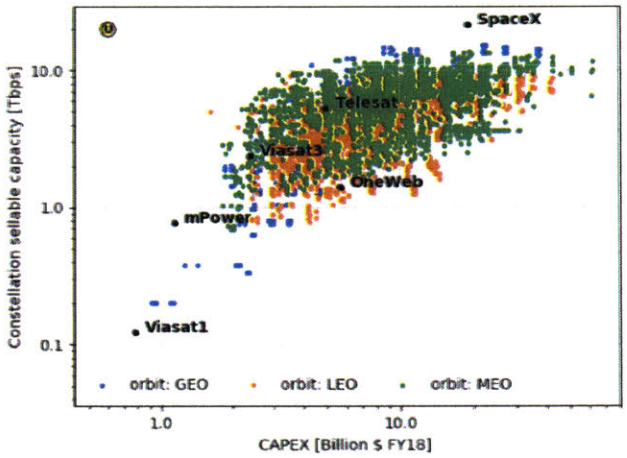


Figure 7-10: Tradespace exploration results for satellite networks. Each point represents a different architecture. The yellow dot at the upper left corner is the *utopia point* (ideal value).

Figure 7-10, however, lacks two elements that are required to understand the tradespace fully: first, the OpEx and replenishment costs are not taken into account, and second, the throughput offered by the different architectures relative to demand (i.e., whether the architectures are over-provisioning or under-provisioning capacity), which in turn is a function of the price.

Figure 7-11 shows an alternative view of the tradespace, in which the horizontal axis represents the minimum cost per Mbps/month that each architecture would

need to sell its capacity to achieve 15% IRR. The black line represents the estimated demand for the broadband market[4], hence, architectures over the black line can be interpreted as over-provisioning capacity (and thus will not be able to sell all their capacity at the targeted price), whereas those below the line are under-provisioning. Note that the demand curve represents the "global" broadband + backhaul market demand curve.

Several conclusions can be extracted from this image:

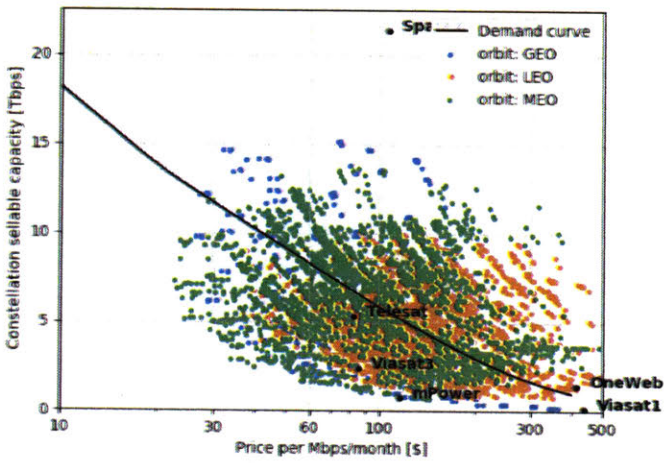


Figure 7-11: Demand curve and tradespace exploration results for satellite networks. Each point represents a different architecture. The black curve represents the estimated global broadband demand[5] as a function of the price per Mbps/month.

- Only a small proportion of the architectures are viable businesses, capable of selling their full capacities at the required price per Mbps/month. The majority of the architectures over-provision capacity, and will face challenges to remain profitable; some of them will be able to raise prices (at the cost of selling less capacity) or target markets more profitable than consumer broadband (e.g., military, mobility, etc.), and still be able to generate enough revenue to survive. Others will not be able to generate such revenue and might have a more cruel fate.

- The farther to the bottom-left of the plot that architectures are positioned, the higher the likelihood of them being profitable (because there is more room to increase prices and adjust as new competition enters the market). In that sense, GEO and MEO architectures present the highest likelihoods of profitability; however, LEO systems can also be competitive if they are designed wisely (as will be explained later in this section).

[4]It is assumed that satellite systems will be able to capture 0.5% of the market for covered areas, 10% of the market for under-served areas, and 30% of the market for uncovered areas.

- If all the currently-proposed satellite systems were to be launched (with their designs unchanged), there will be an overall excess of capacity. In particular, total capacity would be more than 40 Tbps of capacity, an amount over 20 times higher than the current global comms. satellite capacity.

Tradespace results filtered by decision

Figure 7-12 shows the tradespace results colored based on the option chosen for each decision. It can be observed that both MEO and GEO systems with a single equatorial plane (as seen in Fig. 7-12a) dominate the profitable region. In contrast, polar orbits are a poor choice and offer low sellable capacities, since a lot of the capacity is wasted in the polar regions where there are no (or few) customers. Figure 7-12b shows the architectures classified by the number of beams per satellite. For GEO and MEO systems, architectures with 370 and 740 beams respectively dominate over those with a lower number of beams, whereas for LEO architectures 7- and 19-beam satellites seem the most promising.

Figure 7-12c depicts the different architectures colored based on the number of planes in the constellation, while Figure 7-12d shows them colored based on the number of satellites per plane. It can be observed that these two decisions are responsible for great stratification within the tradespace, with architectures with a small number of planes (depicted in blue, purple, and orange in Figure 7-12c) and a small number satellites of planes (depicted in green, purple, and blue in Figure 7-12d) clustering at the bottom-left of the charts (and are therefore the most likely to be profitable), while architectures with larger number of satellites trend towards the upper-right areas of the charts. This indicates that (given the demand assumptions) constellations with a large number of satellites are oversized, which results in higher prices per Mbps/month since there are diminishing returns in launching new satellites, i.e., once full coverage is achieved, each new satellite added will serve users actively for a lower percentage of its orbital period.

Characteristics of most profitable architectures

As mentioned at the beginning of this section, GEO, MEO, and LEO designs can all be competitive if they are designed wisely. In particular, Tables 7.11, 7.12, and 7.13 show the most popular options for the decisions among architectures that indicate the highest likelihoods of being profitable (i.e., those farthest to the bottom-left of the demand curve in Figure 7-11). For each table, the morphological matrix of decisions is depicted, and the different options are colored based on their popularity among Pareto-optimal architectures. The more intense the coloring, the more popular an option is.

It can be observed that popular GEO architectures are those with a single equatorial plane, a small number of satellites (3-10), and payloads with a large number of

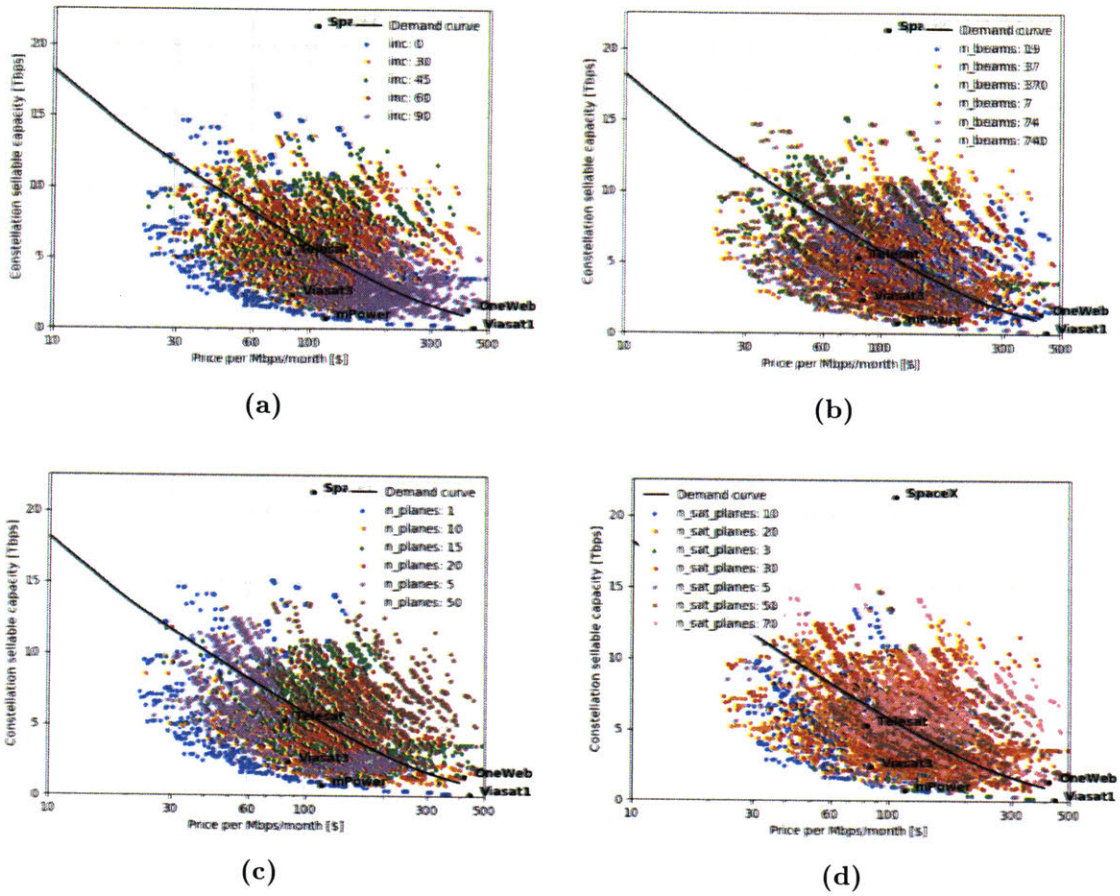


Figure 7-12: Tradespace exploration results for satellite networks by a) inclination of the orbits, b) number of beams per satellite, c) number of planes in the constellation, and d) number of satellites per plane. Each point represents a different architecture.

beams operating at high-frequency bands. These architectures are very similar to the current HTS and very high throughput satellite (vHTS) concepts, with the primary difference of using higher-frequency bands. For MEO architectures, equatorial orbits are also preferred, but with a greater number of satellites per plane (5-30, with 10 being the most popular option). For MEO architectures, there is considerable variability in the frequencies of operation for all types of links. Finally, for LEO constellations, the dominant architectures are relatively small constellations with between 200 and 450 satellites distributed across 10-15 planes of 20-30 satellites each, a preferred orbital inclination of 60°, simple payloads with 7-19 beams, and (surprisingly) without cross-links between satellites.

This last result indicates that the increased sellable throughput and ground segment cost savings from having ISLs onboard are not worth the greater satellite masses (which in turn increases launch and manufacturing costs). However, there are other reasons to carry ISLs which were not featured by the model: first, ISLs allow for

additional markets, e.g., maritime and airliner markets; second, some countries require operators to land all data originated by their citizens within their borders, a task that would be facilitated by ISLs. These factors might explain why the companies are trending towards the development of architectures with ISLs for their constellations.

Table 7.11: Popularity of decision-options for GEO architectures on the Pareto front. The color intensity is proportional to the number of times each option was present in Pareto optimal architectures.

Decision name	Decision Options						
Number of planes	1	5	10	15	20	50	
Number of satellites / plane	3	5	10	20	30	50	70
Orbital inclination	0	30	45	60	90		
Number of payload beams	7	19	37	74	370	740	
Frequency User ↔ Satellite	Ku	Ka	V/Q				
Frequency Satellite ↔ Gateway	Ka	V/Q	E	Optical			
Frequency Inter-satellite Links	None	E	Optical				

Table 7.12: Popularity of decision-options for MEO architectures on the Pareto front. The color intensity is proportional to the number of times each option was present in Pareto optimal architectures.

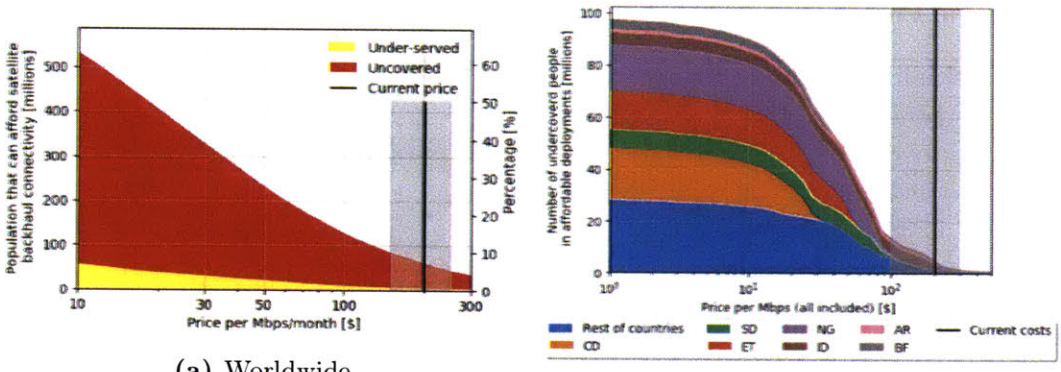
Decision name	Decision Options						
Number of planes	1	5	10	15	20	50	
Number of satellites / plane	3	5	10	20	30	50	70
Orbital inclination	0	30	45	60	90		
Number of payload beams	7	19	37	74	370	740	
Frequency User ↔ Satellite	Ku	Ka	V/Q				
Frequency Satellite ↔ Gateway	Ka	V/Q	E	Optical			
Frequency Inter-satellite Links	None	E	Optical				

Table 7.13: Popularity of decision-options for LEO architectures on the Pareto front. The color intensity is proportional to the number of times each option was present in Pareto optimal architectures.

Decision name	Decision Options						
Number of planes	1	5	10	15	20	50	
Number of satellites / plane	3	5	10	20	30	50	70
Orbital inclination	0	30	45	60	90		
Number of payload beams	7	19	37	74	370	740	
Frequency User ↔ Satellite	Ku	Ka	V/Q				
Frequency Satellite ↔ Gateway	Ka	V/Q	E	Optical			
Frequency Inter-satellite Links	None	E	Optical				

Impact in terms of bringing uncovered and under-served populations online

In this section, the impact of satellite networks in terms of bringing uncovered and under-served populations online is analyzed. Figure 7-13 shows the uncovered and under-served population that can afford satellite connectivity vs. price per Mbps/month, both at a worldwide scale and when focusing on the 37 countries shown in Table 6.1. Given that current prices for the satellite backhaul and broadband markets are ~$200 per Mbps/month, it is estimated that up to 55 million uncovered and under-served worldwide (or ~8% of the global uncovered and under-served population) could afford satellite backhauled connectivity, out of them, 6 million are within the 37 countries of interest (~5% of the total uncovered and under-served population of those countries). Moreover, when the currently proposed system launch in the next 2-5 years and prices go down to $80-$120 per Mbps/month, the number of people that would be able to afford satellite connectivity would increase to 115 million (~15%) worldwide and 17 million in the countries of interest. Finally, the future generation will enable prices to go even further down to $30-$40 per Mbps/month, making satellite connectivity affordable for 255 million of uncovered and under-served worldwide, and 60 million in the 37 countries.

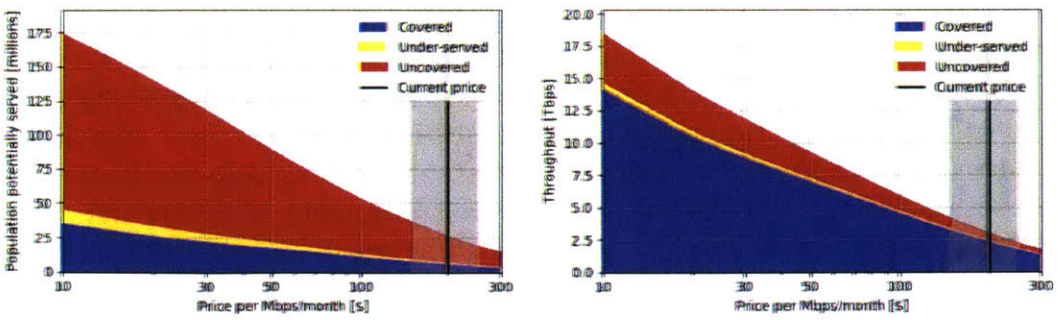


(a) Worldwide

(b) 37 countries

Figure 7-13: Uncovered and under-served population that can afford satellite connectivity vs. price per Mbps/month.

Notice that these two figures depict the population that could afford the service, but do not take into account other factors such as market uptake (see Section 10.2.2), and satellite saturation effects. Figure 7-14a shows the population served (i.e., taking into account the demand model), whereas Figure 7-14b shows how capacity is split across the three market segments considered division of throughput. Note how even representing a small fraction of the users, those in connected areas will still account for most of the throughput.



(a) Worldwide **(b)** 37 countries

Figure 7-14: Estimates of a) population that will be subscribed to satellite-backhauled mobile broadband and b) throughput required vs. price per Mbps/month.

Analysis of main effects

In this section, the main effects for each of the decisions with respect to sellable capacity and price per Mbps/month required for profitability are analyzed. Considering that a full enumeration of the tradespace is unfeasible from a computational standpoint, the main effect (as a percentage point value) of each of the variables and each of the decision options is computed as:

$$Y_{(d_k=\mathcal{O})} = \frac{\mathbb{E}\left[y\right]_{\mathcal{A}|d_k=\mathcal{O}} - \mathbb{E}\left[y\right]_{\mathcal{A}|d_k\neq\mathcal{O}}}{\mathbb{E}\left[y\right]} \cdot 100 \qquad [\%] \tag{7.58}$$

where:

- $Y_{(d_k=\mathcal{O})}$ is the main effect of variable y for cases where decision d_k is assigned option $\mathcal{O}$,
- $\mathbb{E}\left[y\right]_{\mathcal{A}|d_k=\mathcal{O}}$ is the expected value of the quantity of interest, y, over those architectures $\mathcal{A}$ where the value of decision d_k *equals* option $\mathcal{O}$,
- $\mathbb{E}\left[y\right]_{\mathcal{A}|d_k=\mathcal{O}}$ is the expected value of the quantity of interest, y, over those architectures $\mathcal{A}$ where the value of decision d_k *does not equal* option $\mathcal{O}$, and
- $\mathbb{E}\left[y\right]$ is the expected value of the quantity of interest, y, over all architectures.

This main effect value is similar to the *property variable sensitivity* proposed by Simmons in reference [36], with the main difference being that here the percentage difference is computed instead of the average difference.

Figure 7-15 shows the main effect of each of the decisions on the price per Mbps/month. Each subplot corresponds to a decision, and each row shows the main effect of selecting a particular option. The red line in the center of each box indicates the average

effect on the price per Mbps/month of selecting that particular option; the boxes represent the 95% confidence interval, obtained by bootstrapping with 10,000 replicas. As a first-order approximation, the statistical significance of the main effect can be determined by looking at the confidence intervals: if two boxes overlap, the difference in the main effect of the two variables is not statistically significant. Similarly, if a box crosses the 0% vertical line (these boxes are indicated by in lighter colors), there is insufficient statistical evidence to assert that the main effect is different from 0. Finally, the reader should note that with this main effects analysis one cannot distinguish between first-order effects and higher-order interaction effects. For example, having 740 beams per satellite has a similar main effect to selecting a GEO orbit, in part because GEO satellites have 740-beam payloads (whereas LEO do not).

From Figure 7-15, it can be observed that MEO and GEO networks would be able to offer lower prices per Mbps/month than LEO networks, which are, on average, 27% times more expensive. As for the number of planes and the number of satellites per plane, the smaller the number of planes and of satellites per plane, the lower the price per Mbps/month. This merely confirms that there are diminishing returns with large constellations, as most of the value is delivered by the initial satellites. It is also interesting to note that the higher the number of beams per satellite, the lower the cost per Mbps/month. This supports the hypothesis that it is more cost-effective to have a small number of highly-capable satellites (similar to the current vHTS proposals in GEO and MEO) than many low-complexity satellites (as in current LEO mega-constellations proposals). Finally, with regard to the frequency of operation, there is a reduction of $\sim$8% in the price per Mbps/month when transitioning feeder links into higher frequency-bands (i.e., moving from Ka-band to Q/V or E band), and a similar effect is observed for the user frequency, where the Q-band offers the lowest price per Mbps/month. Finally, E-band ISLs result in a 10% increase in the price per Mbps/month, whereas using the optical band allows for a reduction of $\sim$ 5%. However, this latter effect is not statistically significant as compared to not having ISLs at all.

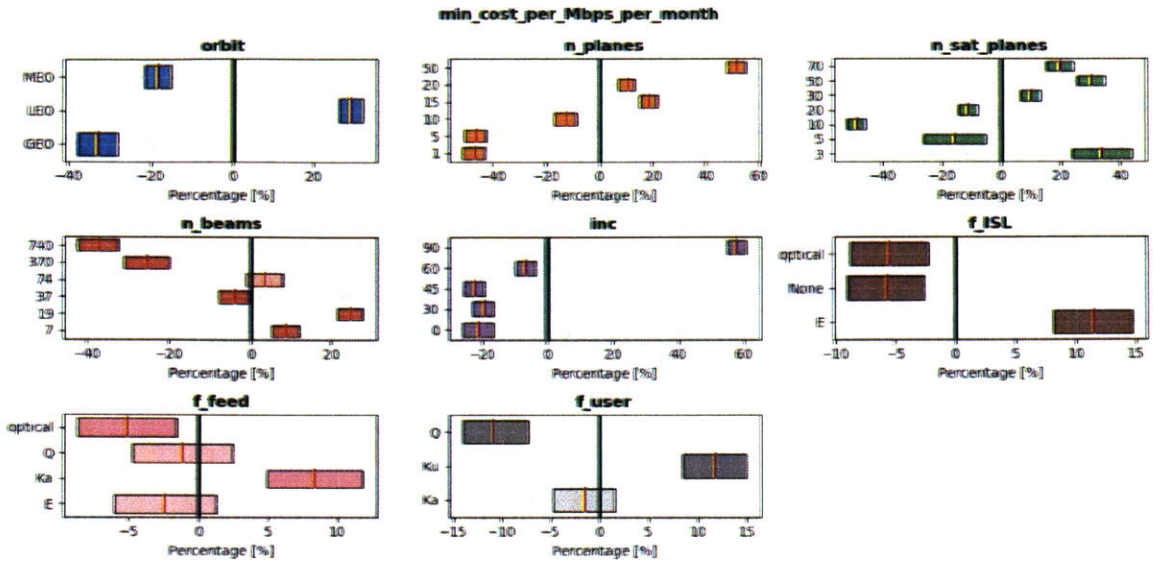


Figure 7-15: Main effect of decision for satellite networks on price per Mbps/month. Each subplot corresponds to a different decision. The patches represent the 95% confidence interval for the main effect and were computed by bootstrapping (10,000 replicas).

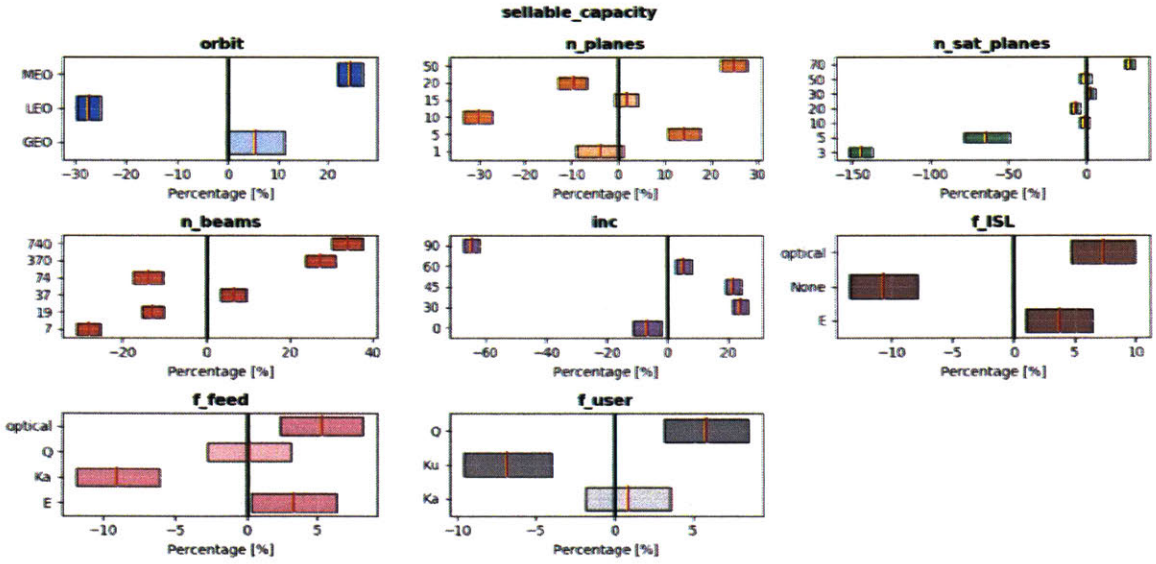


Figure 7-16: Main effect of decision for satellite networks on sellable capacity. Each subplot corresponds to a different decision. The patches represent the 95% confidence interval for the main effect and were computed by bootstrapping (10,000 replicas).

Figure 7-16 shows the main effect on the sellable capacity of the main satellite network decisions. Out of all decisions, the ISL frequency and user frequency play the least important roles and are not significant when computing the sellable capacity

of a network. The type of orbit has a large impact, with MEO networks providing on average 25% more throughput than LEO networks. This is mainly a consequence of the design constraints imposed on such networks (it would not make sense to have a LEO network composed of 6 tonnes vHTS). As expected, other decisions such as the number of beams, the number of planes, and the number of satellites per plane, all have a strong effect on the sellable capacity: the more beams/planes/satellites, the higher the network's sellable capacity. Finally, in terms of orbit inclination, polar orbits (90°) result in $\sim$ 60% less sellable capacity, whereas inclinations of 30°and 45°result in moderate gains of around 20%.

Global sensitivity analysis

The results presented in the previous section hinge on a large number of model parameters for the cost, financial, and satellite-sizing models, which have degrees of uncertainties associated with them. In this section, the variance of the results as a function of such uncertainties is analyzed through global sensitivity analysis. In particular, the following parameters are considered: satellite and payload weights, lifetime, cost of satellite, backbone transit, and CPE terminal, launch costs, percentage of maximum sellable capacity sold; user demand (per user allocation as described in Section 7.2.3); and learning rate (i.e., the mass production cost-discounts). Table 7.14 shows the variation range (i.e., uncertainty) associated with each of the parameters. These values were derived from: a) model errors, for the satellite-sizing and satellite cost models, b) literature, for the learning rate, launch costs, and lifetime, and c) estimates obtained from interviewing industry experts[6], for the remaining of the parameters.

Table 7.14: Variability ranges for parameters of the space architecture models

Parameter	Range	Units
Payload weight	±40%	Percentage
Satellite weight	±40%	Percentage
Lifetime	±20%	Percentage
Satellite cost	±70%	Percentage
Backbone transit cost	[3 - 20]	$/Mbps/month
CPE terminal cost	[300 - 50,000]	$/terminal
Launch cost	±40%	Percentage
Percentage max. capacity sold	60 - 100	% of sellable throughput
User demand	±50%	Percentage
Learning rate	[0.75 - 0.95]	[-]

To understand how these parameters affect the metrics obtained, the sensitivities of two quantities of interest (minimum price per Mbps/month to achieve profitability

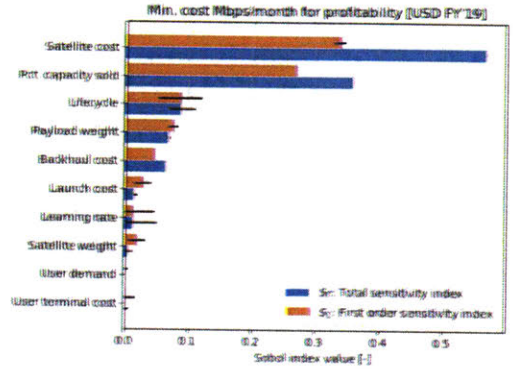
[6]A satellite consultant with more than 20 years of experience and a system architect at a major satellite operator were consulted

and population that can afford the service) to these parameters are analyzed using global sensitivity analysis via Sobol indices [273]. This type of analysis quantifies the influence of each parameter on the final value of a quantity of interest, y, using two indices: the Sobol index (S_1) quantifies the reduction of variance in the output when a variable d_i is fixed; and the total-order Sobol index (S_T) quantifies the expected variance if only d_i was to vary (i.e., the other variables are fixed). The values of the first and total Sobol indices are given by:

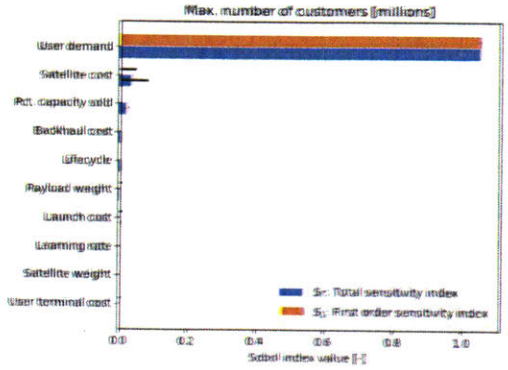
$$S_{1_i} = \frac{V(\mathbb{E}_{d_{-i}}(y|d_i))}{V(y)}, \qquad S_{T_i} = \frac{\mathbb{E}_{d_{-i}} V(y|d_{-i}))}{V(y)}, \tag{7.59}$$

where $V(\mathbb{E}_{d_{-i}}(y|d_i))$ is the variance of the expected value of the quantity of interest, y, when the value of variable d_i is fixed (i.e., it measures the first-order interactions of variable d_i), and $\mathbb{E}_{d_{-i}} V(y|d_{-i}))$ is the expected variance if all the values but variable d_i are fixed (i.e., it measures the higher-order interactions of d_i with other parameters). A Saltelli sampler [274] is used to generate sample values for the parameters, which are then passed into the models described in this Chapter for evaluating the quantity of interest. Finally,the Sobol indices are computed using the `python` package `SALib` [275].

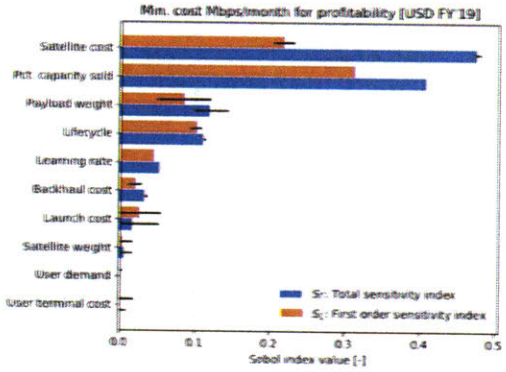
Figure 7-17 shows the values of the Sobol indices for the two quantities of interest for representative GEO, MEO, and LEO architectures. It can be observed that for both GEO and MEO architectures, the parameters which have the most impact on the variability of the minimum price per Mbps/month are satellite cost and percentage of capacity sold. Moreover, the large differences between the values of S_T and S_i indicate that there are higher-order interaction effects between parameters, especially for satellite cost. This is not surprising since satellite cost depends directly on other parameters such as the satellite and payload weights. In contrast, for LEO architectures, the learning rate (i.e., the cost discount obtained from the mass-production of satellites) has the greatest importance, which is also expected, due to the large number of satellites that need to be deployed for LEO constellations. If the learning rate is very high, discounts due to economies of scale will not materialize and the CapEx cost of a constellation would be higher, which increases the price per Mbps/month required for profitability.



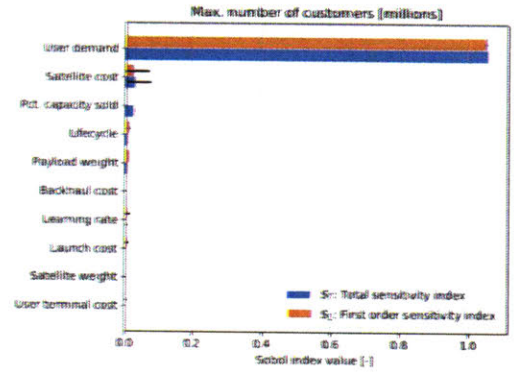
(a) Min. price Mbps/month (GEO)



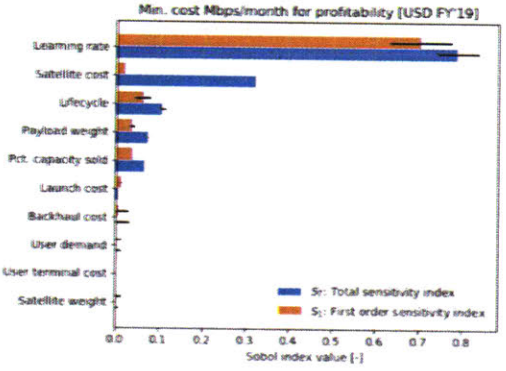
(b) Population afforded (GEO)



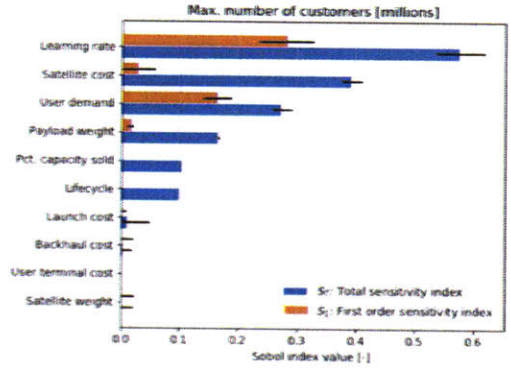
(c) Min. price Mbps/month (MEO)



(d) Population afforded (MEO)



(e) Min. price Mbps/month (LEO)



(f) Population afforded (LEO)

Figure 7-17: Global sensitivity analysis for representative space architectures. The error bars represent the 95% confidence interval for the S_i and S_{T_i} indices, and were computed by bootstrapping (10,000 replicas).

Looking at the sensitivity of the population that can afford connectivity to the model parameters (Figures 7-17b, 7-17d, and 7-17f), for both GEO and MEO architectures user demand (i.e., the data rate allocated per user) is responsible for most of the variability in the results. This is important, as one of the key assumptions made in this dissertation is a relatively low data rate per user and moderate CAGR. Moreover, given the small differences between S_T and S_1, it can be concluded that,

for these two types of architectures, the effects of high-order interactions between variables is limited. As before, the LEO architecture presents a different behavior; in this case, the learning rate, satellite cost (primarily due to interactions), and user demand have the most importance. The reason for this is similar to the reason before; the number of users who can afford connectivity is tied to the costs of deployment, and the parameters that directly affect CapEx costs are also those that the quantity of interest would be most sensitive to.

In light of these results, if higher-fidelity analyses were to be conducted, the main model parameters whose uncertainty values would need to be reduced are satellite cost, percentage of capacity sold, and user demand for GEO and MEO constellations, and the learning rate (and in turn the cost per satellite) for LEO constellations.

7.6 Chapter conclusions

This chapter analyzed the following three research questions

Research Question 7.1

What are the characteristics of dominant space architectures?

Our analysis concludes that GEO and MEO satellite constellations are the most affordable and viable space architectures to extend connectivity to uncovered and under-served regions. For GEO networks, the dominant architectures have a small number of satellites (3-10) carrying highly-capable payloads (> 1 Tbps of throughput), and use higher frequency bands for user and feeder links. For MEO satellites, a larger number of moderate-capacity spacecraft is preferred. Finally, some LEO designs can also be competitive; the model predicts that the best designs would feature a relatively small number of satellites (200-450) with simple payloads, positioned at inclined orbits.

Research Question 7.2

Taking into account both technical and economic factors, what is the potential impact of space systems in terms of connecting additional populations?

It is envisioned that the next generation of satellites (to be launched within the next 2-5 years) will bring down prices to $\sim$ \$100/Mbps/month, which represents an estimated demand of $\sim$5 Tbps (currently the global satellite capacity for data networks is below 2 Tbps), which according to the demand model, would be affordable for up to 120 million of currently uncovered and under-served people worldwide, with 40 million of them adopting the service. However, further reductions in price may be achievable for the future generations of satellites (8-10 years); if prices drop below the \$35/Mbps/month threshold, the number of additional people currently uncovered

and under-served for whom satellite backhaul could be affordable would rise to 300 million worldwide, with 81 million adopting the service.

Research Question 7.3

Which technical decisions play the most important role in the design of space systems? Which technologies, if further developed, would yield the most benefit for such systems?

There are two important technology trends that could allow for lower costs per capacity unit for future satellite systems. First, transitioning to higher frequency bands (for both the user and feeder links) enable cost reductions through increasing in sellable capacity. Second, further miniaturization and integration of components enable payload weights to be further reduced, which in turn could enable lower satellite costs.

Chapter 8

Aerial networks

Aerial connectivity has the advantages of allowing for broader coverage areas than terrestrial networks, encountering less interference caused by obstacles (buildings, ground elevations), and having shorter deployment times. This makes aerial networks especially suitable for rolling out basic connectivity services in remote and sparsely-populated areas. As already mentioned, there are two main categories of aerial systems: high-altitude platforms (HAPs), which fly at the 20-35km range (where lower wind speeds are registered), above the commercial airline airspace, and low-altitude platforms (LAPs), which fly below 1km altitude.

HAPs have gained much attention in the last two decades, specifically after 1997, when the ITU coined the term *high-altitude platform* at the World Radiocommunication Conference (WRC-97) and allotted a frequency band for high-altitude telecommunication stations. In that sense, several commercial efforts to deploy HAP-based networks are underway, with some systems in development and others already in testing.

Most of the concepts proposed consist of unmanned platforms which can fly for long periods of time (weeks to several months). In particular, three categories of platforms can be distinguished:

- **Aircraft**: an unmanned high-altitude, long-endurance solar or fuel-propelled aircraft is used as the aerial platform. These aircraft have typical takeoff weights between 75 - 500 kg, and can support payloads with masses between 2 - 20 kg and power consumption below 200 W. This approach is being pursued by Facebook, Airbus, Aurora Sciences, or HAPSMobile.
- **Airships**: the aerial platform consists of a large aerostat with a rigid or semi-rigid outer shell, which holds the solar panels, payload, and electric propellers. Proposals in this category consist of large systems which generate multiple kilowatts of power, and are capable of hosting payloads with masses in the hundreds

to thousands of kilograms. The main projects involving airships are Thales' Stratobus, StratXX's X-station, and Lockheed Martin's HALE-D.

While airships can host much higher-mass payloads, given the low data rates required to provide communications in sparsely-populated regions, this is superfluous. And since airships are considered to be between two to three times more expensive than high-altitude UAVs [81], with no advantage in terms of their station-keeping capabilities, they will not be considered in this section.

- **Balloons**: stratospheric balloons are used to create a communications network in the sky. Unlike in the other two categories, balloons are not propelled and therefore station-keeping capabilities are limited[1]. The main proponent of this approach is Google's Loon project, which started in 2011 and is expected to begin trial services in Peru and Kenya in 2020.

For LAP systems, the main proposals consist of tethered aerostats (balloons and blimps), which can fly at altitudes between 200m and 1km. The main advantage of such systems is that they remain fixed at a position even in inclement weather conditions since they are literally attached to the ground. As of 2019, various companies (e.g., Raven aerostats, Altaeros) have proposed the use of LAPs to provide connectivity to uncovered and under-served regions, although the predominant uses of active deployments have been for research, policing, disaster relief, and military operations.

This chapter presents the models used to evaluate the performance of aerial networks, as well as the results of several analyses due to identify the optimal space architectures. Its objective is to provide answers to the same three research questions formulated in Chapter 7, but as addressed to aerial systems.

Research Question 8.1

What are the characteristics of dominant aerial architectures?

Research Question 8.2

Taking into account both technical and economic factors, what is the potential impact of aerial systems in terms of connecting additional populations?

[1]Balloons can be moved up or down into a layer of wind blowing in a given direction, providing the balloons with limited control capabilities.

Research Question 8.3

Which technical decisions play the most important role in the design of aerial systems? Which technologies, if further developed, would yield the most benefit for such systems?

8.1 Model overview

As in the previous chapter, the performance and economics of aerial networks are analyzed using a system model composed of a performance model and an economic model, illustrated in Figure 8-1.

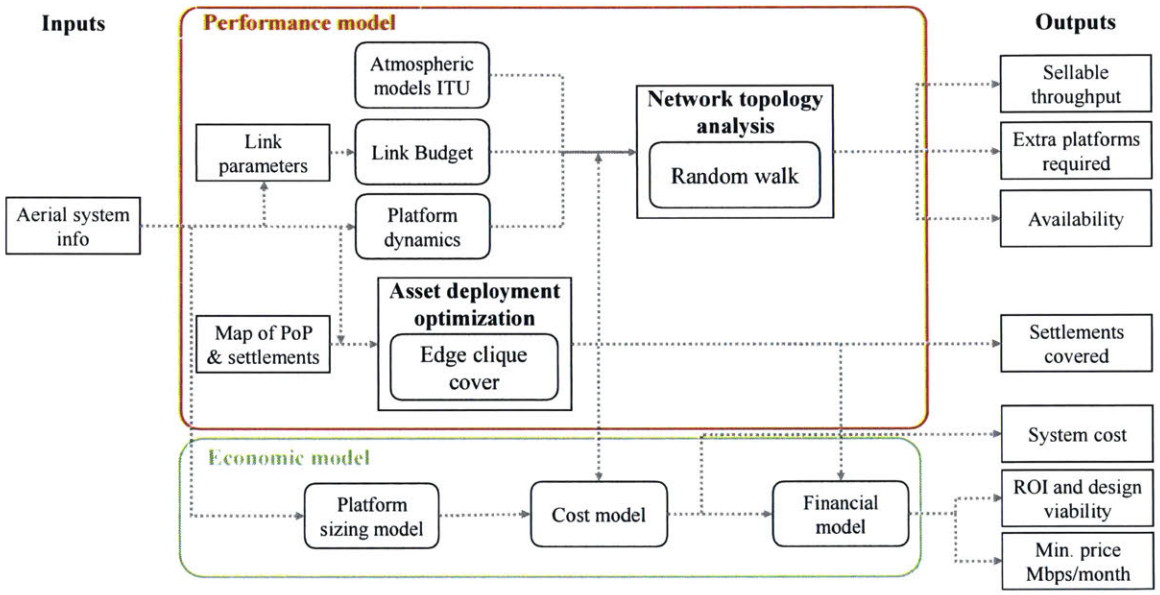


Figure 8-1: Performance and cost model overview for aerial systems.

The performance model evaluates the number and locations of the aerial assets, the number of extra platforms required to ensure continuous coverage, and the payload dimensions required to satisfy the demand (at each frequency band). The number and locations of the assets is determined by transforming the problem into a clique edge covering problem in graph theory; whereas the performance of the payload is assessed using the atmospheric and link budget models described in Section 7.2.1 and Section 7.2.2 respectively, together with a network topology model which uses a Gaussian random walk heuristic to determine the number of additional platforms required to achieve continuous coverage and the target availability of 99%.

The economic model estimates the masses of the aerial platforms and uses the

values as inputs for the platform cost model. The computed cost is used together with the cost of the ground infrastructure to determine the total system cost (CapEx) and the operating expenses (OpEx). Finally, these parameters are fed into a financial model which determines which deployments are able to achieve the desired returns (IRR of at least 15%) and the price per Mbps/month that would be required to achieve profitability.

8.2 Aerial network performance modeling

This section describes the different models used to assess the performance of aerial networks.

8.2.1 Atmospheric model and link budget model

As per the ITU recommendations, aerial platforms can use the following bands for operations:

- **S-band**: ITU recommendation M.1456 establishes that HAP systems can operate in the IMT-2000 band, specifically in the bands 1885-1980 MHz, 2010-2025 MHz, and 2110-2170 MHz for Regions 1 and 3, and 1885-1980 MHz and 2110-2160 MHz for Region 2 [276].

- **Ka-band**: ITU recommendation F.1569 establishes that HAP systems can operate in the bands 27.5-28.35 GHz (DL) and 31-31.3 GHz (UL) [277].

- **V-band**: ITU recommendation SF.1481 establishes that the bands 47.2-47.5 GHz and 47.9-48.2 GHz are designated for use by HAPs [278].

Taking these recommendations into consideration, the ITU atmospheric model for slant-paths (described in Section 7.2.1) is used together with the link budget model (described in Section 7.2.2) to estimate the achievable data rates.

Finally, it is worth mentioning that both the S and Ka frequency bands overlap partially with the allocations for 4G and 5G technologies. This allows aerial systems to connect directly to the users' handsets (whereas for space concepts, CPEs are always necessary), which reduces the overall cost of the system; however lower data rates are to be expected when a direct-to-user connection is used instead of a platform-to-CPE connection.

8.2.2 Location of aerial assets

Since aerial networks can be strategically deployed to target only the settlements that are currently uncovered or under-served, and only where there is a business opportunity, it is especially important to determine the optimal sites for deployment.

High altitude platforms

The method to select the optimal locations of aerial platforms is based on graph theory and is similar to the one published in [62][2] to determine the pointing directions of satellite beams. The inputs for the method are the location of the settlements, and the output is the set of locations where HAPs should be placed. The algorithm is as follows:

1. First, the radius for coverage of each of the HAPs is estimated.

2. Second, the settlement map is transformed into a graph $\mathcal{G}(V, E)$, where each node represents a settlement and an edge is drawn between two nodes if the corresponding two settlements are separated by a distance smaller than twice the radius of coverage. In other words, the presence of an edge between two settlements indicates that they can be covered by the same HAP.

3. The maximum cliques of $\mathcal{G}$ are identified. A clique is a sub-graph of $\mathcal{G}$ such that all nodes are interconnected, that is, a fully-connected sub-graph $\mathcal{G}$. A maximum clique is a clique that cannot be extended by adding a new node (i.e., a maximum clique cannot be a sub-graph of any other clique). The presence of a maximum clique indicates that a HAP would (in theory) be able to cover all the nodes (settlements) in the clique simultaneously.

   Thus, assuming that the number of HAPs to be deployed shall be minimized, the problem can be re-stated as finding the minimum number of cliques that would cover all the nodes of $\mathcal{G}$. This problem is commonly referred to in the literature as the *edge covering problem*, and is NP-hard; therefore, no known methods exist to produce an optimal solution in polynomial time. In this thesis, a greedy approximation which yields good-enough results is employed, as follows:

   (a) The maximum cliques are sorted by number of people covered in descending order.

   (b) A greedy algorithm loops through the list of cliques and determines if the next clique is selected (i.e., if a HAP will be placed to cover the nodes that make up the clique). A clique is selected if none of its nodes (settlements) overlaps with any of the nodes in cliques already selected.

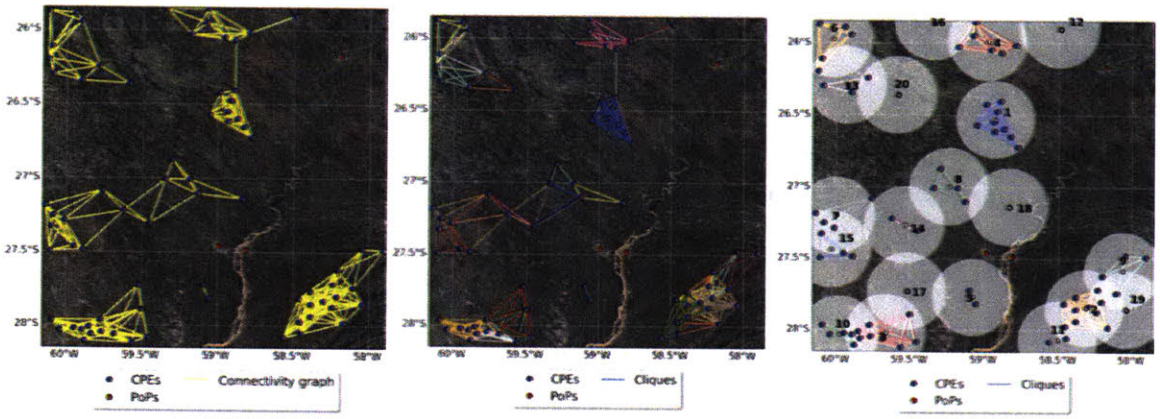
[2]This method is based on a proprietary and patent pending method for beam placement by SES.

(c) For the selected clique, remove its nodes from $\mathcal{G}$ and add the clique to the list of results.

(d) If all the cliques have an overlapping node, or if it has been 20 iterations since the last clique was selected, return to step 3 and recompute the maximum cliques (note that since nodes have been removed from $\mathcal{G}$ the maximum cliques will change). Otherwise, keep iterating in the list (step 3.b).

4. Repeat the procedure until all the nodes of $\mathcal{G}$ are covered by cliques.

Figure 8-2 illustrates this procedure step-by-step for a scenario with 57 settlements (blue dots) and 3 PoPs (red dots). The yellow lines depict graph $\mathcal{G}$, and the maximum cliques of $\mathcal{G}$, each painted with a different color, are shown in Figure 8-2b. Finally, Figure 8-2c presents the locations of the HAPs (small white dots), identified after running the algorithm, as well as their regions of coverage (shaded regions) and selected-clique. The number next to each of the HAPs indicates the iteration of the algorithm in which the clique was selected. Notice that there is a substantial imbalance in the populations covered by the HAPs due to the greedy nature of the algorithm; those assigned first (e.g., 1-4) cover a large number of settlements while those assigned last (e.g., 16-20) cover a single settlement.



(a) (b) (c)

Figure 8-2: Step-by-step illustration of the methodology used to determine the locations of HAPs. a) PoPs, settlements and graph $\mathcal{G}$, as described in step 2; b) maximum cliques of graph $\mathcal{G}$ (each clique is colored in a different color); c) selected cliques, locations of HAPs, and their areas of coverage. The number next to each HAP denotes the iteration of in which the clique was selected.

8.2.3 Network topology analysis

Once the asset locations are determined, a network topology analysis is conducted. This analysis has a threefold objective:

- To determine the number of extra platforms required to guarantee *continuous* coverage over the regions of interest.

- To estimate the number of extra platforms required to guarantee service availability (considering link outage probabilities and the network topology).

- To size the ground segment (i.e., gateways that support the feeder links).

The number of additional platforms required to guarantee continuous coverage for platforms with station-keeping ability (i.e., propelled or tethered platforms) is assumed to be 25% of the total number of platforms deployed, as these backups should only be needed in case of malfunctioning platforms and during the takeoff and landing operations.

For non-station-keeping platforms, the number of additional platforms required is estimated by simulating the movement of the platforms due to wind drift, assuming that there is some control mechanism that can be employed to keep a platform hovering over a region of a certain radius, and then computing the number necessary to ensure coverage over all the settlements.

In particular, a platform's movement due to the wind is modeled by a Gaussian random walk, where the position of the balloon at each time-step is given by:

$$\begin{bmatrix} x \\ y \end{bmatrix}_{t+1} = \begin{bmatrix} x \\ y \end{bmatrix}_t + T \cdot \begin{bmatrix} v_x \\ v_y \end{bmatrix}_t, \qquad \begin{bmatrix} v_x \\ v_y \end{bmatrix}_t \sim \mathcal{N}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} v_w & 0 \\ 0 & v_w \end{bmatrix} \right), \tag{8.1}$$

where T is the simulation time-step (in seconds), v_w is the wind velocity (in m/s), and x and y are given in meters. Since both x and y are Gaussian distributed random variables with mean 0 and standard deviation $\sigma = v_w T\sqrt{N}$, the distance by which the platform has moved after N steps is characterized by a Rayleigh distribution with the same standard deviation. Using the CDF of the Rayleigh distribution, the radius of the area that contains the platform with 95 % probability after N steps, denoted as r, is computed as:

$$\text{CDF}(x) = 1 - e^{-\frac{x^2}{2\sigma^2}} \quad \rightarrow \quad \text{CDF}(r) = 1 - e^{-\frac{r^2}{2\sigma^2}} = 0.95 \tag{8.2}$$

$$r = \sigma\sqrt{\ln\left((1-0.95)^{-2}\right)} = 2.44\sigma = 2.44 \cdot v_w T\sqrt{N}. \tag{8.3}$$

For the control mechanism, it is assumed that on each step, a balloon can be controlled with probability p and that the control effect allows it to counteract the wind speed (i.e., the platform will remain at the current position). This has the effect of reducing the standard deviation of the Rayleigh distribution to $\sigma' = v_w T\sqrt{N(1-p)}$. Thus, given a maximum allowable distance $r_{\max}$, the value of p that guarantees the balloon will remain within a circle of radius $r_{\max}$ from the target position is given by:

$$p = \max\left(0, 1 - \left(\frac{r_{\max}}{v_w \cdot T}\right)^2 \frac{1}{N \ln\left((1-0.95)^{-2}\right)}\right). \tag{8.4}$$

The simulation is run for one day (N=144), with the radius of coverage of a platform set as 40 km, wind speed $v_w = 10$ m/s (which results in a wind profile at 20 km in line with the representative values of the Scalar Wind Profile Model [279], whose 50th and 99th percentiles absolute wind speeds are 7 m/s and 25 m/s, respectively), and assuming that wind direction is constant within 10-minute intervals ($T = 600$ seconds). Figure 8-3 shows the number of platforms required as a function of the maximum allowable distance r. An analysis of historical data[3] from Google's Loon project revealed that their balloons were able to be confined within an area of radius ~100 km 95 % of the time. This target value will be assumed for the rest of the chapter, which, based on the curve in Figure 8-3, will required five aerial platforms for non-station keeping concepts to meet the desired continuous coverage targets (i.e., four back-up balloon for each "active" HAP).

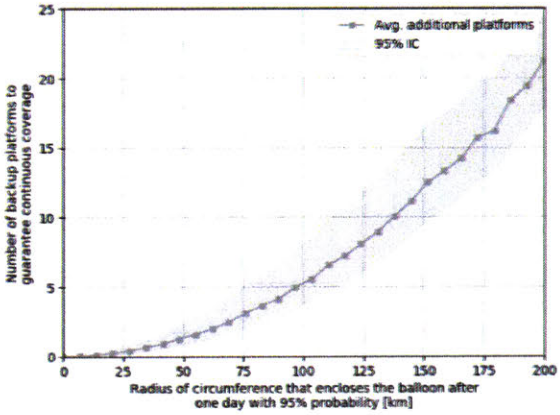


Figure 8-3: Performance and cost model overview for aerial systems.

The availability of the service is computed after the network topology is known. Five types of topologies are considered: bus, ring, star, mesh, and no-crosslinks (i.e., direct-to-ground). For bus and ring topologies, it is assumed that each platform has two ISL terminals, whereas for star and mesh topologies, each platform has three terminals. To design the network topology (i.e., which platforms are linked), a four-step procedure is followed:

1. Construct the connectivity graph between platforms, with platforms as nodes and any pair of platforms within range and LoS is connected by an edge.

[3]Historical data was extracted from https://www.flightradar24.com/ for 94 balloons deployed over Peru and Puerto Rico between August 2018 and September 2019

2. Created a sorted list of nodes by ascending order of their degree (number of neighbors).

3. The *edge betweenness centrality measure* is computed for all the edges. The *edge betweenness centrality measure* is the sum of the fraction of all-pairs shortest paths that pass through a given edge:

$$c_B(v) = \sum_{s,t \in V} \frac{\sigma(s,t|e)}{\sigma(s,t)} \tag{8.5}$$

where V is the set of nodes, $\sigma(s,t)$ is the number of shortest (s,t)-paths, and $\sigma(s,t|e)$ is the number of those paths passing through edge e.

4. Iterate through the nodes in the sorted list from step 2, and for each node, selected the k edges with the highest *edge betweenness centrality* measure if both platforms at the end of the link have ISL terminals available.

The strategy followed in the previous procedure assigns links first to nodes with the fewest neighbors, and the links assigned are those more crucial in connecting pairs of balloons.

Then, the node availability is computed assuming that the link failure probability is 0.5%. If any node is below the 99% availability threshold (as defined in Section 6.3.1), additional platforms (up to 5) are added to act as relays in an alternative path to the closest PoP (which is unreachable from the HAP in the original graph).

Finally, the ground segment is sized by determining the number of ground stations and gateways required for operations. The number of ground stations on each deployment is the smaller value between the number of PoPs present in the final configuration and the number of ground stations needed to transmit the forward traffic required (S); the number of gateways on each ground station depends on the number of platforms within line-of-sight from each ground station, and is given by:

$$N_{GW} = \min\left(n_{\text{PoP}}, \frac{S}{R_{\text{bfeed}}}\right), \tag{8.6}$$

where n_{PoP} is the number of PoPs visible, S is the total throughput of the deployments, and R_{bfeed} is the data rate attainable per feeder link.

8.2.4 Aerial networks sizing model

Given the relatively less-developed status of the aerial platform industry (and the lack of historical data available), few sizing models similar to the one used in Chapter 7 exist. Instead, most approaches use a bottom-up, multi-disciplinary optimization approach, where detailed physics-based models and simulations of the aircraft's sub-

systems (aerodynamics, power, propulsion, stability, etc.) and concept of operations are jointly-considered to produce a detailed design for the platform [98, 280, 281].

However, this approach would be overkill for the purposes of this dissertation, as a) it requires the evaluation of multiple expensive optimization functions, b) detailed knowledge of each of the components and subsystems is needed, and c) the accuracy gains obtained from detailed mass-models are offset by the longer evaluation times, especially given that mass is not a driver for the results of the analyses. Instead, simplified models based on the results of previous research are used, as detailed below.

UAV-based platforms

For UAV-based platforms, the sizing model is inspired by the results obtained in reference [280]. The types of structures considered within the reference are similar to the most recent designs by companies such as Facebook (Aquila), Airbus (Zephyr), Aurora Flight Sciences (Odysseus), and HAPs Mobile (HAWK30).

The model is based on a mass breakdown of the different structural components as a function of the total mass of the aircraft, which is shown in Figure 8-4. The results from [280] show that there is little variation between single-boom and dual-boom designs (in terms of mass breakdown), and these values are considered as guidelines. Moreover, even though the mass breakdown values reported are optimal for platforms operating at 20°latitude, it will be assumed that the mass breakdown does not vary substantially for higher latitudes.

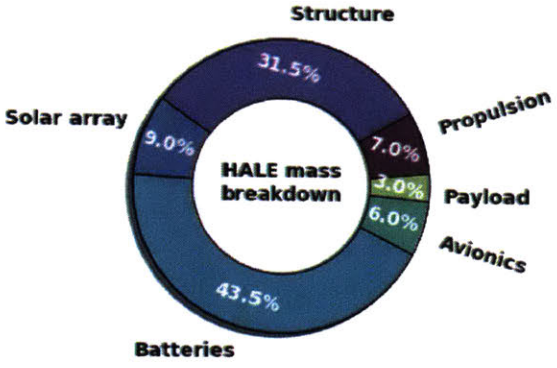


Figure 8-4: Mass breakdown for solar-powered, high-altitude, long-endurance aircraft. Adapted from [280].

Balloon-based platforms

The sizing model for balloon-based assumes that helium is used as the lifting gas, as in the designs from Loon, Stratobus, or HALE-D. The procedure to size the balloon-based platforms attempts to estimate the volume of helium required to provide the

necessary lift over the period of deployment. To do so, the system mass must first be determined.

In addition to the payload mass , there are other subsystems (thermal, solar power, avionics, propulsion) that also contribute to the weight of the system. As with UAV-based platforms, a mass breakdown model is used to estimate the weight contributions from such elements. In this case, it is estimated that the payload mass represents 10% of the total system mass.

Once the system mass is computed, the volume of helium (V_{He}) necessary to provide the required lift can be estimated using the density of helium ($\rho_{\text{He}}(h)$) and air $\rho_{\text{air}}(h)$ at a given altitude h. The density of helium can be estimated using the ideal gas equation, as shown in Eq. 8.7:

$$\rho_{\text{He}}(h) = \frac{P(h) \cdot M}{R \cdot T(h)} \qquad \left[\frac{g}{L}\right], \tag{8.7}$$

where $P(h)$ and $T(h)$ are the atmospheric pressure (in atmospheres) and temperature (in K) at height h, M is the mass number of helium (4.03 g/mol), and R is the universal gas constant (0.08206 L·atm/(mol·K)). Once density is known, the volume and mass of helium are computed as

$$V_{\text{He}} = \frac{m}{\rho_{\text{air}}(h) - \rho_{\text{He}}}, \tag{8.8}$$

$$m_{\text{He}} = V_{\text{He}}\rho_{\text{He}}(h), \tag{8.9}$$

where m is the total mass of the system, and $\rho_{\text{air}}(h)$ is the density of air at height h, which can be extracted from the US standard atmosphere tables in [282].

8.3 Aerial networks economic model

8.3.1 Cost model

As before, detailed parametric cost models for aerial networks are lacking in the literature due to the early stage of development of aerial platforms compared to satellite systems. Instead, simplified CER models based on platform weights and performance are used.

Heavier-than-air platforms (UAVs)

The average of the US Army's performance-based CER and weight-based CER models are used to characterize the CapEx costs of heavier-than-air (HTA) platforms [283].

The performance-based CER model is shown in Eq. 8.10,

$$\text{UAV}_1 = 118.75 \cdot (\text{endurance} \cdot m_{\text{payload}})^{0.587} \cdot e^{-1.951} \qquad [FY'03\$k] \tag{8.10}$$

where UAV_1 is the theoretical first-unit cost (in FY'03 \$k) of air vehicle hardware, normalized for learning (95% slope) and rate (95% slope), *endurance* is the UAV's endurance in flight-hours, and m_{payload} is the weight of the total payload in pounds.

The weight-based CER model is shown in Eq. 8.11:

$$\text{UAV}_1 = 12.55 \cdot (\text{MGTOW})^{0.749} \cdot e^{-0.371} \qquad [FY'03\$K] \tag{8.11}$$

where UAV_1 is the theoretical first-unit cost (in FY'03 \$k) of UAV air vehicle hardware, normalized for learning (95% slope) and rate (95% slope), and MGTOW is the maximum take-off weight in pounds.

The reader should note that these cost models were originally developed for military, fuel-propelled UAVs; however, here they have been applied to solar-powered aircraft, which might change the cost relationships. Because both CER models use measures that are sensitive to the UAV technologies, the validity of these models for solar-powered designed UAVs for communication purposes needs to be verified. Given the scarcity of data and the inherent difficulties in carrying out this validation, sensitivity analyses will be conducted instead.

Lighter than air platforms (balloons)

For lighter-than-air (LTA) platforms, the CapEx costs are computed taking into account the costs of the payload, electronics, solar panels, as well as the volume of helium required to provide lift, as detailed in Table 8.1. Using these values, the total cost per balloon is determined to be \$49,500 (not considering the cost of helium required, as computed in Eq. 8.9). This number is in line with estimates for high-pressure balloons used in current projects[4].

Ground segment

The ground segment cost model is similar to the one developed for satellite concepts in Chapter 7, with the main difference being that the antenna diameters (and the number of dishes co-located in the same facility) are much smaller. In particular, it is assumed that 1.2 m antennas are used and that no new infrastructure (in terms of dedicated buildings and facilities) is required to host them. Thus, the cost model for the gateways (including power amplifiers, modulators, and the rest of the equipment)

[4]Based on a personal interview with a source from industry.

Table 8.1: Parameters of the cost model for lighter than air platforms.

Item	Price	
Balloon shell	\$12,000	per balloon unit
Electronics and batteries	\$10,000	per balloon unit
Solar panels	\$6,000	per balloon unit
Structure	\$1,500	per balloon unit
RF equipment	\$7,000	per balloon unit
Feeder antenna	\$7,500	per balloon unit
User antenna	\$7,500	per balloon unit
Crosslink antenna	\$6,000	per antenna unit
helium	\$7.50	per m^3

is given by:

$$C_{\text{GW-ant}} = 50 + 39 \left(\frac{D_{GW} \cdot N_p}{N_c} \right)^{2.7} \qquad [k\$], \tag{8.12}$$

which results in a cost of \$59.8k per 1.2 m gateway antenna, in line with the costs of commercial antennas for satellite services.

8.3.2 Financial model

The financial model estimates the free cash flow for each of the deployment, as well as the net present value and the internal rate of return (IRR). If the IRR is higher than 15%, a deployment is considered profitable. For the model, the CapEx costs are computed using the cost models described in the previous section, whereas the OpEx costs are estimated as follows:

- For LTA concepts such as balloons and blimps, the OpEx is primarily driven by the re-fill costs of helium required for successive launches, which is given as $\$7.5/m^3$, according to estimates from the USGS National Minerals center[5], and the costs to recover a balloon, procure a new shell costs, in addition to day-to-day maintenance. These expenses amount to 30% of the CapEx.
- For HTA platforms, the OpEx is estimated to be 10% of the CapEx .

In addition, the following assumptions are made:

- The backbone transit cost is estimated to be \$20 per Mbps/month with a CAGR of -5%.

[5] Available at https://www.usgs.gov/centers/nmic/helium-statistics-and-information

- The ARPU is set as 2% of the estimated monthly income, averaged across of all the settlements covered.
- The number of subscribers is assumed to increase progressively up to 60%, and the data-consumption per user is 1GB with an compound annual growth rate of 9%.
- HAPs have an endurance of 135 days and a lifetime of 4 years.

Table 8.2 contains an example of the free cash flow analysis for a deployment in Nigeria.

8.4 Tradespace exploration for aerial networks

8.4.1 Design space

The design space for aerial networks is depicted in Table 8.3 using a morphological matrix. This morphological matrix allows us to represent a wide range of concepts, corresponding to both actual proposals and novel design configurations. For example, the architecture {HAP, Active, Bus, CPE, 19, Ka, E, optical} corresponds to a network of UAVs which connect to CPE and use Ka-band for all the user links (similar to the proposals by Facebook), E-band for the feeder links, and optical frequencies for the cross-links; whereas the architecture {HAP, Passive, Mesh, Direct, 1, S, Ka, optical} corresponds to a network of balloons which connect directly to user handsets using S-band, while using Ka-band for the feeder links and optical crosslinks (similar to Google's Loon project).

The following decisions were considered when studying aerial networks:

- **Altitude**: The aerial platforms considered can either be HAPs with an altitude greater than 20km, or LAPs. In the latter case a distinction is made between platforms flying at 200m or lower, and platforms between 200m and 1km.
- **Station-keeping ability**: Platforms can have passive or active station-keeping capabilities. Active platforms (e.g., planes, propelled blimps, tethered platforms) have the ability to hover over certain regions, whereas passive platforms are subject to wind-drift or have very limited station-keeping abilities (e.g., balloons).
- **Network topology**: The network topology determines the number of cross-link terminals as well as the ground segment requirements. Five topology options are considered, with various numbers of cross-link terminals per platform: in *bus* and *ring* topologies, aerial platforms have 2 crosslink terminals); whereas in the *star* and *mesh* topoplogies, there are 3 crosslink terminals. The

Table 8.2: Financial model for an aerial deployment connecting in Nigeria.

		CAGR	Year 0	Year 1	Year 2	Year 3	Year 4	Year 5	Year 6	Year 7	...	Year 20
Subscribers												
Total population	40,088	0%	40,088	40,088	40,088	40,088	40,088	40,088	40,088	40,088		40,088
Percentage subscribers				21%	36%	51%	60%	60%	60%	60%		60%
Number subscribers				8,419	14,432	20,445	24,053	24,053	24,053	24,053		24,053
Monthly data usage/data sub (GB)	1.00	9%		1.0	1.1	1.2	1.3	1.4	1.5	1.7		5.1
Total Data Usage (GB)				101,023	188,768	291,489	373,792	407,433	444,102	484,071		1,484,068
CapEx												
Balloons	5.0		$ 334,439				$ 334,439					
Ground station	1		$ 59,852									
Total CapEx			**$ 394,291**	**$ -**	**$ -**	**$ -**	**$ 334,439**	**$ -**	**$ -**	**$ -**		**$ -**
OpEx												
Replenishment				$84,439	$84,439	$84,439	$84,439	$84,439	$84,439	$84,439		$84,439
Operations				$100,332	$100,332	$100,332	$100,332	$100,332	$100,332	$100,332		$100,332
Backbone transit	$20.00	-5%		$14,966	$26,567	$38,973	$47,478	$49,164	$50,909	$52,717		$82,966
Total OpEx				**$199,737**	**$211,338**	**$223,744**	**$232,249**	**$233,935**	**$235,680**	**$237,487**		**$267,736**
	ARPU											
Revenue	**$ 1.86**	**0%**		**$ 187,902**	**$ 322,118**	**$ 456,333**	**$ 536,863**	**$ 536,863**	**$ 536,863**	**$ 536,863**		**$ 536,863**
EBITDA				**$ (11,835)**	**$ 110,780**	**$ 232,589**	**$ 304,614**	**$ 302,928**	**$ 301,183**	**$ 299,375**		**$ 269,126**
Free Cash Flow			**($394,291)**	**($11,835)**	**$110,780**	**$232,589**	**$304,614**	**$302,928**	**$301,183**	**$299,375**		**$269,126**
Cumulative FCF			*($394,291)*	*($406,126)*	*($295,346)*	*($62,757)*	*$241,857*	*$544,785*	*$845,967*	*$1,145,343*		*$3,694,860*
NPV (i=15%)			**$874,147**									
IRR			**41%**									

Table 8.3: Morphological matrix for aerial systems.

Decision name	Decision options				
Platform altitude	<200 m LAP	< 1km LAP	HAP		
Platform station-keeping	Active	Passive			
Network topology	None	Bus	Ring	Star	Mesh
User access	Direct (handset)		CPE		
Number payload beams	1	7	19	37	
Frequency user ↔ platform	S	Ku	Ka	V/Q	
Frequency platform ↔ gateway	Ka	V/Q	E		
Frequency crosslinks	None	Ka	E	Optical	

fifth topology is the *no-crosslinks* topology, where aerial platforms do not have crosslink terminals and thus must be within LoS of a gateway antenna to operate.

Bus and ring topology architectures favor long chains of interconnected platforms, whereas star topology architectures try to minimize the number of hops to the nearest gateway station. Finally, in mesh topology architectures, aerial platforms are connected creating a mesh, which minimizes the probability of outages due to link failure.

- **User access**: Both *direct* user access and access though a CPE terminal are considered. In the direct method, a user is connected directly to the aerial platform using their handheld device, whereas in the CPE method additional equipment (e.g., a community-shared access point) is required.

- **Number of beams in the payload**: The number of beams in the payload determines the number of times frequency bandwidth can be reused. The valid values for this decision are $\{1, 7, 19, 37\}$.

- **Frequencies of operation**: The frequencies of operation for the user links ({S, Ku, Ka, Q}), feeder links ({Ka, V, E}), and cross-links ({Ka, E, optical, None}) determine the bandwidth available and the data rates achievable.

8.4.2 Tradespace exploration results

This section presents the results of the tradespace exploration analysis for aerial networks, in particular, how different aerial network concepts can be used to provide connectivity to uncovered and under-served regions in 37 developing countries in Africa, South America, and Southeast Asia (shown in Table 6.1). The methods and models described in Sections 8.1 - 8.3 are applied to each of the countries, and results are then analyzed from a triple standpoint. First, an overview of the results and the most effective designs is provided; second, a main effects analysis is conducted to determine the relative importance of each of the decisions; finally, the robustness of the results is studied through sensitivity analysis.

Overview of results

Figure 8-5 shows the tradespace exploration results for aerial networks. The horizontal axis represents the CapEx costs to set up service, and the vertical axis shows the total number of people in the 37 countries studied that would be able to afford the service. Each point corresponds to the aggregated CapEx and population of all profitable deployments in an architecture (i.e., assuming a unique combination of decision-options from Table 8.3, only profitable deployments are considered). The *utopia point* at the top-left corner denotes the location of desired solutions (i.e., large population at low CapEx costs).

It can be observed that passive station-keeping platforms (i.e., balloons without active controls to allow for static placement) are clearly cheaper than active station-keeping platforms (i.e., UAVs). This is not unexpected, as the CapEx costs of active station-keeping platforms (~$1,200,000) greatly exceed the CapEx costs of passive station-keeping platforms (~$70,000, including helium costs).

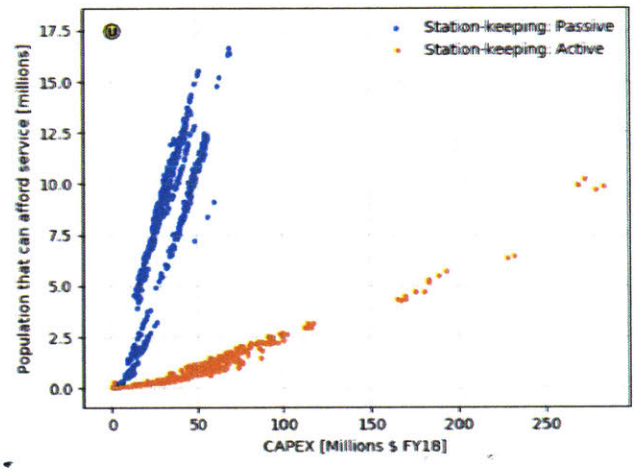


Figure 8-5: Tradespace exploration results for aerial networks. Each point represents a different architecture. The yellow dot at the upper left corner represents the *utopia point* (ideal value). Architectures have been color-coded according to their station-keeping capabilities.

In addition, Figure 8-5 reveals that a relatively small number of people would benefit from aerial network deployments. For passive station-keeping platforms (i.e., stratospheric balloons), only about ~16 million people would be able to afford connectivity, even with the best architectures; for active station-keeping platforms (i.e., UAVs), the number of people impacted would be only ~10 million. For context, the total uncovered and under-served population for the 37 countries considered is 159 million.

In contrast, the required CapEx investment is relatively low, below $100 million and $300 million for balloon-based and UAV-based platforms, respectively. Beyond

CapEx, OpEx costs constitute a significant portion of the total cost of ownership (especially for balloon-based architectures). Taking both CapEx and OpEx into account, Figure 8-6 plots the number of people who would be able to afford service against the price per Mbps/month required to achieve profitability, together with the demand curve[6]. It can be observed that passive station-keeping architectures can achieve profitable deployments with prices as low as $50 per Mbps/month, whereas for active station-keeping architectures, the minimum price is closer to $60 per Mbps/month. These numbers are similar to those obtained for the analysis of next generation of satellite systems (to be launched in the next 2-5 years) in Chapter 7, but are considerably higher than the $30 per Mbps/month attainable by future space concepts. Moreover, the prices per Mbps/month calculated for by aerial concepts have already excluded settlements in which the deployments are not profitable, whereas for the proposed space concepts, estimated prices are likely to be applicable wherever coverage is available.

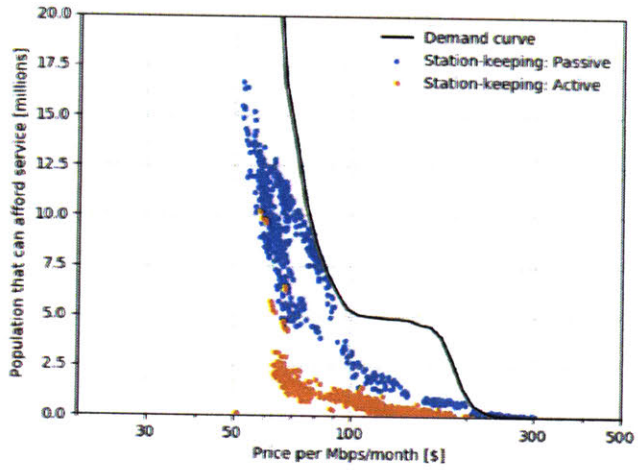


Figure 8-6: Tradespace exploration results for aerial networks. Each point represents a different architecture. Architectures have been color-coded according to their station-keeping capabilities.

Figure 8-7 shows the same tradespace, but this time architectures have been color-coded based on the chosen option for each decision. First, Figure 8-7a shows that HAPs are (as a group) more effective than LAPs in terms of population covered, (since they have larger radii of coverage), which allow for fewer deployments to be required for a given population size. Next, it can be observed in Figure 8-7b that balloon HAP networks are more economical than UAV HAP networks, which results in them being more affordable for larger populations. This is mainly due to the lower CapEx costs per balloon deployed (~$70,000 including helium costs vs. ~$1,200,000 on average for a UAV), which, even after taking into account the higher number of

[6]Computed assuming that uncovered populations are to be provided with 1GB/month and 60% of the market can be captured, and under-served populations are to be provided 3GB/month and 30% of the market can be captured

additional platforms to guarantee continuous coverage (4 for balloons vs. 0.25 for UAVs respectively) and the greater OpEx costs, still results in significant savings. In terms of network topology (Figure 8-7c), architectures with a low number of cross-link terminals (no-crosslinks, bus, and ring) are preferred. Finally, architectures with S- and Ka-band frequencies where direct-to-user links can be established perform best (due to lowest costs), as shown in Figure 8-7d.

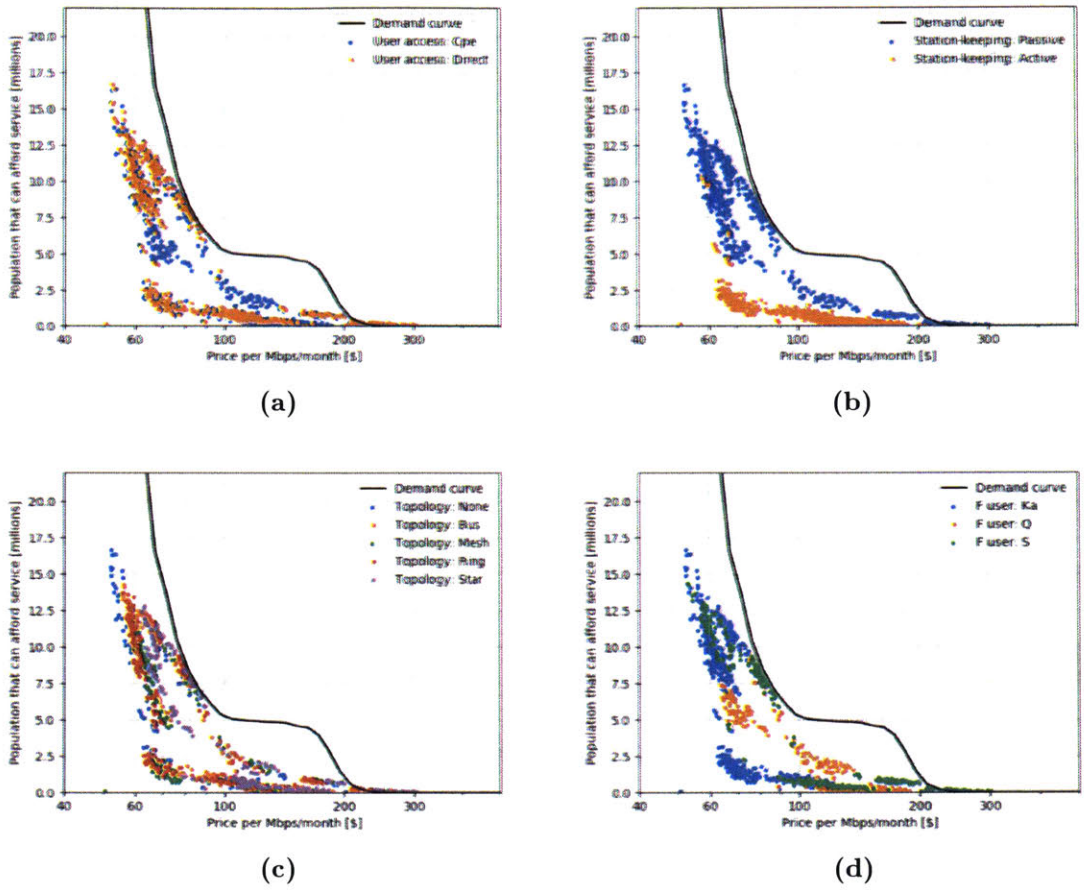


Figure 8-7: Tradespace exploration results for aerial networks by a) platform altitude, b) station-keeping capabilities, c) network topology, and d) user links frequency. Each point represents a different architecture.

Analysis of main effects

The analysis of main effects is conducted using the same method as in Section 7.5.3. Figure 8-8 shows the main effect of each of the decisions on the price per Mbps/month. Each subplot corresponds to a different decision factor from Table 8.3, and each row (with a colored box) represents a different option for the decision. The vertical red line in the center of the box indicates the mean value of the main effect, and the colored box around the mean is the 95% confidence interval for the main effect. In that

sense, the first row of the Station-keeping subplot in Figure 8-8 can be interpreted as: "Active station-keeping enabled platforms have a minimum price per Mbps/month for profitability which is 13% (95% CI 8% - 18%) higher than architecture without active station-keeping capabilities". The degree of overlap of the boxes for two options indicates the statistical significance of the difference in the main effect from selecting either option (i.e., if the boxes substantially overlap, as it happens in the feed-frequency subplot, the difference in main effect is not statistically significant).

From the figure, it can be seen that passive station-keeping architectures can offer lower prices (due to the significantly lower CapEx costs of balloon systems), and that direct platform-to-handset connections are capable of offering lower prices as well. In terms of the number of beams, platforms with multi-beam payloads achieve lower prices than those with a single omnidirectional beam, as this allows for higher throughput. In terms of ISL frequencies, non-crosslinked architectures allow for lower prices per Mbps, whereas options with E-band cross-links are the least preferred. For the user-link frequencies, the Ka-band is preferred (as it allows to reduce prices up to 40% when compared against other frequencies), and finally, for the feeder links, the confidence interval overlapping makes the decision non-statistically significant.

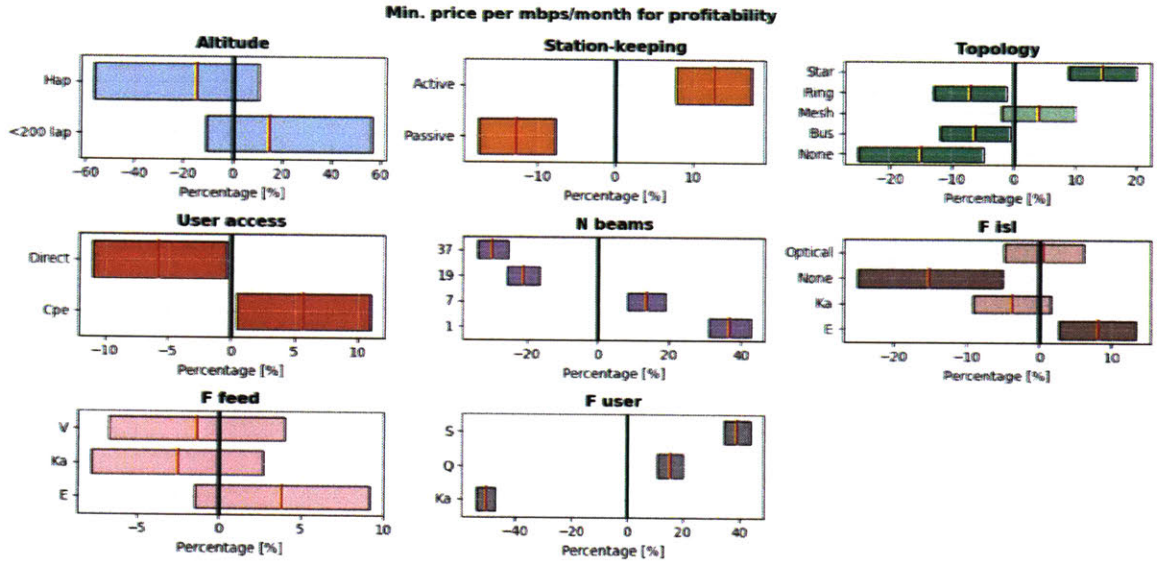


Figure 8-8: Main effect of each decision for aerial networks on price per Mbps/month. Each subplot corresponds to a different decision. The patches represent the 95% confidence interval for the main effect and were computed using bootstrapping (10,000 replicas). Options with a negative main effect (i.e., lower minimum price per Mbps/month for profitability) are preferred.

Figure 8-9 illustrates the main effect on the number of users that can be served profitably for different decisions. It can be seen that HAP concepts are capable of bringing as many as two times more users online (on average) compared to LAP concepts, and passive station-keeping platforms are about 1.5 times more effective

than active station-keeping platforms. In the same sense, since most of the profitable deployments are achieved using a low number of platforms with direct LoS to a PoP, topologies where there are no cross-links between platforms are more effective. Finally, direct-to-user strategies using Ka-band (where more spectrum is available) prove to be superior, allowing to serve an average of 20% more users than CPE-based architectures.

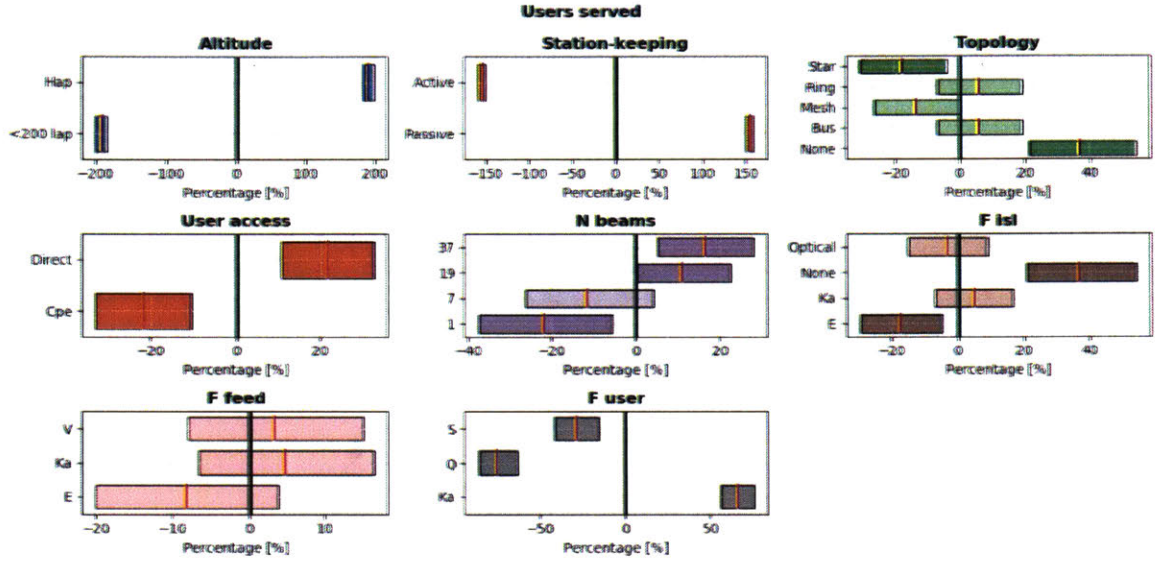


Figure 8-9: Main effect of each decision for aerial networks on number of users that can be served. Each subplot corresponds to a different decision. The patches represent the 95% confidence interval for the main effect and were computed using bootstrapping (10,000 replicas). Options with a positive main effect (i.e., greater number of users served) are preferred.

To summarize, the decisions that have a more beneficial main effect on the number of users served and the minimum price per Mbps/month required to achieve profitability led to architectures where the CapEx and OpEx costs are minimized as much as possible. Thus, the lack of cross-links and the use of direct-to-user strategies, along with a preference for the largest bandwidth available (by using Ka-band and a large number of beams), is unsurprising. Overall, in order to provide affordable coverage to uncovered and under-served regions, very high-capability systems are not required; instead, low-cost, moderate-throughput concepts are best-suited for these goals.

Detailed architecture results

In this section, the specifics of six representative architectures are analyzed in greater depth. These architectures were selected to have good coverage of the range of potential designs, and their parameters are depicted in Table 8.4. In this aspect, some of the architectures are comparable to existing proposals of various companies. For

example, the low-throughput balloon network is similar to Loon's current LTE proposal, whereas the moderate- and high-throughput balloon networks are similar to an evolved Loon proposal with 5G technology onboard, and the high data rate UAV system is similar to the proposals of Facebook's Aquila and Softbank's HAPSMobile.

Table 8.4: Representative architectures used for further analysis

	UAVhigh	UAVmod	UAVlow	Balloonhigh	Balloonmod	Balloonlow
Altitude	HAP	HAP	HAP	HAP	HAP	HAP
Station keeping	active	active	active	passive	passive	passive
Topology	bus	bus	star	bus	bus	mesh
Num. beams	37	7	1	37	7	1
User access	CPE	CPE	Direct	Direct	Direct	Direct
Freq. User	Ka	Ka	S	Ka	Ka	S
Freq. Feeder	E	E	Ka	E	E	Ka
Freq. ISL	Optical	E	E	Optical	E	E

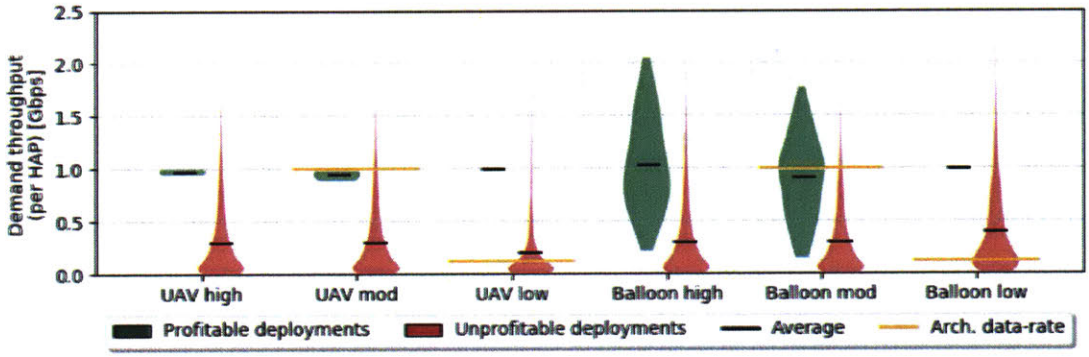


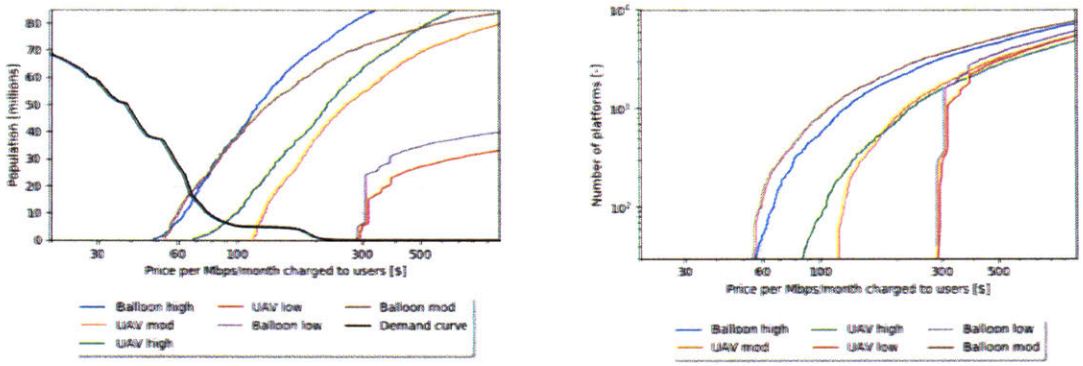
Figure 8-10: Distribution of requested throughput per platform for profitable and unprofitable deployments. The horizontal black lines denote the average requested throughput for each group.

First, to understand the requirements in terms of throughput per platform, the demands per potential deployment are analyzed (i.e., for each deployment, what is the requested data rate per platform). The distributions of requested throughput per platform for the six architectures are depicted in Figure 8-10. Notably, profitable deployments have, on average, a significantly higher throughput per platform than unprofitable deployments ($\sim$ 1 Gbps for profitable deployments versus $\sim$ 300 Mbps for unprofitable ones). Moreover, the distribution of requested throughput per platform for unprofitable deployments is very skewed towards low data rates, the most common requirement being less than 200 Mbps. Finally, it is interesting to compare the requested throughput per platform against the data rates offered; both the low-throughput UAV and balloon architectures are throughput limited (i.e., the requested throughput is higher than what the platform can provide), and consequently the number of profitable deployments is meager (only one deployment is profitable in both cases). In contrast, all the moderate- and high-throughput architectures can

offer data rates higher than what is demanded, and therefore the number of profitable deployments is higher.

In summary, the story told by this image is that HAPs systems are only cost-effective in scenarios where the populations covered are large (and where moderate to high data rates are required), but they are highly inefficient when the populations covered are small.

Another interesting analysis studies how the serviceable population evolves with the price per Mbps/month charged. That is, assuming that a company was able to charge a given price per Mbps/month, what population would they serve (equivalent to a supply curve). Figure 8-11a shows these curves, together with the demand curve. Figure 8-11b depicts the number of platforms deployed as a function of the price per Mbps/month charged (i.e., analogously, assuming that a company was able to charge a given price per Mbps/month, how many platforms would be deployed). Several conclusions can be drawn from these two figures:



(a) Population affordable

(b) Number of platforms

Figure 8-11: Evolution of a) population affordable and b) number of platforms deployed as a function of the price per Mbps/month for selected architectures.

- For all types of platforms, the market equilibrium (i.e., where the supply meets the demand, depicted with a dot in Figure 8-11a) is achieved at a point that provides coverage to less than 15% of the population, which shows the limited impact of aerial systems in terms of their ability to in extend connectivity to uncovered and under-served regions.
- As one tries to increase the number of people to be covered, the price per Mbps/month required for profitability grows rapidly. This growth is more steeper for station-keeping-enabled platforms (UAVs), where providing coverage to half of the population in the 37 countries analyzed would require prices northwards of $250 per Mbps/month, than for balloon networks, where this value would be $100/month. Furthermore, if 80% of the population were to be covered, the required price per Mbps/month for both types of systems would be comparable of those of fiber.

- Looking at the chart of the number of platforms deployed as a function of price per Mbps/month (Figure 8-11b), one can observe a rapid growth of the number of platforms deployed when prices per Mbps/month northwards of $100 can be charged to users. This is due because the Pareto principle applies to this case; a small number of platforms (< 300) are capable of achieving profitability highly populated settlements at relatively low prices per Mbps/month, but the number of platforms required to serve the rest of the smaller settlements grows exponentially (i.e., thousands of platforms would be required). This result is similar to the conclusion obtained from analyzing Figure 8-10, where it was observed that profitable deployments (which all have a low number of platforms, as explained below) requested a significantly higher throughput (because of larger populations covered) than unprofitable deployments.

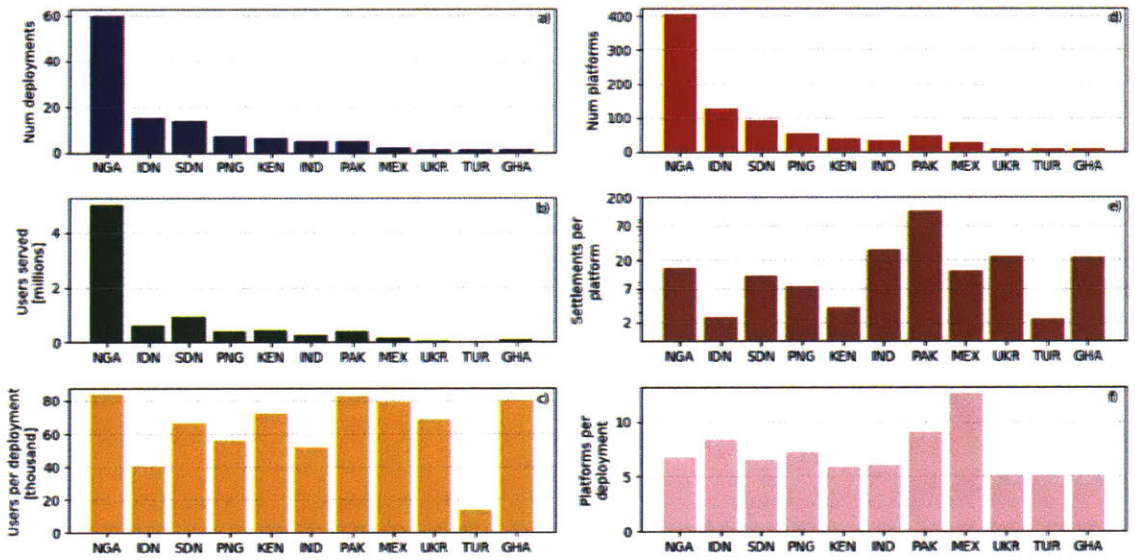


Figure 8-12: Summary statistics for profitable deployments for the moderate-throughput balloon architecture. a) Number of profitable deployments, b) total number of users served, c) average number of users served per deployment, d) number of platforms deployed, e) average number of users per platform, and f) average number of platforms per deployment.

Finally, Figure 8-12 shows different statistics for the profitable deployments for the moderate-throughput balloon architecture (*Balloon mod*), grouped by country. For this architecture, profitable deployments could only be achieved in 11 out of the 37 countries analyzed. It can be seen in Figure 8-12a that the number of profitable deployments varies greatly from country to country, with more than half of the profitable deployments being located in Nigeria, where up to 4 million users could be served using this technology (Fig. 8-12b). In most countries, the average number of users served per deployment ranges from 40,000 to 80,000 (Fig. 8-12c), with the notable exception of Turkey, where the number of users served in its only profitable

deployment is fewer than 15,000. Similarly, the number of settlements covered per platform ranges from 7 to 30 for most countries (Fig. 8-12e). Finally, the total number of platforms deployed (taking into account that for each "active" platform there need to be an additional four platforms to ensure continuous coverage as explained in Section 8.2.2) is below 100 for most countries (Fig. 8-12d), with an average number of platforms per deployment between 5 - 10 (Fig. 8-12f). Given that so many platforms are required for back-up, improving station-keeping technologies could reduce the costs and size of the deployments significantly.

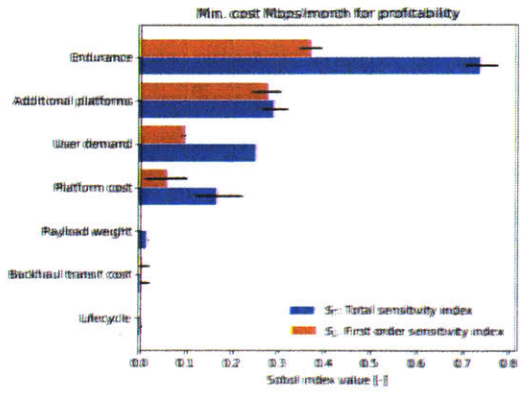
Sensitivity analysis

The sensitivity analysis is conducted using the SOBOL analysis technique, similar to the one described in Chapter 7. In this case, the quantities of interest are the minimum cost per Mbps/month required to achieve profitability, and the the maximum number of customers that could be served by aerial networks. Table 8.5 shows the variation (i.e., uncertainty) associated with each of the parameters considered in the global sensitivity analysis. As in the previous chapter, the ranges were selected from the literature and interviews with industry experts.

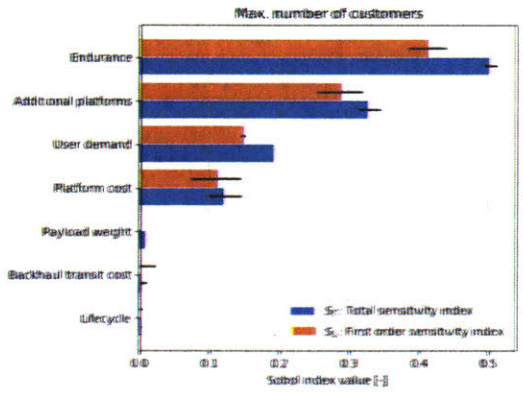
Table 8.5: Variability ranges for parameters of the aerial architecture models

Parameter	Range	Units
Payload weight	±40%	Percentage
Platform cost	±40%	Percentage
Lifecycle	[2-4]	Years
Endurance	±40%	Percentage
Backbone transit cost	[5, 30]	$/Mbps/month
Additional platforms	[50%-200%]	Percentage
Used demand	±40%	Percentage

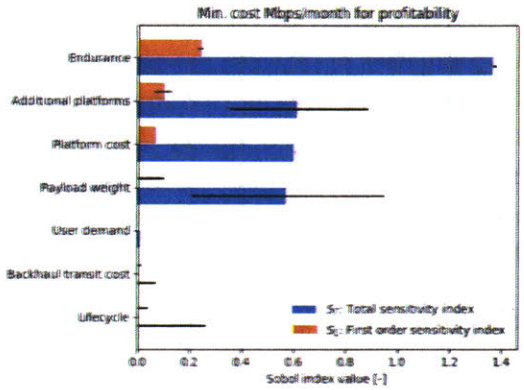
Figure 8-13 shows the values of the Sobol indices for the two quantities of interest for two representative balloon- and UAV-based architectures. It can be observed that for balloon-based and UAV-based networks, the endurance of the platform is the parameter that has the most impact in the variability of both the minimum price per Mbps/month and the population that can afford the service, followed by the number of additional platforms required for service. It can be seen that while for the maximum number of customers, the first-order Sobol sensitivity indices values are of similar magnitude than the total sensitivity indices, for the minimum cost per Mbps/month, there are larger differences between the first-order and total sensitivity indices, which indicates that higher-order interactions have an important effect in this later quantity of interest. Furthermore, the differences between the two indices are larger for UAV networks than for balloon networks, which is evidence of the tighter couplings between variables in UAV systems. Finally, note that while in balloon based networks the user demand is the third most important factor, for UAV-based systems this variable is not very relevant (fifth position).



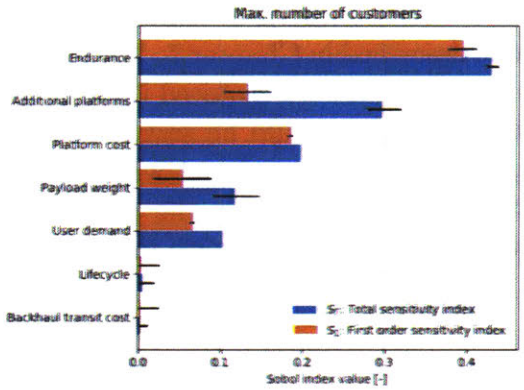
(a) Min. price Mbps/month (Balloon)



(b) Number of customers (Balloon)



(c) Min. price Mbps/month (UAV)



(d) Number of customers (UAV)

Figure 8-13: Global sensitivity analysis for representative aerial architectures. The error bars represent the 95% confidence interval for the S_i and S_{T_i} indices, and were computed by bootstrapping (10,000 replicas).

8.5 Chapter conclusions

To summarize, the following questions are answered in this chapter:

Research Question 8.1

What are the characteristics of dominant aerial architectures?

Given the low average revenue per user (ARPU) per user, the dominant architectures are those that can achieve broad coverage at very low costs. Among the different concepts studied, the Pareto front is composed in its entirety of balloon-based HAPs with limited station-keeping capabilities. UAV-based platforms (are sub-optimal as their CapEx costs (over 1 million dollars per unit) are a large hurdle, and the deployments will rarely work out to be affordable by the uncovered and under-served populations.

Overall, in order to provide affordable coverage to uncovered and under-served regions, very high-capability systems are not required; rather, low-cost, moderate-throughput concepts are best-suited for these goals. Thus, direct-to-handset architectures without crosslinks and using the S- or Ka-band for the user links are preferred.

Research Question 8.2

Taking into account both technical and economic factors, what is the potential impact of aerial systems in terms of connecting additional populations?

Aerial systems will have a limited impact on their ability to extend connectivity to uncovered and under-served regions. For the 37 countries studied and at current costs, the analyses estimate that *at most* 17 million (representing 11% of the total combined uncovered and under-served population of those countries), will be covered by profitable deployments when using balloon-based networks. UAV networks fare even worse, as their CapEx costs are extremely high and do not scale with regard to capacity provided. If deployed, their impact would be minimal, reaching at most 11 million people.

Research Question 8.3

Which technical decisions play the most important role in the design of aerial systems? Which technologies, if further developed, would yield the most benefit for such systems?

The two factors that would increase the number of people covered by balloon-based HAP networks the most would be better station-keeping technology and increased flight endurance. Better station keeping technology would allow for a further reduction in the number of additional platforms required to satisfy the availability requirements and provide continuous coverage, thus significantly reducing the CapEx costs; increased flight endurance would reduce the OpEx costs, which for LTA HAPs are significantly higher than for other systems (for the balloon networks considered, the lifetime OpEx costs are far higher than CapEx costs, whereas for space and terrestrial systems these regularly represent $< 100\%$ of CapEx).

Similarly, for UAV networks, technologies leading to reduced CapEx and OpEx costs would be the most beneficial. Given that these costs are dominated by platform weight and endurance respectively, improvements in battery technology would likely have the most impact on these kind of networks, since batteries make up more than 40% of the total system mass and greatly determine the overall endurance of the system.

Chapter 9

Terrestrial networks

Terrestrial networks for backhauling include both fiber optic and wireless networks. Fiber optic is the most widely-used technology for linking large settlements to PoP or Internet exchange points (IXPs), as well as for international connectivity using submarine cables. In constraint, wireless relay links that use microwave and mm-Wave technologies are common to backhaul cell phone towers in remote locations. Since terrestrial concepts are relatively well established, we need not conduct extensive tradespace exploration. Instead, a fiber-backhaul model and a mm-Wave point-to-point wireless backhaul model are used as baseline concepts for comparison against space-borne and air-borne concepts.

This chapter provides an answer to the following research question:

Research Question 9.1

Taking into account both technical and economic factors, what is the potential impact of terrestrial networks in terms of connecting additional populations?

In particular, the goal of this chapter is to estimate the costs of deployment of fiber optic and mm-Wave backhaul solutions in a greenfield scenario to connect currently uncovered and under-served regions in 37 countries (Table 6.1). This estimation is carried out using a two-step process: first, a quasi-optimal layout of the backhaul solution (fiber optic or mm-Wave) is computed; then, the optimal deployment strategy is analyzed and a techno-economic analysis similar to that presented in the previous two chapters is performed. Finally, the overall impact of these technologies is assessed

9.1 Terrestrial networks modeling

This section describes the models used to estimate the requirements and costs of terrestrial networks. Since terrestrial networks are only used as a baseline, the objective of these models is not to find the optimal architecture, instead, the models represent first-order approximations that are good enough to form comparisons. In particular, the focus is on network layout, which is the main driver of the network cost. Thus, for fiber optic networks, the outputs of the model are the number of kilometers that need to be laid down and the type of the equipment (such as optical cross-connects, switches, passive splitters, etc.) required; whereas for mm-Wave networks, the outputs of the model are the number and height of the towers required.

Finally, the models presented below have been designed with low computational complexity in mind, since they must allow for evaluation over large regions (such as at a country level). Therefore, the models are simplified and have a limited degree of fidelity (i.e., they do not represent solutions that could be used in the real world for network planning purposes). Normally the planning and designing of backhaul networks is done using specialized software which focuses on a very particular region, and it typically takes several months to produce a valid design, since (in addition to the network layout) other factors and constraints such as the network topology, routing, equipment placement, and availability concerns all have to be taken into account.

9.1.1 Fiber optic layout modeling

The objective of the fiber layout optimization is to produce a layout that connects the maximum number of uncovered and under-served settlements in the most cost-effective manner. It is assumed that fiber can only be routed over roads and railway tracks, and that a PoP can be found in any covered settlements with more than 50,000 inhabitants.

The model is based on the connectivity graph, where each unconnected settlement and PoP is a node and the roads and railway tracks are the edges. Given this connectivity graph, the objective is to connect the nodes where there is Internet presence (PoP) to a set of destination nodes (settlements) using a minimum-cost tree, assuming that the cost of selecting an edge (i.e., building fiber along that road) is proportional to its length. Although this problem may resemble a minimum spanning tree problem at first glance (which can be solved efficiently in polynomial time), it is in fact more complex, since only a subset of nodes needs to be connected. Formally, the problem at hand is defined as a Steiner tree problem [284], which is NP-complete [285].

Given the large number of nodes and edges of the connectivity graph, it is computationally impossible to solve it optimally. Instead, a heuristic algorithm is used to determine the fiber layout. This algorithm is an 11-step process, as detailed below:

1. **Determine the points of the graph**: this includes PoP with an existing Internet connection, as well as the settlements that are currently under-served and unconnected.

2. **Filter roads**: the road network for each country is extracted from the OpenStreetMaps database [286]. To reduce the number of edges in the graph, only main roads are considered. These include those with tags `motorway`, `highway`, `primary`, `secondary`, `tertiary`, `trunk`, `road`, and their respective links. Residential roads are not considered as they would significantly increase the complexity of the algorithm.

3. **Identify road intersections (cross-roads) and segments**: un this step the road network is pre-processed to identify intersections and perform road segmentation (i.e., create new roads such that the extreme points of all roads are either intersection points, settlements, or PoPs).

4. **Simplify the road network**: to further reduce the size of the final graph, several steps are performed to remove dead-ends (roads that lead to no settlements), merge contiguous roads without intersection, and reduce the number of roads in very dense regions (e.g., roads in metropolitan areas).

5. **Build the connectivity graph**: the connectivity graph is created with nodes as all the crossroads, PoPs, and settlements, and an undirected edge exists between two nodes if there is a road between them. The edges' weight equals the length of the road they represent. Finally, a dummy "Internet backbone" node with a zero-weight outflowing edge is connected to each of the PoPs.

6. **Divide the connectivity graph into sub-graphs**: to reduce computational complexity, the connectivity graph is divided into sub-graphs using a Voronoi tessellation, with the PoPs as centers of the Voronoi cells. Each of these sub-graphs is processed individually, and then the sub-graph solutions are merged into a global solution.

7. **Solve each of the sub-graph problems**: for each sub-graph, determine a potential layout by solving the Steiner tree problem. Since the problem is NP-hard, I use the well-known (and fast) $2-2/t$ *factor* approximation, obtained by computing the minimum spanning tree of the subgraph of the metric closure of the graph induced by the terminal nodes [287]. (The metric closure of G is the complete graph in which each edge is weighted by the shortest path distance between the nodes in G).

8. **Find the global layout**: solve the Steiner tree problem (using the same approximation method) over the sub-graph induced by the nodes contained in the solutions from the sub-graph problems (i.e., the solutions obtained in step 7). This allows for further reduction in the length of fiber required in the solution layout.

9. **Size the network and compute costs**: the sizing of the network is done by solving the max-flow, min-cost problem on the final graph, where the demand for each settlement is estimated based on its population. Once the throughput of each of the links is known, the number of wavelengths required, as well as the equipment necessary, is determined. Finally, the cost is estimated using the cost model described in Section 9.1.2.

10. **Determine deployment order**: evaluate the deployment strategy (i.e., how fiber segments would be grouped and in what sequence would they be deployed), and conduct an economic analysis to determine the profitability of each deployment.

Figure 9-1 illustrates how this methodology is applied to a simplified case study for a region in Argentina with 57 settlements (blue dots) and 3 PoPs (red dots). Figure 9-1a shows the road network and the location of the PoPs and unconnected settlements. In Figure 9-1b, road intersections are highlighted (gray square-boxes), and each road is colored using a different color. Note some continuous roads have intersection points in the middle without being any real intersection of two roads. Figure 9-1c shows the graph once it has been segmented (each road starts and ends in an intersection, a settlement, or a PoP) and simplified (dead-ends removed and roads merged), as described in steps 3 and 4. Originally, the graph contained 3,258 roads and 4,470 intersection points; after segmentation and simplification, it contains 1,803 roads and 1,151 intersections. Figure 9-1d shows the connectivity graph (in yellow) over the simplified road network, while Figure 9-1e shows the Voronoi tessellation of the graph around each PoP. In Figure 9-1f, the solutions for individual sub-graphs are depicted, whereas in Figure 9-1g the global layout is shown. Note how in the global layout only 2 PoPs are used, and the total fiber length has been reduced from 3,062 km (Fig. 9-1f) to 2,941 km (Fig. 9-1g). Figure 9-1h shows the final layout and network sizing over the road network; and finally, Figure 9-1i shows the deployment order by decreasing profitability. In total, 30 deployments are identified to connect 36 out of the 57 settlements using 2 PoPs. Note that deployments are incremental and that each new deployment can connect multiple settlements.

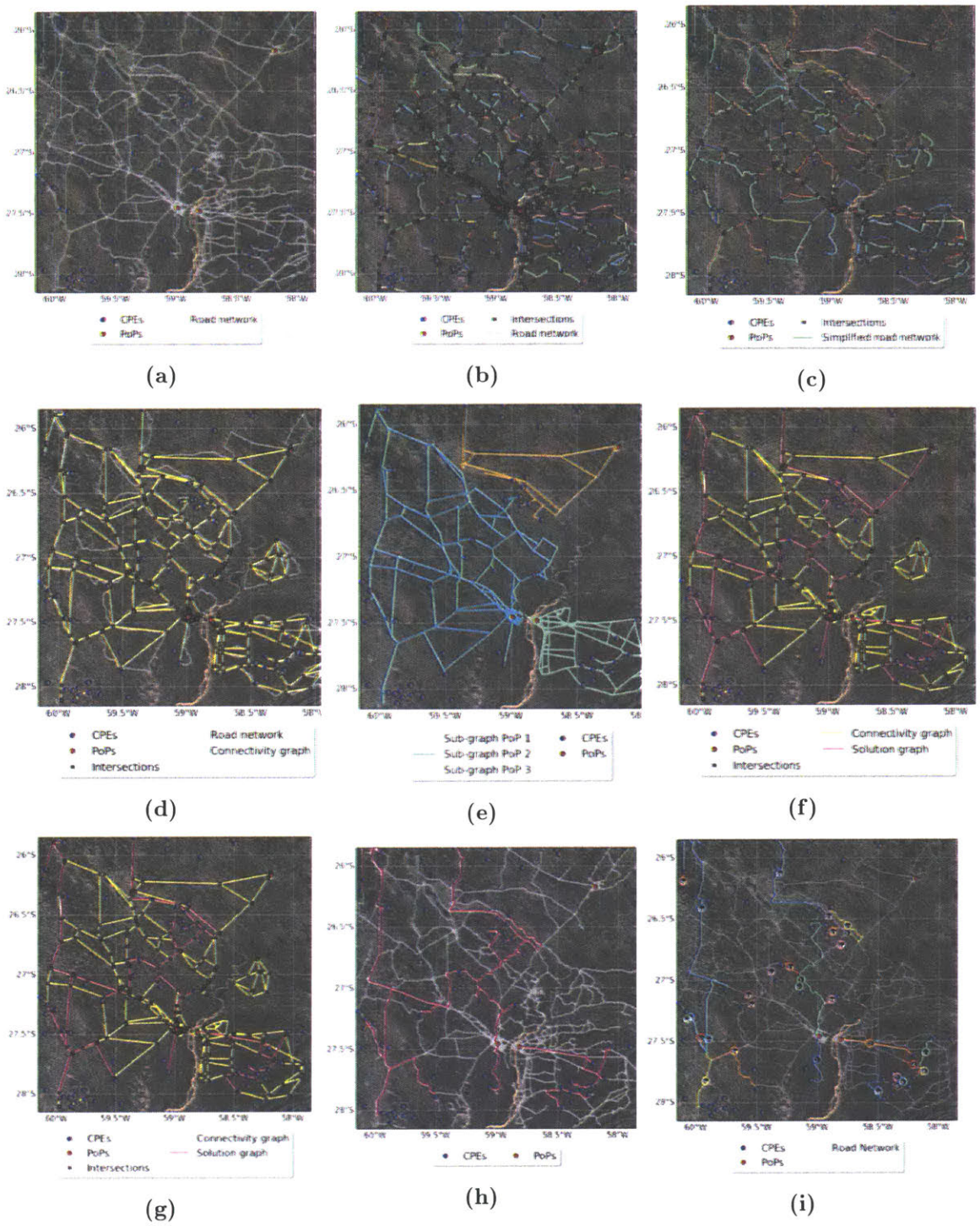


Figure 9-1: Step-by-step illustration of the methodology to determine fiber layout. a) Road network and settlements at the end of steps 1 and 2, b) segmented road network at the end of step 3, c) simplified road network after step 4, d) connectivity graph from step 5, e) Voronoi tessellation of the graph after step 6, f) Steiner tree solution for each of the subgraphs from step 7, g) proposed fiber layout after solving the Steiner Tree problem in step 8, h) network sized with costs after step 9, i) deployment order after performing techno-economic analysis in step 10.

9.1.2 Fiber cost and financial model

Cost model

The fiber cost model used is the same as the one presented in reference [288]. In particular, in greenfield areas where there is no pre-existing fiber plant, the major CapEx components are those due to fiber deployment (i.e., material, trenching, and right-of-way costs), optical amplification, dispersion compensation, and regeneration, as well as switching, routing, and grooming at nodes. The cost model provides relative costs (only for CapEx; OpEx components are ignored) for these different components of the network for 10 Gbps and 40 Gbps line rates, as shown in Table 9.1.

Table 9.1: Relative cost of fiber optic network elements for 10 Gbps and 40 Gbps line rates. Reproduced from [288].

	Relative Cost	
Network Element	10 Gbps	40 Gbps
Tunable medium-reach transceiver	0.3α	0.75α
Tunable long-reach WDM transceiver	0.4α	α
Tunable WDM transponder	40α	100α
Optical terminal chassis (per wavelength)	2.5α	2.5α
WAN amplifier and dispersion compensation (per wavelength)	2α	3.1α
WAN OXC port with amplification and dispersion compensation	8α	9.1α
WAN router port	120α	300α
Access fiber deployment (per km wavelength)	0.2α	0.2α
MAN fiber deployment (per km wavelength)	0.2α	0.2α
WAN fiber deployment (per km wavelength)	0.1α	0.1α
Access network EDFA pump power (per 100 mW)	2α	2α

Finally, since the model uses relative costs for each of the components, to estimate absolute costs, there is a need to establish the value of α in Table 9.1. Several previous studies have used values between \$10,000 and \$60,000 as the cost per kilometer of the fiber deployed[289, 290]. Using \$30,000 as an average value, and considering that a single fiber can contain up to 20 wavelengths, it is derived a value of \$12,500 for α.

Financial model

The financial model for the fiber deployments is similar to the ones presented in Chapters 7 and 8. Here, a time horizon of 20 years is assumed, along with a (very optimistic) adoption rate of 80%, an interest ratio of 8%, and a desired IRR of 15%. CapEx costs are extracted directly from the cost model, whereas the yearly OpEx costs (which include monitoring, maintenance, power, and backbone networks peering costs) are estimated to be 10% of the CapEx costs. The ARPU is the weighted average (by population) of the 2% of the monthly incomes in all the settlements covered by a deployment.

Table 9.2 shows an example of a deployment which connects 9 settlements in Nigeria using 81.3 km of fiber and 4 optical cross-connects (OXC)text. It can be seen that such a deployment would not be profitable before year 10, and after taking into account the 20-year time horizon, the IRR is -7%, so this deployment is deemed "non-viable" and would not be deployed (i.e., there is no profitable way of connecting these 9 settlements using fiber).

9.1.3 mm-Wave layout optimization

The process to optimize mm-Wave backhaul deployments is similar to the one for fiber optic, in the sense that once a connectivity graph is constructed, a Steiner tree "optimal" solution can be found using steps 5 - 8 of the algorithm described in Section 9.1.1. But to build the connectivity graph the physics of RF propagation under line-of-sight (LoS) conditions need to be considered (which is not the case for fiber optic networks). This procedure is detailed below:

1. **Determine points of interest**: this step is similar to the fiber optic procedure, and PoP and unconnected settlements are identified.
2. **Identify potential locations for placing towers**: potential locations for towers are identified following two criteria:
   (a) Maximize the number of settlements, PoP, and other tower locations within LoS. In other words, prioritize locations that allow for densely-connected networks. This criteria typically results in candidate towers being placed in local elevation peaks.
   (b) Minimize the distance to roads and existing infrastructure. Specifically, towers can be placed at most 2 km away from existing roads.
3. **Compute connectivity graph**: the LoS connectivity graph can be computed using terrain elevation and clutter data following the procedure below, once potential tower locations have been identified.

RF-LoS differs from optical LoS (more commonly referred to as the view-shed) in that for RF propagation, in addition to terrain and elevation, other effects such as the Fresnel zone interference and the RF-horizon need to be considered, as illustrated in Figure 9-2. In particular, two towers (T_x and R_x) will have RF-LoS if the following condition is met:

$$h(T_x) + m \cdot d(p, T_x) \geq e(p) + F_1(d(p, T_x), d(p, R_x)) + E_B(d(p, T_x), d(p, R_x)), \quad (9.1)$$

where $h(R_x), h(T_x)$ are the heights (above sea level) of the transmitter and receiver respectively, $e(p)$ is the elevation (above sea level) at point p (extracted from SRTM's 30-arcsecond digital elevation map [291]) , $d(A, B)$ represents the distance between

Table 9.2: Financial model for a fiber deployment connecting 10 settlements in Nigeria.

		CAGR	Year 0	Year 1	Year 2	Year 3	Year 4	Year 5	Year 6	Year 7	...	Year 20
Subscribers												
Total population	12,985	0%	12,985	12,985	12,985	12,985	12,985	12,985	12,985	12,985		12,985
Percentage subscribers			20%	37%	48%	67%	80%	80%	80%	80%		80%
Number subscribers			2,597	4,804	6,233	8,700	10,388	10,388	10,388	10,388		10,388
Monthly data usage/data sub (GB)	3.00	10%		3.0	3.3	3.6	4.0	4.4	4.8	5.3		18.3
Total Data Usage (GB)				172,960	246,819	378,970	497,751	547,527	602,279	662,507		2,287,154
CapEx												
Fiber deployment [km]	81.3		$2,033,500									
Equipment	4		$50,000									
Transceivers and amplifiers	10 + 1		$73,000									
Total CapEx			**$2,156,500**	**$0**	**$0**	**$0**	**$0**	**$0**	**$0**	**$0**		**$0**
OpEx												
Maintenance	10%	0%		$215,650	$215,650	$215,650	$215,650	$215,650	$215,650	$215,650		$215,650
Backbone transit	$20.00	-5%		$38,436	$52,106	$76,005	$94,835	$99,103	$103,563	$108,223		$191,792
Total OpEx				**$254,086**	**$267,756**	**$291,655**	**$310,485**	**$314,753**	**$319,213**	**$323,873**		**$407,442**
	ARPU											
Revenue	**$ 1.78**	**5%**		**$ 102,623**	**$ 139,789**	**$ 204,879**	**$ 256,863**	**$ 269,706**	**$ 283,191**	**$ 297,351**		**$ 560,699**
EBITDA				**$(151,463)**	**$(127,967)**	**$ (86,776)**	**$ (53,623)**	**$ (45,047)**	**$ (36,022)**	**$ (26,522)**		**$ 153,257**
Free Cash Flow			**($2,156,500)**	**($151,463)**	**($127,967)**	**($86,776)**	**($53,623)**	**($45,047)**	**($36,022)**	**($26,522)**		**$153,257**
Cumulative FCF			*($2,156,500)*	*($2,307,963)*	*($2,435,930)*	*($2,522,705)*	*($2,576,328)*	*($2,621,375)*	*($2,657,397)*	*($2,683,919)*		*$786,199*
NPV (i=15%)			**($2,128,273)**									
IRR			**-7%**									

points A and B, and p is any point in the geodesic that connects the transmitter (T_x) and the receiver (R_x). m is the slope of the line between the transmitter and receiver, $F_n(d_1, d_2)$ corresponds to the radius of the n-th Fresnel zone at a point located at distances d_1 and d_2 from points 1 and 2 respectively, and $e_B(d_1, d_2)$ is the Earth's bulge at that same point. Note that this inequality needs to be satisfied for every point between the transmitter and receiver.

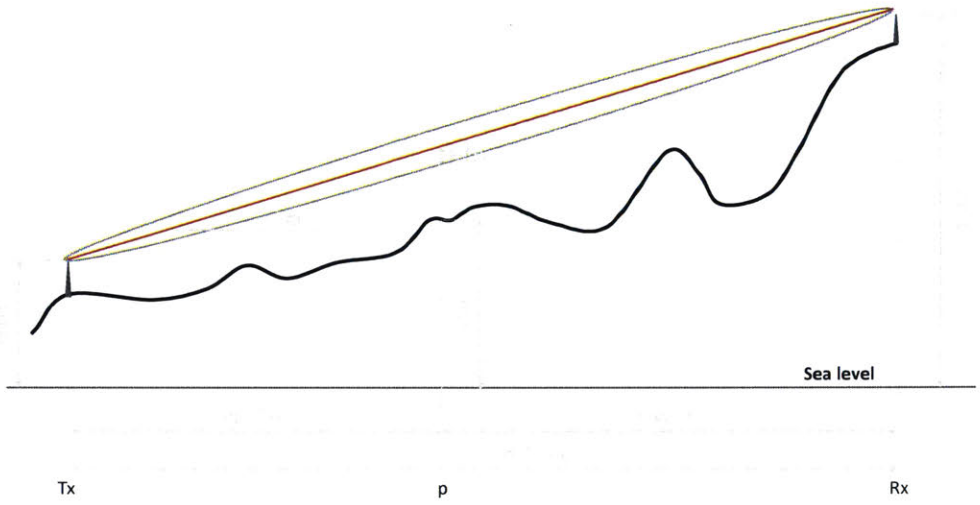


Figure 9-2: Elements involved when assessing LoS conditions between a transmitter (Tx) and a receiver (Rx) antenna. The contributions of a) elevation terrain, b) the Earth's bulge, and c) Fresnel interference are all considered.

The terms for the slope, Fresnel zone diameter, and the Earth's bulge are given by:

$$m = \frac{h(R_x) - h(T_x)}{d(T_x, R_x)} \tag{9.2}$$

$$F_n(d_1, d_2) = \sqrt{\frac{n\lambda \cdot d_1 \cdot d_2}{d_1 + d_2}} \tag{9.3}$$

$$E_B(d_1, d_2) = \frac{3d_1 \cdot d_2}{8R_E} \tag{9.4}$$

9.1.4 mm-Wave cost and financial model

Cost model

The cost of a tower is mainly driven by height, and comprises site acquisition, foundations, transportation, construction labor, tower materials, electrical and lighting

costs. The CapEx model for tower deployment is given in Table 9.3. The numbers are aggregated from average costs given by different tower manufacturers [1].

Table 9.3: CapEx costs for guyed-towers construction.

Tower height	15m	20 m	30 m	40 m	60 m	80 m
Site acquisition			$2,500			
Transportation	$750	$1,000	$1,000	$1,500	$1,500	$2,000
Tower material	$3,500	$7,0000	$10,000	$13,000	$18,000	$23,000
Building costs	$5,000	$10,000	$15,000	$20,000	$27,000	$35,000
Lightning			$3,000			
Total Tower cost	$14,750	$23,500	$ 31,500	$40,000	$52,000	$65,500

In addition to the tower costs, the power, RF equipment, and PoP equipment ($2,500) costs need to be taken into account. In places where grid power is available, the power equipment costs are assumed to be $1,000; whereas in places where grid power is not available, solar cells are used, with an aggregated cost of $7,500 (which includes panels, batteries, cabinet and rest of equipment). Finally, the RF equipment runs up to $2,000 for each RF link, which includes two 1 m antennas and two radios, while the radio access network (RAN) equipment costs are $10,000.

In terms of yearly OPEX, the spectrum license ($50 per link), backhaul cost ($20 Mbps/month), tower maintenance (10% of its cost), and power consumption ($65/month for the grid-connected towers) are considered.

Financial model

As in the case of fiber optic networks, once the tower layout is obtained, how these towers are to be deployed needs to be studied. The first deployment connects a PoP with one or more towers (which can be daisy-chained) to cover a set of settlements at the minimum cost per person. Posterior deployments might originate in towers previously deployed or from PoPs; however, their costs per person will be higher than those of previous deployments. Thus, the first step of the financial model is to produce a list of deployments sorted by increasing order of cost per person. Such a list can be efficiently obtained by exploring the layout graph using a breadth-first search algorithm, together with memoization.

Following this, a free-cash-flow is computed for each of the deployments, and only the deployments that guarantee an IRR higher than 15% are actually considered. The financial model assumes that up to 60% of the population could become subscribers of the mm-Wave backhauled service, and that the monthly data-volume consumption per user is 3 GB (with an annual growth of 10%). The CapEx and OpEx variables are extracted directly from the deployment information (i.e., depending on the number

[1] Extracted from https://www.rohnnet.com/ and https://www.3starinc.com/

of towers and links). and the ARPU is estimated to be 2% of the average monthly income of the settlements connected in each deployment.

Table 9.4 shows the free-cash flow analysis for the 3-tower deployment to connect the 9 villages in Nigeria (same as in Section 9.1.2). In this case, the deployment is highly profitable, returning a 40% IRR.

9.2 Results

An analysis to determine the potential impact of fiber optic and mm-Wave backhaul networks was conducted for 37 countries (shown in Table 6.1) for which settlement data (population, estimated income) and connectivity coverage information is available. For each country, it was assumed that Internet connection through a fiber PoP is available in any settlement with more than 50,000 inhabitants that has broadband coverage (such that at least 60% of the settlement area is covered by 4G networks). The objective is then to extend connectivity to uncovered and under-served regions, defined as those that do not have 3G or 4G connectivity.

Network layouts are obtained by running the algorithm described in Section 9.1.1 for each of the countries, whereas the viability of the deployments is analyzed using the cost and financial models presented in 9.1.2.

9.2.1 Fiber optic networks

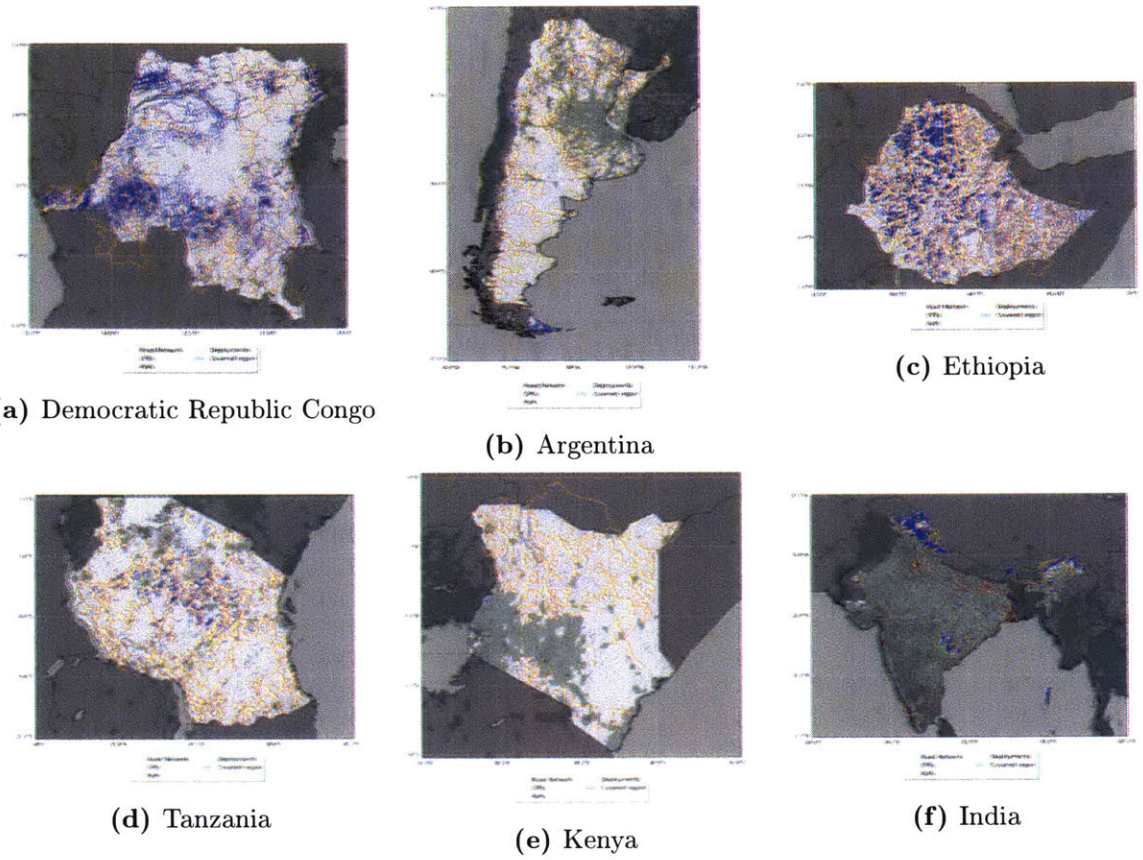
Figure 9-3 shows the fiber deployment layouts for 6 representative countries. In all cases, fiber originates in very densely-populated locations and expands throughout the uncovered rural regions. Within the figure, several differences can be observed between countries. On one hand, in India, most of the country is close enough to some very densely populated cities, such that uncovered regions are scarce. This results in deployments being concentrated in very specific regions. On the other hand, in underdeveloped countries such as Ethiopia or the Democratic Republic of Congo, uncovered populations are much more sparsely distributed, and thus, deployments are also spread out across the whole country.

Table 9.5 contains statistics on the number of settlements that can be connected using fiber, the length of fiber that would need to be deployed, the number of deployments required, the number of deployments that would be profitable (under current costs), as well as the total cost of the deployments. First, note that for most countries, the number of settlements that end up being connected by fiber ranges from 50% to 75% of the total number of settlements; the exceptions include the Philippines, Haiti, Indonesia, and Bangladesh (which have significantly lower numbers). Because the algorithm requires the presence of a road or railroad between a PoP and a settlement, there are two main reasons why a settlement might not be connected: the lack of

Table 9.4: Financial model for a 3 tower deployment in Nigeria.

		Growth	Year 0	Year 1	Year 2	Year 3	Year 4	Year 5	Year 6	Year 7	Year 8	Year 9	Year 10
Subscribers													
Total population	12,985	0%	12,985	12,985	12,985	12,985	12,985	12,985	12,985	12,985	12,985	12,985	12,985
Percentage subscribers		0%	20%	34%	59%	73%	90%	90%	90%	90%	90%	90%	90%
Number subscribers			2,597	4,415	7,661	9,479	11,687	11,687	11,687	11,687	11,687	11,687	11,687
Data/subscriber (GB)	3.00	10%		3.0	3.3	3.6	4.0	4.4	4.8	5.3	5.8	6.4	7.1
Total Data Usage (GB)				158,936	303,382	412,907	559,970	615,967	677,564	745,321	819,853	901,838	992,022
CapEx													
Towers			$95,000										
RAN			$23,000										
Power			$7,500										
Total CapEx			**$125,500**	**$0**	**$0**	**$0**	**$0**	**$0**	**$0**	**$0**	**$0**	**$0**	**$0**
OpEx													
Maintenance	10%	0%		$12,550	$12,550	$12,550	$12,550	$12,550	$12,550	$12,550	$12,550	$12,550	$12,550
Backbone transit	$20.00	0%		$35,319	$67,418	$91,757	$124,438	$136,882	$150,570	$165,627	$182,189	$200,408	$220,449
Spectrum license	$50.00	0%		$1,800	$1,800	$1,800	$1,800	$1,800	$1,800	$1,800	$1,800	$1,800	$1,800
Total OpEx				**$49,669**	**$81,768**	**$106,107**	**$138,788**	**$151,232**	**$164,920**	**$179,977**	**$196,539**	**$214,758**	**$234,799**
	ARPU												
Revenue	**$ 1.56**	0%		**$ 82,647**	**$ 143,417**	**$ 177,448**	**$ 218,771**	**$ 218,771**	**$ 218,771**	**$ 218,771**	**$ 218,771**	**$ 218,771**	**$ 218,771**
EBITDA				**$ 32,978**	**$ 61,649**	**$ 71,341**	**$ 79,983**	**$ 67,540**	**$ 53,851**	**$ 38,794**	**$ 22,232**	**$ 4,013**	**$ (16,028)**
Free Cash Flow			**($125,500)**	**$32,978**	**$61,649**	**$71,341**	**$79,983**	**$67,540**	**$53,851**	**$38,794**	**$22,232**	**$4,013**	**($16,028)**
Cumulative FCF			*($125,500)*	*($92,522)*	*($30,874)*	*$40,467*	*$120,450*	*$187,990*	*$241,841*	*$280,636*	*$302,868*	*$306,881*	*$290,853*
NPV (i=15%)			**$102,888**										
IRR			**40%**										

road infrastructure, or the lack of quality road data.



(a) Democratic Republic Congo

(b) Argentina

(c) Ethiopia

(d) Tanzania

(e) Kenya

(f) India

Figure 9-3: Layouts of fiber deployment in 6 representative countries. Blue regions represent uncovered and under-served settlements, orange lines are the fiber layout, and green areas are regions which have broadband coverage (3G/4G).

Finally, it can be observed that at current costs, the number of profitable deployments is very small, almost zero in most countries. The only three countries studied with a significant number of profitable deployments are Nigeria, Ethiopia, and the Democratic Republic of Congo, where fiber deployment could bring online an additional 6 million people online.

Figure 9-4 plots the population size benefited by the profitable deployments as a function of the cost per kilometer of the deployment (including fiber, labor, and right of way). It can be observed that at current costs (approximately $30,000 per kilometer), just slightly over 10 million people (out of 159 million, which is the total uncovered and under-served population of the 37 countries considered) could be connected. If a 3x reduction in cost were to be achieved, this number of people would double; a 10x reduction would result in 30 million people being connected. The Figure also shows that the three countries that would benefit most from new fiber deployments are Nigeria, Ethiopia, and the Democratic Republic of Congo.

Table 9.5: Summary of fiber deployments in 28 countries.

ISO2	# sett.		# PoPs		km fiber	cost	# people[million]		# sett
	total	connected	total	connected		[million $]	total	prof.	prof.
AF	76	45	39	7	2114.3	57.6	1.9	0.4	3
AR	3589	1656	158	72	92691.3	2725.1	1.5	0	0
BD	242	10	519	2	112	4.7	0	0	0
BF	2975	2533	68	36	27393.1	1186.8	4.9	0	0
BR	24310	7750	1130	214	392608.7	11318.5	2.8	0	5
CD	38030	25324	305	140	114668.2	7730.1	37.7	0	0
CI	646	362	81	30	9067.5	568.5	0.7	0	0
CM	1709	1337	91	42	21782.8	856.4	2	0.2	2
DZ	381	216	233	39	11565	340.9	0.2	0	0
EG	308	166	394	9	3313.5	120.7	0.1	0	0
ET	29288	23185	356	171	116181.8	7632.6	23.5	1.4	532
HT	484	68	33	7	1368.8	49.5	0.2	0	0
ID	3007	117	743	22	18429.1	487.3	5	0	0
IN	13040	6612	3134	106	47265.4	6060.8	3.2	0	80
KE	1822	1391	240	41	19563.8	776.9	3	0.1	12
MM	3983	1567	163	39	34801.1	1201.2	2.1	0	0
MW	2173	1515	47	24	7136.2	499.9	0.7	0	0
MZ	10284	7343	101	53	57355.5	3014.9	8.1	0	0
NG	14280	9297	753	175	70745.1	4037.9	19.9	2	1188
PE	1515	839	120	49	61149	1733.7	1.4	0	0
PH	662	62	386	22	8364.8	223.6	0.7	0	0
PK	31633	21573	322	54	60273.9	5263.3	3.6	0	0
SD	5268	3276	105	55	32573.7	1489.9	7	0	0
TH	112	23	216	2	551.3	18.6	0	0	0
TR	270	30	307	5	777.2	25.8	0.2	0	1
TZ	7999	5728	222	99	45974.4	2362.8	6.4	0	1
UG	478	321	180	26	4127.3	176.9	0.7	0.2	2
ZA	372	301	341	35	14132	429.4	0.3	0	2

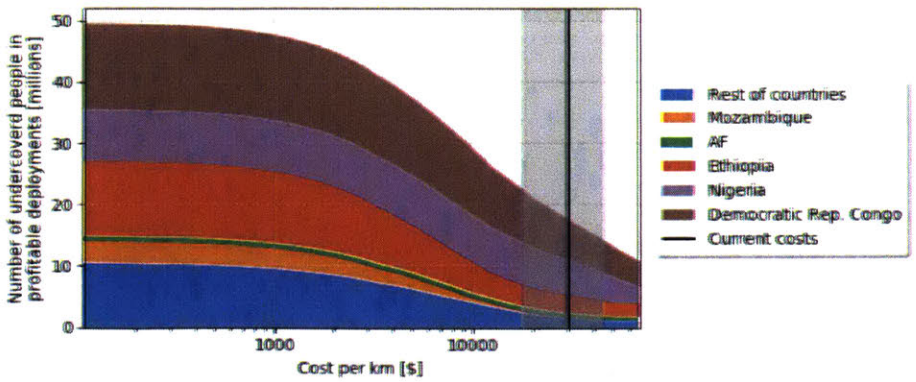


Figure 9-4: Number of people in profitable deployments as a function of the average cost per kilometer of fiber. This cost includes fiber lay-down, labor, and right of way.
NG: Nigeria, **CD**: Democratic Republic of Congo, **ET**: Ethiopia, **AF**: Afghanistan.

Figure 9-5 shows the percentage of uncovered and under-served people that could be served via profitable fiber deployments, versus the average cost per kilometer of deployment. In practice, this is a sensitivity analysis where the impact of fiber technology is assessed against cost. Since the cost model used in this dissertation was developed based on fiber deployment costs in developed countries, where labor and right of way costs might be higher than in developing countries, this image can be used to assess the impact of fiber if the reader believes that the costs herein presented are an overestimation of the real costs. In Figure 9-5, it can be observed that at current costs, approximately 7% of the overall uncovered and under-served population in the countries analyzed could be served via fiber (mostly in Ethiopia, Nigeria, and the Democratic Republic of Congo), whereas it would take cost reductions by more than an order of magnitude to achieve 25% of the population, after which the curve then flattens out since the equipment costs alone are unaffordable.

Since such drastic cost reductions seem unrealistic (at least in the short term), I conclude that fiber will play at most a minor role in serving as backhaul for currently unconnected regions. This does not mean that no new fiber deployments will be carried out in these countries (for example, to increase the capacity of current networks or to replace existing mm-Wave backhaul links), but it is unlikely that new fiber projects will be undertaken to expand connectivity further than the status quo (or for that to happen, governments will have to subsidize such deployments or establish policies so that Internet service providers have incentives to deploy new infrastructure.).

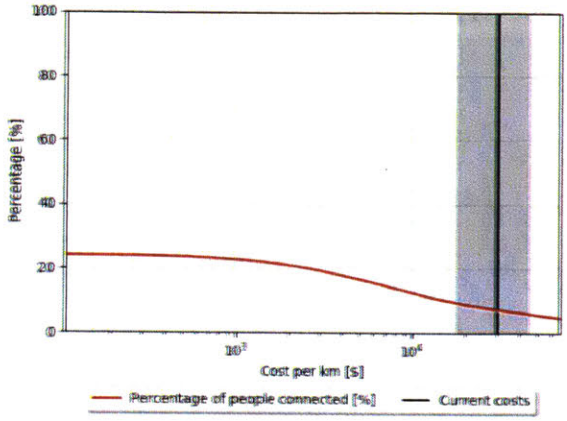


Figure 9-5: Percentage of people in profitable deployments, as a function of the average cost per kilometer of fiber. This cost includes fiber lay-down, labor, and right of way.

9.2.2 mm-Wave networks

The results presented in this section are analogous to those for the fiber optic case, presented in the previous section. A similar analysis was conducted for the same 37 countries, using the identical set of settlements and fiber PoP. The main difference is that here the mm-Wave backhaul models, as described in Sections 9.1.3 and 9.1.4, were used.

Table 9.6 summarizes the deployment information for the countries analyzed. Looking at the cost column, it can be observed that mm-Wave is cheaper than fiber for all countries. This is not surprising, as mm-Wave is a less CapEx-intensive technology compared to fiber. Also comparing the number of settlements/ people that would be connected against the number of in profitable deployments and the number of settlements, it can be seen that even though most of the settlements could be connected using mm-Wave towers, the technology would not be profitable in most of the countries (with exceptions like Nigeria), thus there is little incentive for companies to deploy additional assets to the uncovered and under-served locations.

Figure 9-6 shows the number of people covered as a function of the average cost per tower, broken down by country. It can be observed that at current costs, approximately 16 million additional people would be covered by profitable deployments, half of them living in Nigeria. In addition, countries like India, Indonesia, Kenya, and Turkey, could all deploy mm-Wave networks to offer connectivity to parts of currently uncovered regions. Moreover, if costs were to be reduced by a factor of 3, the number of connected people would more than double, and deployments would also be profitable in vast regions of countries such as Ethiopia and the Democratic Republic of Congo.

Another interesting factor to analyze is the mix of tower heights within the de-

Table 9.6: Summary of mm-Wave deployments in 35 countries.

ISO2	# sett.		# PoPs		# towers	cost	# people[million]		# sett
	total	connected	total	connected		[million $]	total	prof.	prof.
AR	3589	293	158	20	409	21.9	1.5	0	0
BD	242	149	519	8	78	3.6	0	0	0
BF	2975	2631	68	48	1279	61.5	4.9	1.5	55
BR	24310	14418	1130	260	9500	474.4	2.8	0.7	9
CD	38030	30726	305	248	10427	491.2	37.7	2	182
CI	646	385	81	33	422	24.3	0.7	0	0
CM	1709	1365	91	49	1091	51.6	2	0.3	22
DZ	381	97	233	25	86	3.7	0.2	0	0
EG	308	20	394	1	21	0.9	0.1	0	0
ET	29288	22730	356	254	6943	323.6	23.5	6.1	955
GH	540	393	91	28	319	18.8	0.5	0.1	6
HT	484	441	33	19	167	9.5	0.2	0	0
ID	3007	285	743	52	390	22.9	5	1.5	41
IN	13040	7278	3134	147	2311	101.6	3.2	1	75
KE	1822	811	240	55	534	28.6	3	0.7	76
LK	35	35	85	3	17	1.1	0	0	0
LS	271	246	5	3	126	6.6	0.1	0	0
MM	3983	1247	163	40	822	39	2.1	0.6	3
MW	2173	1975	47	31	585	25.9	0.7	0	0
MX	3572	1975	507	31	585	13.1	1.2	0	0
MZ	10284	8456	101	77	2818	134.5	8.1	0	0
NG	14280	12585	753	244	4448	193.5	19.9	9.7	2741
PE	1515	744	120	42	886	50.6	1.4	0.1	6
PG	1953	127	38	10	68	3.2	2	0.1	6
PH	662	352	386	78	358	16.5	0.7	0.3	7
PK	31633	20663	322	51	3777	170.5	3.6	0.5	388
RW	5	2	31	2	9	0.5	0	0	0
SD	5268	3027	105	46	1190	58.5	7	2.2	260
TH	112	43	216	10	48	2.3	0	0	0
TR	270	88	307	8	83	3.7	0.2	0.1	14
TZ	7999	467	222	45	131	3.3	6.4	0.6	369
UA	1404	314	151	11	138	6.6	0.5	0	11
UG	478	443	180	49	255	11	0.7	0.3	18
ZA	372	204	341	29	456	19.7	0.3	0	0
ZM	9284	6580	61	45	2849	140.6	3.8	0.1	2

ployments. In our analysis, towers of up to 80 m height were considered, but in current deployments, towers are normally of lower height (< 40 m). Figure 9-7 shows a histogram of tower height distribution across deployments (for all 37 countries in our analysis). It can be observed that the majority of the towers deployed are 80 m towers, which suggests that despite their higher costs, taller structures might be more cost-efficiency in terms of expanding connectivity.

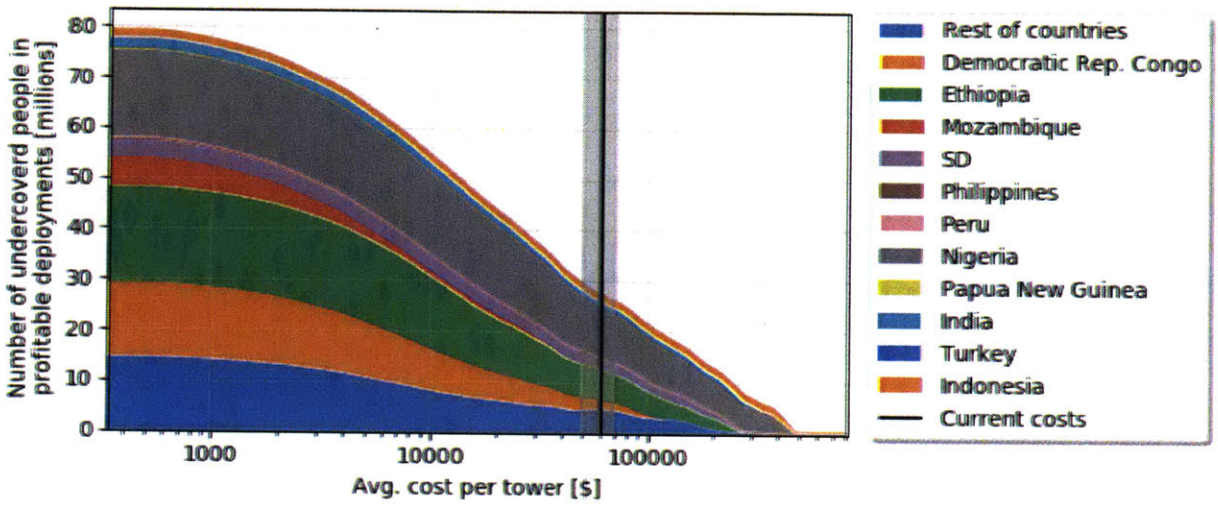


Figure 9-6: Number of people in profitable deployments as a function of the average cost per tower. This cost includes tower construction, labor, and spectrum rights.

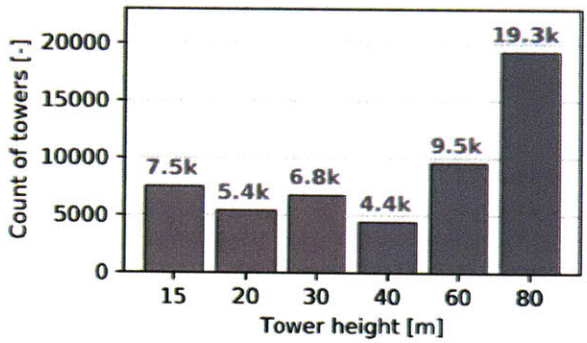


Figure 9-7: Histogram of tower heights in deployments.

Figure 9-8 shows the percentage of the uncovered and under-served population reached by profitable deployments (orange line), as well as the maximum population served (red line), as functions of the average cost per tower. The gap between people reached and people served is due to the maximum capacity of the mm-Wave link: even though a given settlement might be reached, the requested capacity might not be 100% met due to the limits of the mm-Wave link. Note that the gap decreases as the average cost per tower goes down, since it might then be profitable to deploy multiple towers to provide multiple paths to a given settlement, and thus increasing the total capacity delivered. In terms of population served, connectivity could be brought to approximately 10% of the population using mm-Wave backhaul links (at current costs). This value would increase to 17% if costs were halved, and further to 28% if the costs were reduced by a factor of 5. Note that even at the extreme low end of the cost axis, there will still be a significant portion of the population that

would not be connected. This is due to two reasons: first, the cost of equipment might be too expensive for certain settlements; and second, the algorithm limits the number of hops required to reach settlements from an existing PoP. If the number of hops is higher than 8, it is considered that the settlement cannot be connected using mm-Wave backhaul.

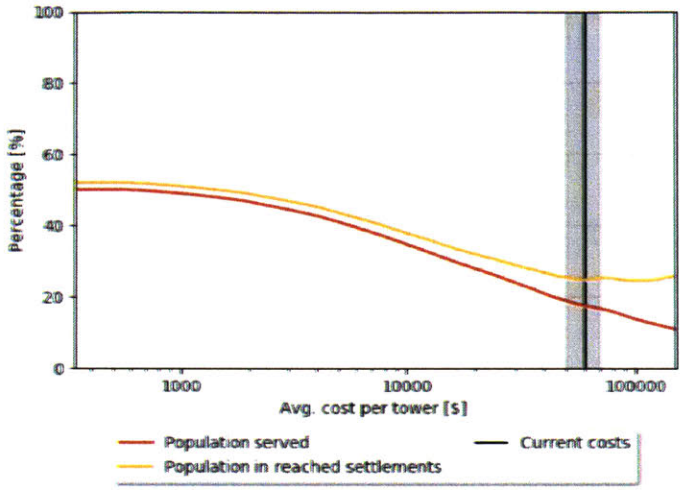


Figure 9-8: Percentage of people in profitable deployments as a function of the average cost per tower. This cost includes tower construction, labor, and spectrum rights.

9.3 Chapter conclusions

This chapter studied backhaul terrestrial networks as a means of expanding connectivity for a set of 37 countries. The main research question was:

Research Question 9.1

Taking into account both technical and economic factors, what is the potential impact of terrestrial networks in terms of connecting additional populations?

From the results, the following conclusions can be extracted.

- At current costs, neither fiber optic nor mm-Wave backhauling networks are capable of bringing a significant number of people online. In particular, fiber optic networks could bring up to approximately 7% of the uncovered and underserved online; for mm-Wave backhaul networks, this number is closer to 17%.
- Fiber networks will have a very limited impact in connecting new regions even if large cost reductions were to be achieved. In particular, even with a 30× cost

reduction, only 25% of the uncovered and under-served population in the 37 countries studied would be able to afford the service.

- mm-Wave networks have greater potential than fiber optic networks in terms of connecting new regions; with just a 2×, they could provide service profitably to 30% of the population in these countries. This is not a surprising result since up to 80% of the current rural deployments in the countries under consideration already use this technology to provide backhaul [292].

Chapter 10

Space, Aerial, and Terrestrial networks comparison

Chapters 7, 8, and 9 analyzed, as isolated systems, the impact and capabilities of space, aerial, and terrestrial networks, respectively. In this chapter, the impact of each of these technologies is assessed as placed within a competitive market. That is, assuming that broadband is a commodity (where, given the fungibility of a product, the cheapest option is always preferred), the potential impact of each technology in terms of extending connectivity to uncovered and under-served populations is assessed.

In particular, this chapter aims to answer the following research question:

Research Question 10.1

Given capacity limits, market uptake, and achievable prices per Mbps/month, what is the impact of space and aerial technologies in terms of expanding connectivity to uncovered and under-served regions?

10.1 Overview of comparison methodology

The comparison of technologies is carried out for the 37 countries in Africa, Southeast Asia, and South America that were shown in Table 10.1. Only uncovered and under-served settlements (i.e., settlements with no connectivity or only 2G connectivity) are considered as the target population. The methodology to estimate the number of people who would be connected is a two-step procedure, as follows:

1. For each settlement, the different technologies (fiber, aerial, mm-Wave, and space) are ranked by monthly price per user. To compute the price per user,

Table 10.1: Countries considered in the comparative analysis of technologies.

ISO2	Country name	ISO2	Country name	ISO2	Country name
AF	Afghanistan	ID	Indonesia	PG	Papua New Guinea
AR	Argentina	IN	India	PH	Philippines
BD	Bangladesh	KE	Kenya	PK	Pakistan
BF	Burkina Faso	LK	Sri Lanka	RW	Rwanda
BR	Brazil	LS	Lesotho	SD	Sudan
CD	Congo (DRC)	MM	Myanmar	TH	Thailand
CI	Cote d'Ivore	MW	Malawi	TR	Turkey
CM	Cameroon	MX	Mexico	TZ	Tanzania
DZ	Algeria	MY	Malaysia	UA	Ukraine
EG	Egypt	MZ	Mozambique	UG	Uganda
ET	Ethiopia	NG	Nigeria	ZA	South Africa
GH	Ghana	PE	Peru	ZM	Zambia
HT	Haiti				

it is assumed that for uncovered settlements each user is to be provided with a data tonnage of 1 GB/month, whereas for under-served settlements 3GB/month are required. Moreover, a conservative 9% CAGR in demand is assumed [293], which results in a data rate allocation of 30 kbps and 100 kbps for uncovered and under-served users, respectively. Then, the price per Mbps/month of each technology is divided by the number of subscribers 1 Mbps would support (using the data rate allocations). Furthermore, the cost of the CPE (if required) is also included in the price per user (distributed evenly across subscribers).

Finally, the price per user for each technology is compared to the 2% of the average monthly income of each settlement (as per the ITU recommended threshold), to determine which technologies are affordable in each settlement.

2. Settlements in each country are ordered in descending order of income, and the most affordable technology is assigned to the settlements sequentially (so that wealthy settlements are assigned first). In doing so, the market uptake and the capacity limits of some of the technologies are also taken into account (as will be further explained in Section 10.2.2).

Besides, three scenarios with different time horizons will be considered in the analysis:

- **Current scenario**: Current costs are used as a reference for the analysis. It is assumed that if fiber or mm-Wave technology were to be deployed, the total capacity offered would be unlimited or 1.6 Gbps, respectively, whereas for satellite networks, the maximum capacity-density is 50 kbps/km^2, which limits the maximum throughput that can be delivered to any country. Only fiber, mm-Wave, and space technologies are considered, as there are currently no viable aerial communication networks.

- **Next-generation scenario (2-5 years)**: For this scenario, the cost (and thus price) reductions that might be achieved in the next 2-5 years are considered. In the space industry it is estimated that monthly cost per Mbps could drop to the $70- $100 range (see Chapter 9). For aerial systems, it will be assumed that the best performing LTA HAP network is used, with prices set at below $50 Mbps/month for profitable settlements, as obtained from the analysis in Chapter 8.

- **Future scenario (8-10 years)**: For this scenario, advanced concepts for space and aerial networks are considered. According to the results from Figure 7-11 in Chapter 7, MEO networks are the most cost-effective space architectures for providing coverage to most of the uncovered and under-served populations. In particular, a vHTS MEO network of 15 satellites in equatorial orbit operating in Q/V-band is considered, which would allow for monthly prices to be set at ~$30 per Mbps/month. For aerial networks, a passive station-keeping, Ka-band system with no crosslinks, a single beam, and direct-to-user links, which was identified in Chapter 8 as the most cost-effective architecture that covers the largest population, is considered. This architecture allows for prices in profitable deployments to be set at below $50 per Mbps/month.

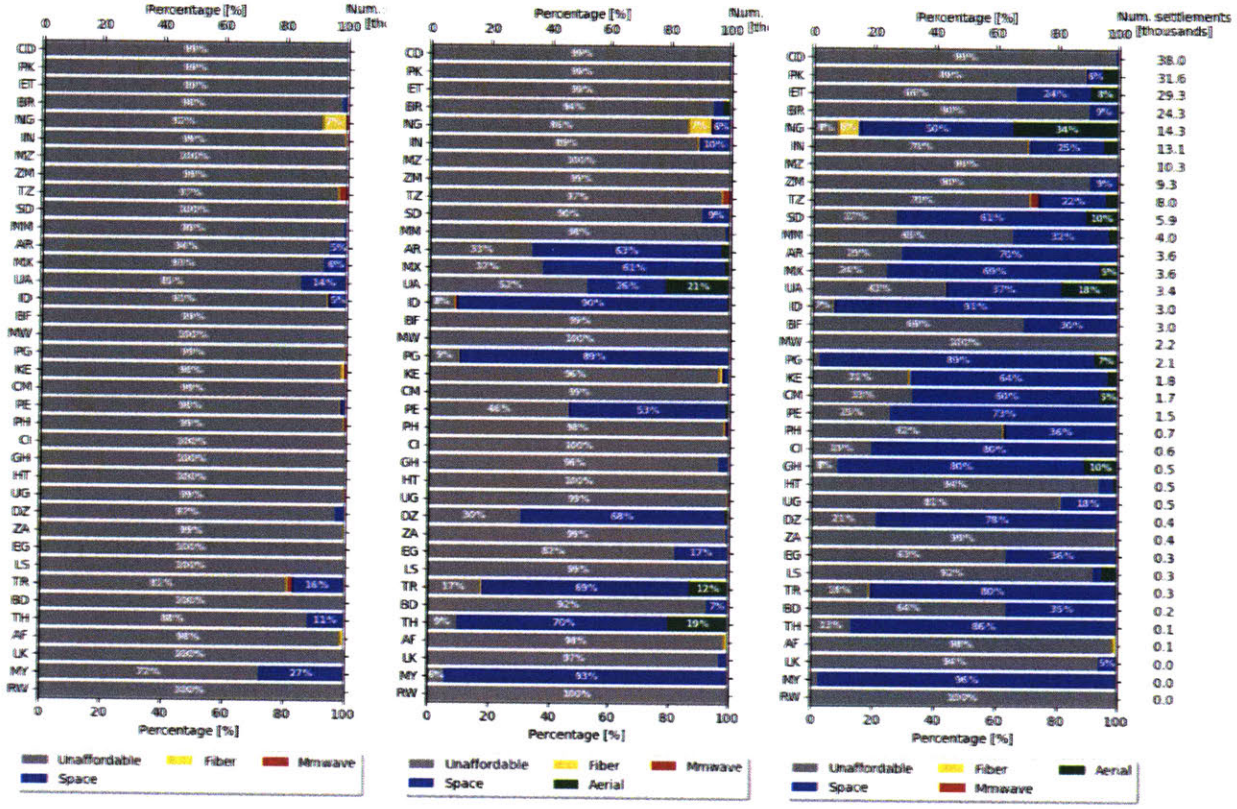
10.2 Results

10.2.1 Percentage of population that can afford connectivity services by technology

This section presents the results of the first step of the comparative methodology, namely, to determine the most affordable technology for each settlement. It is important to emphasize that the results in this section only evaluate **affordability**, and do not take into account the number of people that would be brought online by each of these technologies (which will be analyzed separately in Section 10.2.2). In other words, the results presented in this section only consider the relationship between price and demand (i.e., given the estimated price per Mbps/month for each technology, what is the number of people that can afford it), but do not account for factors like market uptake and the limited supply of capacity in some cases (e.g., space systems).

Figure 10-1 shows the distribution of the most affordable technology as a percentage of all settlements in each country. Results are shown for the current, next-generation, and future scenarios. It can be observed that at current prices, space technology is unaffordable in most of the countries considered, while mm-Wave and fiber networks both have a limited impact in terms of the percentage of settlements afforded. In the next-generation scenario, however, there is a big jump in the percentage of settlements where satellite networks have become the most affordable option, even

though this only applies to countries with higher incomes. In particular, countries in South America (Argentina, Mexico, Peru) and Southeast Asia (Myanmar, Thailand, Indonesia) would benefit the most from cheaper satellite connectivity, whereas many sub-Saharan countries would remain unconnected. Finally, in the future scenario, there is a dramatic increase in the number of settlements in which space technology has become affordable. However, even at prices ~8 times lower than the current scenario, there would be countries where bringing people online will remain a huge challenge, due to their extremely low incomes per capita (below $500), (e.g., the Democratic Republic of Congo, Malawi, and Mozambique).

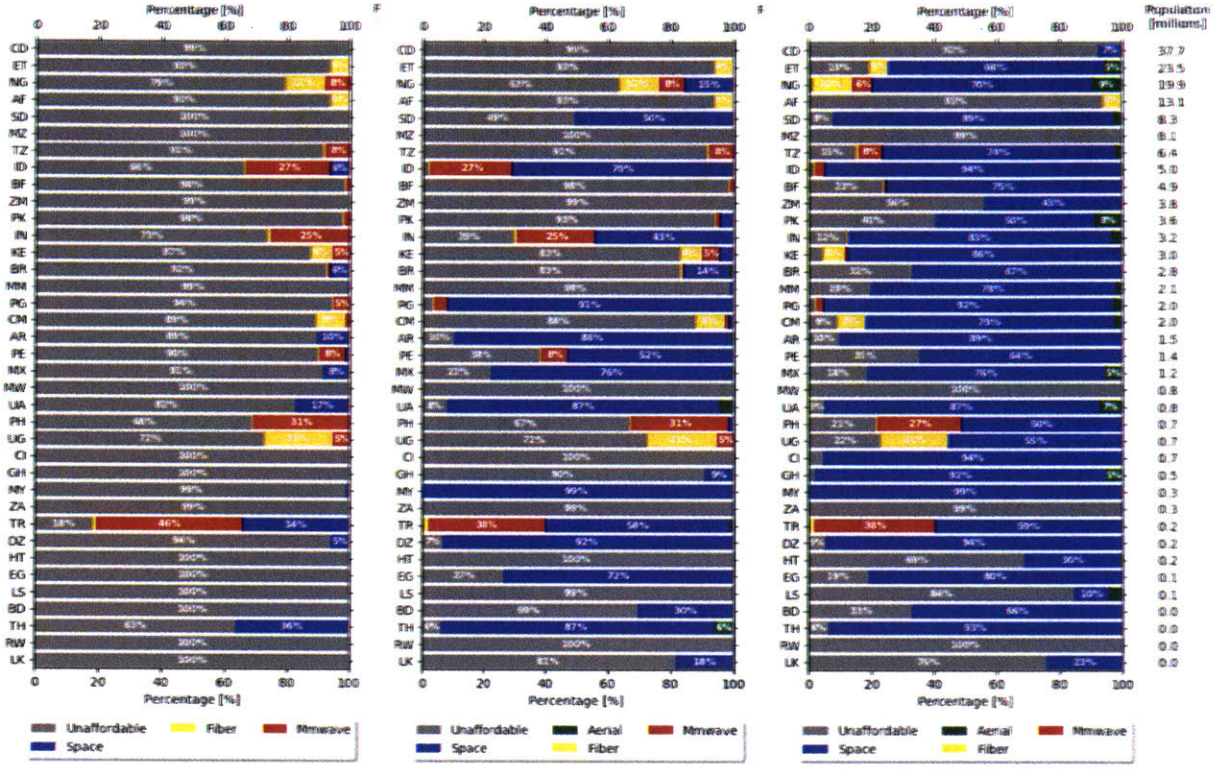


(a) Current scenario **(b)** Next-generation (2-5 years) **(c)** Future scenario (8-10 years)

Figure 10-1: Percentage of uncovered and under-served settlements for all countries considered, broken down by most affordable technology, under the a) current, b) next-generation, and c) future scenarios. Countries are ordered by number of settlements in descending order.

Although the number of settlements that can afford connectivity serviced per technology provides an initial sizing up of the problem at hand, a more insightful overview is obtained when considering the population sizes of those settlements. Figure 10-2 is the counterpart to Figure 10-1, where percentages of the total population have replaced percentages of total settlements. These values are computed after grouping settlements by the most affordable technology, adding up their populations, and then dividing that by the total uncovered and under-served population in each country.

For the current scenario the graph shows how even though the number of settlements in which mm-Wave and fiber are affordable is rather low (below 5% for most of the countries), they amount to a much more significant percentage of the population (more than 20% of the population in several countries). This is because being and mm-Wave, being very CapEx intensive technologies, are only viable in very highly populated settlements. Also, it can be observed that mm-Wave is generally preferred over fiber (as one would expect, given the lower costs).



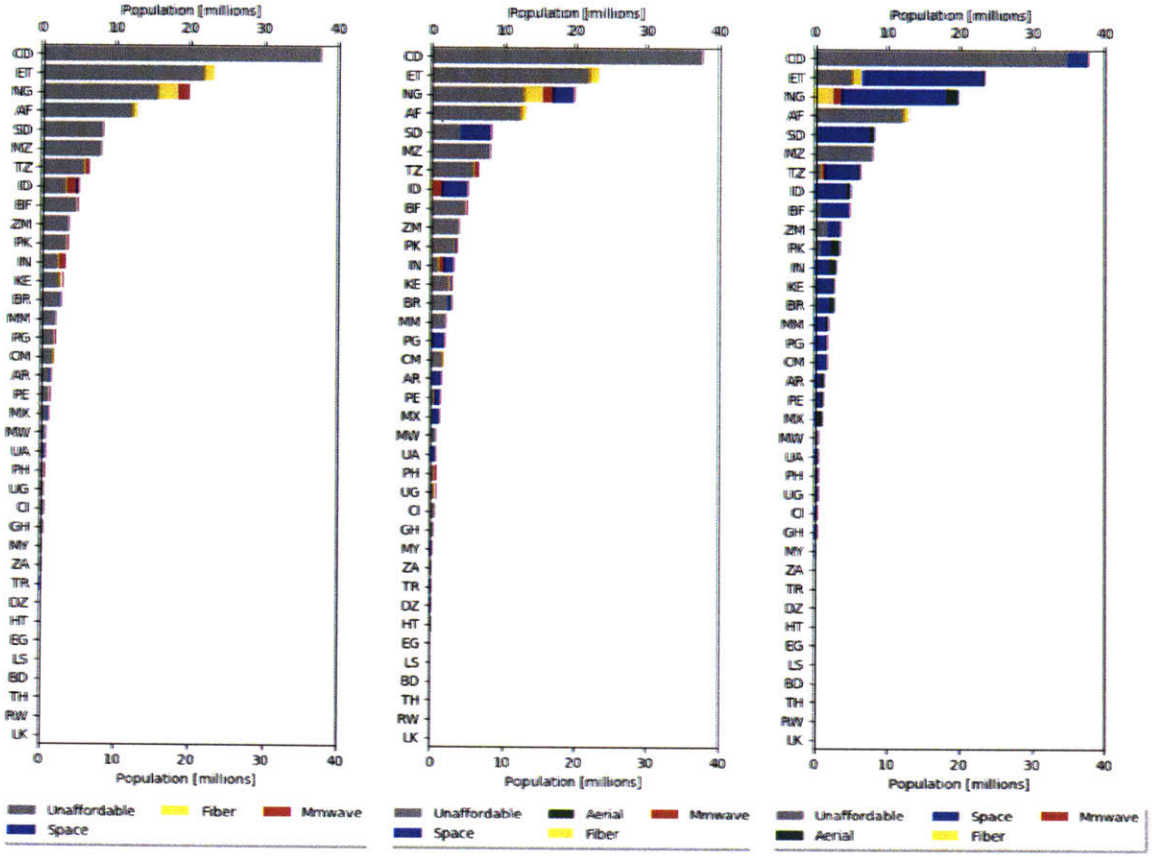
(a) Current scenario **(b)** Next generation (2-5 years) **(c)** Future scenario (8-10 years)

Figure 10-2: Percentage of uncovered and under-served population for all countries considered, broken down by most affordable technology, under the a) current, b) next-generation, and c) future scenarios. Countries are ordered by the total uncovered and under-served population in descending order.

Looking at the next-generation scenario in Figure 10-2b, it is again evident that mm-Wave and fiber deployments are only profitable in very high populated settlements. Notably, while the percentage of settlements in which fiber and mm-Wave deployments are most affordable for countries like India, Indonesia, or Turkey is tiny (Figure 10-1), this figure shows that they account for more than 20% of the uncovered and under-served population in these countries. Also, note that satellite-backhauled networks would be able to provide affordable connectivity to almost all the population of the high-income countries mentioned above.

Finally, for the future scenario (Figure 10-2c), the results show that satellite tech-

nology would be the most affordable technology for most of the people in the majority of the countries studied (more than 80% in multiple countries) in the majority of the countries, again except for the lowest income countries like Malawi, Mozambique, and the Democratic Republic of Congo. These results are highly promising for the space industry, which could increase its market share if (as the analysis in this chapter suggests) they are able to lower the price per Mbps/month substantially.



(a) Current scenario **(b)** Next generation (2-5 years) **(c)** Future scenario (8-10 years)

Figure 10-3: Uncovered and under-served population for all countries considered, broken down by most affordable technology, under the a) current, b) next-generation, and c) future scenarios. Countries are ordered by the total uncovered and under-served population in descending order.

Figure 10-3 shows the total uncovered and underserved population by most-affordable technology for each one of the 37 countries studied. Note how there are large differences between countries in the number of uncovered and under-served. in particular, the Democratic Republic of Congo, Ethiopia, and Nigeria account for 50% of the uncovered and under-served in the 37 countries.

Finally, Figure 10-4 shows the most affordable technology by percentage of the population, this time aggregated across all countries, under the three scenarios con-

sidered. It can be observed that at current prices, only 8% of those uncovered and under-served (approximately 13 million people in the 37 countries of interest) can afford connectivity and that the most effective technology would be mm-Wave deployments. The introduction of a new generation of satellites in the next 2-5 years could bring that number up to slightly higher than 19%, but it is worth noting that space networks will barely cannibalize any market share from other technologies such as mm-Wave or fiber, at least not when considering only uncovered and under-served populations. Finally, if space connectivity prices continue to drop, as is expected to happen in the next 8-10 years, the number of people that could afford connectivity through space networks will drastically increase to approximately 56% of the uncovered and under-served population in the 37 countries studied.

(a) Current scenario

(b) Next generation (2-5 years)

(c) Future scenario (8-10 years)

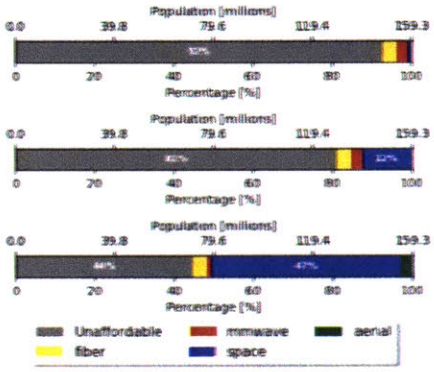


Figure 10-4: Overall percentage of uncovered and under-served population for all countries considered, broken down by most affordable technology, under the a) current, b) next-generation, and c) future scenarios.

Detailed results for selected countries

Figure 10-5 shows maps of the most affordable technology for Nigeria and Ethiopia. These maps were constructed by dividing each country's area into a uniform grid (0.1 deg $\times$0.1 deg) and coloring each cell according to the technology that the most number of people in the settlements contained within that cell can afford. Magenta regions represent areas where broadband connectivity (3G or 4G) is currently available.

In Nigeria (top row), it can be observed that there is currently connectivity around the large coastal cities, and also in large urban centers such as Abuja, Zaria, and Kano. At current costs, fiber and mm-Wave deployments would be most affordable in some regions in the north of the country (representing 20% of the uncovered and under-served population), whereas satellite backhaul would not be affordable in any area. The launch of new constellation systems in the next 2-5 years could afford Internet connectivity to more spots in the country (adding up an extra 18% of the population); in the future scenario, lower prices will result in practically all regions of the country being able to afford satellite connectivity. This is in contrast with Ethiopia (lower

row), where the only region currently with connectivity is the capital, Addis Ababa, and the recently-proposed satellite systems would have minimal impact in terms of connectivity. However, in the future scenario, most of the uncovered and under-served regions of the country will benefit from lower satellite connectivity prices.

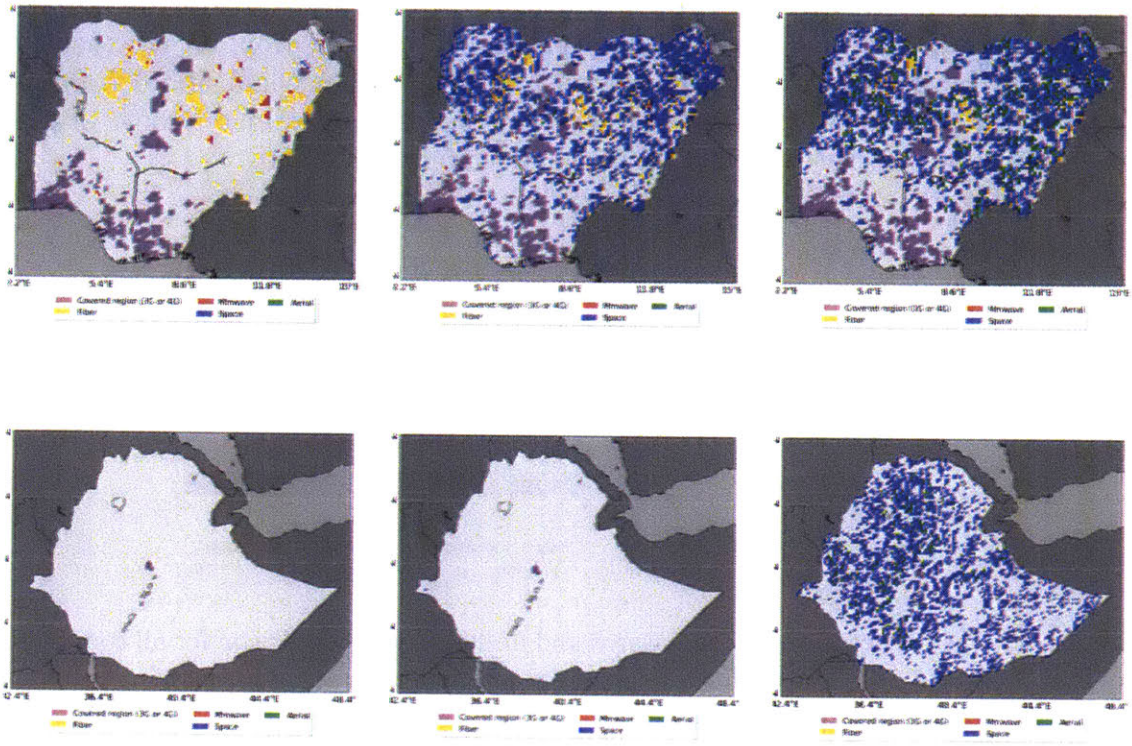


Figure 10-5: Regions by most affordable technology for Nigeria (top row) and Ethiopia (bottom row), under the current (left column), next-generation (middle column), and future (right column) scenarios. Each country's surface was divided into a uniform grid (0.1 deg $\times$0.1 deg) and each cell was assigned the color of the technology that would be affordable to the most people living within it.

10.2.2 Users brought online per technology

In the previous section, the population that could afford each of the technologies was analyzed. However, the analysis ignored non-price factors such as the market uptake rate or capacity saturation limits of space concepts. This section focuses on the second step of the methodology described at the beginning of this chapter, namely, to derive estimates of the potential number of users that would be brought online by each technology. In this regard, two main aspects need to be taken into account:

1. First, while aerial and terrestrial networks allow for rapid, incremental, and

localized deployments (by targeting only those settlements where a positive return on investment is likely), space networks need to be fully launched before becoming operational. This implies that unlike for aerial and terrestrial networks, where the cost per Mbps/month can be different for each settlement (depending on the number of potential users, CapEx and OpEx costs of the deployment), in space systems the cost per Mbps/month can be considered as uniform across all settlements.

2. Second, satellite constellations are capacity-constrained systems, as there are limits to the frequency reuse and a maximum bandwidth per beam (and thus data rate), which in turn limits the total throughput that can be supplied to a given area. In contrast, fiber optic deployments (given the low demands of the deployments considered herein) can be assumed to have infinite capacity. Finally, a fraction of the aerial and mm-Wave deployments might also be capacity-limited, for example, if the population of the settlements covered is very large; however, the capacity constraints have already been taken into account when defining the deployments and determining the price per Mbps/month for each settlement (as explained in Chapters 8 and 9).

Taking these into future account, the process to estimate the number of users that would be brought online is as follows:

1. Compute the maximum capacity that the space network can supply to a given country. These values are approximated as the country land-area multiplied by the throughput-density of the satellite network. For space networks, a maximum throughput density of 50 kbps/km^2, 100 kbps/km^2, and 150 kbps/km^2 are assumed to be achievable in the current, next-generation, and future scenarios, respectively.

2. For each settlement order the different technologies (space, aerial, terrestrial, and mm-Wave) from most affordable to least.

3. Sort settlements by descending order of income.

4. Iterate through the list of settlements to assign the most affordable technology per settlement, so that wealthier settlements are assigned first. Every time that satellite backhauling is identified as the most affordable technology for a settlement, the capacity required by the settlement is subtracted from the total capacity supplied to the country. If there is not enough remaining capacity to satisfy the demand, then the second-most affordable technology (if any) is assigned to the settlement.

5. Finally, the market uptake (or market adoption rate) is assumed to depend on the price of the service as compared to ITU 2% threshold. In particular, the adoption curve is defined using a sigmoid price-response function such that when the price is exactly 2% of the average monthly income, the market adoption rate

is 50% (note that most of the deployments considered herein are "greenfield"[1] deployments, and given the lack of competition it is therefore assumed that the market uptake will be high)[294]. Specifically, the market uptake (α) is computed as

$$\alpha = 1 - \frac{1}{1 + \exp\left(-0.5\left(\frac{\log(x)}{\log(1.5)} - 1\right)\right)}, \tag{10.1}$$

where x is the price of the service as a percentage of the monthly average income of the settlement. This price-response function is depicted in Figure 10-6. The parameters of this function were chosen manually so that they provide a good fit to the macroeconomic curve "Internet penetration vs. the average cost of a mobile broadband data-plan as a percentage of the GNI per capita".

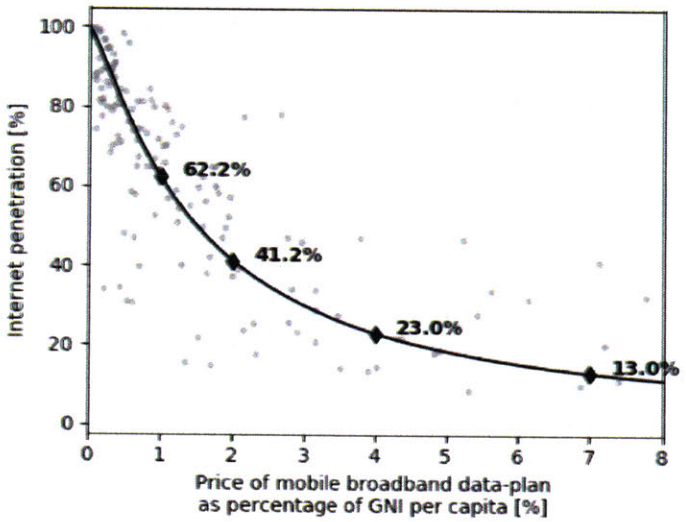


Figure 10-6: Market uptake as a function of the price of service.

Figure 10-7 shows the percentage of the population that is estimated to be brought online under the three scenarios considered. Several key differences can be observed between these results and those regarding affordability from the previous section (Figure 10-4):

- For the current scenario, the difference between the population that can afford connectivity services and the number of users brought online is very small. This can be explained by the observation that most of the affordable deployments are fiber and mm-Wave deployments, which are effective to cover large settlements, not capacity limited, and capable of offering very low prices per user which ensures high market uptake.

[1] A greenfield deployment is the installation of a network where none had existed before.

- For the next-generation scenario, it was estimated that up to 19% of those uncovered and under-served could afford connectivity services (Figure 10-4b); however, given the capacity constraints and the market adoption model considered, only half of them are estimated to actually become Internet users, bringing the aggregated potential impact of all the technologies considered in this dissertation down to just over 11% of the total population of these countries.

- For the future scenario, the greatest differences can be observed. Given the ever-lower satellite connectivity prices and the ubiquity of service, it was estimated that over 56% of the population on the countries studied could afford connectivity. However, after taking into account the market uptake estimates and limited satellite capacity, this value decreases to 32%.

(a) Current scenario

(b) Next generation (2-5 years)

(c) Future scenario (8-10 years)

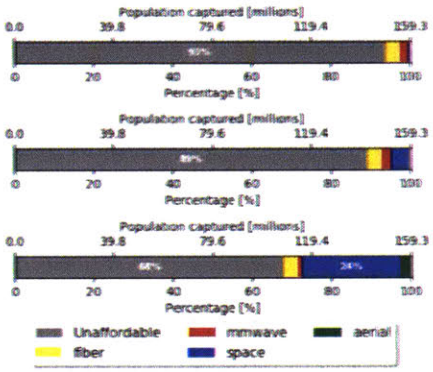
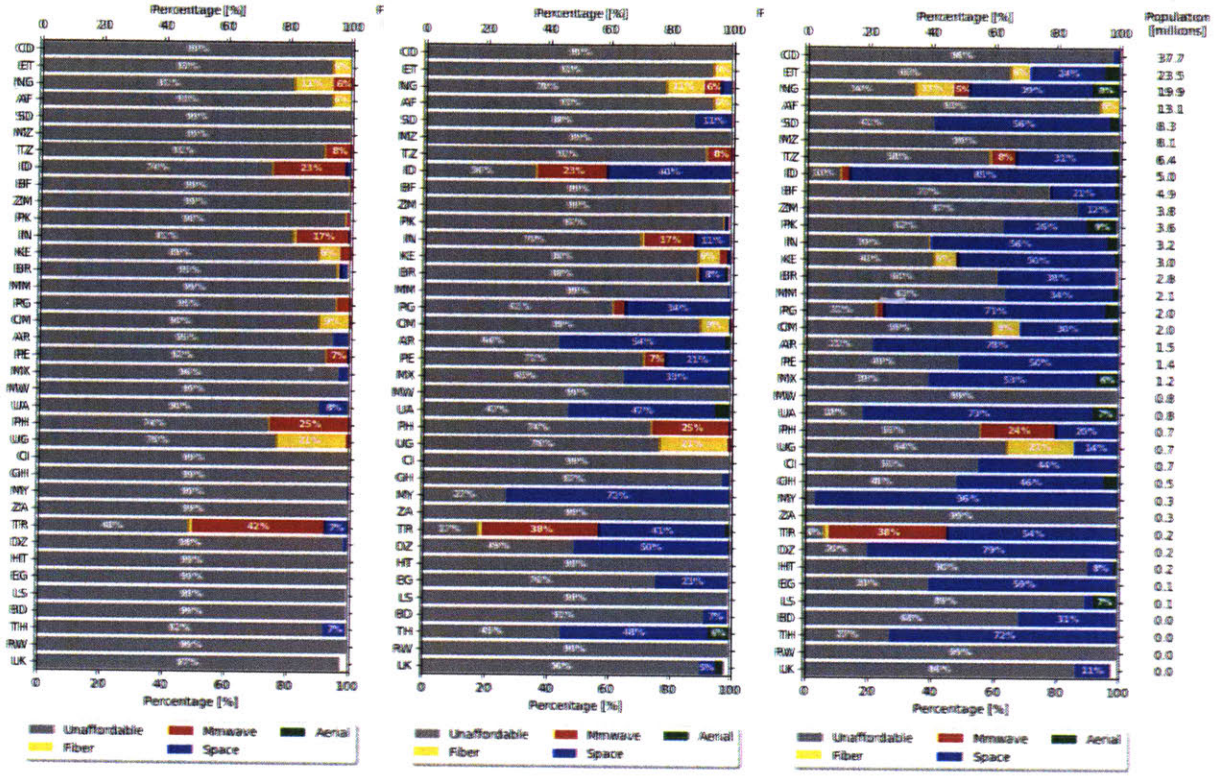


Figure 10-7: Overall percentage of uncovered and under-served population for all countries considered, broken down by most affordable technology, under the a) current, b) next-generation, and c) future scenarios.

Figure 10-8 shows detailed estimates for the number of users that would be brought online at the country-level. This image is analogous to Figure 10-2 (in the same way that Figure 10-7 was analogous to Figure 10-4), and by comparing both images, several characteristic behaviors can be observed.

- Countries with relatively small uncovered and under-served populations and relatively high incomes, such as Indonesia, Turkey, Ukraine, Argentina, Algiers, and Thailand, present the smallest differences between the population that can afford connectivity services and users brought online. This is because the high incomes result in higher market uptake and the smaller populations do not saturate the satellite capacity.

- Sub-Saharan countries with large uncovered and under-served populations, such as Nigeria, Ethiopia, and Tanzania, suffer the largest losses in terms of users actually brought online. The reasons for this are a combination of low GNI (which translates into low market uptake) and saturation of the satellite capacity (especially in Ethiopia and Nigeria).



(a) Current scenario **(b)** Next generation (2-5 years) **(c)** Future scenario (8-10 years)

Figure 10-8: Uncovered and under-served users brought online for each country broken down by most affordable technology, under the a) current, b) next-generation, and c) future scenarios. Countries are ordered by the total uncovered and under-served population, in descending order.

10.2.3 Influence of cost of customer premise equipment (CPE)

Another interesting factor to analyze is how the cost of the CPE[2] might influence the number of people brought online by space systems, given that advanced electronically-steered phased-array antennas or mechanically-steered gimballed antennas will be required to connect to NGSO constellations. Although current prices for such terminals are within the $20,000 - $50,000 range, it is envisioned that prices could decrease below the $1,000 threshold over the next few years.

For GEO networks the price of CPE terminals can be as low as $500, with the fixed-pointing parabolic dish antenna accounting for 20%-30% of this value. Thus, if the cheaper prices for the phased-array antennas required for MEO and LEO con-

[2]This CPE is assumed to be comprised of an antenna, modem, wireless access point, and additional electronics so that the mobile network operator can handle operational aspects (such as authentication, subscriber consumption, logging). Moreover, it is assumed that a single CPE can serve up to 5,000 subscribers. If the number of subscribers in a settlement is higher than this value, multiple CPEs will be deployed.

stellations do not materialize, GEO constellations will have an edge when trying to connect uncovered and under-served populations.

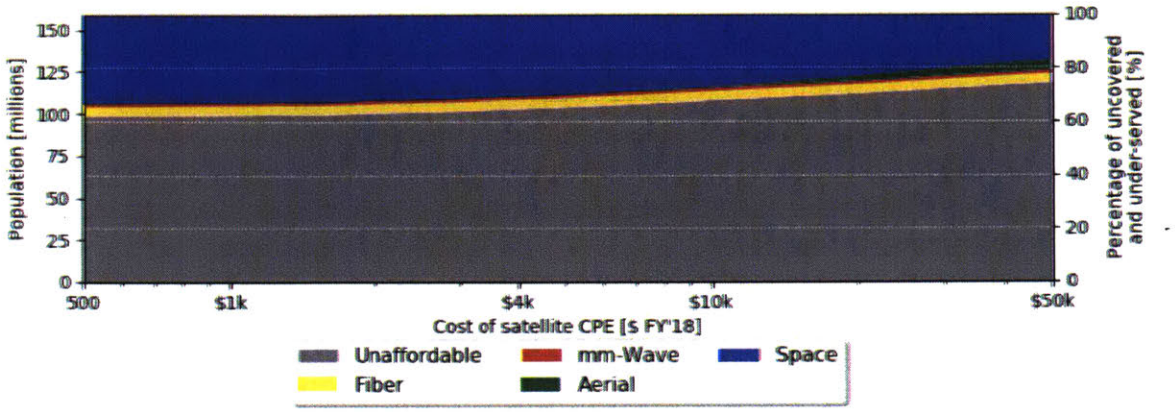


Figure 10-9: Users brought online by different technologies versus cost of satellite CPE. It is assumed that a CPE can serve at most 5,000 subscribers.

Figure 10-9 shows the influence of the cost of the CPE on the number of people connected. Note that if the CPE cost is $50,000, the total number of people connected by satellite is slightly under 20%, but if CPE costs fall below $2,000, this percentage increases to over 30%, showing that the total number of users brought online is very sensitive to the CPE cost. This result highlights the importance of the user terminals for NGSO constellations; if manufacturers can lower the prices of the antennas, LEO and MEO constellations would have a larger impact in terms of bringing uncovered and under-served populations online, as they would then be capable of providing higher throughput densities than GEO networks. Otherwise, GEO networks would be the best option given that their current CPE costs are significantly lower.

10.3 Chapter conclusions

This chapter conducted a comparative assessment of the potential impact of space, aerial, and terrestrial networks, and in doing so answered the following research question:

Research Question 10.1

Given capacity limits, market uptake, and achievable prices per Mbps/month, what is the impact of space and aerial technologies to expand connectivity to uncovered and under-served regions?

Under the current scenario, the impact of space and aerial systems in terms of expanding connectivity would be rather modest; the current cost of satellite tech-

nology ($\sim$ \$200 Mbps/month) are affordable for less than 1% of the uncovered and under-served population in the countries of interest. In contrast, it is estimated that fiber optic and mm-Wave deployments could bring online 7% of the uncovered and under-served population. Looking at the next 2-5 years, after large LEO constellations and vHTS GEO and MEO satellites are launched, the impact of space and aerial systems might still be limited too; aerial concepts would remain too expensive, and space systems would only have an impact in high-income countries in South America and Southeast Asia, where around 8 million people ($\sim$5% of the uncovered and under-served population in the 37 countries studied) would connect to the Internet by satellite backhauled services. Finally, in the future scenario (8-10 years), where satellite connectivity prices are expected to drop to the \$30-\$50 per Mbps/month range, satellite networks would be a viable technology for up to 38.4 million people (24%) in these countries.

Finally, a critical piece of technology that has a substantial impact on the number of people brought online by NGSO constellations is the cost per CPE. If this cost is \$50,000, the total number of people connected by satellite is slightly under 20% of the uncovered and under-served population in the 37 countries, but if costs fall below \$2,000, this percentage increases to over 30%.

Chapter 11

Conclusion

Internet connectivity has radically transformed the way we live and work. The progress in connectivity in the last two decades has brought about a whole new digital economy, and numerous advancements in healthcare, education, government, and business. However, 46.4% of the world's population does not currently benefit from these advancements, reason being that they remain unconnected. Due to the sparse distribution of those uncovered, several space and aerial concepts have been proposed in the last five years to help bring the remaining uncovered population online.

The main research objective of this dissertation was to analyze the potential impact of these concepts from a techno-economic perspective. Given the scale of the problem from the large tradespaces under consideration to the different spatial and temporal scales in which these concepts operate, this evaluation would have required the development of computationally-expensive simulation models. However, due to limited computing resources, the number of design-configurations that can be evaluated was heavily constrained; thus, a secondary research objective was to assess the efficiency of Bayesian optimization as a method to conduct tradespace exploration.

This final chapter summarizes the key findings and contributions of this dissertation. Opportunities for future research are also outlined, and finally, I close with some final remarks regarding the future of connectivity.

11.1 Thesis summary

Chapter 1 began by presenting the motivation for these two research objectives; first, highlighting the importance of connecting the other half, as stated by international bodies such as the International Telecommunications Union and the UN Broadband Commission, and then analyzing the problem from a System Architecture perspective. Within this context, Chapter 2 conducted a literature review on the three main topics

of interest of this thesis: space and aerial systems for expanding Internet connectivity, concept selection methods, and tradespace exploration for System Architecture problems. This review allowed for the identification of relevant research gaps, namely, the need for comprehensive techno-economic analyses that compare the impact of space and aerial concepts in terms of expanding connectivity, and the need for a method to conduct tradespace exploration that can operate effectively in scenarios where there exists a tight budget for function evaluations.

The rest of the dissertation was divided into two parts, each providing answers to one of the two research objectives.

In the first part, the suitability of Bayesian optimization (BO) as a tradespace exploration method for System Architecture problems (SAP) was studied. Chapter 3 presented the theory underlying Bayesian optimization using a Gaussian process-based example to demonstrate how it is used in continuous one-dimensional problems. With a firm understanding of the fundamentals of BO, three different ways to adapt BO to SAPs were identified: a) use a model well suited for combinatorial problems, b) change the covariance function, and c) use a tailored distance metric.

Chapters 4 and 5 analyzed the performance of the alternative various of Bayesian optimization as compared to other state-of-the-art optimization algorithms for single- and multi-objective problems, respectively. These chapters concluded with specific recommendations on the scenarios where BO would yield time-saving gains, as well as what was identified as the best BO method for each type of decision pattern based on its asymptotic performance.

In the second part of the dissertation, the problem of how to extend global connectivity to uncovered and under-served communities using space and aerial networks was addressed. First, Chapter 6 provided a general overview of the current state of connectivity, showing that among the unconnected, a large fraction lives in areas where there is broadband connectivity (*usage gap*), whereas only a small fraction lives in places with no or low-quality connectivity (*coverage gap*). This led to further analysis of the barriers preventing wider adoption, concluding that the *usage gap* can be attributed mainly to socio-cultural barriers (which are beyond the scope of this thesis), whereas techno-economic obstacles mainly caused the *coverage gap*. Being aware of how the complex interplay between infrastructure and affordability issues gives rise to the *coverage gap*, I developed a five-step techno-economic methodology to evaluate the technical performance and assess the economic viability of different aerospace communications concepts which might help to expand global connectivity to uncovered and under-served regions.

Chapters 7, 8, and 9 systematically applied this methodology to space, aerial, and terrestrial systems, respectively. These three chapters followed a similar structure: first, the specific technical and economic models used to assess the performance and cost of different types of architectures were outlined; then, the models were exercised to explore the tradespace of possible architectures and identify the best designs; fi-

nally, main-effects analysis was conducted to identify the decisions that were major performance and cost drivers, and sensitivity analysis was performed to identify the model parameters that had the most influence on the results and assess the degree of validity of the models.

Finally, the second part of this dissertation concluded by comparing the potential impact of each concept in the context of a competitive market. This comparison was made by conducting a study with 37 countries for which population and income data was readily available and determining the most affordable technology for each settlement under consideration. The study provided estimates for the total population that would stand to be impacted by each of the concepts under three time-horizons: current, next-generation (i.e., concepts in 2-5 years), and future (in 8-10 years) scenarios.

11.2 Executive summary of results

This section summarizes the results obtained in answering the two research questions introduced in Chapter 1.

Research Question 1

What kind of novel space and aerial systems can complement existing infrastructure and contribute towards expanding global connectivity at an affordable cost? What is the potential impact of such systems in terms of connecting additional populations?

The analysis for satellite networks conducted in Chapter 7 identified that **GEO and MEO constellations are most cost-effective in providing connectivity to uncovered and under-served regions**. In particular, equatorial constellations of 3-10 satellites in GEO with high-capable payloads and constellations of 10-20 satellites in MEO with moderate-capacity payloads are the preferred architectures. Reduced costs will be achieved if higher frequency bands (Q/V and EHF) are used and with further miniaturization and digitalization of payloads.

Regarding aerial systems, the high CapEx and OpEx costs render many of the current proposals infeasible, including station-keeping high altitude platforms such as blimps and UAVs. Similarly, low altitude platforms will also have a very small impact, since their smaller coverage radii and the stringent LoS requirements, coupled with relatively high costs (as compared to wireless backhaul-networks), limit the ability of these systems to create networks that would extend great distances from existing PoPs. Lighter-than-air HAP systems can be competitive in certain regions as a result of their lower CapEx costs, although the required prices per Mbps/month for profitability ($\sim$\$50) are still higher than those obtained by satellite networks. In the long run, the success of such systems will be greatly dependent on the development of

technologies that would allow for some degree of station-keeping control, as otherwise, the large number of additional platforms required just to provide continuous coverage will offset the gains obtained from the reduced costs.

In terms of the potential impact that these systems might have, the following results were obtained:

- **Space concepts**: Prices per Mbps/month of ~$100 are expected to be achieved by the next generation (i.e., 2-5 years) of constellations (including O3b mPower, Viasat-3). At these prices, satellite networks would be affordable for up to 120 million of currently uncovered and under-served people worldwide, with 40 million of them adopting the service. However, further reductions in price may be achievable for the future generations of satellites (8-10 years); if prices drop below the $35/Mbps/month threshold, the number of additional people currently uncovered and under-served for whom satellite backhaul could be affordable would rise to 300 million worldwide, with 81 million adopting the service. These price reductions will be enabled by transitioning to higher-frequency bands (Q/V-band), and further miniaturization and integration of components.
- **Aerial concepts**: Aerial systems will have a limited impact on their ability to extend connectivity to uncovered and under-served regions. Considering the 37 countries studied, balloon-based networks will be able to offer affordable coverage to *at most* 17 million (representing 11% of the total combined uncovered and under-served population of those countries), whereas UAV networks will have a minimal impact, reaching at most 11 million people.

For comparative purposes, terrestrial networks were also analyzed, with the following results

- **Terrestrial concepts**: At current CapEx costs, both fiber and mm-Wave networks will have limited impact in providing additional coverage (7% of the uncovered and under-served might be connected by fiber, while ~17% might be served by mm-Wave backhaul networks). Because fiber costs are fairly fixed, equipment costs will continue to prevent a large fraction of the uncovered and under-served population from affording fiber service in the mid-to-long term. mm-Wave backhaul networks, on the other hand, may have a more significant impact (further expanding coverage to 40% of the currently uncovered and under-served) if: a) taller towers (60-80 m) can be deployed economically, and b) further CapEx reductions can be achieved (by a factor of 5×).

Research Question 2

In the context of System Architecture, how should Bayesian optimization be used to conduct tradespace exploration when computationally expensive models are required and the number of point evaluations is limited?

Bayesian optimization can be adapted to combinatorial problems (both single- and multi-objective) by using different models (e.g., random forest, tree Parzen estimator), different kernels (e.g., Matérn, Kendall, Mallows), and/or different distance functions (e.g., Hamming distance, Damerau-Levenshtein distance). The appropriate technique(s) would depend on the type of combinatorial pattern involved, as summarized in Tables 11.1 and 11.2.

Finally, it should be noted that Bayesian optimization is only applicable for problems with a low-to-moderate input dimension ($d < 20$) and for cases where the function evaluation takes more than 5 seconds (for single-objective problems) or 30 seconds (for multi-objective problems). Overall, the speedups attainable when compared to traditional optimization methods range from a factor of 3 to 15.

Table 11.1: Summary of recommendations for single-objective combinatorial patterns.

	Method	Distance function	Speedup
Assigning	BO-RBF	Hamming	>10x
Combining	BO-RBFN	Hamming	4.3x
Downselecting	BO-RBF	Hamming	15.3x
Partitioning	BO-RBF	Damerau-Levenshtein	4.9x
Permuting	BO-RBFN	Hamming	14.1x

Table 11.2: Summary of recommendations for multi-objective combinatorial patterns.

	Method	Distance function	Speedup
Assigning	BO-RBF	Hamming	8.1x
Combining	BO-TREE, BO-RBF, RS	Hamming	5x
Downselecting	BO-RBF	Hamming	6.6x
Partitioning	SPEA2, GA-EAS	-	2.8x
Permuting	BO-RBF, BO-MT3/2	Hamming	3.2x

11.3 Thesis contributions

This thesis has made several contributions and findings that expand the current understanding of how space and aerial concepts can help expand global connectivity, as well as learnings concerning tradespace exploration methodologies when computationally expensive functions are involved. Broadly, these contributions can be divided into two groups: methodological and domain-specific contributions.

11.3.1 Methodological contributions

The methodological contributions include:

- The development of **an adapted Bayesian optimization formulation to aid tradespace exploration in System Architecture problems** where computationally-expensive models are required. The work conducted in Chapters 3, 4, and 5 analyzed the performance of different Bayesian optimization techniques for decision patterns common in System Architecture problems. Time savings were quantified in terms of the reduction in the number of function evaluations needed and the actual time required to conduct the tradespace exploration. Overall, in single-objective problems, the method presented herein is beneficial only when function calls take more than 5 seconds, whereas for the multi-objective case, BO should be applied for function calls with run times greater than 30 seconds.

  In particular, for single-objective problems, the analysis concluded that Bayesian optimization using a Gaussian process together with the RBF kernel and the Hamming distance is the best option for the assigning, downselecting, and partitioning patterns, whereas a Bayesian optimization approach using an RBFN as the surrogate model is the best option for the combining and permuting patterns. In a multi-objective settings, a Gaussian process with the RBF kernel and the Hamming distance is the best option for the assigning and downselecting, and permuting patterns; the genetic algorithm-based methods are preferred for the partitioning pattern; and for the combining pattern, the BO methods and RS were superior to GA-based methods, but there was no clear winner among them.

- The development of a **techno-economic methodology to assess the potential impact of space and aerial concepts** in expanding connectivity to uncovered and under-served regions. This five-step methodology, combined with performance and cost models tailored to the different aerospace concepts, can be applied at a global scale as long as the relevant population and income data is available. Furthermore, the methodology provides a systematic procedure to evaluate, analyze, and compare concepts that operate at different spatial and temporal scales (e.g., a GEO satellite with a 15-year lifetime vs. a network of stratospheric LTA balloons with an endurance of a few months), a problem not addressed before within the literature.

- The development of **statistical performance models for space and aerial concepts**: this dissertation has developed two types of statistical models to evaluate the performance of space and aerial concepts. First, to evaluate the total sellable capacity of space networks, a Monte Carlo simulation model which takes into account the intrinsic stochastic behavior of NGSO constellations was developed. This model generates random samples for both orbital positions

and atmospheric attenuation and is capable of quickly evaluating system performance to a high degree of accuracy, which allowed for the detailed evaluation of novel and currently proposed satellite constellations (such as OneWeb, Telesat, SpaceX, O3b mPower, or Viasat-3 systems). Second, for aerial systems, a Gaussian random walk model was developed to evaluate the number of additional platforms required to maintain continuous coverage. These models were then used to evaluate HAP and LAP networks, operating at different altitudes and frequency bands.

11.3.2 Domain-specific contributions

The principal research question of this dissertation evaluated space, aerial, and terrestrial networks first individually as isolated systems, and then within a competitive framework. The main contributions within this area include:

- The **assessment of the potential impact** of space, aerial, and terrestrial technologies in bringing uncovered and under-served populations online through a **comparative analysis in 37 countries**. With regard to the comparative analysis, the results showed that space systems represented the concept with the highest potential to bring uncovered and under-served populations online.

  For the next generation of satellites, the countries that would benefit the most are those with mid-income levels, mainly in South America and Southeast Asia (e.g., Argentina, Peru, Mexico, Thailand, and Myanmar). In particular, the analysis shows that more than 70% of the uncovered and under-served population in those countries will be able to afford satellite-backhauled connectivity. When considering the 37 countries, however, only 12% will be able to afford satellite connectivity. Overall, once market adoption rates and capacity limits are taken into account, the number of users brought online by satellite technology may be further reduced to just 8 million users (5%).

  The future generation of satellite networks (with prices per Mbps/month as low as $30 Mbps/month) would have a greater potential, as approximately 55% of the uncovered and under-served population of the 37 countries would be able to afford Internet plans through space-backhauled networks. However, once the market uptake and the limited capacity density (i.e., data rate that can be supplied to a given area) are accounted for, the final number of new users to the Internet would be ~38 million (24%) for satellite networks, as these systems might be unable to provide the full throughput required in some countries like Nigeria or Ethiopia).

- The **identification of most important technical decisions** through analyses of the characteristics of the most cost-effective architectures for space and aerial networks. In that sense, for space systems the analysis showed that the orbital altitude, number of satellites, and number of beams played the most

important role; whereas, for aerial systems, the altitude, station-keeping capabilities, and payload capacity were the most relevant decisions.

11.4 Future work

Several areas of future work were identified in the completion of this dissertation, regarding both Bayesian optimization methods for System Architecture problems and the study of the impact of space and aerial concepts to expand global connectivity.

Bayesian optimization for System Architecture problems

There are several open problems in the area of Bayesian optimization for SAPs that are worth further study.

- **Fragment architectures**: The discussion presented herein has focused on SAPs where an architecture is composed of a single decision pattern. However, there are multiple scenarios where this is not the case, and an architecture is composed of different fragments, each corresponding to a different decision pattern (as was the case in [295, 296]). Thus, research must be conducted to explore how Bayesian optimization methods can be adapted for mixed problems. Possible approaches might include defining distance functions that incorporate the different decision patterns, using a linear combination of kernels (each suited to one of the decision fragments) instead of a single kernel, or developing new models that can handle simultaneously multiple kinds of decision patterns.
- **Consider additional surrogate models**: In addition to Gaussian processes, random forest, and tree-structured Parzen density estimator, graph-based models are a popular topic within the combinatorial Bayesian optimization community. Example applications have been studied for combining patterns [218], but their applicability to other types of decision patterns remains an open question.
- **Benchmark test problems**: The analysis presented here used a library of test problems (mostly synthetic) that capture the essence of each of the decision patterns, and allowed for variables such as the input dimensions and the degree of interaction between input variables to be modified arbitrarily. The results obtained, are therefore conditioned and must be interpreted within the context of the test problems in the library. It would be desirable to have a larger collection of real-world problems available for each of the decision patterns, which could be used to further analyze and validate the performance of Bayesian optimization.
- **Effects of dimensionality and degree of interaction**: The study presented herein analyzed visually the effects of higher dimensionality and degree of interaction for the downselecting and partitioning problems. Even though it was

observed that the larger the dimensionality of the problem and the degree of interaction among variables, the worse Bayesian optimization tends to perform, the limited number and diversity of the test problems prevented us from providing clear guidelines with regard to these variables. A deeper study taking into account a larger number of problems would allow for clear recommendations to be prescribed, to help practitioners assess whether Bayesian optimization is the appropriate method for the problem at hand.

Space and aerial concepts to expand global connectivity

Concerning space and aerial concepts, there are also several avenues of further research worth exploring:

- **A larger study including more countries** would extend the results from this dissertation, which focused only on a subset of 37 countries in sub-Saharan Africa, Southeast Asia, and South America (for which data was readily available). All these countries have in common the fact that they are currently classified as underdeveloped or developing countries with low to mid incomes; this played an important role in the results since the low average revenues per user attainable (between $0.50 and $4 per month) heavily conditioned the economic viabilities of the architectures. Thus, it would be interesting to expand the analysis to a larger group of countries, including developed nations, and study whether the conclusions would hold. To that, the data tonnage and average data rate used to characterize the user consumption would have to be adapted accordingly.

- **Analyze the potential impact of space and aerial concepts in reducing the usage gap**: The analysis in this document has focused on studying the impact of space and aerial concepts in reducing the *coverage gap* (i.e., providing connectivity in areas where it is currently not available). However, the majority of the offline population (72%) resides in places where broadband *is* currently available (the *usage gap*), and thus it can be concluded that readiness, relevance, and (perhaps most importantly) affordability barriers are preventing a significant portion of the population from becoming active Internet users. Beyond looking to socio-political measures, a study to understand how space and aerial concepts might benefit these communities could prove to be an important area for future research.

  Conducting this kind of analysis would require not only a techno-economic methodology similar to the one presented in this dissertation but also market models that include the effects of competition from existing ISPs and MNOs. The analysis conducted in this dissertation has shown that satellites, while limited in capacity, can achieve lower monthly costs per Mbps than terrestrial systems in low- and medium-density regions if new infrastructure has to be

built. However, it is unclear if this result would hold in regions where existing infrastructure is already in place, and if so, what potential impact these architectures might have in high-density areas.

- **Model fidelity**: Finally, this body of research and the results presented would also benefit from increased model fidelity and reduced uncertainties with regard to some of the parameter values. In particular, it would be interesting to develop more advanced sizing and cost parametric models for aerial networks, since the ones presented in this study are rather simple due to the lack of historical data. Furthermore, the viabilities of the LEO constellations should be revisited if larger-than-expected cost reductions are achieved through massive-scale satellite manufacturing, if the launch costs are further reduced through rocket-reuse, or if market uptake is significantly higher than the values assumed here.

Appendix A

Cost of a mobile- and fixed-broadband data plans in different countries

Table A.1: Cost of mobile broadband (1 GB/month) and fixed broadband data plans

ISO2	Name	Popul. [million]	GNIpc [$'FY19]	Mobile broadband [$'FY19]	%GNIpc	Fixed broadband [$'FY19]	%GNIpc
AF	Afghanistan	35.53	$550	$1.79	3.9%	$52.78	115.2%
AL	Albania	2.87	$4,860	$3.89	1.0%	$33.17	8.2%
DZ	Algeria	41.32	$4,060	$5.15	1.5%	$33.51	9.9%
AO	Angola	29.78	$3,370	$11.33	4.0%	$82.43	29.4%
AG	Antigua and Barbuda	0.10	$15,810	$23.90	1.8%	$177.15	13.4%
AR	Argentina	44.27	$12,370	$7.60	0.7%	$15.51	1.5%
AM	Armenia	2.93	$4,230	$2.36	0.7%	$46.79	13.3%
AW	Aruba	0.11	$23,630	$12.70	0.6%	N/A	N/A
AU	Australia	24.60	$53,190	$4.57	0.1%	$52.77	1.2%
AT	Austria	8.80	$49,260	$3.31	0.1%	$50.70	1.2%
AZ	Azerbaijan	9.85	$4,050	$3.04	0.9%	$36.97	11.0%
BS	Bahamas	0.40	$30,210	$6.89	0.3%	$72.77	2.9%
BH	Bahrain	1.49	$21,890	$2.75	0.2%	$96.29	5.3%
BD	Bangladesh	164.67	$1,750	$2.49	1.7%	$36.33	24.9%
BB	Barbados	0.29	$15,240	$11.11	0.9%	$111.86	8.8%
BY	Belarus	9.50	$5,670	$5.18	1.1%	$10.46	2.2%
BE	Belgium	11.38	$45,340	$11.61	0.3%	$45.68	1.2%
BZ	Belize	0.37	$4,720	$10.84	2.8%	$58.58	14.9%
BJ	Benin	11.18	$870	$5.22	7.2%	$83.94	115.8%
BT	Bhutan	0.81	$3,080	$1.25	0.5%	N/A	N/A
BO	Bolivia	11.05	$3,370	$6.88	2.4%	$47.83	17.0%
BA	Bosnia and Herzegovina	3.51	$5,690	$4.93	1.0%	$35.47	7.5%
BW	Botswana	2.29	$7,750	$12.08	1.9%	$91.64	14.2%
BR	Brazil	209.29	$9,140	$5.25	0.7%	$48.00	6.3%

Continued on next page

Table A.1 Cost of mobile and fixed broadband data plans – cont. from previous page

ISO2	Name	Popul. [million]	GNIpc [$'FY19]	Mobile broadband [$'FY19]	%GNIpc	Dixed broadband [$'FY19]	%GNIpc
BN	Brunei Darussalam	0.43	$31,020	$6.34	0.2%	$123.29	4.8%
BG	Bulgaria	7.08	$8,860	$6.49	0.9%	$30.27	4.1%
BF	Burkina Faso	19.19	$660	$6.11	11.1%	$201.94	367.2%
BI	Burundi	10.86	$280	$2.00	8.6%	N/A	N/A
KH	Cambodia	16.01	$1,380	$1.49	1.3%	$62.29	54.2%
CM	Cameroon	24.05	$1,440	$3.48	2.9%	$58.75	49.0%
CA	Canada	36.71	$44,860	$13.16	0.4%	$57.66	1.5%
CV	Cape Verde	0.55	$3,450	$5.17	1.8%	N/A	N/A
CF	Central African Republic	4.66	$480	$7.61	19.0%	N/A	N/A
TD	Chad	14.90	$670	$16.50	29.6%	N/A	N/A
CL	Chile	18.05	$14,670	$4.79	0.4%	$45.50	3.7%
CN	China	1386.40	$9,470	$9.89	1.3%	$41.29	5.2%
CO	Colombia	49.07	$6,190	$8.23	1.6%	$44.81	8.7%
KM	Comoros	0.81	$1,320	$9.32	8.5%	$114.34	103.9%
CD	Democratic Republic of Congo	81.34	$490	$6.73	16.5%	N/A	N/A
CG	Congo	5.26	$1,640	$7.25	5.3%	N/A	N/A
CR	Costa Rica	4.91	$11,510	$5.81	0.6%	$73.51	7.7%
HR	Croatia	4.12	$13,830	$4.59	0.4%	$26.74	2.3%
CY	Cyprus	1.18	$26,300	$18.67	0.9%	$53.01	2.4%
CZ	Czech Republic	10.59	$20,260	$7.65	0.5%	$25.94	1.5%
DK	Denmark	5.76	$60,190	$3.35	0.1%	$51.11	1.0%
DJ	Djibouti	0.96	$2,180	$37.92	20.9%	$97.41	53.6%
DM	Dominica	0.07	$7,210	$16.64	2.8%	$71.43	11.9%
DO	Dominican Republic	10.77	$7,370	$7.90	1.3%	$29.66	4.8%
EC	Ecuador	16.62	$6,120	$10.00	2.0%	$62.45	12.2%
EG	Egypt	97.55	$2,800	$1.47	0.6%	$13.58	5.8%
SV	El Salvador	6.38	$3,820	$10.00	3.1%	$39.32	12.4%
GQ	Equatorial Guinea	1.27	$7,050	$50.31	8.6%	N/A	N/A
EE	Estonia	1.32	$20,940	$5.25	0.3%	$27.63	1.6%
SZ	Eswatini	1.37	$3,850	$12.14	3.8%	$68.78	21.4%
ET	Ethiopia	104.96	$790	$3.23	4.9%	$125.19	190.2%
FJ	Fiji	0.91	$5,860	$7.79	1.6%	$64.49	13.2%
FI	Finland	5.51	$47,750	$6.22	0.2%	$52.23	1.3%
FR	France	67.11	$41,080	$6.79	0.2%	$31.14	0.9%
GA	Gabon	2.03	$6,800	$5.86	1.0%	$103.59	18.3%
GM	Gambia	2.10	$700	$5.29	9.1%	N/A	N/A
GE	Georgia	3.73	$4,130	$3.33	1.0%	$18.70	5.4%
DE	Germany	82.69	$47,180	$8.88	0.2%	$36.68	0.9%
GH	Ghana	28.83	$2,130	$3.69	2.1%	N/A	N/A
GR	Greece	10.75	$19,600	$19.74	1.2%	$58.22	3.6%
GD	Grenada	0.11	$9,780	$14.11	1.7%	$60.37	7.4%

Continued on next page

Table A.1 Cost of mobile and fixed broadband data plans – cont. from previous page

ISO2	Name	Popul. [million]	GNIpc [$'FY19]	Mobile broadband [$'FY19]	%GNIpc	Dixed broadband [$'FY19]	%GNIpc
GT	Guatemala	16.91	$4,410	$6.48	1.8%	$41.27	11.2%
GN	Guinea	12.72	$830	$3.34	4.8%	N/A	N/A
GW	Guinea-Bissau	1.86	$750	$8.82	14.1%	N/A	N/A
GY	Guyana	0.78	$4,760	$7.89	2.0%	$123.11	31.0%
HT	Haiti	10.98	$800	$5.18	7.8%	$207.39	311.1%
HN	Honduras	9.27	$2,330	$11.94	6.1%	$68.50	35.3%
HK	Hong Kong	7.39	$50,310	$11.96	0.3%	$72.68	1.7%
HU	Hungary	9.79	$14,590	$6.81	0.6%	$18.37	1.5%
IS	Iceland	0.34	$67,950	$8.57	0.2%	$76.66	1.4%
IN	India	1339.18	$2,020	$0.73	0.4%	$28.23	16.8%
ID	Indonesia	263.99	$3,840	$2.30	0.7%	$54.85	17.1%
IR	Iran	81.16	$5,470	$1.47	0.3%	$8.20	1.8%
IQ	Iraq	38.27	$5,030	$8.00	1.9%	$51.95	12.4%
IE	Ireland	4.81	$59,770	$8.14	0.2%	$65.12	1.3%
IL	Israel	8.71	$40,850	$5.01	0.1%	$13.02	0.4%
IT	Italy	60.54	$33,540	$2.77	0.1%	$29.48	1.1%
CI	Ivory Coast	24.29	$1,610	$7.02	5.2%	$73.87	55.1%
JM	Jamaica	2.89	$4,990	$5.90	1.4%	$62.57	15.0%
JP	Japan	126.79	$41,340	$15.36	0.4%	$37.15	1.1%
JO	Jordan	9.70	$4,210	$3.28	0.9%	$77.45	22.1%
KZ	Kazakhstan	18.04	$7,830	$2.13	0.3%	$16.14	2.5%
KE	Kenya	49.70	$1,620	$3.16	2.3%	$75.18	55.7%
KI	Kiribati	0.12	$3,140	$6.25	2.4%	N/A	N/A
KR	Korea	51.47	$30,600	$14.10	0.6%	$32.29	1.3%
KW	Kuwait	4.14	$33,690	$2.01	0.1%	$54.60	1.9%
KG	Kyrgyzstan	6.20	$1,220	$1.67	1.6%	$108.22	106.4%
LA	Laos	6.86	$2,460	$4.71	2.3%	$239.25	116.7%
LV	Latvia	1.94	$16,740	$7.12	0.5%	$18.68	1.3%
LB	Lebanon	6.08	$7,690	$17.00	2.7%	$37.60	5.9%
LS	Lesotho	2.23	$1,380	$3.62	3.1%	$108.49	94.3%
LR	Liberia	4.73	$600	$5.00	10.0%	N/A	N/A
LY	Libya	6.37	$6,330	$8.05	1.5%	N/A	N/A
LT	Lithuania	2.83	$17,350	$4.94	0.3%	$16.84	1.2%
LU	Luxembourg	0.60	$69,420	$11.45	0.2%	$71.15	1.2%
MO	Macao	0.62	$78,320	$4.84	0.1%	$43.72	0.7%
MK	Macedonia	2.08	$5,450	$2.73	0.6%	$29.50	6.5%
MG	Madagascar	25.57	$440	$6.24	17.0%	$66.64	181.7%
MW	Malawi	18.62	$360	$4.86	16.2%	N/A	N/A
MY	Malaysia	31.62	$10,460	$5.90	0.7%	$47.92	5.5%
MV	Maldives	0.44	$9,310	$7.46	1.0%	$81.55	10.5%
ML	Mali	18.54	$830	$9.22	13.3%	$160.53	232.1%
MT	Malta	0.47	$26,220	$16.23	0.7%	$53.70	2.5%
MR	Mauritania	4.42	$1,190	$3.12	3.1%	$768.16	774.6%
MU	Mauritius	1.26	$12,050	$5.94	0.6%	$64.06	6.4%
MX	Mexico	129.16	$9,180	$8.11	1.1%	$33.32	4.4%
FM	Micronesia	0.11	$3,580	$12.87	4.3%	$103.60	34.7%
MD	Moldova	3.55	$2,990	$2.91	1.2%	$11.28	4.5%
MN	Mongolia	3.08	$3,580	$6.42	2.2%	$17.97	6.0%
ME	Montenegro	0.62	$8,400	$4.67	0.7%	$33.87	4.8%

Continued on next page

Table A.1 Cost of mobile and fixed broadband data plans – cont. from previous page

ISO2	Name	Popul. [million]	GNIpc [$'FY19]	Mobile broadband [$'FY19]	%GNIpc	Dixed broadband [$'FY19]	%GNIpc
MA	Morocco	35.74	$3,090	$2.10	0.8%	$52.74	20.5%
MZ	Mozambique	29.67	$440	$2.52	6.9%	$69.11	188.5%
MM	Myanmar	53.37	$1,310	$0.66	0.6%	$37.56	34.4%
NA	Namibia	2.53	$5,250	$11.02	2.5%	$383.83	87.7%
NR	Nauru	0.01	$11,240	$28.13	3.0%	N/A	N/A
NP	Nepal	29.30	$960	$2.25	2.8%	$16.47	20.6%
NL	Netherlands	17.13	$51,260	$10.97	0.3%	$59.23	1.4%
NZ	New Zealand	4.79	$40,820	$11.54	0.3%	$58.77	1.7%
NI	Nicaragua	6.22	$2,030	$9.20	5.4%	$63.84	37.7%
NE	Niger	21.48	$380	$3.20	10.1%	$118.80	375.2%
NG	Nigeria	190.89	$1,960	$2.74	1.7%	$84.16	51.5%
NO	Norway	5.28	$80,790	$14.53	0.2%	$73.92	1.1%
OM	Oman	4.64	$15,110	$13.44	1.1%	$150.63	12.0%
PK	Pakistan	197.02	$1,580	$1.79	1.4%	$24.87	18.9%
PW	Palau	0.02	$16,910	$8.34	0.6%	N/A	N/A
PS	Palestine	4.68	$3,710	$2.06	0.7%	$31.79	10.3%
PA	Panama	4.10	$14,370	$7.78	0.6%	$108.38	9.1%
PG	Papua New Guinea	8.25	$2,530	$15.60	7.4%	$571.67	271.1%
PY	Paraguay	6.81	$5,680	$8.06	1.7%	$210.83	44.5%
PE	Peru	32.17	$6,530	$5.04	0.9%	$45.94	8.4%
PH	Philippines	104.92	$3,830	$5.62	1.8%	$53.14	16.6%
PL	Poland	37.97	$14,150	$2.05	0.2%	$18.27	1.5%
PT	Portugal	10.30	$21,680	$12.02	0.7%	$61.15	3.4%
PR	Puerto Rico	3.33	$21,100	$14.93	0.8%	$63.45	3.6%
QA	Qatar	2.64	$61,190	$4.62	0.1%	$140.58	2.8%
RO	Romania	19.58	$11,290	$3.80	0.4%	$14.42	1.5%
RU	Russia	144.50	$10,230	$1.53	0.2%	$9.77	1.1%
RW	Rwanda	12.21	$780	$2.80	4.3%	N/A	N/A
KN	Saint Kitts and Nevis	0.06	$18,640	$15.05	1.0%	$74.71	4.8%
LC	Saint Lucia	0.18	$9,460	$9.78	1.2%	$71.02	9.0%
WS	Samoa	0.20	$4,190	$19.65	5.6%	N/A	N/A
ST	Sao Tome and Principe	0.20	$1,890	$5.33	3.4%	N/A	N/A
SA	Saudi Arabia	32.94	$21,540	$6.08	0.3%	$95.72	5.3%
SN	Senegal	15.85	$1,410	$3.48	3.0%	N/A	N/A
RS	Serbia	7.02	$6,390	$4.33	0.8%	$19.24	3.6%
SC	Seychelles	0.10	$15,600	$19.55	1.5%	$122.63	9.4%
SL	Sierra Leone	7.56	$500	$9.31	22.3%	$135.62	325.5%
SG	Singapore	5.61	$58,770	$4.93	0.1%	$50.43	1.0%
SK	Slovakia	5.44	$18,330	$6.10	0.4%	$21.62	1.4%
SI	Slovenia	2.07	$24,840	$5.26	0.3%	$31.72	1.5%
SB	Solomon Islands	0.61	$2,000	$15.48	9.3%	N/A	N/A
ZA	South Africa	56.72	$5,750	$7.83	1.6%	$55.25	11.5%

Continued on next page

Table A.1 Cost of mobile and fixed broadband data plans – cont. from previous page

ISO2	Name	Popul. [million]	GNIpc [$'FY19]	Mobile broadband [$'FY19]	%GNIpc	Dixed broadband [$'FY19]	%GNIpc
SS	South Sudan	12.58	$460	$5.53	14.4%	N/A	N/A
ES	Spain	46.59	$29,450	$5.95	0.2%	$42.38	1.7%
LK	Sri Lanka	21.44	$4,060	$0.78	0.2%	$5.65	1.7%
VC	St. Vincent	0.11	$7,940	$11.09	1.7%	N/A	N/A
SD	Sudan	40.53	$1,560	$0.68	0.5%	$35.86	27.6%
SR	Suriname	0.56	$4,990	$16.52	4.0%	N/A	N/A
SE	Sweden	10.06	$55,040	$6.22	0.1%	$55.99	1.2%
CH	Switzerland	8.45	$83,580	$21.56	0.3%	$80.00	1.1%
TJ	Tajikistan	8.92	$1,010	$3.68	4.4%	$29.16	34.6%
TZ	Tanzania	57.31	$1,020	$4.37	5.1%	$181.80	213.9%
TH	Thailand	69.04	$6,610	$6.01	1.1%	$25.58	4.6%
TL	Timor-Leste	1.30	$1,820	$5.29	3.5%	N/A	N/A
TG	Togo	7.80	$650	$9.80	18.1%	N/A	N/A
TO	Tonga	0.11	$4,300	$25.52	7.1%	N/A	N/A
TT	Trinidad and Tobago	1.37	$16,240	$29.23	2.2%	$60.77	4.5%
TN	Tunisia	11.53	$3,500	$2.87	1.0%	$24.28	8.3%
TR	Turkey	80.75	$10,380	$3.95	0.5%	$15.96	1.8%
TM	Turkmenistan	5.76	$6,740	$17.76	3.2%	$76.15	13.6%
TV	Tuvalu	0.01	$5,430	$14.37	3.2%	N/A	N/A
UG	Uganda	42.86	$620	$4.36	8.4%	N/A	N/A
UA	Ukraine	44.83	$2,660	$2.70	1.2%	$5.00	2.3%
AE	United Arab Emirates	9.40	$41,010	$15.96	0.5%	$157.10	4.6%
GB	United Kingdom	66.02	$41,340	$9.65	0.3%	$39.58	1.1%
US	USA	325.15	$62,850	$18.46	0.4%	$67.69	1.3%
UY	Uruguay	3.46	$15,650	$3.93	0.3%	$46.50	3.6%
UZ	Uzbekistan	32.39	$2,020	$3.27	1.9%	$21.26	12.6%
VU	Vanuatu	0.28	$2,970	$6.88	2.8%	$138.54	56.0%
VN	Vietnam	95.54	$2,400	$2.15	1.1%	$69.63	34.8%
YE	Yemen	28.25	$960	$15.73	19.7%	$22.17	27.7%
ZM	Zambia	17.09	$1,430	$4.20	3.5%	N/A	N/A
ZW	Zimbabwe	16.53	$1,790	$19.40	13.0%	$128.71	86.3%

Appendix B

Summary of results for experiments using Bayesian Optimization in single- and multi-objective problems

This section presents tables with numerical values of the results of the experiments to assess the feasibility of Bayesian optimization as a method to conduct tradespace exploration for single- and multi-objective problems. For each problem and optimization method, the minimum function value attained (or maximum hypervolume for multi-objective problems), the median value across the (*med.*), the standard deviation (*std.*) across all 20 runs are reported, as well as the relative error between the median value and the optimal ($\Delta opt.$) value:

$$\Delta opt. = \frac{med. - opt.}{opt.} \cdot 100 \qquad [\%] \tag{B.1}$$

B.1 Single-objective problems

Table B.1: Summary of results for the single-objective combining problem.

Problem	Metric	RS	SMAC	HyperOpt	BOCS	GA-EAS	GA-$\mu+\lambda$	GA-μ,λ	BO-TREE	BO-TREE-RT	BO-ARBF	BO-RBF	BO-MT 5/2	BO-MT 3/2	BO-RBFN
SNW-area	min.	-16.2	-16.2	-16.2	-16.2	-16.2	-16.2	-16.1	-16.2	-16.2	-16.2	-16.2	-16.2	-16.2	-16.2
	med.	-15.4	-15.6	**-16.1**	-15.5	-14.9	-14.7	-14.5	-15.5	-15.6	-15.8	-15.1	-15.4	-15.5	-14.8
	std.	0.700	0.500	0.700	0.800	1.20	1.00	1.40	0.400	0.600	1.00	1.10	0.900	0.800	1.00
	Δopt. [%]	4.90	3.70	0.600	4.30	8.00	9.30	10.5	4.30	3.70	2.50	6.80	4.90	4.30	8.60
SNW_thorughput	min.	-14.7	-14.7	-14.7	-14.7	-14.7	-14.7	-14.7	-14.7	-14.7	-14.7	-14.7	-14.7	-14.7	-14.7
	med.	-12.6	-13.1	-13.2	**-14.7**	-11.8	-11.5	-11.4	-13.4	-13.2	-12.3	-13.7	-13.1	-12.9	-13.7
	std.	1.20	0.900	1.20	0.000	1.20	1.30	1.50	1.40	1.10	1.20	1.00	1.40	1.30	1.00
	Δopt. [%]	14.3	10.9	10.2	0.000	19.7	21.8	22.4	8.80	10.2	16.3	6.80	10.9	12.2	6.80
NoC-energy	min.	-9.97	-9.97	-9.97	-9.97	-9.97	-9.97	-9.97	-9.97	-9.97	-9.97	-9.97	-9.97	-9.97	-9.97
	med.	-9.86	**-9.97**	**-9.97**	**-9.97**	-9.84	-9.80	-9.82	**-9.97**	**-9.97**	**-9.97**	**-9.97**	**-9.97**	**-9.97**	**-9.97**
	std.	0.070	0.070	0.050	0.050	0.230	0.420	0.290	0.000	0.000	0.000	0.000	0.160	0.160	0.290
	Δopt. [%]	1.10	0.000	0.000	0.000	1.30	1.70	1.50	0.000	0.000	0.000	0.000	0.000	0.000	0.000
NoC-runtime	min.	-5.12	-5.12	-5.12	-5.12	-5.12	-5.12	-5.12	-5.12	-5.12	-5.12	-5.12	-5.12	-5.12	-5.12
	med.	**-5.12**	**-5.12**	**-5.12**	**-5.12**	**-5.12**	**-5.12**	**-5.12**	**-5.12**	**-5.12**	**-5.12**	**-5.12**	**-5.12**	**-5.12**	**-5.12**
	std.	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	Δopt. [%]	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
SW-SW-LLVM-perf	min.	-269	-269	-270	-270	-270	-269	-269	-270	-270	-270	-270	-270	-270	-270
	med.	-267	-267	-267	**-270**	-268	-262	-259	-268	-267	**-270**	**-270**	-269	-269	**-270**
	std.	2.70	2.50	4.30	1.00	5.40	5.90	6.10	1.10	1.90	1.20	0.800	3.00	3.60	0.800
	Δopt. [%]	1.20	1.20	1.10	0.000	0.700	3.10	3.90	0.700	1.10	0.000	0.000	0.300	0.300	0.000
SW-SW-LLVM-mem	min.	-29.0	-29.0	-29.0	-29.0	-29.0	-29.0	-29.0	-29.0	-29.0	-29.0	-29.0	-29.0	-29.0	-29.0
	med.	**-29.0**	**-29.0**	**-29.0**	**-29.0**	**-29.0**	**-29.0**	**-29.0**	**-29.0**	**-29.0**	**-29.0**	**-29.0**	**-29.0**	**-29.0**	**-29.0**
	std.	0.000	0.000	0.000	0.000	0.200	0.000	0.300	0.000	0.000	0.200	0.000	0.000	0.000	0.000
	Δopt. [%]	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
rs-6d-c3-thr	min.	-222000	-232000	-223000	-222000	-232000	-222000	-222000	-222000	-232000	-232000	-232000	-232000	-232000	-232000
	med.	-214000	-217500	-216000	-219000	-215000	-205500	-199500	-219000	-219000	-217500	-219000	**-220000**	**-220000**	**-220000**
	std.	7687	5132	12110	4260	43365	47711	41180	2765	29682	42985	40661	38658	33420	32530
	Δopt. [%]	7.80	6.20	6.90	5.60	7.30	11.4	14.0	5.60	5.60	6.20	5.60	5.20	5.20	5.20
rs-6d-c3-lat	min.	-34733	-34733	-33989	-32738	-34733	-34733	-34733	-34733	-34733	-34733	-34733	-34733	-34733	-34733
	med.	-30856	-32430	-30843	-32507	-31396	-27074	-21488	**-33486**	-32773	-32773	-32658	-32658	-33147	**-33486**
	std.	2730	1988	3019	895	11547	11298	11397	1558	5045	5290	8237	9071	3027	7211
	Δopt. [%]	11.2	6.60	11.2	6.40	9.60	22.1	38.1	3.60	5.60	5.60	6.00	6.00	4.60	3.60
sol-6d-c2_0	min.	-114000	-115000	-115000	-115000	-115000	-104000	-113000	-115000	-115000	-115000	-115000	-115000	-115000	-115000
	med.	-81163	-98084	-99278	-107000	-76306	-68430	-69996	-114000	-111500	-112000	-114000	-108500	**-114500**	-115000
	std.	13094	14729	15393	21257	19809	20066	19844	3554	13527	19040	6008	17628	11970	15889
	Δopt. [%]	29.4	14.7	13.7	7.00	33.6	40.5	39.1	0.900	3.00	2.60	0.900	5.70	0.400	0.000
sol-6d-c2_1	min.	-40328	N/A	-36025	-34303	-40328	-40328	-34395	N/A	N/A	N/A	N/A	N/A	N/A	N/A
	med.	-29394	N/A	-26865	-25933	**-30406**	-22172	-14833	N/A	N/A	N/A	N/A	N/A	N/A	N/A
	std.	6379	N/A	4424	7947	9253	10393	8958	N/A	N/A	N/A	N/A	N/A	N/A	N/A
	Δopt. [%]	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
wc+sol-3d-c4-thr	min.	-63734	-63734	-63734	-63113	-63734	-63734	-63734	-63734	-63734	-63734	-63734	-63734	-63734	-63734
	med.	-61058	-63424	-63113	-46272	-60713	-37968	-37968	**-63734**	**-63734**	**-63734**	**-63734**	**-63734**	**-63734**	**-63734**
	std.	8815	4786	7675	5147	12894	12360	14249	3220	3225	0.000	0.000	0.000	0.000	0.000
	Δopt. [%]	4.20	0.500	1.00	27.4	4.70	40.4	40.4	0.000	0.000	0.000	0.000	0.000	0.000	0.000
wc+sol-3d-c4-lat	min.	-93904	-93904	-93904	-93904	-93904	-93904	-93904	-93904	-93904	-93904	-93904	-93904	-93904	-93904
	med.	-85576	**-93904**	-90593	**-93904**	-84021	-75359	-77432	**-93904**	**-93904**	**-93904**	**-93904**	**-93904**	**-93904**	**-93904**
	std.	4015	4148	4565	17816	29862	30558	33217	0.000	4413	3682	3311	3652	4283	0.000
	Δopt. [%]	8.90	0.000	3.50	0.000	10.5	19.7	17.5	0.000	0.000	0.000	0.000	0.000	0.000	0.000
wc-5d-c5-thr	min.	-20536	-20591	-20536	-20415	-20591	-20534	-20534	-20591	-20591	-20591	-20591	-20591	-20591	-20591
	med.	-19080	-19322	-20084	-18972	-18646	-16726	-16825	-18972	-18972	-20536	-18972	**-20564**	-20543	-19753
	std.	1112	832	1159	449	2051	1788	1696	773	912	800	1136	940	941	1345
	Δopt. [%]	7.30	6.20	2.50	7.90	9.40	18.8	18.3	7.90	7.90	0.300	7.90	0.100	0.200	4.10
wc-5d-c5-lat	min.	-405	-405	-405	-405	-405	-401	-390	-405	-405	-405	-405	-405	-405	-405
	med.	-380	**-405**	-385	**-405**	-384	-375	-355	**-405**	**-405**	**-405**	**-405**	-385	-388	**-405**
	std.	9.20	9.30	11.4	1.90	55.4	57.1	62.8	1.40	28.2	6.20	0.800	10.2	10.9	27.6
	Δopt. [%]	6.10	0.000	5.00	0.000	5.10	7.40	12.4	0.000	0.000	0.000	0.000	5.00	4.10	0.000
wc-c1-3d-c1-thr	min.	-23075	-22887	-23075	-22124	-23075	-22799	-23075	-23075	-23075	-23075	-23075	-23075	-23075	-23075
	med.	-22119	-22213	-22255	-20391	-22213	-21504	-21677	-22887	-21784	-22799	-22981	**-23075**	-22981	-22323
	std.	624	484	620	810	795	861	990	362	759	689	588	566	662	812
	Δopt. [%]	4.10	3.70	3.60	11.6	3.70	6.80	6.10	0.800	5.60	1.20	0.400	0.000	0.400	3.30
wc-c1-3d-c1-lat	min.	-9421	-9421	-9421	-9421	-9421	-9421	-8358	-9421	-9421	-9421	-9421	-9421	-9421	-9421
	med.	-7769	-7754	-7825	**-9421**	-952	-951	-852	**-9421**	-8029	-8112	-8126	-8194	-8358	-9260
	std.	2677	2677	2630	297	3748	3470	3322	325	3338	3491	4105	3643	3659	4090
	Δopt. [%]	17.5	17.7	16.9	0.000	89.9	89.9	91.0	0.000	14.8	13.9	13.7	13.0	11.3	1.70
wc-c3-3d-c1-thr	min.	-11930	-11930	-11930	-10985	-11224	-11048	-11014	-11930	-11224	-11930	-11930	-11930	-11930	-11930
	med.	-10884	-11076	-11030	-10818	-10847	-10612	-10589	-11047	-10934	-11074	**-11160**	-11107	-11107	-11141
	std.	376	349	345	108	378	277	256	405	234	451	382	278	329	322
	Δopt. [%]	8.80	7.20	7.50	9.30	9.10	11.0	11.2	7.40	8.30	7.20	6.50	6.90	6.90	6.60
wc-c3-3d-c1-lat	min.	-53482	-57645	-55522	-53482	-55522	-55522	-53482	-57645	-57645	-57645	-57645	-57645	-57645	-57645
	med.	-47772	-52823	-50570	-52823	-48962	-46120	-40744	-54559	-52776	-53596	**-55522**	-53596	-53596	**-55522**
	std.	6675	2954	2861	3974	20689	19737	17534	1982	14052	16430	23031	20829	20111	23020
	Δopt. [%]	17.1	8.40	12.3	8.40	15.1	20.0	29.3	5.40	8.40	7.00	3.70	7.00	7.00	3.70

Table B.2: Summary of results for the single-objective downselecting problem.

Problem	Metric	RS	SMAC	BOCS	HyperOpt	GA-EAS	GA-$\mu + \lambda$	GA-μ, λ	BO-TREE	BO-ARBF	BO-RBF	BO-MT 5/2	BO-MT 3/2	BO-RBFN
NK(N=10, K=1)	min.	0.356	0.332	0.332	0.332	0.332	0.332	0.332	0.332	0.332	0.332	0.332	0.332	0.332
	med.	0.390	**0.332**	**0.332**	0.366	0.344	0.389	0.398	**0.332**	**0.332**	**0.332**	0.363	0.356	**0.332**
	std.	0.020	0.023	0.000	0.024	0.020	0.036	0.030	0.000	0.000	0.000	0.032	0.035	0.000
	Δopt. [%]	17.5	0.000	0.000	10.2	3.60	17.2	19.9	0.000	0.000	0.000	9.30	7.20	0.000
NK(N=10, K=2)	min.	0.356	0.356	0.356	0.368	0.356	0.356	0.356	0.356	0.356	0.356	0.356	0.356	0.356
	med.	0.396	0.368	**0.356**	0.387	0.380	0.387	0.408	0.368	**0.356**	**0.356**	0.391	0.395	**0.356**
	std.	0.021	0.016	0.006	0.021	0.029	0.023	0.032	0.011	0.010	0.014	0.019	0.022	0.000
	Δopt. [%]	11.2	3.40	0.000	8.70	6.70	8.70	14.6	3.40	0.000	0.000	9.80	11.0	0.000
NK(N=10, K=5)	min.	0.242	0.242	0.277	0.242	0.242	0.242	0.242	0.242	0.242	0.242	0.242	0.242	0.242
	med.	0.280	0.279	0.281	0.281	0.279	0.308	0.311	0.277	**0.269**	0.277	0.277	0.277	0.257
	std.	0.024	0.018	0.012	0.027	0.018	0.028	0.027	0.011	0.014	0.014	0.016	0.027	0.016
	Δopt. [%]	15.7	15.3	16.1	16.1	15.3	27.3	28.5	14.5	11.2	14.5	14.5	14.5	6.20
NK(N=15, K=1)	min.	0.278	0.266	0.266	0.278	0.267	0.266	0.281	0.266	0.266	0.266	0.276	0.272	0.266
	med.	0.311	0.291	**0.266**	0.304	0.292	0.304	0.313	0.276	**0.266**	**0.266**	0.299	0.305	**0.266**
	std.	0.015	0.016	0.002	0.018	0.018	0.021	0.019	0.013	0.011	0.006	0.021	0.023	0.012
	Δopt. [%]	17.0	9.40	0.000	14.3	9.80	14.3	17.7	3.80	0.000	0.000	12.4	14.7	0.000
NK(N=15, K=2)	min.	0.261	0.256	0.225	0.255	0.225	0.225	0.273	0.225	0.225	0.225	0.225	0.225	0.225
	med.	0.305	0.290	0.255	0.298	0.296	0.310	0.323	0.256	0.280	**0.248**	0.298	0.285	**0.248**
	std.	0.022	0.019	0.023	0.020	0.025	0.028	0.022	0.024	0.023	0.025	0.034	0.030	0.020
	Δopt. [%]	35.4	28.7	13.2	32.3	31.4	37.6	43.4	13.6	24.3	10.1	32.3	26.5	10.1
NK(N=15, K=5)	min.	0.277	0.219	0.272	0.272	0.219	0.272	0.264	0.219	0.219	0.219	0.260	0.260	0.219
	med.	0.344	0.315	0.319	0.316	0.331	0.342	0.346	0.299	**0.296**	**0.296**	0.320	0.316	0.313
	std.	0.027	0.028	0.017	0.026	0.035	0.034	0.048	0.029	0.028	0.034	0.035	0.034	0.040
	Δopt. [%]	57.4	44.1	45.9	44.5	51.4	56.4	58.3	36.8	35.4	35.4	46.4	44.5	43.2
NK(N=20, K=1)	min.	0.318	0.314	0.284	0.305	0.291	0.320	0.326	0.288	0.284	0.284	0.290	0.290	0.284
	med.	0.368	0.336	**0.284**	0.347	0.330	0.359	0.369	0.326	0.311	0.288	0.383	0.374	0.312
	std.	0.021	0.020	0.006	0.022	0.019	0.019	0.023	0.026	0.021	0.013	0.029	0.026	0.016
	Δopt. [%]	25.4	14.5	-3.20	18.3	12.5	22.4	25.8	11.1	6.00	-1.80	30.5	27.5	6.30
NK(N=20, K=2)	min.	0.299	0.276	0.247	0.261	0.267	0.326	0.315	0.255	0.255	0.247	0.306	0.306	0.247
	med.	0.364	0.329	0.265	0.338	0.325	0.372	0.378	0.303	0.273	**0.247**	0.376	0.375	0.255
	std.	0.027	0.025	0.024	0.030	0.029	0.029	0.025	0.025	0.023	0.019	0.027	0.026	0.015
	Δopt. [%]	31.1	18.5	-4.50	21.7	17.1	34.0	36.2	9.10	-1.70	-11.0	35.4	35.1	-8.20
NK(N=20, K=5)	min.	0.278	0.245	0.218	0.246	0.269	0.286	0.278	0.201	0.247	0.201	0.247	0.247	0.201
	med.	0.340	0.319	0.313	0.332	0.331	0.358	0.358	0.299	0.292	**0.269**	0.332	0.329	0.308
	std.	0.030	0.029	0.035	0.039	0.036	0.034	0.033	0.036	0.035	0.046	0.034	0.034	0.047
	Δopt. [%]	44.3	35.4	32.9	40.9	40.5	52.0	52.0	26.9	24.0	14.2	40.9	39.7	30.8
NK(N=25, K=1)	min.	0.300	0.264	0.241	0.279	0.255	0.292	0.282	0.260	0.241	0.241	0.300	0.297	0.241
	med.	0.349	0.314	0.256	0.317	0.301	0.339	0.339	0.294	0.262	**0.242**	0.345	0.347	0.254
	std.	0.024	0.022	0.009	0.019	0.024	0.034	0.036	0.019	0.022	0.008	0.035	0.034	0.007
	Δopt. [%]	29.3	16.3	-5.20	17.4	11.5	25.6	25.6	8.90	-2.90	-10.3	27.8	28.6	-5.90
NK(N=25, K=2)	min.	0.311	0.286	0.273	0.295	0.289	0.308	0.284	0.270	0.262	0.258	0.295	0.311	0.259
	med.	0.366	0.333	0.295	0.328	0.325	0.360	0.360	0.311	0.285	**0.270**	0.363	0.367	0.269
	std.	0.019	0.019	0.020	0.018	0.020	0.026	0.034	0.023	0.020	0.007	0.036	0.025	0.008
	Δopt. [%]	27.9	16.4	3.10	14.6	13.6	25.8	25.8	8.70	-0.400	-5.60	26.9	28.3	-6.00
NK(N=25, K=5)	min.	0.309	0.306	0.306	0.278	0.277	0.277	0.277	0.288	0.230	0.219	0.276	0.276	0.228
	med.	0.359	0.329	0.342	0.325	0.351	0.359	0.372	0.344	0.311	**0.278**	0.360	0.368	0.273
	std.	0.020	0.022	0.022	0.028	0.030	0.042	0.046	0.026	0.036	0.033	0.042	0.043	0.034
	Δopt. [%]	34.4	23.1	28.0	21.6	31.4	34.4	39.2	28.7	16.4	4.00	34.7	37.7	2.20

Table B.3: Summary of results for the single-objective assigning problem.

Problem	Metric	RS	SMAC	HyperOpt	GA-EAS	GA-$\mu+\lambda$	GA-μ,λ	BO-TREE	BO-TREE-FT	BO-ARBF	BO-RBF	BO-MT 5/2	BO-MT 3/2	BO-RBFN
GAP(par515-1)	min.	256	245	249	244	258	256	251	247	240	240	250	N/A	240
	med.	267	251	257	252	264	267	253	252	**240**	**240**	257	N/A	242
	std.	4.70	3.10	3.20	4.40	5.70	7.20	2.80	3.00	0.000	0.000	6.20	N/A	3.10
	Δopt. [%]	11.5	4.60	7.30	5.00	10.2	11.5	5.60	5.20	0.000	0.000	7.30	N/A	0.800
GAP(par515-4)	min.	272	253	263	252	262	267	254	256	251	251	259	N/A	251
	med.	277	263	270	261	277	280	262	261	**251**	**251**	268	N/A	255
	std.	2.90	4.20	4.00	5.80	6.80	7.10	4.50	2.60	0.000	0.000	7.80	N/A	3.30
	Δopt. [%]	10.6	4.80	7.80	4.00	10.6	11.8	4.40	4.00	0.000	0.000	6.80	N/A	1.60
GAP(par520-2)	min.	292	256	277	259	282	283	268	265	248	248	275	N/A	248
	med.	306	280	290	273	302	303	282	280	**248**	**248**	295	N/A	255
	std.	6.90	10.1	9.40	7.50	11.4	12.3	7.20	9.20	2.40	0.000	11.2	N/A	7.80
	Δopt. [%]	23.6	12.9	16.9	10.1	22.0	22.2	13.7	13.1	0.000	0.000	19.2	N/A	2.80
GAP(par824-3)	min.	418	402	402	387	403	390	401	397	372	372	402	N/A	379
	med.	429	412	414	398	419	426	413	411	**372**	**372**	417	N/A	391
	std.	4.20	4.70	5.70	7.30	9.00	11.4	4.70	5.40	0.200	0.000	9.40	N/A	7.00
	Δopt. [%]	15.3	10.9	11.3	7.00	12.8	14.5	11.2	10.5	0.000	0.000	12.2	N/A	5.10
NK(N=15, K=1, L=4)	min.	0.279	0.182	0.196	0.144	0.202	0.202	0.195	0.158	0.137	0.137	0.234	N/A	0.137
	med.	0.339	0.230	0.272	0.241	0.330	0.321	0.231	0.232	0.175	**0.151**	0.311	N/A	0.149
	std.	0.028	0.027	0.036	0.054	0.052	0.039	0.021	0.038	0.035	0.019	0.034	N/A	0.017
	Δopt. [%]	50.9	2.40	21.1	7.30	46.9	42.9	2.90	3.30	-22.1	-32.8	38.5	N/A	-33.7
NK(N=15, K=2, L=4)	min.	0.250	0.170	0.228	0.193	0.240	0.247	0.211	0.151	0.154	0.150	0.204	N/A	0.143
	med.	0.317	0.250	0.285	0.265	0.306	0.322	0.256	0.255	0.216	**0.165**	0.286	N/A	0.186
	std.	0.022	0.029	0.021	0.040	0.034	0.037	0.025	0.040	0.037	0.015	0.045	N/A	0.030
	Δopt. [%]	44.0	13.6	29.5	20.4	39.0	46.3	16.3	15.9	-1.90	-25.0	29.9	N/A	-15.5
NK(N=15, K=4, L=4)	min.	0.246	0.186	0.224	0.231	0.242	0.280	0.227	0.203	0.170	0.160	0.177	N/A	0.159
	med.	0.316	0.259	0.282	0.276	0.309	0.327	0.283	0.282	0.241	**0.214**	0.277	N/A	0.209
	std.	0.028	0.038	0.023	0.026	0.031	0.031	0.032	0.032	0.035	0.026	0.041	N/A	0.043
	Δopt. [%]	63.8	34.2	46.2	43.0	60.2	69.5	46.7	46.2	24.9	10.9	43.6	N/A	8.30
NK(N=15, K=1, L=5)	min.	0.212	0.104	0.144	0.141	0.204	0.149	0.133	0.154	0.108	0.104	0.133	N/A	0.104
	med.	0.259	0.178	0.200	0.217	0.269	0.266	0.172	0.183	0.133	**0.108**	0.222	N/A	0.117
	std.	0.027	0.031	0.034	0.045	0.047	0.053	0.026	0.024	0.027	0.004	0.050	N/A	0.020
	Δopt. [%]	95.7	34.5	51.1	64.0	103	101	30.0	38.3	0.500	-18.4	67.7	N/A	-11.6
NK(N=15, K=2, L=5)	min.	0.183	0.146	0.168	0.128	0.162	0.162	0.137	0.118	0.106	0.118	0.156	N/A	0.106
	med.	0.275	0.204	0.232	0.240	0.267	0.278	0.204	0.207	0.161	**0.150**	0.214	N/A	0.147
	std.	0.031	0.036	0.035	0.041	0.058	0.052	0.042	0.033	0.039	0.021	0.042	N/A	0.027
	Δopt. [%]	112	57.5	79.1	85.3	106	114	57.5	59.8	24.3	15.8	65.2	N/A	13.5
NK(N=15, K=5, L=5)	min.	0.211	0.158	0.166	0.177	0.214	0.214	0.194	0.190	0.162	0.110	0.130	N/A	0.199
	med.	0.265	0.255	0.258	0.257	0.274	0.286	0.244	0.241	0.211	**0.202**	0.209	N/A	0.253
	std.	0.025	0.039	0.032	0.038	0.035	0.037	0.029	0.021	0.031	0.040	0.043	N/A	0.023
	Δopt. [%]	79.9	73.1	75.2	74.5	86.0	94.2	65.7	63.6	43.3	37.1	41.9	N/A	71.8

Table B.4: Summary of results for the single-objective partitioning problem.

Problem	Metric	RS	SMAC	HyperOpt	GA-EAS	GA-$\mu + \lambda$	GA-μ, λ	BO-TREE	BO-TREE-FT	BO-RBF	BO-ARBF	BO-MT 5/2	BO-MT 3/2	BO-RBFN
PAR(N=7, K=1)	min.	-1.26	-1.26	-1.26	-1.26	-1.26	-1.26	-1.26	-1.26	-1.26	-1.26	-1.26	-1.26	-1.26
	med.	**-1.26**	**-1.26**	**-1.26**	**-1.26**	-0.900	-0.900	**-1.26**	**-1.26**	**-1.26**	**-1.26**	**-1.26**	**-1.26**	**-1.26**
	std.	0.080	0.000	0.000	0.210	0.210	0.280	0.000	0.000	0.000	0.000	0.110	0.080	0.000
	Δopt. [%]	0.000	0.000	0.000	0.000	28.6	28.6	0.000	0.000	0.000	0.000	0.000	0.000	0.000
PAR(N=7, K=1)	min.	-2.29	-2.29	-2.29	-2.29	-2.29	-2.29	-2.29	-2.29	-2.29	-2.29	-2.29	-2.29	-1.66
	med.	**-2.29**	**-2.29**	-2.22	**-2.29**	**-2.29**	**-2.29**	**-2.29**	**-2.29**	**-2.29**	**-2.29**	**-2.29**	**-2.29**	-0.130
	std.	0.020	0.020	0.030	0.180	0.160	0.240	0.000	0.000	0.020	0.020	0.020	0.020	0.830
	Δopt. [%]	0.000	0.000	3.10	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	94.3
PAR(N=10, K=1)	min.	-1.70	-2.17	-2.17	-1.70	-1.70	-1.70	-1.70	-1.70	-1.70	-1.70	-1.70	-1.70	-1.70
	med.	-1.67	**-2.17**	**-2.17**	-1.70	-1.70	-1.67	-1.70	-1.70	-1.70	-1.70	-1.70	-1.70	-1.70
	std.	0.034	0.000	0.264	0.162	0.329	0.352	0.000	0.000	0.027	0.024	0.298	0.152	0.000
	Δopt. [%]	2.00	-27.4	-27.4	0.000	0.000	2.00	0.000	0.000	0.000	0.000	0.000	0.000	0.000
PAR(N=10, K=2)	min.	-2.39	-2.51	-2.51	-2.39	-2.39	-2.06	-2.39	-2.39	-2.39	-2.39	-2.39	-2.39	-2.39
	med.	-1.74	**-2.08**	**-2.08**	-1.92	-1.20	-1.31	**-2.08**	**-2.08**	**-2.08**	**-2.08**	-1.81	-1.74	-2.39
	std.	0.249	0.216	0.300	0.343	0.406	0.323	0.220	0.166	0.248	0.274	0.383	0.405	0.148
	Δopt. [%]	27.2	12.9	12.9	19.8	49.7	45.0	13.0	12.9	12.9	12.9	24.1	27.2	0.000
PAR(N=10, K=4)	min.	-3.71	-4.42	-4.42	-3.75	-4.01	-4.01	-4.01	-4.01	-4.01	-4.01	-4.01	-4.01	-4.01
	med.	-2.83	**-3.98**	-3.53	-3.30	-2.92	-2.66	-3.23	-3.62	-3.73	-3.44	-2.97	-3.08	-4.01
	std.	0.278	0.170	0.385	0.260	0.433	0.557	0.430	0.345	0.288	0.423	0.399	0.322	0.258
	Δopt. [%]	29.5	0.800	12.0	17.8	27.2	33.7	19.5	9.90	7.10	14.4	26.0	23.3	0.000
PAR(N=15, K=2)	min.	-3.83	-6.51	-5.47	-6.23	-4.20	-4.49	-4.83	-5.48	-6.69	-6.23	-5.84	-6.07	-6.69
	med.	-3.24	-4.87	-4.09	-4.64	-3.20	-3.09	-3.94	-4.30	**-5.92**	-4.89	-4.52	-4.55	**-5.92**
	std.	0.233	0.650	0.525	0.683	0.571	0.514	0.479	0.585	0.560	0.777	0.785	0.777	0.407
	Δopt. [%]	38.9	8.20	22.9	12.6	39.7	41.7	25.7	18.9	-11.7	7.70	14.8	14.2	-11.7
PAR(N=15, K=4)	min.	-4.18	-6.96	-6.09	-6.70	-6.00	-4.29	-6.68	-6.56	-8.24	-7.72	-6.54	-5.93	-8.24
	med.	-3.31	-5.95	-4.49	-5.04	-3.91	-2.99	-4.82	-5.03	**-6.64**	-5.65	-4.71	-4.09	-6.92
	std.	0.357	0.703	0.571	0.786	0.884	0.799	0.836	0.693	1.04	1.12	0.746	0.619	0.893
	Δopt. [%]	38.8	-10.1	16.8	6.70	27.6	44.6	10.7	6.90	-23.0	-4.60	12.8	24.2	-28.2

Table B.5: Summary of results for the single-objective permuting problem.

Problem	Metric	RS	GA-EAS	GA-$\mu + \lambda$	GA-μ, λ	BO-TREE	BO-TREE-RT	BO-RBF	BO-ARBF	BO-MT 5/2	BO-MT 3/2	BO-RBFN
UNI(10)	min.	4.00	0.000	4.00	2.00	0.000	3.00	0.000	0.000	0.000	0.000	0.000
	med.	5.00	2.00	5.00	5.00	2.50	4.00	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**
	std.	0.750	1.37	0.940	1.51	1.07	0.700	0.000	0.000	0.000	0.000	0.000
	Δopt. [%]	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
UNI(15)	min.	9.00	2.00	5.00	8.00	7.00	7.00	0.000	0.000	N/A	N/A	0.000
	med.	**10.0**	5.50	**10.0**	9.50	9.00	9.00	**0.000**	**0.000**	N/A	N/A	**0.000**
	std.	0.560	2.09	1.35	1.11	0.640	0.840	0.000	0.000	N/A	N/A	0.000
	Δopt. [%]	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
UNI(20)	min.	14.0	8.00	13.0	13.0	12.0	12.0	0.000	0.000	N/A	0.000	0.000
	med.	15.0	11.0	15.0	15.0	14.0	15.0	**0.000**	**0.000**	N/A	**0.000**	**0.000**
	std.	0.700	1.61	1.20	1.10	0.900	0.700	0.000	2.29	N/A	0.000	4.41
	Δopt. [%]	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
UNI(30)	min.	23.0	17.0	22.0	23.0	22.0	21.0	0.000	0.000	N/A	N/A	11.0
	med.	25.0	20.0	25.0	26.0	24.0	25.0	**0.000**	15.0	N/A	N/A	22.5
	std.	1.00	1.90	1.30	1.20	0.800	1.20	6.02	3.56	N/A	N/A	3.20
	Δopt. [%]	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
TSP(ulysses16)	min.	9055	8119	8648	8228	8694	8713	7324	7698	N/A	N/A	7039
	med.	9606	8714	9649	9818	9570	9424	**8390**	8586	N/A	N/A	7566
	std.	333	500	502	678	323	403	556	615	N/A	N/A	337
	Δopt. [%]	40.0	27.0	40.7	43.1	39.5	37.4	22.3	25.2	N/A	N/A	10.3
TSP(ulysses16)	min.	1802	1646	1783	1783	1831	1865	1580	1682	N/A	N/A	1126
	med.	2076	1860	2034	1988	2030	2032	**1772**	1834	N/A	N/A	1467
	std.	106	140	137	114	83.0	75.0	114	92.0	N/A	N/A	149
	Δopt. [%]	121	98.5	117	112	116	116	89.1	95.7	N/A	N/A	56.6
TSP(gr21)	min.	4881	4399	4848	4774	4811	5106	4162	4395	N/A	N/A	3623
	med.	5656	5143	5694	5732	5526	5756	**4742**	4926	N/A	N/A	4412
	std.	332	295	316	331	326	231	238	256	N/A	N/A	434
	Δopt. [%]	13.2	2.90	14.0	14.7	10.6	15.2	-5.10	-1.40	N/A	N/A	-11.7
TSP(gr17)	min.	3042	2533	2855	2789	3017	3064	2098	2610	2563	2551	2222
	med.	3319	2989	3315	3498	3424	3278	3034	2920	3016	**2888**	3028
	std.	185	235	196	300	153	132	323	441	285	216	441
	Δopt. [%]	23.3	11.1	23.2	30.0	27.2	21.8	12.7	8.50	12.1	7.30	12.5
TSP(bays29)	min.	4502	3721	3835	4330	4230	4014	3881	3812	N/A	N/A	2949
	med.	4722	4279	4633	4842	4647	4688	4204	**4175**	N/A	N/A	3294
	std.	151	286	315	250	161	218	201	214	N/A	N/A	238
	Δopt. [%]	133	111	129	139	130	132	108	106	N/A	N/A	63.1
FSP(reC27_30_15)	min.	2677	2643	2682	2716	2700	2725	2623	2637	N/A	N/A	2527
	med.	2796	2724	2779	2835	2776	2772	**2658**	2693	N/A	N/A	2580
	std.	44.0	44.0	61.0	51.0	29.0	26.0	24.0	30.0	N/A	N/A	41.0
	Δopt. [%]	4.70	2.00	4.10	6.20	4.00	3.80	-0.400	0.900	N/A	N/A	-3.40
FSP(hel2_20_10)	min.	151	141	149	150	150	147	142	143	146	N/A	140
	med.	153	149	154	155	152	153	**146**	**146**	150	N/A	**146**
	std.	1.60	3.70	2.50	2.40	1.00	2.10	1.90	1.50	2.50	N/A	2.50
	Δopt. [%]	5.90	2.80	6.20	6.90	5.20	5.50	0.700	0.700	3.40	N/A	0.700
FSP(car2_13_4)	min.	7520	7376	7617	7596	7376	7376	7305	7376	7365	7273	7166
	med.	7888	7586	7940	7973	7638	7794	7712	7940	**7505**	7617	7402
	std.	123	220	185	160	147	136	212	279	219	229	236
	Δopt. [%]	6.90	2.80	7.60	8.10	3.60	5.70	4.60	7.60	1.70	3.30	0.400
FSP(car7_7_7)	min.	6643	6643	6590	6681	6590	6590	6590	6590	6590	6590	6590
	med.	6760	6776	6887	6978	**6590**	6681	**6590**	6616	**6590**	**6590**	**6590**
	std.	87.0	115	202	195	19.0	109	23.0	60.0	41.0	52.0	0.000
	Δopt. [%]	2.60	2.80	4.50	5.90	0.000	1.40	0.000	0.400	0.000	0.000	0.000
QAP(nug16a)	min.	1802	1736	1778	1792	1834	1834	1632	1696	1678	1754	1660
	med.	1904	1846	1901	1892	1886	1884	**1695**	1767	1814	1818	1709
	std.	32.0	44.0	46.0	48.0	26.0	16.0	39.0	36.0	44.0	44.0	32.0
	Δopt. [%]	18.3	14.7	18.1	17.5	17.1	17.0	5.30	9.80	12.7	12.9	6.10
QAP(nug21)	min.	2856	2724	2834	2904	2934	2878	2536	2672	2848	2856	2540
	med.	3031	2897	3030	3023	3036	3003	**2786**	2827	2871	2866	2651
	std.	67.0	84.0	91.0	70.0	50.0	55.0	112	68.0	39.0	32.0	70.0
	Δopt. [%]	24.3	18.8	24.3	24.0	24.5	23.2	14.3	16.0	17.8	17.6	8.70
QAP(nug27)	min.	6346	6120	6106	6268	6232	N/A	5948	6020	N/A	N/A	5578
	med.	6496	**6235**	6458	6485	6388	N/A	6247	6255	N/A	N/A	5816
	std.	73.0	112	125	113	73.0	N/A	114	124	N/A	N/A	124
	Δopt. [%]	24.1	19.1	23.4	23.9	22.0	N/A	19.4	19.5	N/A	N/A	11.1
QAP(tho30)	min.	186194	177930	181196	183994	184468	N/A	177768	178890	N/A	N/A	163904
	med.	193839	188318	193382	193274	193029	N/A	**181014**	186957	N/A	N/A	170729
	std.	3506	4239	4747	4137	2937	N/A	3234	4815	N/A	N/A	2908
	Δopt. [%]	29.3	25.6	29.0	28.9	28.7	N/A	20.7	24.7	N/A	N/A	13.9
QAP(tai35a)	min.	2778066	2718748	2739172	2763250	2795622	N/A	2725608	2721784	N/A	2657670	2572362
	med.	2824711	2771789	2818025	2811133	2818290	N/A	2758379	2801022	N/A	**2723840**	2609397
	std.	18147	25653	26895	21075	12339	N/A	21012	27767	N/A	29812	28392
	Δopt. [%]	16.6	14.4	16.4	16.1	16.4	N/A	13.9	15.6	N/A	12.5	7.70

B.2 Multi-objective problems

Table B.6: Summary of results for the single-objective combining problem.

Problem	Metric	RS	GA-EAS	NSGA-II	SPEA-2	BO-TREE	BO-TREE-RT	BO-RBF	BO-MT 3/2	BO-MT 5/2	BO-ARBF
SNW	max.	165	157	158	157	165	165	165	166	166	165
	med.	151	122	140	131	145	145	**152**	148	146	**152**
	std.	7.00	17.6	17.5	15.4	11.3	11.3	11.0	10.5	11.0	11.0
	Δopt. [%]	9.10	26.5	16.1	21.4	13.0	13.0	9.00	11.3	12.4	9.00
NoC	max.	20.5	20.5	20.5	20.5	20.5	20.5	20.5	20.5	20.5	20.5
	med.	20.4	20.4	20.0	20.0	**20.5**	**20.5**	**20.5**	**20.5**	**20.5**	**20.5**
	std.	0.200	0.500	0.600	0.600	0.100	0.100	0.100	0.100	0.100	0.100
	Δopt. [%]	0.300	0.300	2.20	2.20	-0.200	-0.200	-0.200	-0.200	-0.200	-0.200
SW-LLVM	max.	3848	3863	3835	3854	3841	3841	3832	3828	3825	3832
	med.	3804	**3814**	3806	3811	3800	3800	3791	3798	3798	3791
	std.	28.0	60.0	66.0	43.0	23.0	23.0	43.0	35.0	31.0	43.0
	Δopt. [%]	1.50	1.30	1.50	1.40	1.60	1.60	1.90	1.70	1.70	1.90
rs-6d-c3	max.	8.507E+08	8.187E+08	8.756E+08	9.454E+08	8.852E+08	8.852E+08	8.712E+08	8.864E+08	8.506E+08	8.712E+08
	med.	7.293E+08	3.980E+08	6.501E+08	6.730E+08	6.983E+08	6.983E+08	6.744E+08	6.368E+08	7.248E+08	6.744E+08
	std.	1.186E+08	1.685E+08	1.717E+08	1.898E+08	1.308E+08	1.308E+08	1.084E+08	1.353E+08	1.125E+08	1.084E+08
	Δopt. [%]	42.1	68.4	48.4	46.6	44.6	44.6	46.5	49.5	42.5	46.5
sol-6d-c2	max.	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
	std.	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
	Δopt. [%]	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
wc+sol-3d-c4	max.	5.531E+09	5.555E+09	5.511E+09	5.511E+09	5.531E+09	5.531E+09	5.555E+09	5.555E+09	5.555E+09	5.555E+09
	med.	5.158E+09	5.131E+09	4.711E+09	4.805E+09	5.327E+09	5.327E+09	5.082E+09	5.323E+09	5.296E+09	5.082E+09
	std.	4.758E+08	6.758E+08	1.081E+09	9.659E+08	2.768E+08	2.768E+08	4.205E+08	3.552E+08	3.206E+08	4.205E+08
	Δopt. [%]	7.20	7.60	15.2	13.5	4.10	4.10	8.50	4.20	4.70	8.50
wc-5d-c5	max.	4910895	4781693	4916508	4880276	4928457	4928457	4878828	4960184	4949131	4878828
	med.	4727141	4598116	4607532	4629590	4750790	4750790	**4766579**	4714689	4738659	**4766579**
	std.	83980	389630	369936	341764	76098	76098	65080	77053	82546	66693
	Δopt. [%]	6.00	8.50	8.30	7.90	5.50	5.50	5.20	6.20	5.70	5.20
wc-c1-3d-c1	max.	1.017E+07	8636724	9564906	9489882	1.050E+07	1.050E+07	1.054E+07	1.054E+07	1.047E+07	1.054E+07
	med.	9317042	6194995	7373332	8139126	9316117	9316117	9727979	9751330	**9869961**	9757031
	std.	879759	919160	1346215	1286067	860983	860983	403022	385357	459016	397074
	Δopt. [%]	17.4	45.1	34.6	27.8	17.4	17.4	13.7	13.5	12.5	13.5
wc-c3-3d-c1	max.	2.143E+07	1.676E+07	2.186E+07	2.144E+07	2.011E+07	2.011E+07	2.088E+07	2.091E+07	2.082E+07	2.088E+07
	med.	1.764E+07	1.024E+07	1.366E+07	1.369E+07	1.782E+07	1.782E+07	1.775E+07	1.810E+07	1.827E+07	1.775E+07
	std.	2763579	3734067	4725102	4617100	2158320	2158320	1995976	1897032	1970420	1995976
	Δopt. [%]	25.6	56.8	42.4	42.3	24.8	24.8	25.2	23.7	22.9	25.2

Table B.7: Summary of results for the single-objective downselecting problem.

Problem	Metric	RS	GA-EAS	NSGA-II	SPEA-2	BO-TREE	BO-TREE-FT	BO-RBF	BO-MT 3/2	BO-MT 5/2	BO-ARBF
SYN-DWN-01	max.	0.401	0.414	0.415	0.410	0.420	0.420	0.420	0.415	0.415	0.420
	med.	0.377	0.390	0.377	0.384	0.415	0.415	**0.420**	0.397	0.395	**0.420**
	std.	0.011	0.017	0.013	0.013	0.010	0.010	0.006	0.013	0.012	0.006
	Δopt. [%]	10.2	7.10	10.2	8.60	1.20	1.20	0.000	5.50	6.00	0.000
SYN-DWN-02	max.	0.405	0.414	0.416	0.415	0.419	0.419	0.419	0.420	0.416	0.419
	med.	0.390	0.391	0.383	0.389	**0.407**	**0.407**	0.403	0.393	0.392	0.404
	std.	0.009	0.019	0.016	0.014	0.010	0.010	0.013	0.016	0.015	0.013
	Δopt. [%]	8.80	8.50	10.4	9.00	4.80	4.80	5.70	8.10	8.30	5.50
SYN-DWN-03	max.	0.378	0.390	0.386	0.371	0.401	0.401	0.411	0.375	0.384	0.412
	med.	0.347	0.363	0.351	0.349	0.374	0.374	**0.400**	0.354	0.359	**0.400**
	std.	0.014	0.016	0.015	0.012	0.012	0.012	0.011	0.013	0.016	0.013
	Δopt. [%]	15.0	11.1	14.0	14.5	8.40	8.40	2.00	13.3	12.1	2.00
SYN-DWN-04	max.	0.435	0.440	0.422	0.442	0.465	0.465	0.455	0.436	0.437	0.452
	med.	0.386	0.401	0.395	0.394	0.428	0.428	**0.432**	0.380	0.379	**0.432**
	std.	0.017	0.016	0.015	0.020	0.023	0.023	0.018	0.025	0.028	0.020
	Δopt. [%]	24.4	21.5	22.7	22.9	16.2	16.2	15.4	25.6	25.8	15.4
SYN-DWN-05	max.	0.488	0.474	0.477	0.477	0.533	0.533	0.539	0.489	0.500	0.542
	med.	0.436	0.441	0.439	0.429	**0.465**	**0.465**	0.437	0.446	0.449	0.450
	std.	0.023	0.020	0.016	0.020	0.024	0.024	0.037	0.023	0.028	0.036
	Δopt. [%]	21.4	20.5	20.8	22.6	16.2	16.2	21.2	19.6	19.0	18.9
SYN-DWN-06	max.	0.396	0.425	0.449	0.414	0.423	0.423	0.462	0.404	0.411	0.464
	med.	0.377	0.392	0.381	0.378	0.398	0.398	**0.423**	0.378	0.363	0.439
	std.	0.013	0.019	0.026	0.016	0.016	0.016	0.023	0.024	0.021	0.022
	Δopt. [%]	19.5	16.3	18.7	19.3	15.0	15.0	9.70	19.3	22.5	6.30
SYN-DWN-07	max.	0.421	0.400	0.418	0.418	0.417	0.417	0.453	0.405	0.403	0.454
	med.	0.378	0.376	0.372	0.375	0.393	0.393	**0.434**	0.378	0.380	0.428
	std.	0.018	0.014	0.017	0.017	0.014	0.014	0.024	0.018	0.022	0.026
	Δopt. [%]	18.5	19.0	19.8	19.2	15.3	15.3	6.50	18.5	18.1	7.80

Table B.8: Summary of results for the single-objective assigning problem.

Problem	Metric	RS	GA-EAS	NSGA-II	SPEA-2	BO-TREE	BO-TREE-FT	BO-RBF	BO-MT 3/2	BO-MT 5/2	BO-ARBF
SYN-ASG-01	max.	0.482	0.522	0.506	0.489	0.538	0.545	0.585	0.492	0.493	0.573
	med.	0.437	0.464	0.446	0.438	0.478	0.480	**0.549**	0.454	0.436	0.542
	std.	0.021	0.035	0.030	0.034	0.022	0.024	0.048	0.031	0.031	0.033
	Δopt. [%]	29.2	24.8	27.7	29.0	22.5	22.2	11.0	26.4	29.3	12.1
SYN-ASG-02	max.	0.540	0.542	0.503	0.529	0.501	0.503	0.604	0.558	0.524	0.558
	med.	0.446	0.463	0.455	0.461	0.468	0.474	**0.514**	0.461	0.456	0.498
	std.	0.026	0.036	0.025	0.026	0.019	0.021	0.046	0.034	0.032	0.033
	Δopt. [%]	26.8	24.0	25.3	24.4	23.2	22.2	15.7	24.4	25.2	18.3
SYN-ASG-03	max.	0.541	0.582	0.556	0.583	0.578	0.597	0.666	0.563	0.565	0.647
	med.	0.485	0.520	0.491	0.480	0.528	0.535	**0.585**	0.501	0.506	0.582
	std.	0.027	0.036	0.034	0.038	0.021	0.030	0.050	0.027	0.031	0.042
	Δopt. [%]	28.0	22.8	27.1	28.8	21.6	20.6	13.2	25.7	24.9	13.6
SYN-ASG-04	max.	0.556	0.574	0.572	0.556	0.586	0.572	0.624	0.571	0.577	0.641
	med.	0.495	0.493	0.495	0.479	0.493	0.536	**0.549**	0.508	0.503	0.541
	std.	0.032	0.036	0.029	0.034	0.032	0.033	0.040	0.029	0.030	0.049
	Δopt. [%]	26.6	26.9	26.6	29.0	26.9	20.5	18.6	24.7	25.4	19.8
SYN-ASG-05	max.	165194	170119	168226	166805	168847	168759	182796	168446	169430	182477
	med.	159698	163244	161666	161859	164533	164905	**179908**	163399	163496	180533
	std.	2028	2807	2457	2251	2297	2072	1507	3268	3524	1522
	Δopt. [%]	11.2	9.20	10.1	10.0	8.50	8.30	-0.100	9.10	9.10	-0.400
SYN-ASG-06	max.	175715	181062	177576	181513	180411	181641	194853	181817	177168	194963
	med.	171189	175292	173238	173870	174919	175856	**193286**	172217	172266	192748
	std.	2673	3128	2671	2858	2865	2371	1738	3934	3519	1485
	Δopt. [%]	11.2	9.00	10.1	9.80	9.20	8.70	-0.300	10.6	10.6	-0.000
SYN-ASG-07	max.	294974	301284	301268	297291	304860	304206	347414	307648	322368	347752
	med.	281146	289804	286685	287744	291308	290108	**339799**	287167	283865	339203
	std.	5560	5238	6498	6023	6563	7223	7144	11700	12877	7605
	Δopt. [%]	18.0	15.5	16.4	16.1	15.1	15.4	0.900	16.3	17.2	1.10
SYN-ASG-08	max.	407975	420271	428496	412449	416991	420030	459628	417856	415678	463101
	med.	398252	408443	405553	401376	402592	406237	**447196**	404135	403758	444432
	std.	4118	4670	7558	5822	5721	5912	7657	7419	7778	9878
	Δopt. [%]	14.1	11.9	12.5	13.4	13.2	12.4	3.60	12.9	12.9	4.20

Table B.9: Summary of results for the single-objective partitioning problem.

Problem	Metric	RS	GA-EAS	NSGA-II	SPEA-2	BO-TREE	BO-TREE-FT	BO-RBF	BO-MT 3/2	BO-MT 5/2	BO-ARBF
SYN-PAR-01	max.	N/A	11.6	11.6	11.6	11.3	11.6	10.8	10.8	10.8	10.8
	med.	N/A	10.7	10.6	**10.8**	**10.8**	**10.8**	9.78	9.80	10.1	9.82
	std.	N/A	0.740	0.520	0.500	0.490	0.490	0.730	0.590	0.490	0.590
	Δopt. [%]	N/A	8.10	8.60	7.20	7.30	7.60	15.9	15.8	13.5	15.6
SYN-PAR-02	max.	21.8	21.1	22.1	21.8	21.7	22.1	20.7	21.4	20.9	20.5
	med.	20.0	20.0	19.6	**20.3**	19.6	19.5	19.0	18.7	18.7	19.0
	std.	1.20	1.10	1.20	1.20	1.10	1.20	0.900	1.00	0.800	0.800
	Δopt. [%]	9.60	9.60	11.4	8.20	11.4	11.9	14.1	15.5	15.5	14.1
SYN-PAR-03	max.	41.7	44.0	42.8	43.6	41.6	41.2	42.3	41.8	41.7	42.3
	med.	39.6	39.4	40.1	**40.4**	36.9	37.4	39.2	37.9	38.5	39.2
	std.	1.20	2.40	1.30	1.60	2.10	2.10	2.10	2.00	2.20	2.10
	Δopt. [%]	13.4	13.8	12.3	11.6	19.3	18.2	14.2	17.1	15.8	14.2
SYN-PAR-04	max.	70.8	86.5	76.5	80.2	76.8	71.4	75.8	80.8	76.1	75.8
	med.	65.2	**71.2**	64.0	67.6	64.3	64.1	67.0	66.7	67.4	67.0
	std.	2.70	7.50	5.80	5.10	5.40	2.90	3.80	5.00	3.80	3.80
	Δopt. [%]	24.9	18.0	26.3	22.2	26.0	26.2	22.9	23.2	22.4	22.9
SYN-PAR-05	max.	129	137	138	141	129	138	128	129	131	132
	med.	121	**125**	119	120	119	118	118	119	117	119
	std.	5.20	8.00	9.60	11.0	6.20	7.10	7.20	7.70	5.70	6.00
	Δopt. [%]	23.4	20.6	24.5	23.5	24.6	25.2	24.8	24.7	25.6	24.2
SYN-PAR-06	max.	124	132	130	123	122	123	121	123	115	121
	med.	104	**118**	109	111	111	109	110	108	107	111
	std.	6.60	9.80	7.10	6.30	6.90	5.40	6.80	7.70	4.60	5.60
	Δopt. [%]	33.9	25.0	30.8	29.9	29.5	30.8	30.5	31.5	32.4	29.8
SYN-PAR-07	max.	167	179	185	178	179	179	169	173	173	173
	med.	154	159	160	**166**	151	151	155	162	158	152
	std.	7.90	9.90	11.9	9.90	12.1	9.70	8.60	9.50	8.20	9.00
	Δopt. [%]	36.7	34.7	34.1	31.9	38.0	37.9	36.4	33.7	35.3	37.5

Table B.10: Summary of results for the single-objective permuting problem.

Problem	Metric	RS	GA-EAS	NSGA-II	SPEA-2	BO-TREE	BO-TREE-RT	BO-RBF	BO-MT 3/2	BO-MT 5/2	BO-ARBF
SYN-PER-01	max.	18.0	31.0	24.0	23.0	26.0	23.0	41.0	38.0	36.0	41.0
	med.	12.5	17.0	14.5	13.0	15.0	14.0	**33.0**	22.5	17.0	34.0
	std.	2.55	4.72	4.47	3.93	3.70	3.94	4.70	8.50	6.85	4.20
	Δopt. [%]	78.1	70.2	74.6	77.2	73.7	75.4	42.1	60.5	70.2	40.4
SYN-PER-02	max.	24.0	21.0	24.0	22.0	23.0	20.0	56.0	36.0	35.0	44.0
	med.	12.0	15.5	11.5	12.5	14.0	13.0	**28.5**	18.0	18.0	33.5
	std.	3.58	3.30	4.10	4.26	3.94	2.51	8.40	8.10	7.88	6.60
	Δopt. [%]	78.2	71.8	79.1	77.3	74.5	76.4	48.2	67.3	67.3	39.1
SYN-PER-03	max.	19.0	25.0	17.0	18.0	17.0	N/A	25.0	34.0	28.0	33.0
	med.	11.0	**16.0**	12.0	10.5	12.0	N/A	**16.0**	13.5	12.5	16.5
	std.	3.16	4.58	3.13	3.03	2.50	N/A	4.10	5.99	5.56	4.40
	Δopt. [%]	78.8	69.2	76.9	79.8	76.9	N/A	69.2	74.0	76.0	68.3
SYN-PER-04	max.	1.686E+08	1.634E+08	1.633E+08	1.621E+08	1.656E+08	1.646E+08	1.662E+08	1.750E+08	1.688E+08	1.661E+08
	med.	1.583E+08	1.577E+08	1.571E+08	1.575E+08	1.574E+08	1.580E+08	1.589E+08	1.640E+08	1.628E+08	1.590E+08
	std.	4052679	3777528	4124006	2435933	4046382	3534889	4133883	3891700	4456981	5190414
	Δopt. [%]	8.80	9.10	9.50	9.20	9.30	9.00	8.50	5.50	6.20	8.40
SYN-PER-05	max.	247336	248096	248096	246772	241132	242464	255112	259944	249016	255884
	med.	229084	234716	233144	229818	230718	229474	**244040**	238258	235464	238008
	std.	6561	8114	6099	6869	5838	5620	7826	8451	5302	8788
	Δopt. [%]	12.5	10.3	10.9	12.2	11.8	12.3	6.80	9.00	10.0	9.10
SYN-PER-06	max.	3.146E+09	3.110E+09	3.104E+09	3.167E+09	3.095E+09	N/A	3.163E+09	3.200E+09	3.151E+09	3.163E+09
	med.	3.013E+09	3.028E+09	2.998E+09	3.007E+09	3.020E+09	N/A	3.018E+09	3.073E+09	3.058E+09	3.006E+09
	std.	4.203E+07	4.190E+07	3.595E+07	5.094E+07	4.074E+07	N/A	5.370E+07	5.689E+07	4.414E+07	6.145E+07
	Δopt. [%]	6.50	6.10	7.00	6.70	6.30	N/A	6.40	4.70	5.10	6.70

Appendix C

Ranking metrics for optimization methods

This Appendix describes the three ranking metrics used to rank the different optimization methods in Chapters 4 and 5. These methods derive from reference [236] and were originally developed in the context of algorithm selection for classification tasks in machine learning research, which, in turn, is an algorithm performance comparison problem. Within the context of this dissertation, an analogy can be made between the "optimization methods - test problems" and the "classification algorithm - dataset", since it is also one of the goals of this research to select the most appropriate optimization method for different tasks, namely optimizing problems belonging to the assigning, combining, downselecting, partitioning, and permuting patterns.

C.1 Average Ranking

The most straightforward ranking metric considered. For each problem, each optimization method is ordered according to the median of the values of the 20 runs and assigned a rank starting by 1 (if several methods achieve the same value, the minimum of the ranks that would have been assigned to all the tied values is assigned to each method). Let r_m^p be the rank of method m on problem p. Then, the *average rank* is computed for each method as

$$\bar{r}_m = \frac{1}{P}\sum_{p=1}^{P} r_m^p, \qquad \text{(smaller values are better)} \tag{C.1}$$

where P is the number of problems. Then, the final ranking is computed by sorting the average ranks and assigning ranks to the optimization method.

C.2 Success Rate Ratio

This method computes the *success rate* between pairs of optimization methods. The first step is to create a table whose entries are the *success rate ratio* between two methods (m_1 and m_2) for a given problem p, which is defined as:

$$\mathrm{SRR}^p_{m_1,m_2} = \frac{1 - \Delta^p_{m_1}}{1 - \Delta^p_{m_2}} \tag{C.2}$$

where Δ^p_m is the percentage error between the value obtained by the method m and the optimal solution of problem p (which can be known *a priori* or be estimated with a large number of evaluations using a genetic algorithm). If the value of Δ^p_m is superior to 100%, it is limited to 1% (as otherwise the sign of SRR will be inverted).

Then, the *pairwise mean success rate ratio* between each pair of methods, which estimates the general goodness of method m_1 when compared with method m_2, is computed as

$$\mathrm{SRR}_{m_1,m_2} = \frac{1}{P}\sum_{p=1}^{P} \mathrm{SRR}^p_{m_1,m_2}. \tag{C.3}$$

Finally, the *overall mean success rate ratio* for each optimization method is computed as

$$\mathrm{SRR}^p_m = \frac{\sum_{i=1}^{M} \mathrm{SRR}_{m,i}}{(M-1)} \qquad \text{(higher values are better)}, \tag{C.4}$$

where M is the number of methods.

C.3 Significant Wins

The significant wins ranking uses pairwise hypothesis tests between two optimization methods to determine whether one significantly outperforms the other. In particular, for each problem and pair of methods, a paired t-test is used to compare whether their performance differs. In the t-test, the null hypothesis is that the difference between the mean performance of both algorithms is zero (i.e., they have the same performance), and the alternative hypothesis is that the difference is not equal to zero. Then, based on the result of the statistical test, a value ($SW^p_{m_1,m_2}$) is assigned:

$$SW^p_{m_1,m_2} = \begin{cases} 1 & \text{if the null hypothesis is rejected and } \mathcal{P}^p_{m_2} << \mathcal{P}^p_{m_1} \\ -1 & \text{if the null hypothesis is rejected and } \mathcal{P}^p_{m_1} << \mathcal{P}^p_{m_2} \\ 0 & \text{if the null hypothesis holds} \end{cases}, \tag{C.5}$$

where $\mathcal{P}_m^p$ is the performance of method m over problem p, and $a << b$ denotes that b is significantly better than a. For single-objective problems, the performance $\mathcal{P}_m$ is the median of the minimum values obtained as a result of the optimization process, whereas for multi-objective problems is the median of the hypervolume, and thus significantly *better* implies significantly *smaller* or significantly *bigger*, respectively.

Then, for each pair of methods, the *pairwise estimate of the probability of winning* is computed as

$$P(w)_{m_1,m_2} = \frac{\sum_{p=1}^{P} H\left(SW_{m_1,m_2}^p\right)}{P}, \tag{C.6}$$

where $H(\cdot)$ is the step function with $H(0) = 0$. In other words, the pairwise estimate of probability counts the fraction of problems for which method m_1 is superior to methods m_2.

Finally, for each method the *overall pairwise estimate of the probability of winning* is

$$P(w)_m = \frac{\sum_{i=1}^{M} P(w)_{m,i}}{(M-1)} \qquad \text{(higher values are better)}, \tag{C.7}$$

which is used to create a ranking of optimization methods.

Appendix D

Launch vehicles database

Table D.1: Launch vehicles database.

Name	Country	Cost [k$]	FY [FY]	$payl_{LEO}$ [kg]	$payl_{GEO}$ [kg]	f_D [m]	f_H [m]	f_w [m]	H_{cyl} [m]	D_{cone} [m]	PAF_D [m]	PAF_H [m]	E_H [m]
Falcon 9 (reusable)	USA	50,000	2019	15,000	5,500	4.60	11.00	0.25	6.70	1.45	1.57	0.90	13.00
Falcon 9 (expendable)	USA	62,000	2018	22,800	8,300	4.60	11.00	0.25	6.70	1.45	1.57	0.90	13.00
Falcon Heavy (reusable)	USA	90,000	2018	30,000	8,000	4.60	11.00	0.25	6.70	1.45	1.57	0.90	13.00
Falcon Heavy (semi-reusable)	USA	95,000	2018	56,000	23,000	4.60	11.00	0.25	6.70	1.45	1.57	0.90	13.00
Falcon Heavy (expendable)	USA	150,000	2018	63,800	26,700	4.60	11.00	0.25	6.70	1.45	1.57	0.90	13.00
Delta 2	USA	74,910	2010	5,144	1,800	2.70	7.50	0.15	5.50	0.70	1.50	0.60	8.30
Delta IV Heavy	USA	400,000	2018	28,790	14,210	5.08	19.00	0.25	13.70	2.50	1.50	0.90	19.81
Atlas V 501	USA	137,000	2018	8,123	3,460	3.75	11.10	0.25	5.83	0.91	1.59	0.90	13.00
Atlas V 551	USA	179,000	2018	18,814	8,900	4.57	17.21	0.25	12.19	1.47	1.59	0.90	19.00
Vulcan 501 + Centaur V	USA	100,000	2018	20,000	10,250	4.57	17.21	0.25	12.19	1.47	1.59	0.90	19.00
Vulcan 564 + ACES	USA	150,000	2018	40,000	17,230	4.57	17.21	0.25	12.19	1.47	1.59	0.90	19.00
Antares 230	USA	85,000	2018	6,600	-	3.45	6.80	0.00	4.50	1.38	0.60	1.70	9.90
Ariane 44L	Europe	153,220	2010	10,200	4,790	3.75	9.70	0.20	6.50	1.20	1.57	1.30	11.50
Ariane 5 ECA	Europe	178,000	2018	2,100	9,600	4.57	15.58	0.25	10.00	0.80	2.60	1.60	17.60
Ariance 6 A62	Europe	94,000	2018	-	4,500	4.57	18.00	0.25	11.13	0.57	1.78	0.90	20.60
Ariance 6 A64	Europe	117,000	2018	20,000	12,000	4.57	18.00	0.25	11.13	0.57	1.78	0.90	20.60
Proton M	Russia	65,000	2018	23,000	5,000	4.35	16.94	0.25	12.10	0.40	2.20	2.20	19.60
GSLV	India	47,000	2018	5,000	2,500	3.05	6.40	0.17	3.00	2.20	2.80	1.39	7.70
LVM3	India	60,000	2018	8,000	4,000	3.05	6.40	0.17	3.00	2.20	2.80	1.39	7.70
H2A 202	Japan	90,000	2018	10,000	4,100	3.70	10.20	0.15	5.80	1.10	1.19	1.00	12.00
H2A 204	Japan	106,000	2018	15,000	6,000	3.70	10.20	0.15	5.80	1.10	1.19	1.00	12.00
H2B	Japan	112,500	2018	16,500	8,000	3.70	10.20	0.15	5.80	1.10	1.19	1.00	12.00
H3-24L	Japan	65,000	2018	-	6,500	3.70	10.20	0.15	5.80	1.10	1.19	1.00	12.00
Long March 2C	China	30,645	2010	3,200	1,000	3.00	5.80	0.16	3.60	1.82	1.60	0.90	8.36
Long March 2E	China	68,100	2010	9,200	3,370	3.00	5.80	0.16	3.60	1.82	1.60	0.90	8.36
Long March 3B	China	81,720	2010	13,600	5,200	3.85	7.33	0.20	4.55	1.85	1.21	0.90	9.56
Zenit 3	Ukraine	115,770	2010	15,876	5,250	3.75	8.53	0.15	4.93	1.30	1.80	0.77	10.50
Dnepr	Russia	20,430	2010	4,400	-	2.70	5.16	0.15	3.00	0.40	2.70	0.55	6.10
Soyuz	Russia	51,070	2010	7,000	1,350	3.80	6.55	0.15	4.39	2.50	1.57	0.90	8.30

Appendix E

Performance of Bayesian optimization in tradespace exploration of space networks

In this section, the performance of Bayesian optimization when used to conduct tradespace exploration of space networks is assessed. In particular, results when using three different optimization methods (random sampling (RS), BO-TREE, and BO-RBF) are reported. In all case, the problem was frames as a multi-objective combining pattern, with the decisions and options those shown in Table 7.10.

A total of 100 function evaluations were allowed, and each optimization method was run 20 times, each with a different random seed. For each of the seeds, the first ten function evaluations are common to all methods (i.e., they serve as a starting population common to all three methods). To evaluate the performance of the different algorithms, the probability distribution function (PDF) of the tradespace point-distributions a) when using the optimization method and b) when having a full enumeration of the tradespace, were compared. In particular, the metric used to assess the performance of the algorithms is the ratio between both PDFs:

$$\eta = \frac{\frac{1}{20}\sum_{r=0}^{20} \text{PDF}_{\text{algorithm}^r}}{\text{PDF}_{\text{full-enumeration}}} \tag{E.1}$$

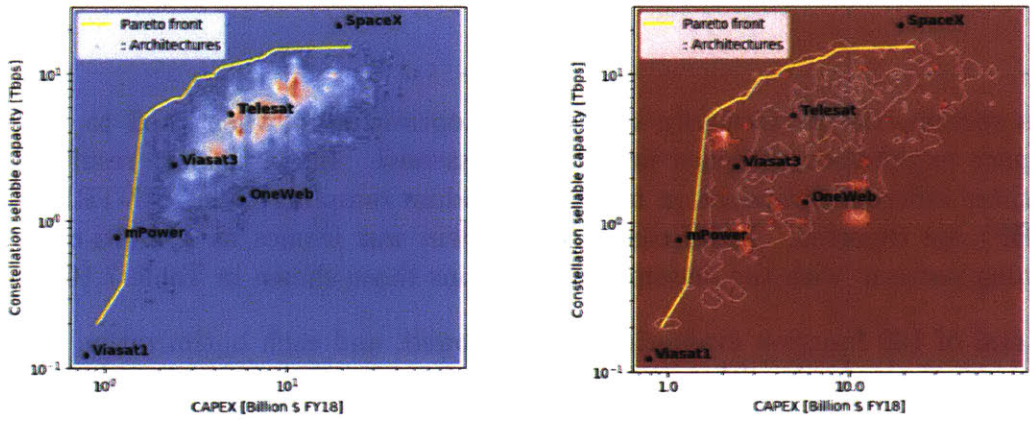
where $\frac{1}{20}\sum_{r=0}^{20} \text{PDF}_{\text{algorithm}^r}$ is the average PDF of the distribution of points in the tradespace across the 20 runs of each algorithm.

In those regions of the tradespace where $\eta > 1$, there is an unconventionally high distribution of points (i.e., the optimization method focused its search on that region of the tradespace) and they are over-sampled, whereas in those regions where $\eta < 1$ are being under-sampled. In particular, it would be desirable for the different optimization methods to over-sample the regions of the tradespace close to the Pareto-

front.

E.1 Sellable capacity vs. CapEx to begin service

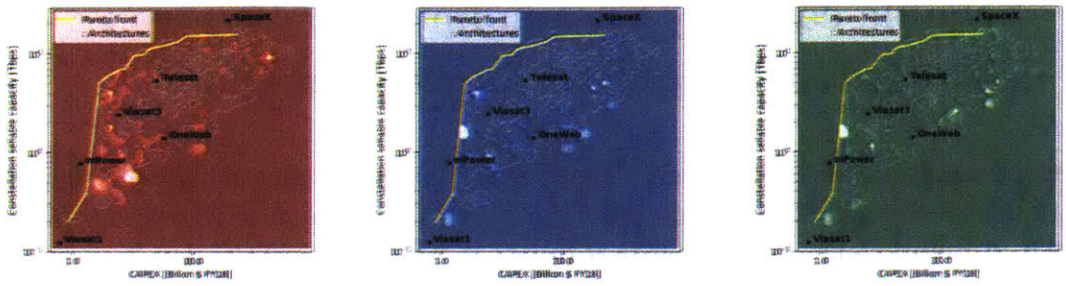
First, the tradespace formed by the sellable capacity versus CapEx costs to set up service. The point-distribution PDF of full-enumerated tradespace is depicted in Figure E-1a, whereas Figure E-1b shows the PDF of the starting ten samples (common to all algorithms). Figure E-2 shows the value of η after 25, 50, 75, and 100 function evaluations for each optimization method. Bright colors indicate that the optimization method over-sampled that region of space (i.e., the optimization method trends to evaluate points in that region), whereas dark colors denote regions where there were barely samples evaluated.



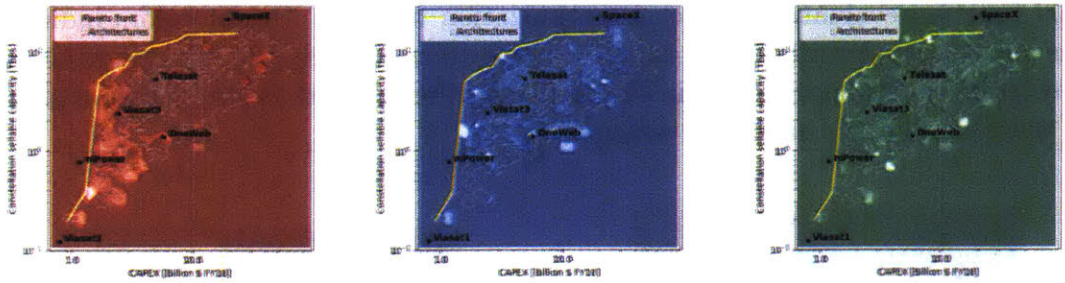
(a) **(b)**

Figure E-1: a) Point-distribution PDF for the sellable capacity vs. CapEx to begin service tradespace; b) PDF of starting ten samples common to all algorithms

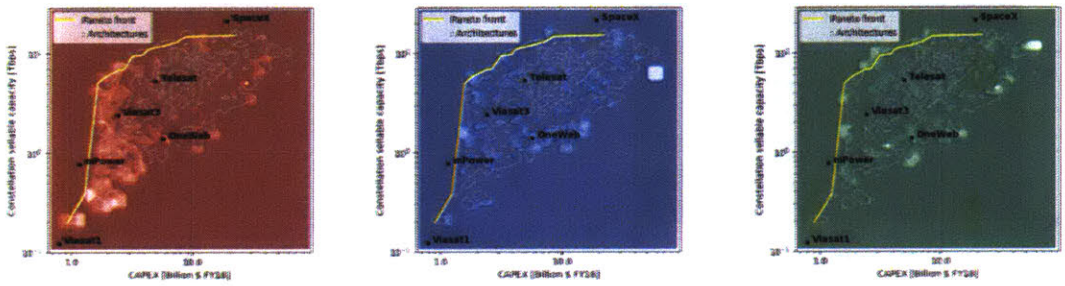
It can be observed that BO-RBF performs the best, in the sense that it is the optimization method that over-samples the most the region of the tradespace close to the Pareto-front (depicted with a yellow line). In particular, from iteration 50 onward (i.e., after having evaluated 50 functions or more), it can be seen that most of the points concentrate in a vertical band very close to the Pareto front. Note that this does not mean that the method does not sample point on other regions, but that the density of point in this particular region is relatively high (as compared to the number of points that would be sampled if random evaluation were to be used). On the other hand, BO-TREE presents a performance comparable to random sampling, which makes it unsuitable for this particular problem.



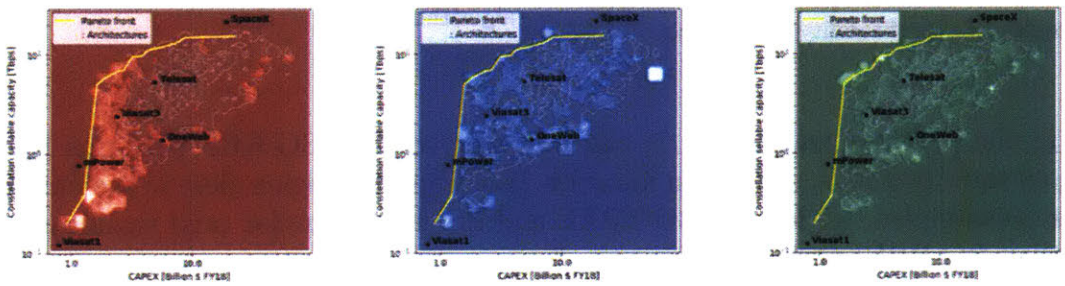
(a) BO-RBF $N_f = 25$ (b) BO-TREE $N_f = 25$ (c) RS $N_f = 25$



(d) BO-RBF $N_f = 50$ (e) BO-TREE $N_f = 50$ (f) RS $N_f = 50$



(g) BO-RBF $N_f = 75$ (h) BO-TREE $N_f = 75$ (i) RS $N_f = 75$

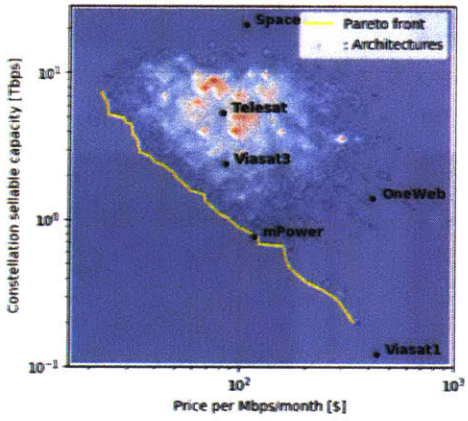


(j) BO-RBF $N_f = 100$ (k) BO-TREE $N_f = 100$ (l) RS $N_f = 100$

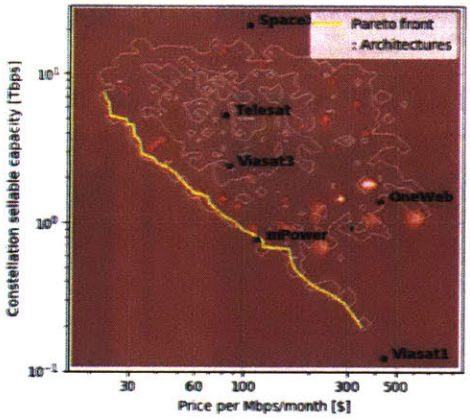
Figure E-2: Performance of Bayesian optimization when used to conduct tradespace exploration of space networks. The value of η is displayed when using the BO-RBF (left column, in red tones), BO-TREE (middle column, in blue tones), and RS (right column, green tones) optimization methods.

E.2 Sellable capacity vs. minimum price per Mbps/month required for profitability

The second study involves the alternative view of the tradespace where the sellable capacity is plotted against minimum price per Mbps/month required for profitability, as shown in Figure 7-11. The point-distribution PDF of full-enumerated tradespace is depicted in Figure E-3a, whereas Figure E-1b shows the PDF of the starting ten samples (common to all algorithms). Figure E-4 shows the value of η after 25, 50, 75, and 100 function evaluations for each optimization method. Bright colors indicate that the optimization method over-sampled that region of space (i.e., the optimization method trends to evaluate points in that region), whereas dark colors denote regions where there were barely samples evaluated.

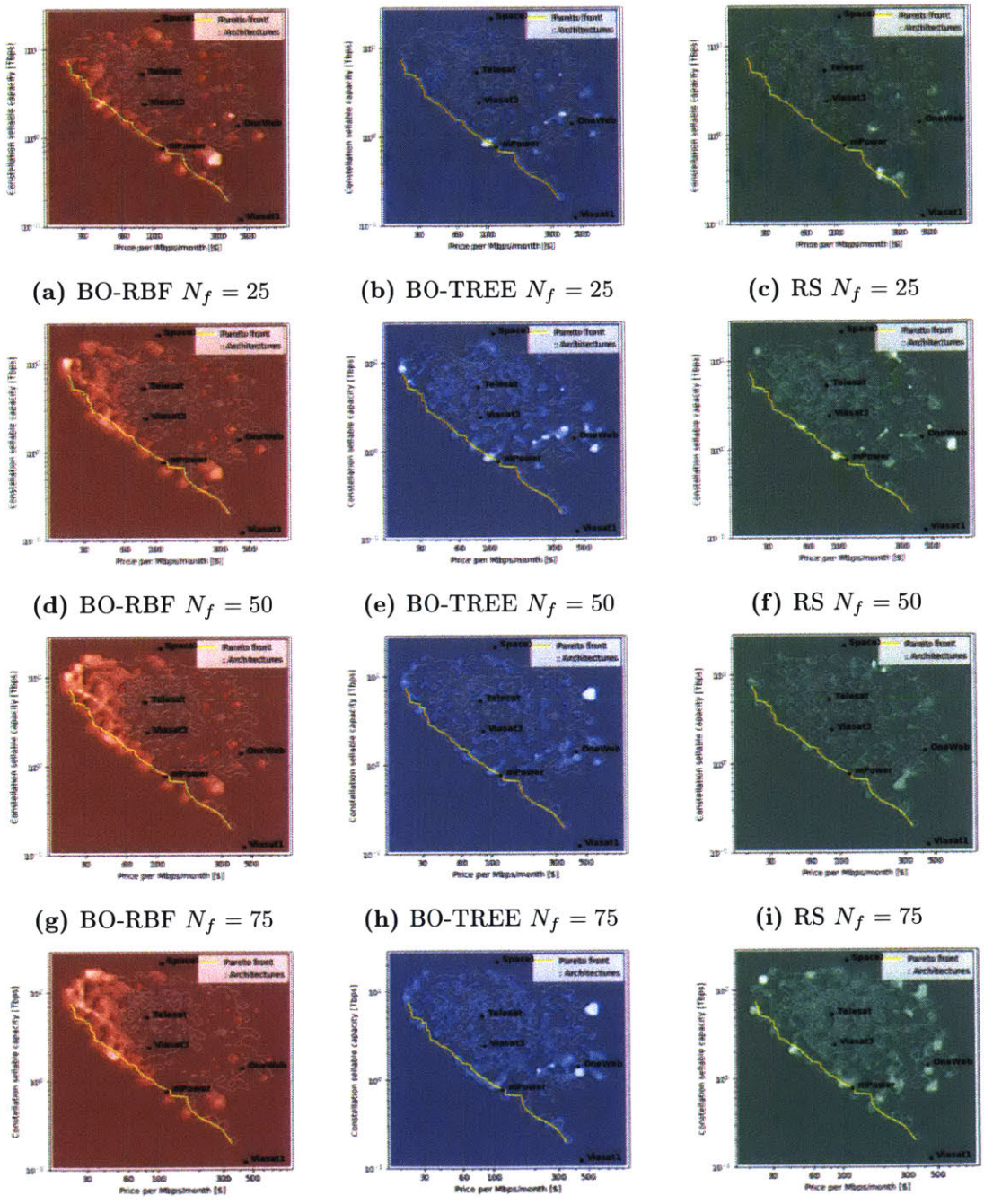


(a) Point-distribution PDF for the sellable capacity vs. minimum price per Mbps/month required for profitability tradespace



(b) PDF of starting ten samples common to all algorithms

As in the previous tradespace, it can be observed that BO-RBF performs best among all three methods, since from iteration 25 onward, the region close to the Pareto front is oversampled, while both BO-TREE and RS show a similar (and significantly worse than BO-RBF) performance.



(a) BO-RBF $N_f = 25$ (b) BO-TREE $N_f = 25$ (c) RS $N_f = 25$

(d) BO-RBF $N_f = 50$ (e) BO-TREE $N_f = 50$ (f) RS $N_f = 50$

(g) BO-RBF $N_f = 75$ (h) BO-TREE $N_f = 75$ (i) RS $N_f = 75$

(j) BO-RBF $N_f = 100$ (k) BO-TREE $N_f = 100$ (l) RS $N_f = 100$

Figure E-4: Performance of Bayesian optimization when used to conduct tradespace exploration of space networks. The value of η is displayed when using the BO-RBF (left column, in red tones), BO-TREE (middle column, in blue tones), and RS (right column, green tones) optimization methods.

Acronyms

ACM Adaptive Coding And Modulation.
ACO Ant Colony Optimization.
ADCS Attitude Determination And Control Subsystem.
AHP Analytic Hierarchy Process.
ALCCA Aircraft Life Cycle Cost Analysis.
ANN Artificial Neural Network.
AR Average Ranking.
ARPU Average Revenue Per User.
ASIC Application-specific Integrated Circuit .
AXI Advanced extensible interface.
BO Bayesian Optimization.
BO4CO Bayesian Optimization For Configuration Optimization.
CAGR Compound Annual Growth Rate.
CapEx Capital Expenditure.
CDF Cumulative Distribution Function.
CDMA Code-division Multiple Access.
CER Cost Estimating Relationship.
CFD Computer Fluid Dynamics.
CPE Customer Premise Equipment.
CSM Concept Selection Methods.
DC Direct Current.
DoE Design Of Experiments.
DSL Digital Subscriber Line.
DSM Desing Structure Matrix.
EDGE Enhanced Data Rates For GSM Evolution.
EHF Extremely High Frequency.
EHI Expected Hypervolume Improvement.
EI Expected Improvement.
EIRP Effective Isotropic Radiated Power.
ENGGEN Engine Cycle Analysis Program.
FLOPS FLight OPtimization System.
FPGA Field-programmable Gate Array.
FSP Flow Shop Problem.
FWD Forward Link.
FY Fiscal Year.
GA Genetic Algorithm.
GAP Generalized Assignment Problem.
GEO Geostationary Orbit.
GINA Generalized Information Network Analysis.
GNI Gross National Income.
GP Gaussian Process.
GPRS General Packet Radio Services.
GPWv4 Gridded Population Of The World.
GSM Global System For Mobile Communications.
HALE High-altitude Long-endurance.
HAP High Altitude Platform.
HoQ House Of Quality.
HSPA+ Evolved High Speed Packet Access.
HTA Heavier-than-air.
HTS High Throughput Satellite.

IRR Internal Rate Of Return.
ISL Inter-satellite Link.
ISP Internet Service Provider.
ITU International Telecommunications Union.
IXP Internet Exchange Point.
JAXA Japan Aerospace Exploration Agency.
LAP Low Altitude Platform.
LEO Low Earth Orbit.
LoS Line-of-sight.
LTA Lighter-than-air.
LTE Long-Term Evolution.
MDO Multidisciplinary Design Optimization.
MEO Medium Earth Orbit.
MIMO Multiple-input Multiple-output.
MNO Mobile Network Operator.
MODCOD Modulation And Coding Scheme.
MPSoC Multi-processor System On Chip.
NGO Non-governmental Organization.
NGSO Non-geostationary Satellite Orbit.
NOAA National Oceanic And Atmospheric Administration.
NSGA-II Non-Dominated Sorting Genetic Algorithm-II.
OBO Output-backoff.
OpEx Operational Expenditure.
OXC Optical Cross-connect.
PCGS Product Concept Generation And Selection.
PDF Probability Distribution Function.
PES Predictive Entroypy Search.
PI Probability Of Improvement.
PoP Point Of Presence.
PSO Particle Swarm Optimization.
QAP Quadratic Assignment Problem.
QFD Quality Function Deployment.
RBF Radial Basis Function.
RBFN Radial Basis Function Network.
RF Radiocommunications Frequency.
RS Random Sampling.
RSM Response Surface Methodology.
SAP System Architecture Problems.
SMAC Sequential Model For Algorithms Configuration.
SMAD Space Mission Analysis And Design.
SPEA2 Strength Pareto Evolutionary Algorithm.
SRR Success Rate Ratio.
SSCM Small Spacecraft Cost Model.
SUR Sequential Uncertainty Reduction.
SW Significant Wins.
TPE Tree-structured Parzen Density Estimator.
TSE Tradespace Exploration.
TSP Traveling Salesman Problem.
TT&C Telemetry, Tracking, And Command.
UAV Unmanned Aerial Vehicle.
UCB Upper Confidence Bound.
UMTS Universal Mobile Telecommunications Service.
USAF United States Air Force.
USCM Unmanned Space Vehicle Cost Model.
USGS United States Geological Survey.
vHTS Very High Throughput Satellite.
WRC World Radiocommunication Congress.

Index of terms

Bibliography

[1] ITU, "Measuring digital development. Facts and figures 2019," 2019.

[2] R. Katz and F. Callorda, "The economic contribution of broadband, digitalization and ICT regulation." http://handle.itu.int/11.1002/pub/811e77bf-en, 2018. Accessed: 2019-03-31.

[3] ITU, "Connect 2020 Agenda for global telecommunication/information and communication technology development.." https://www.itu.int/en/connect2020/Pages/default.aspx. Accessed: 2018-01-13.

[4] U. B. Commission *et al.*, "2025 Targets: "Connecting the Other Half"," 2018.

[5] I. Statistics, "International Telecommunication Union. ITU ICT statistics," 2017.

[6] I. Philbeck, "Connecting the unconnected: Working together to achieve Connect 2020 Agenda Targets ," tech. rep., Broadband Commission and the World Economic Forum at Davos Annual Meeting 2017, 2017.

[7] F.-S. Wu and W.-L. Chu, "Diffusion models of mobile telephony," *Journal of Business Research*, vol. 63, no. 5, pp. 497–501, 2010.

[8] K. Sprague, F. Grijpink, J. Manyika, L. Moodley, B. Chappuis, K. Pattabiraman, and J. Bughin, "Offline and falling behind: Barriers to Internet adoption," *McKinsey & Company, Tech. Rep*, 2014.

[9] Facebook, "State of connectivity 2015: A report on global internet access," tech. rep., Facebook, 2015.

[10] GSMA, "Digital inclusion and mobile sector taxation 2016," tech. rep., GSMA, 2016.

[11] Broadband Commission for Sustainable Development, ITU, UNESCO, "Broadband targets for 2025." https://www.gsmaintelligence.com/research/

?file=b9a6e6202ee1d5f787cfebb95d3639c5&download, 2018. Accessed: 2019-03-30.

[12] R. Thanki, "Measuring the local impact of TVWS broadband," tech. rep., ECS Partners, 2015.

[13] M. Ravallion, S. Chen, and P. Sangraula, "New evidence on the urbanization of global poverty," *Population and Development Review*, vol. 33, no. 4, pp. 667–701, 2007.

[14] C. A. Mattson and A. Messac, "Concept selection using s-pareto frontiers," *AIAA journal*, vol. 41, no. 6, pp. 1190–1198, 2003.

[15] M. Buonanno and D. Mavris, "Aerospace vehicle concept selection using parallel, variable fidelity genetic algorithms," in *10th AIAA/ISSMO Multidisciplinary Analysis and Optimization Conference*, vol. 30, 2004.

[16] C. A. Mattson and A. Messac, "Concept selection in n-dimension using s-pareto frontiers and visualization. 9th AIAA," in *ISSMO Symposium on Multidisciplinary Analysis and Optimization. Atlanta, GA, USA, September*, pp. 4–6, 2002.

[17] A. Moshaiov and G. Avigad, "Concept-based multi-objective problems and their solution by EC," in *Proceedings of the 9th annual conference companion on Genetic and evolutionary computation*, pp. 2865–2868, ACM, 2007.

[18] L. McCullers, "Flight optimization system," *NASA Langley Research Center, Hampton, VA*, 2001.

[19] N. Siegmund, S. S. Kolesnikov, C. Kästner, S. Apel, D. Batory, M. Rosenmüller, and G. Saake, "Predicting performance via automated feature-interaction detection," in *Proceedings of the 34th International Conference on Software Engineering*, pp. 167–177, IEEE Press, 2012.

[20] M. Zuluaga, P. Milder, and M. Püschel, "Streaming sorting networks," *ACM Transactions on Design Automation of Electronic Systems (TODAES)*, vol. 21, no. 4, p. 55, 2016.

[21] O. Almer, N. Topham, and B. Franke, "A learning-based approach to the automated design of MPSoC networks," in *International Conference on Architecture of Computing Systems*, pp. 243–258, Springer, 2011.

[22] L. Gu, "A comparison of polynomial based regression models in vehicle safety analysis," in *ASME Design Engineering Technical Conferences, ASME Paper No.: DETC/DAC-21083*, 2001.

[23] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in neural information processing systems*, pp. 1097–1105, 2012.

[24] S. Jeong, M. Murayama, and K. Yamamoto, "Efficient optimization design method using kriging model," *Journal of aircraft*, vol. 42, no. 2, pp. 413–420, 2005.

[25] R. Garnett, M. A. Osborne, and S. J. Roberts, "Bayesian optimization for sensor set selection," in *Proceedings of the 9th ACM/IEEE International Conference on Information Processing in Sensor Networks*, pp. 209–219, ACM, 2010.

[26] R. Marchant and F. Ramos, "Bayesian optimisation for intelligent environmental monitoring," in *Intelligent Robots and Systems (IROS), 2012 IEEE/RSJ International Conference on*, pp. 2242–2249, IEEE, 2012.

[27] D. J. Lizotte, T. Wang, M. H. Bowling, and D. Schuurmans, "Automatic gait optimization with gaussian process regression.," in *IJCAI*, vol. 7, pp. 944–949, 2007.

[28] R. Martinez-Cantin, N. de Freitas, A. Doucet, and J. A. Castellanos, "Active policy learning for robot planning and exploration under uncertainty.," in *Robotics: Science and Systems*, pp. 321–328, 2007.

[29] H. Wang, J. Shi, and Z. Zhu, "An expected hypervolume improvement algorithm for architectural exploration of embedded processors," in *Design Automation Conference (DAC), 2016 53nd ACM/EDAC/IEEE*, pp. 1–6, IEEE, 2016.

[30] V. Cisco, "Cisco visual networking index: Forecast and trends, 2017–2022," *White Paper*, vol. 1, 2018.

[31] N. Starosielski, *The undersea network.* Duke University Press, 2015.

[32] M. W. Maier, *The art of systems architecting.* CRC press, 2009.

[33] ISO, May, "Systems and software engineering–architecture description," tech. rep., ISO/IEC/IEEE 42010, 2011.

[34] E. Crawley, B. Cameron, and D. Selva, *System architecture: Strategy and product development for complex systems.* Prentice Hall Press, 2015.

[35] C. P. Frank, M. F. Atanian, O. J. Pinon-Fischer, and D. N. Mavris, "A conceptual design framework for performance, life-cycle cost, and safety evaluation of suborbital vehicles," in *Proc., 57th AIAA Science and Technology Forum and Exposition 2016 (AIAA SciTech 2016)*, 2016.

[36] W. L. Simmons, *A framework for decision support in systems architecting.* PhD thesis, Massachusetts Institute of Technology, 2008.

[37] F. Zwicky, *Morphology of propulsive power.* Society for Morphological Research, 1962.

[38] D. Selva, B. Cameron, and E. Crawley, "Patterns in system architecture decisions," *Systems Engineering*, vol. 19, no. 6, pp. 477–497, 2016.

[39] A. M. Ross and D. E. Hastings, "11.4. 3 the tradespace exploration paradigm," in *INCOSE international Symposium*, vol. 15, pp. 1706–1718, Wiley Online Library, 2005.

[40] M. Daskilewicz and B. German, "Aspects of effective visualization of multidimensional design spaces," in *9th AIAA Aviation Technology, Integration, and Operations Conference (ATIO) and Aircraft Noise and Emissions Reduction Symposium (ANERS)*, p. 6929, 2009.

[41] M. E. Fitzgerald and A. M. Ross, "Controlling for framing effects in multi-stakeholder tradespace exploration," *Procedia Computer Science*, vol. 28, pp. 412–421, 2014.

[42] G. B. Shaw, D. W. Miller, and D. E. Hastings, *The generalized information network analysis methodology for distributed satellite systems.* PhD thesis, Citeseer, 1999.

[43] C. D. Jilla, *A multiobjective, multidisciplinary design optimization methodology for the conceptual design of distributed satellite systems.* Thesis, Massachusetts Institute of Technology, 2002.

[44] E. Papapetrou, I. Gragopoulos, and F. Pavlidou, "Performance evaluation of LEO satellite constellations with inter-satellite links under self-similar and poisson tra c," *International Journal of Satellite Communications*, vol. 17, pp. 51–64, 1999.

[45] J. H. J. H.-C. Bau, *Topologies for satellite constellations in a cross-linked space backbone network.* Thesis, Massachusetts Institute of Technology, 2002.

[46] M. Guelman, A. Kogan, A. Kazarian, A. Livne, M. Orenstein, and H. Michalik, "Acquisition and pointing control for inter-satellite laser communications," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 40, no. 4, pp. 1239–1248, 2004.

[47] J.-W. Lee, J.-W. Lee, T.-W. Kim, and D.-U. Kim, "Satellite over satellite (SOS) network: a novel concept of hierarchical architecture and routing in satellite network," in *Local Computer Networks, 2000. LCN 2000. Proceedings. 25th Annual IEEE Conference on*, pp. 392–399, IEEE, 2000.

[48] C. C. Gumbert, *Assessing future growth potential of mobile satellite systems using a cost per billable minute metric.* Thesis, Massachusetts Institute of Technology, 1996.

[49] A. Kelic, *Assessing the technical and financial viability of broadband satellite systems using a cost per T1 minute metric.* Thesis, Massachusetts Institute of Technology, 1998.

[50] O. de Weck, D. Zhang, and C. Jilla, "Architecture trade methodology for LEO personal communication systems," in *20th AIAA International Communication Satellite Systems Conference and Exhibit*, p. 1866, 2002.

[51] O. L. De Weck, R. D. Neufville, and M. Chaize, "Staged deployment of communications satellite constellations in low earth orbit," *Journal of Aerospace Computing, Information, and Communication*, vol. 1, no. 3, pp. 119–136, 2004.

[52] D. Mignolo, E. Re, A. Ginesi, A. B. Alamanac, P. Angeletti, and M. Harverson, "Approaching terabit/s satellite: a system analysis," in *17th Ka and Broadband Communications Conference*, 2011.

[53] P. Inigo, O. Vidal, B. Roy, E. Alberty, N. Metzger, D. Galinier, J. Anzalchi, G. Huggins, and S. Stirland, "Review of terabit/s satellite, the next generation of HTS systems," in *Advanced Satellite Multimedia Systems Conference and the 13th Signal Processing for Space Communications Workshop (ASMS/SPSC), 2014 7th*, pp. 318–322, IEEE, 2014.

[54] E. Cianca, T. Rossi, A. Yahalom, Y. Pinhasi, J. Farserotu, and C. Sacchi, "EHF for satellite communications: The new broadband frontier," *Proceedings of the IEEE*, vol. 99, no. 11, pp. 1858–1881, 2011.

[55] T. Rossi, F. Maggio, M. De Sanctis, M. Ruggieri, S. Falzini, and M. Tosti, "System analysis of smart gateways techniques applied to Q/V-band high throughput satellites," in *Aerospace Conference, 2014 IEEE*, pp. 1–10, IEEE, 2014.

[56] N. Perlot, T. Dreischer, C. M. Weinert, and J. Perdigues, "Optical GEO feeder link design," in *Future Network & Mobile Summit (FutureNetw), 2012*, pp. 1–8, IEEE, 2012.

[57] D. Giggenbach, E. Lutz, J. Poliak, R. Mata-Calvo, and C. Fuchs, "A high-throughput satellite system for serving whole europe with fast internet service, employing optical feeder links," in *Broadband Coverage in Germany. 9th ITG Symposium. Proceedings*, pp. 1–7, VDE, 2015.

[58] Space Exploration Holdings, LLC, "SpaceX Ka-band NGSO constellation FCC filing SAT-LOA-20161115-00118." Available at http://licensing.fcc.gov/myibfs/forwardtopublictabaction.do?file_number=SATLOA2016111500118 (2017/09/12).

[59] Telesat Canada, "Telesat Ka-band NGSO constellation FCC filing SAT-PDR-20161115-00108." Available at http://licensing.fcc.gov/myibfs/forwardtopublictabaction.do?file_number=SATPDR2016111500108 (2017/09/12).

[60] WorldVu Satellites Limited, "OneWeb Ka-band NGSO constellation FCC filing SAT-LOI-20160428-00041." Available at http://licensing.fcc.gov/myibfs/forwardtopublictabaction.do?file_number=SATLOI2016042800041 (2017/09/12).

[61] I. del Portillo, B. Cameron, and E. Crawley, "Ground segment architectures for large LEO constellations with feeder links in EHF-bands," in *Aerospace Conference, 2018 IEEE*, pp. 1–14, IEEE, 2018.

[62] N. Pachler, M. Guerster, I. del Portillo, E. F. Crawley, and B. G. Cameron, "Static beam placement and frequency plan algorithms for LEO constellations," *International Journal of Satellite Communications and Networking (under review)*, 2020.

[63] J. J. G. Luis, N. Pachler, M. Guerster, I. del Portillo, E. F. Crawley, and B. G. Cameron, "Artificial intelligence algorithms for power allocation in high throughput satellites: A comparison," in *2020 IEEE Aerospace Conference (under review)*, 2020.

[64] B. Barritt and W. Eddy, "Temporospatial SDN for aerospace communications," in *AIAA SPACE 2015 Conference and Exposition*, p. 4656, 2015.

[65] B. Barritt, T. Kichkaylo, K. Mandke, A. Zalcman, and V. Lin, "Operating a UAV mesh & internet backhaul network using temporospatial SDN," in *Aerospace Conference, 2017 IEEE*, pp. 1–7, IEEE, 2017.

[66] I. T. Union, "ITU-R : Revised technical and operational parameters for typical IMT-2000 terrestrial systems using high altitude platform stations and CDMA radio transmission technologies - Document, 8-1/307-E," tech. rep., International Telecommunications Union, 1999.

[67] G. M. Djuknic, J. Freidenfelds, and Y. Okunev, "Establishing wireless communications services via high-altitude aeronautical platforms: a concept whose time has come?," *IEEE Communications Magazine*, vol. 35, no. 9, pp. 128–135, 1997.

[68] Y.-C. Lee and H. Ye, "Sky station stratospheric telecommunications system, a high speed low latency switched wireless network," in *17th AIAA International Communications Satellite Systems Conference and Exhibit*, p. 1394, 1998.

[69] I. F. Akyildiz, X. Wang, and N. J. Colella, "HALO (High Altitude Long Operation): a broadband wireless metropolitan area network," in *Mobile Multimedia Communications, 1999.(MoMuC'99) 1999 IEEE International Workshop on*, pp. 271–275, IEEE, 1999.

[70] J. Munk, "[–the wireless solution," in *The 3rd Stratospheric Platform Systems Workshop*, pp. 45–51, 2001.

[71] I. Smith, M. Lee, M. Fortneberry, and R. Judy, "HiSentinel80: flight of a high altitude airship," in *11th AIAA Aviation Technology, Integration, and Operations (ATIO) Conference, including the AIAA Balloon Systems Conference and 19th AIAA Lighter-Than*, p. 6973, 2011.

[72] M. Nakadate, "Development and flight test of SPF-2 low altitude stationary flight test vehicle," in *AIAA 5th ATIO and16th Lighter-Than-Air Sys Tech. and Balloon Systems Conferences*, p. 7408, 2005.

[73] D.-M. Kim, "Korea stratospheric airship program and current results," in *AIAA's 3rd Annual Aviation Technology, Integration, and Operations (ATIO) Forum*, p. 6782, 2003.

[74] Y.-G. Lee, D.-M. Kim, and C.-H. Yeom, "Development of korean high altitude platform systems," *International Journal of Wireless Information Networks*, vol. 13, no. 1, pp. 31–42, 2006.

[75] F. Baurreau, R. Staraj, F. Ferrero, L. Lizzi, J.-M. Ribero, and J.-P. Chessel, "Stratospheric platform for telecommunication missions," in *Antennas and Propagation & USNC/URSI National Radio Science Meeting, 2015 IEEE International Symposium on*, pp. 914–915, IEEE, 2015.

[76] T. E. Noll, S. D. Ishmael, B. Henwood, M. E. Perez-Davis, G. C. Tiffany, J. Madura, M. Gaier, J. M. Brown, and T. Wierzbanowski, "Technical findings, lessons learned, and recommendations resulting from the helios prototype vehicle mishap," tech. rep., NATIONAL AERONAUTICS AND SPACE ADMIN LANGLEY RESEARCH CENTER HAMPTON VA, 2007.

[77] X. Zhu, Z. Guo, and Z. Hou, "Solar-powered airplanes: A historical perspective and future challenges," *Progress in Aerospace Sciences*, vol. 71, pp. 36–53, 2014.

[78] A. Rapinett, "Zephyr: a high altitude long endurance unmanned air vehicle," *Doctor, Department of Physics, University of Surrey*, 2009.

[79] P. Richfield, "Aloft for 5 years: DARPA's vulture project aims for ultra long UAV missions," *Defense News*, vol. 22, no. 30, p. 30, 2007.

[80] A. Colozza and J. Dolce, "Initial feasibility assessment of a high altitude long endurance airship," tech. rep., NASA, 2003.

[81] D. Grace and M. Mohorcic, *Broadband communications via high-altitude platforms.* John Wiley & Sons, 2011.

[82] d'Oliveira, Flavio Araripe and Melo, Francisco Cristovão Lourenço de and Devezas, Tessaleno Campos, "High-altitude platforms—present situation and technology trends," *Journal of Aerospace Technology and Management*, vol. 8, no. 3, pp. 249–262, 2016.

[83] S. Karapantazis and F. Pavlidou, "Broadband communications via high-altitude platforms: A survey," *IEEE Communications Surveys & Tutorials*, vol. 7, no. 1, pp. 2–31, 2005.

[84] T. Tozer and D. Grace, "High-altitude platforms for wireless communications," *Electronics & Communication Engineering Journal*, vol. 13, no. 3, pp. 127–137, 2001.

[85] D. Grace, N. Daly, T. C. Tozer, A. G. Burr, and D. A. Pearce, "Providing multimedia communications services from high altitude platforms," *International Journal of Satellite Communications and Networking*, vol. 19, no. 6, pp. 559–580, 2001.

[86] P. Pace, G. Aloi, F. De Rango, E. Natalizio, A. Molinaro, and S. Marano, "An integrated satellite-HAP-terrestrial system architecture: resources allocation and traffic management issues," in *Vehicular Technology Conference, 2004. VTC 2004-Spring. 2004 IEEE 59th*, vol. 5, pp. 2872–2875, IEEE, 2004.

[87] A. Mohammed, A. Mehmood, F.-N. Pavlidou, and M. Mohorcic, "The role of high-altitude platforms (HAPs) in the global wireless connectivity," *Proceedings of the IEEE*, vol. 99, no. 11, pp. 1939–1953, 2011.

[88] G. Araniti, A. Iera, and A. Molinaro, "The role of HAPs in supporting multimedia broadcast and multicast services in terrestrial-satellite integrated systems," *Wireless Personal Communications*, vol. 32, no. 3-4, pp. 195–213, 2005.

[89] S. Karapantazis and F.-N. Pavlidou, "The role of high altitude platforms in beyond 3G networks," *IEEE Wireless Communications*, vol. 12, no. 6, pp. 33–41, 2005.

[90] V. Milas, M. Koletta, and P. Constantinou, "Interference and compatibility studies between satellite service systems and systems using high altitude platform stations," in *ESA Special Publication*, vol. 541, 2003.

[91] R. Miura and M. Oodo, "Wireless communications system using stratospheric platforms: R and D program on telecom and broadcasting system using high altitude platform stations," *Journal of the Communication Research Laboratory, vol. 48, iss. No. 4, p. 33-48*, vol. 48, pp. 33–48, 2001.

[92] B.-J. Ku *et al.*, "Conceptual design of multibeam antennas (MBA) and user antenna for stratospheric communication system (SCS)," in *Proceedings of the second stratospheric platform systems workshop 2000*, pp. 163–170, 2000.

[93] T. Hult and A. Mohammed, "Compact MIMO antennas and HAP diversity for high data rate communications," in *Proc. 64th IEEE Veh. Technol. Conf.*, pp. 25–28, 2006.

[94] P. Likitthanasate, D. Grace, and P. D. Mitchell, "Spectrum etiquettes for terrestrial and high-altitude platform-based cognitive radio systems," *IET communications*, vol. 2, no. 6, pp. 846–855, 2008.

[95] M. Antonini, S. Betti, V. Carrozzo, E. Duca, and M. Ruggieri, "Feasibility analysis of a HAP-LEO optical link for data relay purposes," in *Aerospace Conference, 2006 IEEE*, pp. 7–pp, IEEE, 2006.

[96] E. Cestino, "Design of solar high altitude long endurance aircraft for multi payload & operations," *Aerospace science and technology*, vol. 10, no. 6, pp. 541–550, 2006.

[97] B. S. d. Mattos, N. R. Secco, and E. F. Salles, "Optimal design of a high-altitude solar-powered unmanned airplane," *Journal of Aerospace Technology and Management*, vol. 5, no. 3, pp. 349–361, 2013.

[98] M. J. Burton and W. W. Hoburg, "Solar-electric and gas powered, long-endurance UAV sizing via geometric programming," in *18th AIAA/ISSMO Multidisciplinary Analysis and Optimization Conference*, p. 4147, 2017.

[99] A. Colozza and J. L. Dolce, "High-altitude, long-endurance airships for coastal surveillance," *NASA Technical Report, NASA/TM-2005-213427*, 2005.

[100] C. Nickol, M. Guynn, L. Kohout, and T. Ozoroski, "High altitude long endurance air vehicle analysis of alternatives and technology requirements development," in *45th AIAA Aerospace Sciences Meeting and Exhibit*, p. 1050, 2007.

[101] D. Yu and X. Lv, "Configurations analysis for high-altitude/long-endurance airships," *Aircraft Engineering and Aerospace Technology*, vol. 82, no. 1, pp. 48–59, 2010.

[102] A. Qiantori, A. B. Sutiono, H. Hariyanto, H. Suwa, and T. Ohta, "An emergency medical communications system by low altitude platform at the early stages of a natural disaster in indonesia," *Journal of medical systems*, vol. 36, no. 1, pp. 41–52, 2012.

[103] S. H. Alnajjar, F. Malek, M. S. Razalli, and M. S. Ahmad, "Low-altitude platform to enhance communications reliability in disaster environments," *Journal of Advances in Information Technology*, vol. 5, no. 1, pp. 21–30, 2014.

[104] A. Valcarce, T. Rasheed, K. Gomez, S. Kandeepan, L. Reynaud, R. Hermenier, A. Munari, M. Mohorcic, M. Smolnikar, and I. Bucaille, "Airborne base stations for emergency and temporary events," in *International Conference on Personal Satellite Services*, pp. 13–25, Springer, 2013.

[105] F. C. Commission *et al.*, "The role of deployable aerial communications architecture in emergency communications and recommended next steps," *Washington, DC: FCC*, 2011.

[106] J. M. Sullivan, "Revolution or evolution? the rise of the UAVs," in *Technology and Society, 2005. Weapons and Wires: Prevention and Safety in a Time of Fear. ISTAS 2005. Proceedings. 2005 International Symposium on*, pp. 94–101, IEEE, 2005.

[107] A. Bleicher, "5G goes for the gold," *IEEE Spectrum*, vol. 55, no. 1, pp. 32–33, 2018.

[108] A. Al-Hourani, S. Kandeepan, and A. Jamalipour, "Modeling air-to-ground path loss for low altitude platforms in urban environments," in *Global Communications Conference (GLOBECOM), 2014 IEEE*, pp. 2898–2904, IEEE, 2014.

[109] B. Van der Bergh, A. Chiumento, and S. Pollin, "LTE in the sky: trading off propagation benefits with interference costs for aerial nodes," *IEEE Communications Magazine*, vol. 54, no. 5, pp. 44–50, 2016.

[110] N. Ahmed, S. S. Kanhere, and S. Jha, "On the importance of link characterization for aerial wireless sensor networks," *IEEE Communications Magazine*, vol. 54, no. 5, pp. 52–57, 2016.

[111] R. Amorim, H. Nguyen, P. Mogensen, I. Z. Kovács, J. Wigard, and T. B. Sørensen, "Radio channel modeling for UAV communication over cellular networks," *IEEE Wireless Communications Letters*, vol. 6, no. 4, pp. 514–517, 2017.

[112] A. Al-Hourani, S. Kandeepan, and S. Lardner, "Optimal LAP altitude for maximum coverage," *IEEE Wireless Communications Letters*, vol. 3, no. 6, pp. 569–572, 2014.

[113] A. Al-Hourani, S. Chandrasekharan, A. Jamalipour, L. Reynaud, and S. Kandeepan, "Optimal cluster head spacing for energy-efficient communication in aerial-backhauled networks," in *Global Communications Conference (GLOBECOM), 2015 IEEE*, pp. 1–6, IEEE, 2015.

[114] M. Mozaffari, W. Saad, M. Bennis, and M. Debbah, "Efficient deployment of multiple unmanned aerial vehicles for optimal wireless coverage," *IEEE Communications Letters*, vol. 20, no. 8, pp. 1647–1650, 2016.

[115] K. Sundaresan, E. Chai, A. Chakraborty, and S. Rangarajan, "SkyLiTE: End-to-end design of low-altitutde UAV networks for providing LTE connectivity," *arXiv preprint arXiv:1802.06042*, 2018.

[116] L. Gupta, R. Jain, and G. Vaszkun, "Survey of important issues in UAV communication networks," *IEEE Communications Surveys & Tutorials*, vol. 18, no. 2, pp. 1123–1152, 2016.

[117] S. Chandrasekharan, K. Gomez, A. Al-Hourani, S. Kandeepan, T. Rasheed, L. Goratti, L. Reynaud, D. Grace, I. Bucaille, T. Wirth, *et al.*, "Designing and implementing future aerial communication networks," *IEEE Communications Magazine*, vol. 54, no. 5, pp. 26–34, 2016.

[118] Y.-C. Liu, A. Chakrabarti, and T. Bligh, "Towards an 'ideal' approach for concept generation," *Design Studies*, vol. 24, no. 4, pp. 341–355, 2003.

[119] G. E. Okudan and S. Tauhid, "Concept selection methods–a literature review from 1980 to 2008," *International Journal of Design Engineering*, vol. 1, no. 3, pp. 243–277, 2008.

[120] S. Pugh and D. Clausing, *Creating innovative products using total design: the living legacy of Stuart Pugh.* Addison-Wesley Longman Publishing Co., Inc., 1996.

[121] C. A. Mattson and A. Messac, "Pareto frontier based concept selection under uncertainty, with visualization," *Optimization and Engineering*, vol. 6, no. 1, pp. 85–115, 2005.

[122] S. Pugh, "Total design: integrated methods for successful product engineering," *1990*, 1991.

[123] A. M. Gonçalves-Coelho, A. J. Mourão, and Z. L. Pereira, "Improving the use of QFD with axiomatic design," *Concurrent Engineering*, vol. 13, no. 3, pp. 233–239, 2005.

[124] S. Ter Harr, D. Clausing, and S. Eppinger, "Integration of quality function deployment and the design structure matrix," *Cambridge, MA*, 1993.

[125] K. N. Otto and E. K. Antonsson, "The method of imprecision compared to utility theory for design selection problems," *Design Theory and Methodology–DTM'93*, pp. 167–173, 1993.

[126] G. Pahl and W. Beitz, "Engineering design, the design council," *London, Springer*, vol. 12, pp. 221–226, 1984.

[127] R. L. Keeney and H. Raiffa, "Decisions with multiple objectives: Preferences and value tradeoffs," 1976.

[128] D. L. Thurston, "A formal method for subjective design evaluation with multiple attributes," *Research in engineering Design*, vol. 3, no. 2, pp. 105–122, 1991.

[129] T. Santy, "The analytical hierarchy process: Planning, priority setting, resource allocation," *Decision Making Series, McGraw Hill, New York, USA*, 1980.

[130] T. L. Saaty, "What is the analytic hierarchy process?," in *Mathematical models for decision support*, pp. 109–121, Springer, 1988.

[131] E. Marsh, "Hierarchical decision making in machine design," 1993.

[132] M. Mullens, R. Armacost, and R. Nippani, "A two-stage approach to concept selection using the analytic hierarchy process," *International Journal of Industrial Engineering*, vol. 2, no. 3, pp. 199–208, 1995.

[133] K. N. Otto and K. L. Wood, "Estimating errors in concept selection," in *Proc. of ASME Design Theory and Methodology Conference*, pp. 397–412, 1995.

[134] S. Takai and K. Ishii, "Modifying Pugh's design concept evaluation methods," in *DETC2004-57512. ASME design engineering technical conferences, Salt Lake City, UT*, 2004.

[135] P. Van Laarhoven and W. Pedrycz, "A fuzzy extension of saaty's priority theory," *Fuzzy sets and Systems*, vol. 11, no. 1-3, pp. 229–241, 1983.

[136] D.-Y. Chang, "Applications of the extent analysis method on fuzzy AHP," *European journal of operational research*, vol. 95, no. 3, pp. 649–655, 1996.

[137] W. Lee, H. Lau, Z.-z. Liu, and S. Tam, "A fuzzy analytic hierarchy process approach in modular product design," *Expert Systems*, vol. 18, no. 1, pp. 32–42, 2001.

[138] W. Yan, C.-H. Chen, and M.-D. Shieh, "Product concept generation and selection using sorting technique and fuzzy c-means algorithm," *Computers & Industrial Engineering*, vol. 50, no. 3, pp. 273–285, 2006.

[139] J. H. Holland, "Adaptation in natural and artificial systems. an introductory analysis with application to biology, control, and artificial intelligence," *Ann Arbor, MI: University of Michigan Press*, 1975.

[140] G. Avigad and A. Moshaiov, "Set-based concept selection in multi-objective problems: optimality versus variability approach," *Journal of Engineering Design*, vol. 20, pp. 217–242, June 2009.

[141] A. Moshaiov and Y. Snir, "Tailoring ϵ-MOEA to concept-based problems," *Parallel Problem Solving from Nature-PPSN XII*, pp. 122–131, 2012.

[142] G. Avigad and A. Moshaiov, "Simultaneous concept-based evolutionary multi-objective optimization," *Applied Soft Computing*, vol. 11, pp. 193–207, Jan. 2011.

[143] G. Avigad and A. Moshaiov, "Set-based concept selection in multi-objective problems involving delayed decisions," *Journal of Engineering Design*, vol. 21, pp. 619–646, Dec. 2010.

[144] A. Snir, B. Samina, and A. Moshaiov, "Concept-Based Evolutionary Multi-Criteria Exploration of Design Spaces Under Run-Time Limitation," in *2015 IEEE Symposium Series on Computational Intelligence*, pp. 853–860, IEEE, 2015.

[145] G. Avigad, A. Moshaiov, and N. Brauner, *Interactive Concept-based Search using MOEA: The Hierarchical Preferences Case.* EUSFLAT, Jan. 2005.

[146] G. Avigad and A. Moshaiov, "Interactive evolutionary multiobjective search and optimization of set-based concepts," *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, vol. 39, no. 4, pp. 1013–1027, 2009.

[147] S. Kim, "Gradient-based simulation optimization," in *Proceedings of the 2006 winter simulation conference*, pp. 159–167, IEEE, 2006.

[148] T. Simpson, V. Toropov, V. Balabanov, and F. Viana, "Design and analysis of computer experiments in multidisciplinary design optimization: A review of how far we have come-or not," in *12th AIAA/ISSMO multidisciplinary analysis and optimization conference*, p. 5802, 2008.

[149] W.-L. Loh *et al.*, "On latin hypercube sampling," *The annals of statistics*, vol. 24, no. 5, pp. 2058–2080, 1996.

[150] G. M. Stump, M. Yukish, T. W. Simpson, and J. J. O'Hara, "Trade space exploration of satellite datasets using a design by shopping paradigm," in *2004 IEEE Aerospace Conference Proceedings (IEEE Cat. No. 04TH8720)*, vol. 6, pp. 3885–3895, IEEE, 2004.

[151] A. B. Owen, "Orthogonal arrays for computer experiments, integration and visualization," *Statistica Sinica*, pp. 439–452, 1992.

[152] J. M. Hammersley, "Monte carlo methods for solving multivariable problems," *Annals of the New York Academy of Sciences*, vol. 86, no. 3, pp. 844–874, 1960.

[153] G. Stump, S. Lego, M. Yukish, T. W. Simpson, and J. A. Donndelinger, "Visual steering commands for trade space exploration: User-guided sampling with example," *Journal of Computing and Information Science in Engineering*, vol. 9, no. 4, 2009.

[154] E. Spero, C. L. Bloebaum, B. J. German, A. Pyster, and A. M. Ross, "A research agenda for tradespace exploration and analysis of engineered resilient systems," *Procedia Computer Science*, vol. 28, pp. 763–772, 2014.

[155] L. Cevallos-Torres and M. Botto-Tobar, "Process sampling," in *Problem-Based Learning: A Didactic Strategy in the Teaching of System Simulation*, pp. 13–31, Springer, 2019.

[156] M. Eigen, "Ingo Rechenberg Evolutionsstrategie Optimierung technischer Systeme nach Prinzipien der biologishen Evolution," *mit einem Nachwort von Manfred Eigen, Friedrich Frommann Verlag, Struttgart-Bad Cannstatt*, vol. 45, pp. 46–47, 1973.

[157] D. E. Goldberg, *Genetic algorithms*. Pearson Education India, 2006.

[158] J. Kennedy, "Swarm intelligence," in *Handbook of nature-inspired and innovative computing*, pp. 187–219, Springer, 2006.

[159] J. Kennedy, "Particle swarm optimization," in *Encyclopedia of machine learning*, pp. 760–766, Springer, 2011.

[160] M. Dorigo, M. Birattari, and T. Stutzle, "Ant colony optimization," *IEEE computational intelligence magazine*, vol. 1, no. 4, pp. 28–39, 2006.

[161] D. Selva, B. G. Cameron, and E. F. Crawley, "Rule-based system architecting of earth observing systems: Earth science decadal survey," *Journal of Spacecraft and Rockets*, 2014.

[162] M. S. Net, I. del Portillo, B. Cameron, E. F. Crawley, and D. Selva, "Integrated tradespace analysis of space network architectures," *Journal of Aerospace Information Systems*, 2015.

[163] I. del Portillo, M. Sanchez, B. Cameron, and E. Crawley, "Architecting the ground segment of an optical space communication network," in *Aerospace Conference, 2016 IEEE*, pp. 1–13, IEEE, 2016.

[164] C. P. Frank, R. A. Marlier, O. J. Pinon-Fischer, and D. N. Mavris, "An evolutionary multi-architecture multi-objective optimization algorithm for design space exploration," in *Proc., 57th AIAA/ASCE/AHS/ASC Structures, Structural Dynamics, and Materials Conf*, 2016.

[165] C. Erbas, S. Cerav-Erbas, and A. D. Pimentel, "Multiobjective optimization and evolutionary algorithms for the application mapping problem in multiprocessor system-on-chip design," *IEEE Transactions on Evolutionary Computation*, vol. 10, no. 3, pp. 358–374, 2006.

[166] T. W. Simpson, J. Poplinski, P. N. Koch, and J. K. Allen, "Metamodels for computer-based engineering design: survey and recommendations," *Engineering with computers*, vol. 17, no. 2, pp. 129–150, 2001.

[167] R. Jin, W. Chen, and T. W. Simpson, "Comparative studies of metamodelling techniques under multiple modelling criteria," *Structural and multidisciplinary optimization*, vol. 23, no. 1, pp. 1–13, 2001.

[168] G. E. Box and K. B. Wilson, "On the experimental attainment of optimum conditions," in *Breakthroughs in Statistics*, pp. 270–310, Springer, 1992.

[169] R. H. Myers, D. C. Montgomery, and C. M. Anderson-Cook, *Response surface methodology: process and product optimization using designed experiments*. John Wiley & Sons, 2016.

[170] V. Balabanov, M. Kaufman, D. Knill, D. Haim, O. Golovidov, A. Giunta, R. T. Haftka, B. Grossman, W. Mason, and L. Watson, "Dependence of optimal structural weight on aerodynamic shape for a high speed civil transport," *AIAA Paper*, pp. 96–4046, 1996.

[171] R. Unal, R. Lepsch, W. Engelund, and D. Stanley, "Approximation model building and multidisciplinary design optimization using response surface methods," in *Proc. 6th AIAA/NASA/ISSMO Symp. on Multidisciplinary Analysis and Optimization, Part I*, 1996.

[172] R. Unal, D. O. Stanley, and C. R. Joyner, "Parametric model building and design optimization using response surface methods," *Journal of Parametrics*, vol. 14, no. 1, pp. 81–96, 1994.

[173] A. Khawas, A. Banerjee, and S. Mukhopadhyay, "A response surface method for design space exploration and optimization of analog circuits," in *VLSI (ISVLSI), 2011 IEEE Computer Society Annual Symposium on*, pp. 84–89, IEEE, 2011.

[174] N. R. Costa and J. A. Lourenço, "Exploring pareto frontiers in the response surface methodology," in *Transactions on Engineering Technologies*, pp. 399–412, Springer, 2015.

[175] G. Palermo, C. Silvano, and V. Zaccaria, "ReSPIR: a response surface-based pareto iterative refinement for application-specific design space exploration," *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, vol. 28, no. 12, pp. 1816–1829, 2009.

[176] J. Sobieszczanski-Sobieski and R. T. Haftka, "Multidisciplinary aerospace design optimization: survey of recent developments," *Structural optimization*, vol. 14, no. 1, pp. 1–23, 1997.

[177] A. A. Giunta *et al.*, *Aircraft multidisciplinary design optimization using design of experiments theory and response surface modeling methods.* virginia polytechnic institute and state university Blacksburg, VA, 1997.

[178] J. M. Zurada, *Introduction to artificial neural systems*, vol. 8. West publishing company St. Paul, 1992.

[179] T. W. Simpson, J. Peplinski, P. N. Koch, and J. K. Allen, "On the use of statistics in design and the implications for deterministic computer experiments," *Design Theory and Methodology-DTM'97*, pp. 14–17, 1997.

[180] J. P. Dauby and C. H. Dagli, "A neural network approach to modeling system integration sensitivity for architectural assessment," in *Intelligent Engineering Systems Through Artificial Neural Networks*, ASME Press, 2009.

[181] H. Shi, Y. Gao, and X. Wang, "Optimization of injection molding process parameters using integrated artificial neural network model and expected improvement function method," *The International Journal of Advanced Manufacturing Technology*, vol. 48, no. 9-12, pp. 955–962, 2010.

[182] H. Shi, S. Xie, and X. Wang, "A warpage optimization method for injection molding using artificial neural network with parametric sampling evaluation strategy," *The International Journal of Advanced Manufacturing Technology*, vol. 65, no. 1-4, pp. 343–353, 2013.

[183] D. Verbeeck, F. Maes, K. De Grave, and H. Blockeel, "Multi-objective optimization with surrogate trees," in *Proceedings of the 15th annual conference on Genetic and evolutionary computation*, pp. 679–686, ACM, 2013.

[184] D. G. Krige, "A statistical approach to some basic mine valuation problems on the Witwatersrand," *Journal of the Southern African Institute of Mining and Metallurgy*, vol. 52, no. 6, pp. 119–139, 1951.

[185] T. W. Simpson, T. M. Mauery, J. J. Korte, and F. Mistree, "Kriging models for global approximation in simulation-based multidisciplinary design optimization," *AIAA journal*, vol. 39, no. 12, pp. 2233–2241, 2001.

[186] M. G. Fernández-Godino, C. Park, N.-H. Kim, and R. T. Haftka, "Review of multi-fidelity models," *arXiv preprint arXiv:1609.07196*, 2016.

[187] E. Huang, J. Xu, S. Zhang, and C.-H. Chen, "Multi-fidelity model integration for engineering design," *Procedia Computer Science*, vol. 44, pp. 336–344, 2015.

[188] D. R. Jones, M. Schonlau, and W. J. Welch, "Efficient global optimization of expensive black-box functions," *Journal of Global optimization*, vol. 13, no. 4, pp. 455–492, 1998.

[189] J. Močkus, "On Bayesian methods for seeking the extremum," in *Optimization Techniques IFIP Technical Conference*, pp. 400–404, Springer, 1975.

[190] H. J. Kushner, "A new method of locating the maximum point of an arbitrary multipeak curve in the presence of noise," *Journal of Basic Engineering*, vol. 86, no. 1, pp. 97–106, 1964.

[191] N. Srinivas, A. Krause, S. M. Kakade, and M. Seeger, "Gaussian process optimization in the bandit setting: No regret and experimental design," *arXiv preprint arXiv:0912.3995*, 2009.

[192] E. Kaufmann, O. Cappé, and A. Garivier, "On Bayesian upper confidence bounds for bandit problems," in *Artificial Intelligence and Statistics*, pp. 592–600, 2012.

[193] J. Villemonteix, E. Vazquez, and E. Walter, "An informational approach to the global optimization of expensive-to-evaluate functions," *Journal of Global Optimization*, vol. 44, no. 4, p. 509, 2009.

[194] S. L. Scott, "A modern Bayesian look at the multi-armed bandit," *Applied Stochastic Models in Business and Industry*, vol. 26, no. 6, pp. 639–658, 2010.

[195] P. Hennig and C. J. Schuler, "Entropy search for information-efficient global optimization," *Journal of Machine Learning Research*, vol. 13, no. Jun, pp. 1809–1837, 2012.

[196] E. Brochu, M. W. Hoffman, and N. de Freitas, "Portfolio allocation for bayesian optimization," *arXiv preprint arXiv:1009.5419*, 2010.

[197] K. Swersky, J. Snoek, and R. P. Adams, "Multi-task Bayesian optimization," in *Advances in neural information processing systems*, pp. 2004–2012, 2013.

[198] D. Yogatama and G. Mann, "Efficient transfer learning method for automatic hyperparameter tuning," in *Artificial Intelligence and Statistics*, pp. 1077–1085, 2014.

[199] R. Bardenet, M. Brendel, B. Kégl, and M. Sebag, "Collaborative hyperparameter tuning," in *International Conference on Machine Learning*, pp. 199–207, 2013.

[200] D. Ginsbourger, R. Le Riche, and L. Carraro, "A multi-points criterion for deterministic parallel global optimization based on Gaussian processes," *hal-00260579f*, 2008.

[201] R. Lam, K. Willcox, and D. H. Wolpert, "Bayesian optimization with a finite budget: An approximate dynamic programming approach," in *Advances in Neural Information Processing Systems*, pp. 883–891, 2016.

[202] F. Hutter, H. H. Hoos, and K. Leyton-Brown, "Sequential model-based optimization for general algorithm configuration.," *LION*, vol. 5, pp. 507–523, 2011.

[203] J. Snoek, H. Larochelle, and R. P. Adams, "Practical Bayesian optimization of machine learning algorithms," in *Advances in neural information processing systems*, pp. 2951–2959, 2012.

[204] J. Bergstra, D. Yamins, and D. D. Cox, "Hyperopt: A python library for optimizing the hyperparameters of machine learning algorithms," in *Proceedings of the 12th Python in Science Conference*, pp. 13–20, Citeseer, 2013.

[205] J. Knowles, "ParEGO: a hybrid algorithm with on-line landscape approximation for expensive multiobjective optimization problems," *IEEE Transactions on Evolutionary Computation*, vol. 10, pp. 50–66, Feb. 2006.

[206] W. Ponweiser, T. Wagner, D. Biermann, and M. Vincze, "Multiobjective optimization on a limited budget of evaluations using model-assisted\mathcal {S}-metric selection," in *International Conference on Parallel Problem Solving from Nature*, pp. 784–794, Springer, 2008.

[207] M. Zuluaga, G. Sergent, A. Krause, and M. Püschel, "Active learning for multi-objective optimization," in *International Conference on Machine Learning*, pp. 462–470, 2013.

[208] M. Zuluaga, A. Krause, and M. Püschel, "ϵ-pal: an active learning approach to the multi-objective optimization problem," *The Journal of Machine Learning Research*, vol. 17, no. 1, pp. 3619–3650, 2016.

[209] V. Picheny, "Multiobjective optimization using Gaussian process emulators via stepwise uncertainty reduction," *arXiv:1310.0732 [math, stat]*, Oct. 2013.

[210] D. Hernandez-Lobato, J. M. Hernandez-Lobato, A. Shah, and R. P. Adams, "Predictive Entropy Search for Multi-objective Bayesian Optimization," *arXiv:1511.05467 [stat]*, Nov. 2015.

[211] M. Emmerich and J.-w. Klinkenberg, "The computation of the expected improvement in dominated hypervolume of Pareto front approximations," *Rapport technique, Leiden University*, vol. 34, 2008.

[212] J. M. Hernández-Lobato, M. A. Gelbart, B. Reagen, R. Adolf, D. Hernández-Lobato, P. N. Whatmough, D. Brooks, G.-Y. Wei, and R. P. Adams, "Designing neural network hardware accelerators with decoupled objective evaluations," in *NIPS workshop on Bayesian Optimization*, p. 10, 2016.

[213] Y.-H. Kim, A. Moraglio, A. Kattan, and Y. Yoon, "Geometric generalisation of surrogate model-based optimisation to combinatorial and program spaces," *Mathematical Problems in Engineering*, vol. 2014, 2014.

[214] M. Zaefferer, J. Stork, and T. Bartz-Beielstein, "Distance measures for permutations in combinatorial efficient global optimization," in *International Conference on Parallel Problem Solving from Nature*, pp. 373–383, Springer, 2014.

[215] A. Moraglio and A. Kattan, "Geometric surrogate model based optimisation for genetic programming: Initial experiments," *University of Birmingham*, 2011.

[216] M. Zaefferer, J. Stork, M. Friese, A. Fischbach, B. Naujoks, and T. Bartz-Beielstein, "Efficient global optimization for combinatorial problems," in *Proceedings of the 2014 Annual Conference on Genetic and Evolutionary Computation*, pp. 871–878, ACM, 2014.

[217] R. Baptista and M. Poloczek, "Bayesian optimization of combinatorial structures," *arXiv preprint arXiv:1806.08838*, 2018.

[218] C. Oh, J. M. Tomczak, E. Gavves, and M. Welling, "Combinatorial bayesian optimization using graph representations," *arXiv preprint arXiv:1902.00448*, 2019.

[219] T. Wagner, M. Emmerich, A. Deutz, and W. Ponweiser, "On expected-improvement criteria for model-based multi-objective optimization," *Parallel Problem Solving from Nature, PPSN XI*, pp. 718–727, 2010.

[220] F. Hutter, H. H. Hoos, K. Leyton-Brown, and T. Stützle, "ParamILS: an automatic algorithm configuration framework," *Journal of Artificial Intelligence Research*, vol. 36, no. 1, pp. 267–306, 2009.

[221] X. Huan and Y. M. Marzouk, "Simulation-based optimal Bayesian experimental design for nonlinear systems," *Journal of Computational Physics*, vol. 232, no. 1, pp. 288–317, 2013.

[222] C. E. Rasmussen and C. K. Williams, *Gaussian processes for machine learning*, vol. 1. MIT press Cambridge, 2006.

[223] D. D. Cox and S. John, "A statistical method for global optimization," in *[Proceedings] 1992 IEEE International Conference on Systems, Man, and Cybernetics*, pp. 1241–1246, IEEE, 1992.

[224] H. J. Kushner, "A New Method of Locating the Maximum Point of an Arbitrary Multipeak Curve in the Presence of Noise," *Journal of Basic Engineering*, vol. 86, pp. 97–106, 03 1964.

[225] J. M. Hernández-Lobato, M. W. Hoffman, and Z. Ghahramani, "Predictive entropy search for efficient global optimization of black-box functions," in *Advances in neural information processing systems*, pp. 918–926, 2014.

[226] M. Emmerich, "Single-and multi-objective evolutionary design optimization assisted by Gaussian random field metamodels," *Dissertation, LS11, FB Informatik, Universität Dortmund, Germany*, 2005.

[227] J. Mercer, "Xvi. functions of positive and negative type, and their connection the theory of integral equations," *Philosophical transactions of the royal society of London. Series A, containing papers of a mathematical or physical character*, vol. 209, no. 441-458, pp. 415–446, 1909.

[228] T. Hastie, R. Tibshirani, J. Friedman, and J. Franklin, "The elements of statistical learning: data mining, inference and prediction," *The Mathematical Intelligencer*, vol. 27, no. 2, pp. 83–85, 2005.

[229] O. E. Kundakcioglu and S. Alizamir, *Generalized assignment problemGeneralized Assignment Problem*, pp. 1153–1162. Boston, MA: Springer US, 2009.

[230] M. Zuluaga, P. Milder, and M. Püschel, "Computer generation of streaming sorting networks," in *DAC Design Automation Conference 2012*, pp. 1241–1249, IEEE, 2012.

[231] P. Jamshidi and G. Casale, "An uncertainty-aware approach to optimal configuration of stream processing systems," in *2016 IEEE 24th International Symposium on Modeling, Analysis and Simulation of Computer and Telecommunication Systems (MASCOTS)*, pp. 39–48, IEEE, 2016.

[232] S. A. Kauffman and E. D. Weinberger, "The NK model of rugged fitness landscapes and its application to maturation of the immune response," *Journal of theoretical biology*, vol. 141, no. 2, pp. 211–245, 1989.

[233] J. E. Beasley, "OR-Library: distributing test problems by electronic mail," *Journal of the operational research society*, vol. 41, no. 11, pp. 1069–1072, 1990.

[234] R. E. Burkard, S. E. Karisch, and F. Rendl, "QAPLIB-A quadratic assignment problem library," *Journal of Global optimization*, vol. 10, no. 4, pp. 391–403, 1997.

[235] G. Reinelt, "TSPLIB—A traveling salesman problem library," *ORSA journal on computing*, vol. 3, no. 4, pp. 376–384, 1991.

[236] P. B. Brazdil and C. Soares, "A comparison of ranking methods for classification algorithm selection," in *European conference on machine learning*, pp. 63–75, Springer, 2000.

[237] J. Bergstra and Y. Bengio, "Random search for hyper-parameter optimization," *Journal of Machine Learning Research*, vol. 13, no. Feb, pp. 281–305, 2012.

[238] K. Deb, A. Pratap, S. Agarwal, and T. Meyarivan, "A fast and elitist multiobjective genetic algorithm: NSGA-II," *IEEE transactions on evolutionary computation*, vol. 6, no. 2, pp. 182–197, 2002.

[239] E. Zitzler, M. Laumanns, and L. Thiele, "SPEA2: Improving the strength Pareto evolutionary algorithm," *TIK-report*, vol. 103, 2001.

[240] The Economist, "The inclusive Internet index." https://theinclusiveinternet.eiu.com, 2019. Accessed: 2019-04-02.

[241] B. Commission *et al.*, "The state of broadband 2019: Broadband as a foundation for sustainable development," *The State of Broadband report, Geneva*, 2017.

[242] GSMA Intelligence, "The mobile economy 2019." https://www.gsmaintelligence.com/research/?file=b9a6e6202ee1d5f787cfebb95d3639c5&download, 2019. Accessed: 2019-03-30.

[243] GSMA, GSMA Intelligence, UK aid, "The state of mobile internet connectivity 2019." https://www.gsma.com/mobilefordevelopment/wp-content/uploads/2019/07/GSMA-State-of-Mobile-Internet-Connectivity-Report-2019.pdf, 2019. Accessed: 2019-03-30.

[244] S. Schmida, J. Bernard, T. Zakaras, C. Lovegrove, and C. Swingle, "Connecting the next four billion: Strengthening the global response for universal Internet access," *USAID, Dial, SSG Advisors*, 2017.

[245] B. Lucini, "Consumer barriers to mobile internet adoption in africa. GSMA intelligence," 2016.

[246] US Air Force Weather Agency, "Version 4 DMSP-OLS nighttime lights time series." http://www.ngdc.noaa.gov/eog/dmsp/downloadV4composites.html, 2013. Accessed: 2018-11-23.

[247] C. D. Elvidge, K. Baugh, M. Zhizhin, F. C. Hsu, and T. Ghosh, "VIIRS night-time lights," *International Journal of Remote Sensing*, vol. 38, no. 21, pp. 5860–5879, 2017.

[248] IEA, IREW, UNSD, WHO, "The energy progress report." https://trackingsdg7.esmap.org/, 2019. Accessed: 2019-03-31.

[249] Ericsson, "Ericsson mobility report: Middle East and Africa." https://www.ericsson.com/assets/local/press-releases/middle-east/2019/ericsson-mobilityreport-mmea-feb2019.pdf, Februare 2019. Accessed: 2019-03-30.

[250] A. C. Clarke and E.-T. Relays, "Wireless world," 1945.

[251] A. C. Clarke, *The Space-station: Its Radio Applications*. O. Verlag, 1945.

[252] A. C. Clark, "Extra-terrestrial relays: Can rocket stations give world-wide radio coverage," *Wireless World*, vol. 305, 1945.

[253] J. N. Pelton, "History of satellite communications," in *Handbook of Satellite Applications*, pp. 27–66, Springer, 2013.

[254] D. K. Sachdev, "Satellite Communication Technology-Challenges for the 1980s," *Journal of Spacecraft and Rockets*, vol. 18, no. 2, pp. 110–118, 1981.

[255] G. Comparetto and N. Hulkower, "Global mobile satellite communications-a review of three contenders," in *15th International Communicatons Satellite Systems Conference and Exhibit*, p. 1138, 1994.

[256] J. Lim, R. Klein, and J. Thatcher, "Good technology, bad management: A case study of the satellite phone industry," *Journal of Information Technology Management*, vol. 16, no. 2, pp. 48–55, 2005.

[257] S. Finkelstein and S. H. Sanford, "Learning from corporate mistakes: The rise and fall of iridium," *Organizational Dynamics*, vol. 29, no. 2, pp. 138–148, 2000.

[258] O. Vidal, G. Verelst, J. Lacan, E. Alberty, J. Radzik, and M. Bousquet, "Next generation high throughput satellite system," in *Satellite Telecommunications (ESTEL), 2012 IEEE First AESS European Conference on*, pp. 1–7, IEEE, 2012.

[259] I. del Portillo, B. G. Cameron, and E. F. Crawley, "A technical comparison of three low earth orbit satellite constellation systems to provide global broadband," *Acta Astronautica*, vol. 159, pp. 123–135, 2019.

[260] ITU, "Propagation data and prediction methods required for the design of Earth-space telecommunication systems," *Recommendation ITU-R*, pp. 618–12, 2015.

[261] E. ETSI, "ETSI EN 302307-2, digital video broadcasting (DVB); second generation framing structure, channel coding and modulation systems for broadcasting, interactive services, news gathering and other broadband satellite applications; Part II: S2-Extensions (S2-X)," tech. rep., Tech. rep., ETSI, Tech. Rep, 2014.

[262] C. for International Earth Science Information Network (CIESIN)—Columbia University, "Gridded population of the world, version 4 (GPWV4): population density," 2016.

[263] I. del Portillo, B. Cameron, and E. Crawley, "Ground segment architectures for large LEO constellations with feeder links in EHF-bands," in *2018 IEEE Aerospace Conference*, pp. 1–14, IEEE, 2018.

[264] W. J. Larson and J. R. Wertz, "Space mission analysis and design," tech. rep., Torrance, CA (United States); Microcosm, Inc., 1992.

[265] E. D. Miller, M. DeSpenza, I. Gavrilyuk, G. Nelson, B. Erickson, B. Edwards, E. Davis, and T. Truscott, "A prototype coarse pointing mechanism for laser communication," in *Free-Space Laser Communication and Atmospheric Propagation XXIX*, vol. 10096, p. 100960S, International Society for Optics and Photonics, 2017.

[266] S. J. Isakowitz, "International reference guide to space launch systems," *NASA STI/Recon Technical Report A*, vol. 91, 1991.

[267] H. Jones, "The recent large reduction in space launch cost," in *48th International Conference on Environmental Systems*, 2018.

[268] E. Mahr, A. Tu, and A. Gupta, "Development of the small satellite cost model 2014 (SSCM14)," in *2016 IEEE Aerospace Conference*, pp. 1–13, IEEE, 2016.

[269] P. Nguyen, N. Lozzi, W. Galang, S. Hu, A. Sjovold, P. Young, S. Book, and J. Schmilz, "Unmanned space vehicle cost model," *US Air Force Space and Missile Systems Center (SMC/FMC)*, vol. 2430, pp. 90245–4687, 1994.

[270] D. of Defense, "UFC 3-701-01 DoD Facilities Pricing Guide, with Change 4," 2018.

[271] F. G. Ortíz-Gómez, R. Martínez-Rodríguez-Osorio, and S. Landeros-Ayala, "Method of optimizing the costs of a satellite network in Ka and Q/V bands in the feeder link," in *35th AIAA International Communications Satellite Systems Conference*, p. 5442, 2017.

[272] J. Lessner, "Advancements in SATCOM for Airborne C2 and ISR Networks." Slide report, November 2017. Slide 9.

[273] I. M. Sobol, "Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates," *Mathematics and computers in simulation*, vol. 55, no. 1-3, pp. 271–280, 2001.

[274] A. Saltelli, "Making best use of model evaluations to compute sensitivity indices," *Computer physics communications*, vol. 145, no. 2, pp. 280–297, 2002.

[275] J. Herman and W. Usher, "SALib: An open-source Python library for Sensitivity Analysis," *The Journal of Open Source Software*, vol. 2, jan 2017.

[276] ITU, "Minimum performance characteristics and operational conditions for high altitude platform stations providing IMT-2000 in the bands 1 885-1 980 MHz, 2 010-2 025 MHz and 2 110-2 170 MHz in Regions 1 and 3 and 1 885-1 980 MHz and 2 110-2 160 MHz in Region 2," 2000.

[277] ITU, "Technical and operational characteristics for the fixed service using high altitude platform stations in the bands 27.5-28.35 GHz and 31-31.3 GHz," 2002.

[278] ITU, "Frequency sharing between systems in the fixed service using high-altitude platform stations and satellite systems in the geostationary orbit in the fixed-satellite service in the bands 47.2-47.5 and 47.9-48.2 GHz," 2002.

[279] D. Johnson, C. Hill, and L. Tyree, "Terrestrial Environment (climatic) Criteria Guidelines document updated in 1993 for use in aerospace vehicle development," in *32nd Aerospace Sciences Meeting and Exhibit*, p. 481, 1993.

[280] D. Colas, N. H. Roberts, and V. S. Suryakumar, "HALE multidisciplinary design optimization part I: Solar-powered single and multiple-boom aircraft," in *2018 Aviation Technology, Integration, and Operations Conference*, p. 3028, 2018.

[281] D. Colas, N. H. Roberts, and V. S. Suryakumar, "HALE multidisciplinary design optimization part II: Solar-powered flying-wing aircraft," in *2018 Aviation Technology, Integration, and Operations Conference*, p. 3029, 2018.

[282] National oceanic and atmospheric administration, "US standard atmosphere," *National Aeronautics and Space Administration, United States Air Force, Washington, DC*, 1976.

[283] J. Cherwonik and A. Wehrley, "Unmanned Aerial Vehicle System Acquisition Cost Research: Estimating Methodology and Database," *Washington, DC: The Office of the Deputy Assistant of the Army for Cost and Economic*, 2003.

[284] F. K. Hwang and D. S. Richards, "Steiner tree problems," *Networks*, vol. 22, no. 1, pp. 55–89, 1992.

[285] M. R. Garey and D. S. Johnson, "The rectilinear Steiner tree problem is NP-complete," *SIAM Journal on Applied Mathematics*, vol. 32, no. 4, pp. 826–834, 1977.

[286] M. Haklay and P. Weber, "Openstreetmap: User-generated street maps," *IEEE Pervasive Computing*, vol. 7, no. 4, pp. 12–18, 2008.

[287] L. Kou, G. Markowsky, and L. Berman, "A fast algorithm for Steiner trees," *Acta informatica*, vol. 15, no. 2, pp. 141–145, 1981.

[288] G. Weichenberg, V. W. Chan, and M. Médard, "Design and analysis of optical flow-switched networks," *IEEE/OSA Journal of Optical Communications and Networking*, vol. 1, no. 3, pp. B81–B97, 2009.

[289] M. Mahloo, P. Monti, J. Chen, and L. Wosinska, "Cost modeling of backhaul for mobile networks," in *2014 IEEE international conference on communications workshops (ICC)*, pp. 397–402, IEEE, 2014.

[290] W. Xie, N.-T. Mao, and K. Rundberget, "Cost comparisons of backhaul transport technologies for 5G fixed wireless access," in *2018 IEEE 5G World Forum (5GWF)*, pp. 159–163, IEEE, 2018.

[291] A. Jarvis, H. I. Reuter, A. Nelson, E. Guevara, *et al.*, "Hole-filled SRTM for the globe version 4," *available from the CGIAR-CSI SRTM 30m Database (http://srtm. csi. cgiar. org)*, vol. 15, pp. 25–54, 2008.

[292] GSMA, "Mobile backhaul options. spectrum analysis and recommendations," *White Paper*, vol. 1, 2018.

[293] C. V. Forecast, "Cisco visual networking index: Forecast and trends, 2017–2022," *White paper, Cisco Public Information*, 2019.

[294] R. L. Phillips, *Pricing and revenue optimization.* Stanford University Press, 2005.

[295] D. Selva, *Rule-based System Architecting of Earth Observation Satellite Systems.* PhD thesis, Massachusetts Institute of Technology Department of Aeronautics and Astronautics, Cambridge, Massachusetts, June 2012.

[296] M. Sanchez Net, I. del Portillo, B. G. Cameron, E. F. Crawley, and D. Selva, "Integrated tradespace analysis of space network architectures," *Journal of Aerospace Information Systems*, vol. 12, no. 8, pp. 564–578, 2015.